12-2016

What's Scale Got to do With It? Models for Urban Tree Canopy

Dexter H. Locke
Clark University

Shawn M. Landry
University of South Florida, landry@usf.edu

J. M. Grove
USDA Forest Service

Rinku R. Chowdhury
Clark University

Journal of Urban Ecology, 2016, 1–16

doi: 10.1093/jue/juw006
Research article

What's scale got to do with it? Models for urban tree canopy

Dexter H. Locke,[1,*] Shawn M. Landry,[2] J. Morgan Grove,[3] and Rinku Roy Chowdhury[1]

[1]Graduate School of Geography, Clark University 950 Main Street, Worcester, MA, 01610-1477 USA, [2]School of Geosciences, University of South Florida, 4202 E. Fowler Ave, NES107, Tampa, FL, 33620 USA and [3]Baltimore Field Station, USDA Forest Service, Suite 350, 5523 Research Park Dr, Baltimore, MD, 21228 USA

*Corresponding author. E-mail:dexter.locke@gmail.com

Abstract

The uneven provisioning of ecosystem services has important policy implications; yet the spatial heterogeneity of tree canopy remains understudied. Private residential lands are important to the future of Philadelphia's urban forest because a majority of the existing and possible tree canopy is located on residential land uses. This article examines the spatial distribution of tree canopy in Philadelphia, PA and its social correlates. How are existing tree canopy and opportunities for additional tree canopy distributed across the city of Philadelphia and with respect to three explanations: (i) population density, (ii) the social stratification luxury effect, and 3) lifestyle characteristics of residents? This study used spatial autoregressive regression (SAR), geographically weighted regression (GWR) and multilevel modeling (MLM) to evaluate population density, social stratification luxury effect, and lifestyle characteristics as explanations of the spatial distribution of existing and possible tree canopy, and simultaneously evaluate the efficacy of different statistical analysis techniques. To control for spatial autocorrelation, SAR models were estimated, GWR models were fit to examine potential spatial non-stationarity and realism of the SAR analyses. The MLMs both controlled for spatial autocorrelation (like SAR) but also allow local variation and spatial non-stationarity (like GWR). The multimodel inferential approach showed the statistical models that included lifestyle characteristics outperformed the social stratification and population density models. Our results cast doubt on findings from previous studies using areal units such as block groups. More sophisticated statistical analyses suggest opportunities for enhancing theory and the need to reconsider frequently used methods.

Key words: urban vegetation, multi-level modeling, mixed effects models, urban ecology, tree canopy

Introduction

In recognition of the benefits of trees in 2009, Philadelphia's Mayor Michael Nutter established ambitious urban forestry goals in the City's *Greenworks Philadelphia* plan. Specifically, the plan called for an increase in tree canopy in each neighborhood to reach 30% of land area by the year 2025 (Greenworks—City of Philadelphia 2009). This goal supports a regional, three-state effort to plant one million trees (Plant One Million 2013).

An analysis of high-resolution land cover for the City (O'Neil-Dunne et al., 2012) and the City's tax map indicates that 26% of the city's land base is residential land, and 23% of the city's overall tree canopy is located on residential land uses. In order to reach the 30% goal, Philadelphia needs to add 8,442 acres of new or expanded tree canopy—all of which could be located on private land. Specifically, there are 10 079 acres of

Submitted: 30 June 2016; **Received (in revised form):** 17 September 2016; **Accepted:** 3 October 2016

private residential land that are non-road, non-building, non-water, not existing tree canopy. These areas could support tree canopy. In the context of the overall city, 24% of all opportunities for additional tree canopy are on these residential lands. Given the preponderance of tree canopy on private residential properties, homeowners in urban areas with tree canopy may be understood as the new 'forest owner' (Grove et al., 2013: 377).

Currently, government agencies' activities such as tree planting and maintenance are generally limited to public properties. Therefore, the City's initiative to increase tree canopy is likely to fail to reach its canopy goal if the new 'forest owner' does not participate in planting and stewardship of this dominant, yet seemingly neglected portion of the urban forest: trees on private residential lands. A critical component to increase participation on residential lands is to better understand the spatial distribution of the existing urban forest, opportunities for increasing the urban forest, and factors that shape these spatial distributions.

The purpose of this article is to examine the spatial distribution of tree canopy in Philadelphia, PA and its social correlates. Two questions are addressed: How are (i) existing tree canopy, (ii) and the opportunities for additional tree canopy distributed across the city of Philadelphia and with respect to three mid-level explanations: (i) population density, (ii) the luxury effect and (iii) lifestyle characteristics of residents? Spatial econometric, geographically weighted regression (GWR), and multi-level models (MLM) are used to evaluate these three social theories, and simultaneously evaluate the efficacy of different statistical analysis techniques. Because each theory relates to different potential management strategies, finding more support for one explanation or another provides a rationale for different types of urban forestry practices.

Literature review

Explanations for the distribution of urban vegetation

Over the past 10 years, there has been a growing academic interest in understanding the human drivers and mechanisms that lead to the abundance and distribution of urban vegetation, including trees (Cook et al., 2012). Most research has focused on several complementary explanations, including: population density; explanations related to social stratification (e.g. spatial mobility and neighborhood turnover, access to power, the luxury effect); and lifestyle and reference group behavior theory often termed 'the ecology of prestige'. Researchers have also recognized the importance of historical lags and landscape legacies (e.g. Luck et al., 2009; Boone et al., 2010; Clarke et al., 2013; Bigsby et al., 2014; Grove et al., 2014; Locke and Baine 2014).

The population density explanation suggests that urbanization displaces native ecosystems (Smith et al., 2005; Marco et al., 2008). Roads, buildings and other impervious surfaces leave less space in places that may otherwise host vegetation such as trees. Empirical studies that examined relationships between population density and tree cover have shown statistically significant negative correlations in cities such as Baltimore, MD (Troy et al., 2007), Tampa, FL (Landry and Pu 2010) and Montreal (Pham et al., 2012b) and positive correlations between population density and residential tree cover in Raleigh, NC (Bigsby et al., 2014). These mixed results suggest the spatial distribution of tree canopy is not driven solely by population density. Indeed, several studies suggest that population density is important but that other potential drivers may need to be considered (e.g. Troy et al., 2007).

Social stratification explanations have largely focused on spatial mobility and neighborhood turnover, access to power, and the luxury effect. Mobility afforded through affluence may lead wealthier families to move to neighborhoods with more perceived amenities (Logan and Molotch 1987) which may include vegetation on both public and private lands (Roy Chowdhury et al., 2011). Empirically testing this explanation requires long-term data on vegetation change, real estate transactions and several other potential confounders. Data availability and the analytical complexity are plausible reasons why there is little empirical research on how urban vegetation may attract members of social groups who can afford to move to greener neighborhoods.

Access to power is another variation on social stratification. Specifically, variations in power and income across neighborhoods may lead to different levels of public investment, including investments in green infrastructure and environmental amenities such as street trees and public open space and parks (Grove et al., 2006b). Members of some socioeconomic and/or demographic groups attract more public investment in local greening initiatives than others (Logan and Molotch 1987: 39; Perkins et al., 2004). Since public investments primarily occur on public lands, this explanation largely pertains to vegetation such as street trees in the public right of way, or trees in public parks. The focus of this article, however, is on existing and possible tree canopy on private residential lands where access to power is a less relevant explanation.

Another popular variation on social stratification is the 'luxury effect'. Unlike access to power, the luxury effect is specific to private lands. Proponents of the 'luxury effect' posit that households with more disposable income will purchase more of everything, 'ceteris paribus', including greater expenditures on yard care and yard care services (Hope et al., 2003, 2006; Martin et al., 2004). If the luxury effect plays a strong role in explaining the distribution of tree canopy, then programs that reduce the cost of planting trees for residents might be effective among lower-income households.

Lifestyle explanations for the distribution of urban vegetation are complementary with population density and all three variations on social stratification explanations. For instance, within a given degree of population density 'and' socioeconomic status there are variations in the abundance of urban vegetation associated with indicators of different lifestyles in Baltimore, MD (Grove et al., 2006a; Troy et al., 2007) and New York City (Grove et al., 2014). The so-called 'Ecology of Prestige' lifestyle-based theory posits that in addition to population density and social stratification, a desire to gain acceptance in a neighborhood may influence households' land management decisions (Logan and Molotch 1987: 107–8; Grove et al., 2006a,b; Troy et al., 2007; Zhou et al., 2009; Grove et al., 2014). The idea stems from reference group behavior theory, which describes the process of evaluation and self-appraisal. Individuals adopt the values, standards or norms of a social group as a frame of reference for their own behavior (Hyman 1942; Merton and Kitt 1950). The theory of an Ecology of Prestige proposes a mechanism for how neighborhood scale social norms particular to a lifestyle group become internalized into households' behaviors.

In addition to population density, the three social stratification theory variants, and the ecological prestige theory, historical lags and landscape legacies are sometimes included. The main idea is that trees take time to grow. Therefore, previous residents and former conditions may provide useful insights into the spatial distribution of tree canopy at present. This article does not formally test the idea of legacy effects. Instead it is taken as a given because of the sound logic and abundant empirical and consistent support elsewhere (i.e. Luck et al. 2009; Boone et al. 2010; Clarke et al. 2013; Bigsby et al., 2014; Grove et al., 2014; Locke and Baine 2014). However, a building age variable is examined to explore how time of development is linked with the urban forest today.

The importance of multi-level modeling

Advances in remote sensing technologies, including LiDAR, make sub-meter land cover mapping in urban areas easier, cheaper and more accurate (MacFaden et al., 2012; O'Neil-Dunne et al., 2012, 2014). Consequently a rapidly growing body of research combines sub-meter resolution land cover, municipal parcel databases, and socioeconomic and demographic variables to map and measure socio-spatial distributions of urban vegetation at fine scales, often with spatial econometric techniques. Examples include articles from Baltimore, MD (Grove et al., 2006a,b; Troy et al., 2007; Zhou et al., 2009; Boone et al., 2012), Tampa, FL (Landry and Chakraborty 2009), Montreal, ON (Pham et al. 2012a,b), Raleigh, NC (Bigsby et al., 2014), Boston, MA (Duncan et al., 2013; Raciti et al., 2014), Seattle, WA (Romolini et al., 2013), New York City (Grove et al., 2014) and New Haven, CT (Locke and Baine 2014). The Plum Island Long-Term Ecological Research site has made abundant use of its 0.5 m resolution land cover map for addressing similar questions about the abundance and distribution of residential lawns (Giner and Rogan 2012; Giner et al., 2013, 2014; Runfola et al., 2013a,b, 2014; Runfola and Hughes 2014). A goal of this body of research is to better understand the new forest owner (Grove et al., 2013: 377) using a combination of land cover, parcel and/or demographic variables.

Although researchers have recognized the importance of using multi-scale data and both spatial and multi-scale analyses, previous research, such as the work mentioned above, have only partially met this challenge. Although these studies have employed techniques such as spatial autoregressive regression (SAR) and/or GWR (e.g. Landry and Chakraborty 2009; Pham et al., 2012a,b; Troy et al., 2012), these data are frequently aggregated to Census geographies such as the block group, tract level or Canadian equivalent due to the availability of those data on population characteristics. However, such geographic units of analysis that include collections of parcels make it impossible to make direct inferences about other levels of behavior, such as individual parcels/households.

A major advantage of sub-meter land cover maps is the ability to summarize tree canopy and the opportunities for additional tree canopy at the parcel level. This is important because 'parcels are a basic unit of decision-making associated with household and firm locational choices and behaviors' (Pickett et al., 2011: 17). These types of analyses also allow for a distinction between public and privately managed lands. Unfortunately, these advances in remote sensing have not been paralleled with advances in statistical techniques to analyze both parcel- and neighborhood-level sources of spatial heterogeneity. As we have noted, previous attempts to examine variations in residential ownership and management choices, preferences and practices at the parcel scale have conducted their analyses at an approximation of the neighborhood scale. The concern with this approach is that neighborhood-scale analyses fail to take full advantage of the high-resolution land cover, resulting in parcel-scale summaries and conflating the actions of multiple households grouped within neighborhood geography. Figure 1 shows parcel- and block group-scale heterogeneity of tree canopy cover.

MLM is one solution to overcoming these pitfalls. The statistical and theoretical reasons for implementing MLM are manifold. Often data are clustered or comprised of dependent observations, as is the case in the data analyzed in this article (see Moran's I in Table 1). Landscapes are nested spatial hierarchies (Wu and David 2002), and ignoring sources of spatial heterogeneity in statistical analyses with relatively coarse spatial scales have been shown to produce contradictory results in simulation studies (Hamil et al., 2016). Stratified sampling designs can also create or induce a clustering within the data gathered. Clustered or autocorrelated data violate standard statistical formulas, underestimate sampling variance, and can lead to artificially small standard errors (Hox 1998; Snijders and Bosker, 2012). This means results are may be spuriously significant due to the elevated probability of Type-I errors.

A substantive reason for MLM is that aggregating up or averaging out can mask geographic variation (see e.g. Belaire et al., 2016). It is also possible, for example, that the relationship between independent variable X and dependent variable Y are different in different places. This is called spatial non-stationarity (Fotheringham et al., 2003), and OLS regression is incapable of modeling these geographically varying relationships. Worse still, a positive relationship in one area and negative relationship in another might cancel each other out when the overall global fixed effects are estimated. Allowing for the simultaneous estimation of both fixed and random effects is an important contribution of MLM, and therefore one option for handling both spatial autocorrelation and non-stationarity (Subramanian, Duncan and Jones 2001). In this case, random simply means allowed to vary across level two units, which are neighborhoods in this case.

Another reason to use MLM is to simultaneously reduce the risk of ecological and atomistic fallacies. The dual fallacies may mask the heterogeneity among cross-level relationships (Hamil et al., 2015). This is particularly relevant because drivers of residential land management practices occur at several scales, or levels, which include the parcel or household, and neighborhood (Robbins and Sharp 2003; Roy Chowdhury et al., 2011; Cook et al., 2012). Rather than constraining the analyses to only one unit, either the parcel or the neighborhood, MLMs include both simultaneously. Additional levels can also be added when needed. Thus, the variation of vegetation in parcels within and among neighborhoods can be examined with appropriately matched explanatory variables at each scale. We therefore fit MLMs to compare with the more traditional single-level OLS and spatial regressions to demonstrate the statistical and substantive benefits of MLM.

Methods

Site description

The City of Philadelphia was the study area for this research (Fig. 1). As of July 1, 2011, the US Census Bureau estimated that 1,536,471 people lived in Philadelphia, PA, making it the fifth largest city in the country with respect to population (United States Census Bureau 2011). The City of Philadelphia (39.954354, −75.164725) is ∼142 square miles (368 sq km). It is flanked by the Delaware River and the state of New Jersey to the East. The Schuylkill River divides the Center City neighborhood from West Philadelphia. Existing and possible tree canopy within residential parcels ranges from 0 to 100%. When aggregated to the block

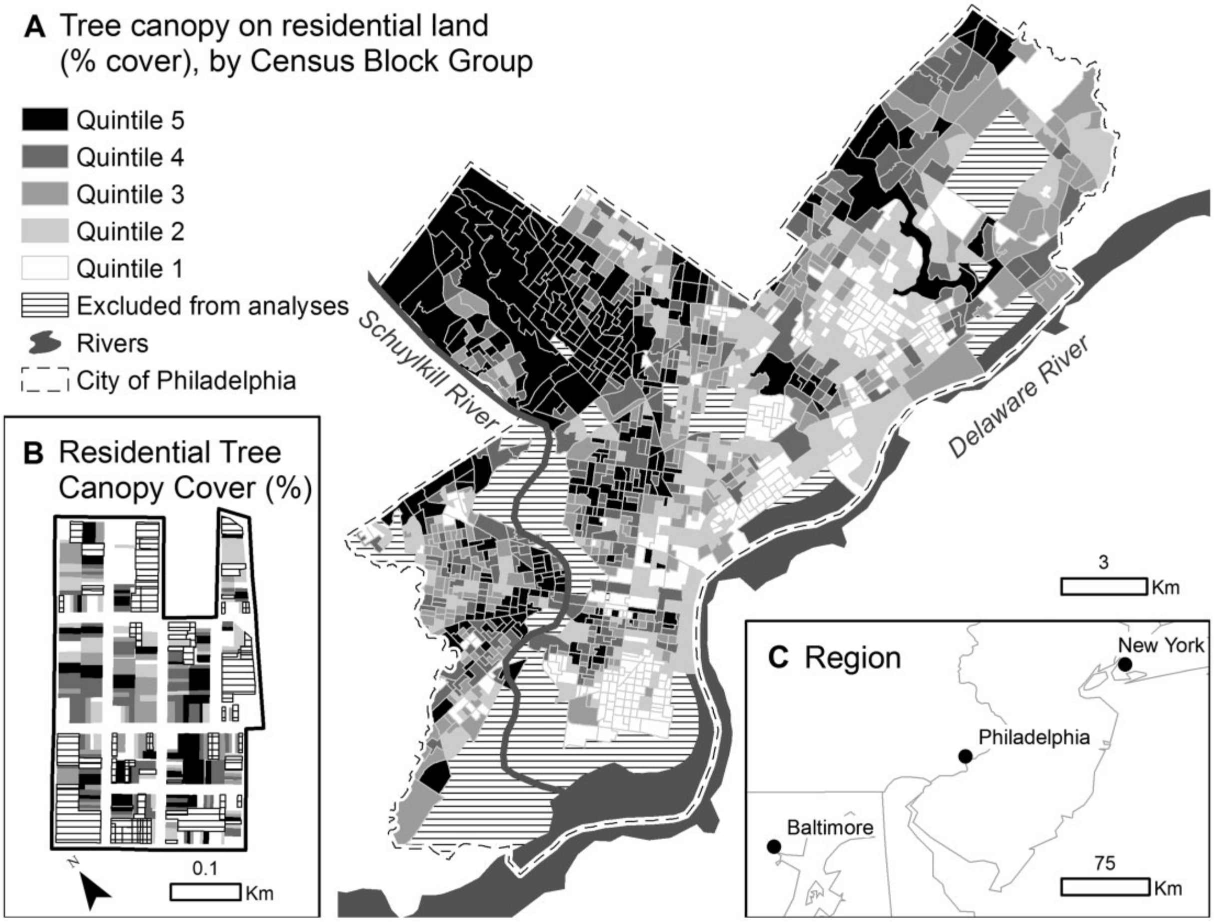

Figure 1. Philadelphia's Census block groups shaded by the percentage of residential parcels covered by tree canopy, grouped by quintile **(A)**. The block group-to-block group variation is comprised of parcel-to-parcel variation in canopy cover, as shown in a sample block group **(B)**. Location of the study area **(C)**.

Table 1. Descriptive statistics for input data.

			Dataset no. 1: Single level ($n = 1,314$)			Dataset no. 2: two-level ($n = 458\,018$)		
Variable set	Variable	Abbreviation	Mean	SD	Moran's I	Mean	SD	Source
Dependent variables	Existing tree canopy, % of land area	UTC_E_P	21.78	12.34	0.68***	18.01	24.63	1, 2
	Possible tree canopy, % of land area	UTC_P_P	76.26	12.27	0.67***	79.77	24.6	1, 2
Population Density	Population density, (pop/mi^2)	PopD	22,876.72	12,850.93	0.50***	22,767.07	11,965	3
	Housing density, (housing units/mi^2)	HseD	10,266.35	7,250.15	0.53***	9,843.41	5,394.6	3
Social stratification (includes Population Density)	Median household income, $s	MHHInc	38,203.07	20,106.44	0.51***	38,885.09	19,594.58	3
	[a](Average) Market Value, year 2014	MktVal14AV	147,008.15	121,244.15	0.84***	133,280.2	117,744.79	4
	Vacant housing units, %	PctVac	0.11	0.07	0.53***	0.1	0.06	3
	[a](Average) building age	AveBldAge	76.20	21.35	0.63***	74.43	22.34	4
	[a](Average) building age^2	AveBldAgeS	6,261.36	4,245.01	0.60***	6,039.13	3,814.99	4
	African American population, %	PctAfam	0.48	0.37	0.92***	0.46	0.38	3
	Detached homes, % area of CBG	PctDH	0.05	0.11	0.52***	0.05	0.11	2
	Row homes, % area of CBG	PctRowHm	0.26	0.19	0.56***	0.28	0.19	2
Lifestyle (includes social stratification and population density)	Average household size	AveHHS	2.54	0.44	0.72***	2.6	0.39	3
	Owner occupied housing units, %	PctOWN	0.50	0.19	0.56***	0.54	0.18	3
	Open space, % area of CBG	PctOpen	0.02	0.07	0.20***	0.02	0.06	5
	Married households, % of families	PctMar	0.27	0.12	0.78***	0.3	0.13	3

[1] Land cover [2] Parcel database [3] 2006–2011 American Community Survey from the US Census Bureau, [4] OPA, [5] Open space shapefile.
[a]Because market value and building age are parcel-level variables they were averaged per block group to make Dataset no. 1, the Census block group file, but were left as provided for Dataset no. 2, hence the parentheses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

group level, residential existing tree canopy ranges from 0 to 67.25%, and possible canopy on residential properties ranges from 30.5 to 97.80%. Homeownership at the block group also varies widely from 0 to 97.08%.

Data and geoprocessing

Two datasets with the same dependent and independent variables (except where noted) were created and analyzed: a single-level Census block group shapefile ($n = 1314$) and a parcel-level shapefile of residential properties ($n = 458\,018$); Table 1, see also Locke (2016) for the data and R script to replicate these analyses). All of the geoprocessing was conducted using ArcGIS v.10.1 (ESRI 2010). Creating these two final datasets required combining five other datasets:

i. A freely available 1 ft^2 resolution land cover raster from 2008 data (Land Cover Philadelphia 2008, 2011) described by (O'Neil-Dunne 2011; O'Neil-Dunne et al., 2012, 2014) contained seven classes including tree canopy, grass/shrub, bare soil, water, buildings, transportation (roads and railroads combined), and other impervious surfaces such as sidewalks and parking lots. 'Overall accuracy of the final land-cover deliverable was 95% [...] Prior to manual corrections, overall accuracy was 94% [...] User's accuracy for the tree canopy class was 97%', both before and after a manual correction process (O'Neil-Dunne et al., 2012: 12)
ii. The city's GIS parcel database, from Philadelphia Water Department was downloaded from via www.opendataphilly.org. The city's building polygons were erased from the parcels so that canopy hanging over buildings would not be included in the analyses. Erasing building footprints maintains comparability with previous research (e.g. Troy et al., 2007) and allows for an examination of the area that is actually manageable: the inverse of the building footprint. The Tabulate Area Tool in ArcGIS 10.1 (ESRI 2012) was used to summarize the land cover dataset's seven classes within each residential parcel without the building footprint area to create two of the dependent variables. The first, existing tree canopy is the percent area of each residential parcel beneath the canopy of a tree. The second, possible tree canopy was calculated as the non-road, non-building, non-water and non-existing tree canopy area, also expressed as a percent of parcel area minus the building footprint. Possible tree canopy has been used by the US Forest Service to estimate land that could possibly support tree canopy (Grove et al., 2006c). The block group level-dataset used the existing and possible tree canopy from only the residential parcels, so public land management activities are not conflated with residential land management activities. The percent of the block group covered with detached homes or with row homes was calculated by querying the parcel database for either detached or row homes, and using Intersect and Dissolve tools in ArcGIS 10.1 (ESRI 2012).
iii. The 5-year 2006–2011 American Community Survey (ACS) data from the US Census Bureau were used to calculate population density (people per mi^2), housing density (housing units per mi^2), percent vacant housing units (vacant lots/total housing units), percent African American population (African American population/total population), percent owner occupied housing units (owner occupied/total housing units), and percent married (married families/total households of all type) for each Census block group. Median household income and average household size were taken directly from the ACS as provided and without modification.
iv. A freely available table provided by the City of Philadelphia's Office of Property Assessment (OPA 2014) contained the assessed market value of each property. Year 2014 assessed market value for each property in dollars was tabularly joined to the residential parcel boundaries from the OPA's table. In 2280 cases (0.5% of total) there were no matches between the database keys in the parcel boundaries, and the containing block group's average values were used instead. Based on visual inspection, these cases were distributed across the study area and not apparently clustered. Assessed market values were averaged per block group for the single-level block group dataset.
v. A table provided by OPA (K. Keene 2014, *Personal Communication*) indicating year built was used to calculate building age as 2008 minus the year built. Only parcels built in year 2007 or earlier were used for comparability with the land cover data, which represents year 2008 conditions (Land Cover Philadelphia 2008, 2011). The file containing year built was tabularly joined to the parcel database; 263 records (0.06%) contained no matches and were set to zero. Almost all of these missing values (245) were in located in the same block group. Since there are 114 block groups, the missing year built data only influenced 0.07% of the block groups, and could not have seriously influenced the results.

This study has two dependent variables: existing tree canopy and possible tree canopy. A multi-model comparison approach is used to test three hypotheses for each dependent variable: population density, social stratification/luxury effect and lifestyle. The analytical approach is described below in detail.

The number of residential parcels per block group ranged from 1 to 975 ($M = 328.5$, $m = 348.6$, SD $= 156.47$). Previous studies exclude block groups with three or fewer parcels because the block group-level estimates are not reliable (Troy et al., 2007; Grove et al., 2014). There were five instances where there was just one residential parcel in a block group and they were retained. Although these singletons do not have any with-in block group variation, they contribute to the level-2 variance in the MLMs, because they contribute to block group to block group variation.

Every attempt was made to ensure that the datasets that represented the same conditions at the time of the land cover (year 2008). Given that the ACS uses a 5-year rolling method, the block group data from 2006 to 2011 is appropriate for 2008 conditions. Unfortunately, the parcel information from the City of Philadelphia, including market value and year built, was not available for 2008. Although the use of year 2014 parcel data is a limitation of our study, it is expected that the relative differences in property values were likely small.

Analytical strategy and statistical model specification

Ordinary least squares regression

Statistical analyses were carried out in four phases. First, for comparison purposes to other similar studies conducted in other areas, a baseline ordinary least squares (OLS) model was fit for each dependent variable and for each theory, using a bi-directional stepwise selection process for parsimony and the block group-level data [Equation (1); see the OLS column in Table 2].

Table 2. Model AIC scores.

DV 1: Existing Tree Canopy	Model types			
		Spatial		
	OLS	SAR	GWR	MLM
Hypotheses 1–3				
Population Density	Model.1	Model.2	Model.3	Model.4
	10,131.21	9,068.368	8,584.335	4,149,251
Socioeconomic status	Model.5	Model.6	Model.7	Model.8
	9,737.777	8,936.927	8,657.545	4,146,155
Lifestyle	Model.9	Model.10	Model.11	Model.12
	9,244.455	8,703.644	8,439.833	4,144,853
DV 2: Possible Tree Canopy				
Hypotheses 4 - 6				
Population Density	Model.13	Model.14	Model.15	Model.16
	10,128.11	9,098.773	8,599.495	4,147,253
Socioeconomic status	Model.17	Model.18	Model.19	Model.20
	9,723.239	8,961.139	8,654.751	4,144,387
Lifestyle	Model.21	Model.22	Model.23	Model 24
	9,207.737	8,702.462	8,453.138	4,142,920

$$y = \beta_0 + \sum_k \beta_k (X - \bar{X}_k) + \epsilon \quad (1)$$

Where y is the dependent variable for each block group, and β_0 the intercept, $\beta_{kj}(X_k - \bar{X}_{jk})$ represents the explanation-grouped set of grand mean-centered independent variables; k begins as 2, 10 or 14 corresponding to the number of population density, social stratification and lifestyle variable sets (see Table 1). Social stratification models include population density variables, and lifestyle models include both social stratification and population density models. Each independent variable is first standardized as grand mean-centered. The use of standardized variables makes interpretation easier because the relative strength of model coefficients can be directly assessed. Evaluating the relative strengths of model coefficients is especially important for the numerous MLM outputs (Aguinis Gottfredson and Culpepper 2013).

Simultaneous autoregressive models

SAR, or spatial regression, was chosen as a method to control for the problems of spatial autocorrelation (i.e. non-random spatial distribution of observation values) common to statistical analysis of geographic data (Anselin and Bera 1998). The SAR models were used because all of the variables exhibited significant spatial autocorrelation (Table 2), and because significant Moran's I values calculated from OLS model residuals' revealed spatial autocorrelation (see results and Tables 3 and 4). To control for spatial autocorrelation, SAR models were estimated for each dependent variable, and for each explanation (see SAR column in Table 3). The decision to estimate the appropriate spatial model, either spatial lag or spatial error, was guided by the Lagrange Multiplier test and the decision tree described by Anselin (2005: 198–200). Spatial lag was the implementation chosen and is shown in Equation (2), where y is the dependent variable (existing or possible tree canopy), X is predicting independent variables grouped by explanation (population density, luxury effect and lifestyle), β is the slope for each grand mean-centered X, ϵ is the error term, u a vector of i.i.d. errors, ρ is the spatial lag coefficient and W is first order queen contiguity spatial weights matrix, which defines polygons sharing edges or vertices as neighbors. The queen contiguity spatial matrix was chosen because many of the block groups form a regular chessboard-like tessellation, and to maintain some comparability with other similar published studies (i.e. Pham et al., 2012b; Grove et al., 2014, among others). For each explanation group, only the variables retained after the bi-directional stepwise selection process described earlier were specified in the spatial models.

$$y = \beta_0 + \rho W y + \sum_k \beta_k (X - \bar{X}_k) + \epsilon \quad (2)$$

Although these models take into account the spatiality of the data, they implicitly assume a single spatial regime or overriding global geographic pattern (ρ). In other words, the spatial models fit global spatial effects but ignore local variation, and do not allow for the possibility for spatial non-stationarity. If non-stationarity is present, SAR models will fail to detect it—by design.

Geographically weighted regression

In the third phase, GWR models were estimated for each dependent variable and for each theory for two principal reasons (see the GWR column in Table 2). The first reason is to understand the realism of the SAR model results. Specifically, detection of non-stationarity using the GWR models would cast doubt on the reliability of the SAR model results. GWR fits a series of roving window regressions across space to reveal local variations—if they are present—using Equation (3), where y_j is the value of the dependent variable in Census block group j described by coordinates (u_j, v_j), β_0 is the local estimated intercept, and β_k represents the slope of the grand mean-centered kth variable at location j (Fotheringham et al., 1998, 2003). The second reason is for demonstration purposes. The MLMs fit in the fourth phase both control for spatial autocorrelation (like the spatial models) but also allow local variation and spatial non-stationarity (like GWR). Although the parameter estimates are not directly comparable, the GWR models were fit as a point of reference when evaluating the realism of spatial lag models and establishing the appropriateness and utility of the MLMs.

Table 3. A comparison of Models 10 and 11 for Exiting Tree Canopy.

Model 10: Spatial Lag (SAR), lifestyle theory for existing tree canopy on private residential properties, per Census block group

Model 11: GWR, lifestyle theory for existing tree canopy on private residential properties, per Census block group

Dimension	Explanatory Variable	Model.10 SAR	Model.11 GWR				
		Global coefficient estimate	Local coefficients			% of significant (95%) blocks	
			Min	Median	Max	% −	% +
Population density	Population density, (pop/mi^2)	−1.17***	−22.45	−1.47	6.65	36.83	0
	Housing density, (housing units/mi^2)						
Socioeconomic Status	Median household income, $s	0.40	−1.86	0.57	7.97	0	17.5
	Market Value, year 2014	1.29***	−13.83	2.32	20.61	1.98	51.29
	Vacant housing units, %	0.87***	−4.66	2.23	5.01	1.6	59.13
	Building age	4.00***	−41.10	5.77	56.77	5.25	52.21
	Building age^2	−3.09***	−86.57	−4.06	97.17	45.36	8.45
	African American population, %	1.57***	−25.30	3.12	17.18	11.26	55.33
	Detached Homes, % area of CBG	3.14***	−6.45	5.68	27.36	0	82.72
	Row Homes, % area of CBG	−0.44	−5.18	−0.64	10.08	12.18	8.3
Lifestyle	Owner Occupied housing units, %	−1.67***	−9.97	−0.83	2.33	18.67	0
	Open Space, % area of CBG	1.06***	−2.14	0.45	3.45	6.01	25.49
	Married Households, % of Families	−0.42	−8.03	−2.52	6.69	39.12	1.37
Model Parameters	Intercept	8.53***	−14.81	23.81	81.49	0	95.13
	ρ	0.60***					
	AIC	8,704		8,440			
	Pseudo-R^2 (lag) or quasi-global R^2 (GWR)	0.717		0.795			
	% of local R^2 > quasi-global R^2			21.23			
	Moran's I of residuals	−0.02					
	Moran's I of residuals of OLS (Model.9)	0.36***					
	N	1,314		1,314			
	Fixed bandwidth size			5,413.16			

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

$$y_j = \beta_0(u_j, v_j) + \sum_k \beta_k(u_j, v_j)(X_j - \bar{X}_k) + \epsilon_j \tag{3}$$

Multi-level models

The MLM phase contains four common steps (Aguinis, Gottfredson and Culpepper 2013). The multi-level dataset, which contains parcel-level and block group-level predictors was used for the MLM analyses. First the null model was fit for each dependent variable. The null model is an unconditional means, one-way random effects analysis of variance, which is used to help evaluate if parcel-level dependent variables within each block group are sufficiently clustered or autocorrelated. Equation (4) shows the first component of the null model

$$y_{ij} = \beta_{0j} + r_{ij} \ (\text{Level 1, null model}) \tag{4}$$

where y_{ij} is parcel i's dependent variable in Census block group j, and β_0 is the mean response across all parcels for block group j, with a parcel-scale residual error term r_{ij}, which is $\sim N(0, \sigma_r^2)$ There are no slopes specified in Equation (4). Instead β_0 is decomposed and estimated as follows:

$$\beta_{0j} = \gamma_{00} + u_{0j} \ (\text{Level 2, null model}) \tag{5}$$

where γ_{00} is the block group-level random effect, with its own residual error term u_{0j}, which is $\sim N(0, \sigma_u^2)$. When Equations (5) is substituted into Equation (4), the resultant null formula is shown in Equation (6):

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \ (\text{combined null model}) \tag{6}$$

As a second MLM step, we fit a random intercept, fixed slope model (RIFSM) for each dependent variable and theory, where Level 1 is Equation (7):

$$y_{ij} = \beta_{0j} + \sum_q \beta_{1jq}(W_{ijq} - \bar{W}_q) + r_{ij} \ (\text{Level 1 RIFSM}) \tag{7}$$

y_{ij} is existing tree canopy (or possible tree canopy) for parcel i in Census block group j, β_{0j} is the intercept for block group j, and the q's are for level-1 variables 1, 2 and 3: Market value in 2014, Building Age and Building Age squared, respectively, denoted with W's. We reserve X's and k's for level two, global variables, to maintain consistency with OLS, SAR and GWR notation. β_{0j} is further decomposed at level-2 into:

$$\beta_{0j} = \gamma_{00} + \sum_k \gamma_{0k}(X_{jk} - \bar{X}_k) + u_{0j} \ (\text{Level 2 RIFSM}) \tag{8}$$

And

$$\beta_{1jq} = \gamma_{q0} \ (\text{Level 2 RIFSM}) \tag{9}$$

Table 4. A comparison of Models 22 and 23 for Possible Tree Canopy

Model 22 Spatial Lag (SAR), lifestyle theory for possible tree canopy on private residential properties, per Census block group

Model 23 GWR, lifestyle theory for possible tree canopy on private residential properties, per Census block group

Dimension	Explanatory variable	Model.22 SAR	Model.23 GWR				
		Global coefficient estimate	Local coefficients			% of significant (95%) blocks	
			Min	Median	Max	% −	% +
Population Density	Population density, (pop/mi^2)	0.81***	−4.95	0.93	23.18	3.50	22.53
	Housing density, (housing units/mi^2)						
Socioeconomic Status	Median household Income, $s	−0.57*	−8.05	−0.62	2.36	19.18	0
	Market value, year 2014	−1.16***	−20.01	−2.01	13.56	46.42	1.98
	Vacant housing units, %	−0.94***	−5.31	−2.36	4.57	63.32	1.75
	Building age	−4.07***	−56.11	−6.66	41.47	53.65	4.87
	Building age^2	3.03***	−131.10	5.11	92.23	8.07	46.04
	African American population, %	−1.45***	−17.17	−3.03	23.53	54.49	10.65
	Detached homes, % area of CBG	−3.14***	−20.37	−5.46	7.35	80.52	0
	Row homes, % area of CBG	0.78*	−9.33	0.98	5.24	0.3	19.03
Lifestyle	Owner occupied housing units, %	1.53***	−1.84	0.68	9.70	0	15.6
	Open space, % area of CBG	−1.31***	−4.05	−0.91	1.01	38.28	0
	Married households, % of Families	0.81	−6.35	2.77	8.89	1.07	43.84
Model Parameters	Intercept	31.93***	−5.87	74.92	115.10	0	98.63
	ρ	0.58***					
	AIC	8,702		8,453			
	Pseudo-R^2 (lag) or quasi-global R^2 (GWR)	0.714		0.791			
	% of local R^2 > quasi-global R^2			21.69			
	Moran's I of residuals	−0.02					
	Moran's I of residuals of OLS (Model.21)	0.35***					
	N	1,314		1,314			
	Fixed bandwidth size			5,411.65			

*p < 0.05; **p < 0.01; ***p < 0.001.

Substituting Equations (7)–(9) leads to the combined equation for the random intercept, fixed slope model:

$$y_{ij} = \gamma_{00} + \sum_q \gamma_{q0}(W_{ijq} - \bar{\mathbf{W}}_q) + \sum_k \gamma_{0k}(X_{ijk} - \bar{\mathbf{X}}_k) + u_{0j} + r_{ij} \quad \text{(combined RIFSM)} \tag{10}$$

In Equation (10) the dependent variable is estimated based on level-1 coefficients γ_{q0}, level-2 coefficients γ_{0k}, intercept γ_{00} with a variance of u_{0j} with a parcel-scale residual error term r_{ij}. Level-2 variables can be considered a cross-level direct effect, or the neighborhood effect operationalized with block groups. The value of k varies to accommodate the number of explanation-grouped level-2 predictors retained after the stepwise OLS process.

The third MLM step involves fitting a random intercept, random slope model (RIRSM). Level one is the same as Equation (7), which is not repeated here for brevity. Likewise, the first part of the level two equation is the same as Equation (8) and not shown. However, the second part is Equation (11):

$$\beta_{1jq} = \gamma_{q0} + u_{1j} \quad \text{(Level 2 RIRSM)} \tag{11}$$

The difference between Equations (9) and (11), is the u_{1j} term, which describes how much block group slopes differ from the pooled slope derived from all block groups. Substituting Equations (7), (8) and (11) creates Equation (12)—the combined random intercept, random slope two-level model.

$$y_{ij} = \gamma_{00} + \sum_k \gamma_{0k}(X_{jk} - \bar{\mathbf{X}}_k) + \sum_q \gamma_{q0}(W_{ijq} - \bar{\mathbf{W}}_q) + u_{0j} + \sum_k u_{1j}(W_{jk} - \bar{\mathbf{W}}_k) + r_{ij} \quad \text{(combined RIRSM)} \tag{12}$$

The four MLM step includes testing a cross-level interaction model (x-level interaction model), which is the most complex two-level mixed effects regression model. The Level 1 component of the cross-level interaction is the same Equation (7). Allowing intercepts and slopes to vary across level two units requires the following equations:

$$\beta_{0j} = \gamma_{00} + \sum_k \gamma_{0k}(X_{jk} - \bar{\mathbf{X}}_k) + u_{0j} \quad \text{(Level 2 x – level interaction model)} \tag{13}$$

$$\gamma_{10} + \sum_k \gamma_{k0}(X_{jk} - \bar{\mathbf{X}}_k) + u_{1j} \quad \text{(Level 2 x – level interaction model)} \tag{14}$$

$$y_{ij} = \gamma_{00} + \sum_k \gamma_{0k}(X_{jk} - \bar{\mathbf{X}}_k) + \sum_q \gamma_{q0}(W_{ijq} - \bar{\mathbf{W}}_q) + \sum_q \sum_k \gamma_{qk}(W_{ijq} - \bar{\mathbf{W}}_q)(X_{jk} - \bar{\mathbf{X}}_k) + u_{0j} + \sum_k u_{1j}(W_{jk} - \bar{\mathbf{W}}_k) + r_{ij} \quad \text{(combined x – level interaction model)} \tag{15}$$

All of the statistics were performed using R version 3.2.2 (14 August 2015)—'Fire Safety' (R Core Team 2015a). The large number of R packages used for the analyses reflect the numerous contributors to the open-source R statistics platform. The foreign package (R Core Team 2015b) was used to read the *.dbf portion of the shapefiles into R for table-only analyses. Variance inflation factors (VIFs) were calculated using the car package's vif() function (Fox and Weisberg 2011). Correlations were performed with the Hmisc package function rcorr() (Harrell and Charles 2015), while spatial models were fit with lagsarlm() in the spdep package (Bivand, Hauke and Kossowski 2013; Bivand and Piras 2015). Moran's I for each Level-1 variable was tested using moran.test() function, and OLS residuals' spatial autocorrelation were examined using the lm.morantest() function. The gwr.sel() and gwr() in the spgwr package (Bivand and Yu 2015) were used to fit the GWR models. MLMs were estimated with lmer() function in the lme4 package version 1.1-9 (Bates et al., 2013, 2014). Finally, model selection was aided with the MuMIn package's model.sel() function (Bartoń 2015), and the fixed-effect's confidence intervals were calculated and conveniently formatted with sjPlot's sjt.lmer() function, using a Wald-like test.

Results

This study used a multimodal inferential approach to evaluate three explanations of the spatial distribution of existing and possible tree canopy: population density, social stratification luxury effect and lifestyle characteristics. Based on an Akaike information criterion (AIC) minimization criterion, the statistical models that included lifestyle variables outperformed the social stratification and population density models (Table 2). Lower AIC values in models with lifestyle variables indicate that this finding was consistent across all model specification types (OLS, SAR, GWR and MLM) and for both dependent variables: existing tree canopy and possible tree canopy.

The models were robust. After the stepwise process, only housing density was dropped as an independent variable in the OLS lifestyle models (Models 9 and 21), indicating some variable importance for the remaining parameters. Multicollinearity was not an issue. VIFs were below 6 in Models 9 and 21, except for average building age and building age squared which are correlated by construction (VIF $= 9.56$, $r = 0.92$, $P < 0.0001$). VIF's below 10 are considered acceptable (O'Brien 2007). The pseudo-R^2 values were > 0.7 for both of the lifestyle SAR models and their residuals were not spatially autocorrelated, even though their OLS counterparts exhibited high and significant residual autocorrelation (Tables 4 and 5). The Moran's I values calculated from the residuals of all of the OLS models ranged from 0.35 to 0.62, and were all highly significant ($P < 0.001$), thus indicating that all OLS models suffered from spatial autocorrelation. The Moran's I values of the corresponding SAR models were much lower and not significant (Tables 4 and 5). These results indicate that the SAR models minimized the effects of spatial autocorrelation and were therefore more appropriate for the dataset used in this study. Therefore, only the SAR models are included in the results since they outperformed their OLS counterparts. The OLS outputs are not shown. Nearly all of the beta coefficients in the spatial lag models (Model 10 and 22) were significant after the stepwise process (Tables 4 and 5). The spatial lag term ρ (Rho) was significant and lower than most explanatory variables in absolute value, suggesting a low probability of omitted variable bias.

Population density

The population density models had the highest AIC within specification type, and for both dependent variables, except for the GWR models (Table 2), signaling the relatively poor fit of population density-only models. Population density GWR models (Models 3 and 15) had lower AICs than their socioeconomic status versions (Models 7 and 19), and the lifestyle GWR models (Models 11 and 23) had even lower AICs. The population density coefficients' absolute values were moderate to modest in the final lifestyle SAR models (Tables 4 and 5), which indicates that population density does explain some of the variation when controlling for other factors such as indicators of social stratification and lifestyle. The signs were as hypothesized, negative for existing tree canopy and positive for possible tree canopy.

The median GWR coefficients for population density in the lifestyle models are on the same order of magnitude and direction as their SAR counterparts. However the GWR estimates for population density show areas of non-stationarity for existing tree canopy. Population density was significantly ($P < 0.05$) and negatively associated with existing tree canopy in 36.83% of local models, and was not a positive predictor in any models at the same significance level (Table 3). Population density had a mixed association with possible tree canopy (statistically significantly negative in $\sim$4% block groups, and positive in 23% of block groups; Table 4). The global SAR models masked the local and largely insignificant relationships between the dependent variables and population density. The local GWR models indicate that population density was strongly associated with existing and possible tree canopy in some areas of the city, but not all areas.

The population density parameter estimates in the MLM models had the same magnitude and direction at the block group level as the SAR and median GWR estimates. In other words, population density alone was negatively associated with existing tree canopy and positively associated with possible tree canopy. Interestingly, population density, which has a negative coefficient for existing tree canopy ($\gamma_{01} = -1.19$), when interacting with market value ($\gamma_{10} = 1.96$), yields a positive $\gamma_{11} = 0.41$. This suggests that high value properties in high-density areas have more tree canopy. The opposite is true for possible tree canopy (Model 24: MktVal14 * PopD (γ_{11}) $= -0.38$). The interaction of population density and building age show results that are not at all surprising; slightly older parcels have more tree canopy even in block groups with higher population density (Model 12: BldAge * PopD (γ_{21}) $= 0.13$). These important interactions would have gone undetected if only global models were used for the analysis.

Social stratification

The models with variables for socioeconomic status fit the data better than their population density counterparts, even when accounting for the additional complexity (Table 2). Within the SAR lifestyle models, the socioeconomic variables had some of the largest absolute values for their coefficients. For existing tree canopy, building age and its squared term had positive and negative coefficients, respectively, signaling an inverted 'U'-shaped relationship of tree cover to time since year built. A similar growth-decline relationship has also found in previous studies conducted in other cities, such as Baltimore and New York City (e.g. Troy et al., 2007; Grove et al., 2014).

The median GWR coefficients for socioeconomic status indicators in the lifestyle models (i.e. Models 11 and 23) also had the same direction and magnitude as their SAR equivalents. However, statistically significant positive and negative relationships were found for each socioeconomic status variable except for median household income and the percentage of the block

Table 5. MLM results for existing and possible tree canopy

	Model. 12			Model. 24		
Level 1, property parcels ($n = 458\,018$)	B	CI	P	B	CI	P
(Intercept) (γ_{00})	16.59	16.24 to 16.94	**<0.001**	80.91	80.58 to 81.24	**<0.001**
MktVal14 (γ_{10})	1.96	1.78 to 2.15	**<0.001**	−1.91	−2.09 to −1.72	**<0.001**
BldAge (γ_{20})	2.92	2.47 to 3.36	**<0.001**	−2.52	−2.96 to −2.07	**<0.001**
BldAgeSq (γ_{30})	−1.62	−2.16 to − 1.08	**<0.001**	1.16	0.63 to 1.70	**<0.001**
Level 2, census block groups ($n = 1314$)						
PopD (γ_{01})	−1.19	−1.80 to − 0.59	**<0.001**	0.87	0.29 to 1.46	**0.004**
MHHInc (γ_{02})	0.87	0.32 to 1.42	**0.002**	−1.2	−1.68 to −0.72	**<0.001**
PctVac (γ_{03})	1.67	1.09 to 2.26	**<0.001**	−1.93	−2.48 to −1.37	**<0.001**
PctAfam (γ_{04})	3.78	3.27 to 4.29	**<0.001**	−3.28	−3.74 to −2.81	**<0.001**
PctDH (γ_{05})	4.63	4.07 to 5.19	**<0.001**	−3.95	−4.44 to −3.46	**<0.001**
PctRowHm (γ_{06})	−1.31	−2.05 to − 0.57	**0.001**	1.3	0.64 to 1.96	**<0.001**
PctOWN (γ_{07})	−2.31	−3.03 to − 1.59	**<0.001**	2.7	2.01 to 3.39	**<0.001**
PctOpen (γ_{08})	0.54	−0.16 to 1.23	0.134	−0.89	−1.36 to −0.42	**<0.001**
PctMar (γ_{09})	−1.88	−2.70 to − 1.05	**<0.001**	1.76	0.95 to 2.57	**<0.001**
Cross-level interactions						
MktVal14 * PopD (γ_{11})	0.41	0.23 to 0.59	**<0.001**	−0.38	−0.56 to −0.20	**<0.001**
MktVal14 * MHHInc (γ_{12})	−0.4	−0.53 to − 0.27	**<0.001**	0.5	0.37 to 0.62	**<0.001**
MktVal14 * PctVac (γ_{13})	−1.78	−1.97 to − 1.60	**<0.001**	1.64	1.45 to 1.82	**<0.001**
MktVal14 * PctAfam (γ_{14})	0.17	−0.10 to 0.43	0.213	0.08	−0.18 to 0.34	0.525
MktVal14 * PctDH (γ_{15})	−0.01	−0.14 to 0.13	0.917	−0.05	−0.18 to 0.08	0.464
MktVal14 * PctRowHm (γ_{16})	−1.63	−1.89 to − 1.37	**<0.001**	1.5	1.24 to 1.76	**<0.001**
MktVal14 * PctOWN (γ_{17})	1.84	1.60 to 2.08	**<0.001**	−1.85	−2.09 to −1.62	**<0.001**
MktVal14 * PctOpen (γ_{18})	−0.29	−0.39 to − 0.19	**<0.001**	0.27	0.17 to 0.37	**<0.001**
MktVal14 * PctMar (γ_{19})	−0.18	−0.51 to 0.15	0.287	0.18	−0.15 to 0.51	0.276
BldAge * PopD (γ_{21})	0.13	−0.27 to 0.54	0.521	−0.13	−0.52 to 0.27	0.539
BldAge * MHHInc (γ_{22})	−0.87	−1.14 to − 0.61	**<0.001**	0.86	0.60 to 1.13	**<0.001**
BldAge * PctVac (γ_{23})	0.44	0.06 to 0.82	**0.025**	−0.42	−0.79 to −0.04	**0.031**
BldAge * PctAfam (γ_{24})	0.51	0.03 to 0.99	**0.038**	−0.5	−0.98 to −0.02	**0.039**
BldAge * PctDH (γ_{25})	0.88	0.58 to 1.18	**<0.001**	−0.74	−1.04 to −0.44	**<0.001**
BldAge * PctRowHm (γ_{26})	−1.55	−2.08 to − 1.03	**<0.001**	1.72	1.19 to 2.24	**<0.001**
BldAge * PctOWN (γ_{27})	−1.81	−2.33 to − 1.30	**<0.001**	1.75	1.24 to 2.26	**<0.001**
BldAge * PctOpen (γ_{28})	0.02	−0.23 to 0.26	0.904	0.17	−0.07 to 0.41	0.17
BldAge * PctMar (γ_{29})	1.79	1.19 to 2.38	**<0.001**	−1.64	−2.23 to −1.04	**<0.001**
BldAgeSq * PopD (γ_{31})	−0.13	−0.47 to 0.22	0.469	0	−0.34 to 0.34	0.994
BldAgeSq * MHHInc (γ_{32})	0.78	0.54 to 1.03	**<0.001**	−0.72	−0.97 to −0.48	**<0.001**
BldAgeSq * PctVac (γ_{33})	−0.16	−0.57 to 0.25	0.441	0.14	−0.26 to 0.54	0.484
BldAgeSq * PctAfam (γ_{34})	−0.15	−0.69 to 0.39	0.586	0.05	−0.48 to 0.59	0.841
BldAgeSq * PctDH (γ_{35})	−0.57	−0.88 to 0.25	**<0.001**	0.46	0.14 to 0.77	**0.005**
BldAgeSq * PctRowHm (γ_{36})	0.79	0.29 to 1.28	**0.002**	−0.89	−1.38 to −0.40	**<0.001**
BldAgeSq * PctOWN (γ_{37})	1.6	1.10 to 2.10	**<0.001**	−1.56	−2.05 to −1.06	**<0.001**
BldAgeSq * PctOpen (γ_{38})	0.08	−0.20 to 0.36	0.576	−0.23	−0.49 to 0.04	0.098
BldAgeSq * PctMar (γ_{39})	−1.91	−2.51 to −1.30	**<0.001**	1.63	1.03 to 2.23	**<0.001**
Variance components						
Within−block group (L1) variance (σ^2)	493.522			491.617		
Intercept (L2) variance (τ_{00})	0.000			17.605		
Slope (L2) variance PopD (τ_{11})	5.483			4.898		
Slope (L2) variance MHHInc (τ_{12})	18.097			9.265		
Slope (L2) variance PctVac (τ_{13})	12.178			1.653		
Slope (L2) variance PctAfam (τ_{14})	12.000			1.800		
Slope (L2) variance PctDH (τ_{15})	5.556			2.878		
Slope (L2) variance PctRowHm (τ_{16})	13.922			10.311		
Slope (L2) variance PctOWN (τ_{17})	18.357			12.996		
Slope (L2) variance PctOpen (τ_{18})	5.611			2.529		
Slope (L2) variance PctMar (τ_{19})	21.851			12.465		
Additional Information						
ICC for null model/full model	0.20/	0.0		0.20/	0.035	
−2 log likelihood (FIML)		4,144,669			4,142,736	
AIC		4,144,853			4,142,920	

P-values < 0.05 are bolded.

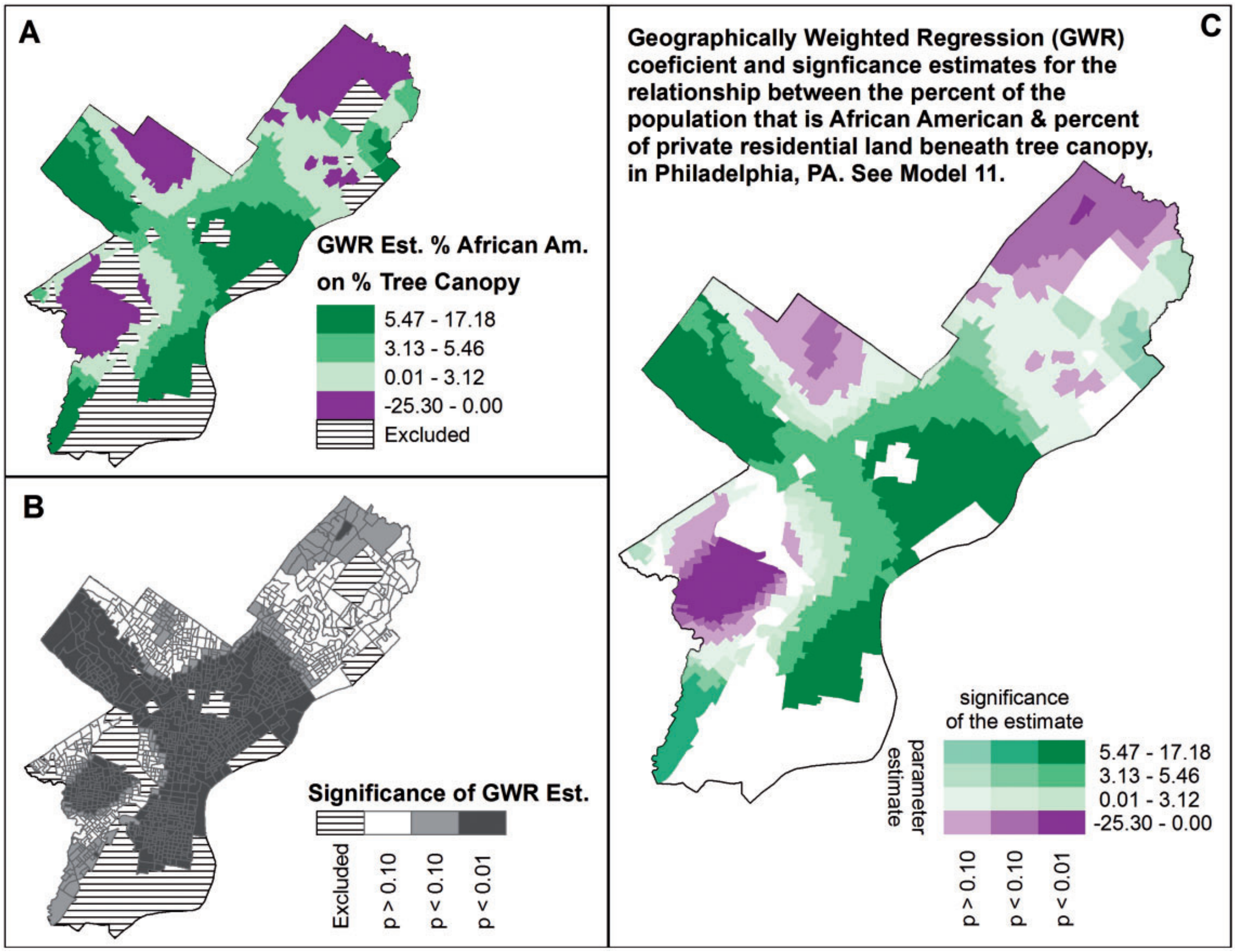

Figure 2. GWR provides the direction and magnitude of the relationship between tree canopy and the percent African American population **(A)**, when controlling for other factors (see Model 11 and Table 3). GWR also provides a significance estimate **(B)**. Following Mennis (2006) we combined this information into one map of sign, magnitude, and significance **(C)**.

group covered by detached homes (Tables 4 and 5). The substantial spatial non-stationarity detected by GWR for all other independent variables reveals the potential bias of the global (SAR) estimate toward the null. In other words, positive associations in one region likely canceled out negative associations in other areas. For example, although there was a significant positive association between median household income and existing tree canopy in 17.5% of local GWR models (Table 3), the global relationship was not significant. The association between percent African American population and existing tree canopy shows that the median GWR estimate is the same direction and order of magnitude as the global estimate, but the range of local GWR beta estimates spanned from -25.30 to 17.18 and was significant ($P < 0.05$) in ~66% of the cases. The percent African American population is significantly positively and negatively associated with existing tree canopy depending on location (Fig. 2). A visual comparison with the Figure 1 shows that both positive and negative associations can be found in areas with the highest percentage of existing tree canopy.

Parcel market value and percentage detached homes, indicators of social stratification, were largely positively associated with existing tree canopy and negatively associated with possible tree canopy. In fact, as previously shown by the results of the market value and population density interaction terms, the negative association between population density and existing tree canopy was mitigated by market value. Percentage of vacant housing units, an indicator of lower socioeconomic status, was positively associated with existing tree canopy and negatively associated with possible tree canopy. This finding was shown in the global SAR, local GWR and MLM results. However, the interaction of market value and percent vacant housing in the MLM model indicates a negative association with existing tree canopy ($\gamma_{13} = -1.78$, 95% CI [−1.97, −1.60]) and positive association with possible tree canopy ($\gamma_{13} = 1.64$, 95% CI [1.45–1.82]). Differences in results between global SAR models and GWR and MLM models show more complex relationships between tree canopy and building age, percent African American population and percent row homes. The use of local specifications (GWR), and cross-level (MLM) models uncovered additional complexity that the global models (OLS, SAR) were unable to detect.

Lifestyle

The OLS, SAR, GWR and MLM models for lifestyle always had the lowest AICs within each specification type even though they had more parameters (Table 2). The SAR lifestyle models for existing tree canopy revealed a significant and positive association with open space (Table 3). Recall that the existing tree canopy dependent variable is derived from private residential lands only. These results suggest that block groups with more open space tend to have well canopied residential lands, while the opposite was true for possible tree canopy. The local GWR estimates confirm the negative association between open space and possible tree canopy (i.e., significantly negative in 38% of blocks; Table 4). Local GWR models show a more complex relationship in that the open space-existing tree canopy relationship was significantly negative in 6.01% of the models and positively associated in 26% of the cases. The MLM model shows that these relationships may be mitigated, in part, by market value. Table 5 shows that the interaction of open space and market value shows a negative association with existing tree canopy ($\gamma_{18} = -0.29$, 95% CI [−0.39 to − 0.19]) and a positive association with possible tree canopy ($\gamma_{18} = 0.27$, 95% CI [0.17–0.37]). When compared with the global SAR model, the local GWR model and the cross-level MLM exposes additional complexity to show that there is not always a positive relationship between open space and private residential canopy.

The results related to another lifestyle variable, percent of households who are married, indicate no significant association with either existing or possible tree canopy in the global SAR models (Tables 4 and 5). The GWR estimates show a significantly negative association with existing tree canopy in about 40% of local models, and a significantly positive association with possible tree canopy in about 44% of models. The MLM model results show similar significant associations between tree canopy and percent married (Table 5).

All three model specification types (global SAR, GWR and level-2 MLM coefficients) indicate that percent owner occupied housing units variable was negatively associated with existing tree canopy and positively associated with possible tree canopy. The cross-level interaction of percent owner occupied housing and market value from the MLM model show the opposite association with tree canopy; a positive association with existing tree canopy ($\gamma_{17} = 1.84$, 95% CI [1.60, 2.08]) and a negative association with possible tree canopy ($\gamma_{17} = -1.85$, 95% CI [−2.09, −1.62])

Because the mean-centered variables allow comparison of the relative importance of predictors, the high absolute values for the lifestyle variables—% owner occupied and % married—further signal their importance. Cross-level interactions are particularly illuminating; results indicate that the relationship between percent owner occupied housing or percentage vacant housing and tree canopy is linked to the market value of homes in an area. Although the level 1 coefficients revealed a positive association between market value and existing tree canopy, the row home—market value interaction ($\gamma_{16} = -1.63$, 95% CI [−1.89, −1.37]) suggests that row homes in high market value areas are negatively associated with tree canopy. Similarly, although building age has a large and positive level-1 association ($\gamma_{20} = 2.92$, 95% CI [2.47, 3.36]) with existing canopy, the negative coefficient for the row home—building age interaction ($\gamma_{26} = -1.55$, 95% CI [−2.08, −1.03]), and building age—owner occupied homes ($\gamma_{27} = -1.81$, 95% CI [−2.33, −1.30]) illustrates that the building age relationship is not the same for all building types or ownership regimes. The global Model 10 had a small and insignificant coefficient for row homes (and existing tree canopy), but the local (GWR) and cross-level (MLM; Model 12) suggests why: there are many local variations and cross-level interactions with market value, age and home ownership that are masked with global estimates. Importantly for the underlying explanations, these variables relate to particular lifestyles.

Discussion

We examined the distribution of existing and possible tree canopy on private residential properties in Philadelphia, PA with respect to three complementary explanations of the distribution of urban vegetation: (i) population density, (ii) the luxury effect variation of social stratification theory and (iii) the ecology of prestige lifestyle-based theory. The multimodel inferential approach used spatial econometric, GWR and MLM to evaluate these three social explanations. Results indicated that the models which included lifestyle variables outperformed all other models, and therefore provide additional support for the ecology of prestige theory. This study adds to the growing body of work that highlights the association between the distribution of vegetation on private residential lands and the lifestyle characteristics and lifestages of neighborhood residents (Grove et al., 2006, 2014; Bigsby et al., 2014; Locke and Grove 2016).

The substantial spatial non-stationarity exposed by the GWR and the cross-level MLMs is important to consider. Although we were not surprised by the non-stationarity in Philadelphia given previous studies (Mennis 2006; Heckert and Mennis 2012; Pearsall and Christman 2012; Foster and Dunham 2014), our local and cross-level interaction models cast some doubt on findings from previous studies using global OLS and/or SAR models. The generalizability of a beta coefficient's sign and direction from one study area to another would seem very tenuous and limited if the local sign and direction are opposite in different areas of the 'same' study area.

The local GWR and MLM analyses may shed some light on the reasons for the mixed results of previous research that examined the relationship between trees and minority populations. Global models used in previous work found that African American populations were negatively associated with tree canopy in some cities (Heynen et al., 2006; Landry and Chakraborty 2009) and positively associated with other cities (Troy et al., 2007; Grove et al., 2014). Our models showed a non-stationary relationship that ranged from negative to positive, even within the same city. Regions with significant relationships (positive or negative) could be used to inform and target a stratified sampling design, with more intensive quantitative and qualitative data collection and analyses to complement the extensive analyses presented in this article (see Grove et al., 2013, 2015). Furthermore, the MLM interaction terms with percent African American indicate a positive association with building age which may support the importance of neighborhood demographic shifts that previous research in Baltimore has suggested (Merse et al. 2009; Boone et al., 2010).

Findings from previous research largely suggested people displace trees (e.g. Troy et al., 2007; Landry and Pu 2010; Pham et al., 2012b). But in our study GWR outputs raise concern about the reliability of the global OLS and SAR models, by revealing that population density was only significantly associated with existing tree canopy across approximately one third (36.83%) of the city's block groups. These GWR results also suggest that population density alone as a proxy for urban development is insufficient to explain the heterogeneity in the spatial distribution of existing or possible tree canopy. Density alone may mask spatial configuration and spatial configuration may have a significant effect on existing and possible canopy cover. Bigsby et al. (2014) used more sophisticated measures of urban form such as percent pervious area, housing density, population density, road density, parcel size, distance to downtown and others, and found that these were important predictors of tree canopy in Raleigh, NC. Another study found that a terrain index calculated from a digital elevation model explained 59% of the variance in tree canopy cover in Cleveland, OH (Berland et al., 2015). The inclusion of percent row homes as a variable in our study, another indicator of urban form, showed only a weak association with tree canopy. Moreover, the relationship with possible tree canopy was both positive and negative (Table 4), which implies that there are opportunities for tree planting on some residential properties even in densely populated neighborhoods. The GWR results revealed spatial heterogeneity that is generally concealed with global models, and point to the importance of other explanations such as urban form and/or topography. Future research should consider incorporating measures of urban form and topography highlighted by these authors, including terrain, percent pervious area, housing density, population density, road density, parcel size, distance to downtown and others, as well as add to cross-site comparative research.

Cross-level interaction terms in the MLMs further exposed important sources of variation to consider for future research. For example, population density at the block group level was negatively associated with tree canopy. However, population density at the block group level interacts with market value at the parcel level to produce a significant coefficient with the opposite sign (Table 5). The MLM model therefore suggests that in densely populated neighborhoods with high market value, there is also relatively high tree canopy coverage. The empirical results provided by local (GWR) and cross-level (MLM) techniques presented here offer a basis for additional theorization as to the explanations for tree canopy and the opportunities for additional planting in Philadelphia and elsewhere.

Although this article focused on the social correlates, ecological outcomes such as tree canopy can also have ecological drivers. A plot-based study of 12 North American cities found that only about one in three trees were planted. Therefore two-thirds of the existing urban forest was from natural regeneration in the studied cities. But on a city-by-city basis, the percentage varied almost as much as is hypothetically possible, from 7.3 to 89% (Nowak 2012). Several studies have found that vacant lots can be sites for volunteer trees and other vegetation (e.g. Heynen et al 2006; McPhearson et al., 2013). The positive relationship between existing tree canopy and percent vacant lots identified by our study suggests that vacant lots in Philadelphia may also be a site of volunteer trees. However, the interaction of market value and percent vacant may point to a more complex relationship. Operating within Philadelphia, the LandCare Program specifically targets vacant lots with the goal to transform these lots into neighborhood amenities (Heckert and Mennis 2012). Cleaning and greening interventions comprised of trash removal, tree planting and grass mowing reduced crime around these properties (Branas et al., 2011). The extent to which these aesthetic and functional improvements may have included the removal of unsightly volunteer trees is unclear. According to their website, the LandCare program has treated over 7000 parcels (phsonline.org/programs/landcare-program). The extent to which ecological processes of natural recruitment drive tree canopy, and whether management programs such as LandCare influence these processes remains unknown. Nevertheless, more attention is needed on the ecological—and even physical as noted above—as well as social aspects of the urban forest, to more fully understand the distribution of urban vegetation.

The point about management is important, because the actors who actually do the work may be associated with different motivations, capacities and interests. Members of different lifestyle groups may have different ideas about what is ideal. All model specification types signaled the importance of variables reflective of lifestyles. Condos and other high-density and rental properties may have hired grounds keepers and maintenance staff—such living arrangements are included in the concept of lifestyle. The positive association between renters and vegetation found in cities such as Montreal (Pham et al., 2012a), suggest that this lifestyle choice can benefit from the management actions of others (i.e. landlords). In Philadelphia, neighborhoods with fewer owner occupied housing units (i.e. more renters) seem to have a greater amount of tree canopy, except in areas with a higher parcel market value where this relationship is reversed. It is possible that these findings indicate difference in vegetation management between landlords and owner-occupied households associated with areas with varying real estate value. Before aggregating parcel-scale land cover summaries into block group boundaries, analysts should consider multi-scalar statistical analyses to avoid missing these types of significant interactions. Unfortunately the parcel market value data came from year 2014, which was an unavoidable limitation of this study.

The relatively recent advent of high-resolution, high-accuracy land cover data (often created with LiDAR data-fusion and OBIA; O'Neil-Dunne et al., 2012, 2014) have allowed analyses of the ecological structure within land managers' property parcels. Although the parcel is an important policy-relevant unit of analysis (Stone 2004; Landry and Pu 2010), most analyses of high-resolution land cover use coarser units of analyses such as census block groups (e.g. Pham et al., 2012a,b; Giner et al., 2013; Grove et al., 2014). Thus, the full benefit of high-quality land cover data have yet to be realized for research applications. Moreover, there are both parcel-scale (i.e. household) drivers of land management, and larger block group (i.e. neighborhood) influences on the urban forest (Roy Chowdhury et al., 2011; Cook et al., 2012). Therefore, methods that simultaneously examine multiple scales and units of analysis (Hamil et al., 2016)—and their interactions—are needed to more fully understand cities as complex systems. For example, we found several instances (e.g. percent vacant and percent open space) where block group level associations were mitigated—or even reversed—when parcel-level characteristics such as market value are considered. In this article, we have shown how MLMs are ideally suited to linking parcel- and neighborhood-level sources of spatial heterogeneity in a human-dominated ecosystem.

Conclusion

This article used several spatial analytical techniques to examine mid-level explanations of the heterogeneity in existing and possible tree canopy: SAR, GWR and MLM. It is important to ask whether or not the more sophisticated statistical analyses led to insights beyond that which could have been obtained with a simpler approach. We believe the greater sophistication was valuable for several reasons. First, we found several examples where the local GWR and MLM uncovered interesting patterns that were undetected by global SAR alone (e.g. population density). The sources of this heterogeneity may be explained in part with improved measures of urban form, topography and land management behaviors associated with different ownership regimes and lifestyles. Second, even when significant relationships were detected using the global SAR, the local GWR uncovered much more complex and even opposite relationships between percent African American population and existing tree canopy (Fig. 2). Future research should investigate the histories and other factors that may have lead to these varied relationships. Finally, the MLM allowed us to uncover interactions between parcel-level market value and block group-level population density, open space and vacancy rates that were dissimilar to the single-level, global analyses of block groups. Vacant lots in high-value areas have less canopy even though block groups with more vacant parcels tended to have more tree canopy. Therefore, block group-level findings from previous studies could be misleading. The methods presented here and the empirical findings they provide may inform advances in our theories describing the socio-spatial distribution of urban tree canopy at multiple scales.

This article adds to the growing body of work showing that the uneven distribution of urban vegetation is not merely related to differences in neighborhood population density or social stratification. These results from Philadelphia reinforce the important association between vegetation on private residential lands and the lifestyle characteristics and lifestages of neighborhood residents. The policy implications of these findings are important when considering that tree planting or canopy goals are being adopted by a growing number of cities. Private residential lands represent a large proportion of the overall land area in many cities like Philadelphia. The stewardship of trees, such as tree planting and management on private residential lands, involves the efforts of actors with different motivations, capacities and interests. Cities that want to achieve their tree canopy goals need to adopt planting and management policies that recognize that urban residents from different lifestyle groups may have different ideas about trees and stewardship.

Data sharing

See Locke (2016) for free access to the data. The file set contains the single level Census Block Group shapefile (1.3 MB), the two-level *.dbf file (~145 MB), the R script (as an *.rtf) used for the article, and the resulting zipped *.R file (1.03 GB zipped, 1.15 GB) created by that script. The two-level parcel polygons are not provided to keep the file size manageable, but are available upon request.

Acknowledgements

The authors declare no conflicts of interest. We thank the USDA Forest Service Northern Research Station, Lynne Westphal, Sarah Low and the Philadelphia Urban Field Station for the Sustainable Science Fellowship. Thank you Dr Hamil Pearsall for providing space in the Spatial Analytics Lab at Temple University (SAL@T) in the summer of 2014, which improved this research. Thanks to Kevin M. Keene for providing the year built data. Thank you Colin Polsky, Director of the Center for Environmental Studies at Florida Atlantic University for introducing some of the authors to MLM and providing training. Our greatest thanks extend to those who plant and steward trees throughout the City of Philadelphia's urban forest, their communities are better because of them.

Funding

The Edna Bailey Sussman Fund, the Libby Fund Enhancement Award and the Pruser Fellowship from Clark University provided funding for this research.

References

Aguinis, H., Gottfredson, R. K., and Culpepper, S. A. (2013) 'Best-Practice Recommendations for Estimating Cross-Level Interaction Effects Using Multilevel Modeling', *Journal of Management*, **39**: 1490–528.

Anselin, L., and Bera, A. K. (1998) 'Spatial dependence in linear regression models with an introduction to spatial econometrics', *Statistics Textbooks and Monographs*, **155**: 237–90.

——, and —— (2005) Exploring spatial data with GeoDaTM: A workbook. Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics. University of Illinois. p. 244 <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>.

Bartoń, K. (2015) MuMIn: multi-model inference. R package version 1.6. 6. <http://CRAN.R-project.org/package=MuMIn>.

Bates, D. et al. (2013) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9 <https://CRAN.R-project.org/package=lme4>.

—— et al. (2014) Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823 <http://arxiv.org/abs/1406.5823>.

Belaire, J. A., Westphal, L. M., and Minor, E. S. (2015) 'Different Social Drivers, Including Perceptions of Urban Wildlife, Explain the Ecological Resources in Residential Landscapes', *Landscape Ecology*, **31**: 401–13.

Berland, A. et al. (2015) 'How Environmental Justice Patterns are Shaped by Place: Terrain and Tree Canopy in Cincinnati, Ohio, USA', *Cities and the Environment (CATE)*, **8**: 1.

Bigsby, K. M., McHale, M. R., and Hess, G. R. (2014) 'Urban Morphology Drives the Homogenization of Tree Cover in Baltimore, MD, and Raleigh, NC', *Ecosystems*, **17**: 212–27.

Bivand, R. S., Hauke, J., and Kossowski, T. (2013) 'Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods', *Geographical Analysis*, **45**: 150–79.

—— and Piras, G. (2015) 'Comparing Implementations of Estimation Methods for Spatial Econometrics', *Journal of Statistical Software*, **63**: 1–36. <http://www.jstatsoft.org/v63/i18/>.

—— and Yu, D. (2015) spgwr: Geographically Weighted Regression. R package version 0.6-28. <http://CRAN.R-project.org/package=spgwr>.

Boone, C. G. et al. (2010) 'Landscape, Vegetation Characteristics, and Group Identity in an Urban and Suburban Watershed: Why the 60s Matter', *Urban Ecosystems*, **13**: 255–71.

Branas, C. C. et al. (2011) 'A Difference-In-Differences Analysis of Health, Safety, and Greening Vacant Urban Space', *American Journal of Epidemiology*, **174**: 1296–306. doi:10.1093/aje/kwr273.

Clarke, L. W., Jenerette, G. D., and Davila, A. (2013) 'The Luxury of Vegetation and the Legacy of Tree Biodiversity in Los Angeles, CA', *Landscape and Urban Planning*, **116**: 48–59.

Cook, E. M., Hall, S. J., Larson, K. L. (2012) 'Residential landscapes as social-ecological systems: A synthesis of multi-scalar interactions between people and their home environment', *Urban Ecosystems*, **15** http://doi.org/10.1007/s11252-011-0197-0.

Duncan, D. T. et al. (2013) 'A Spatially Explicit Approach to the Study of Socio-Demographic Inequality in the Spatial Distribution of Trees across Boston Neighborhoods', *Spatial Demography*, **2014**: 1–29.

ESRI. (2012) *ArcMap 10.1*. Redlands: ESRI.

Forest Service Northern Research Station. (2008) Urban Forestry Bibliography. <http://www.milliontreesnyc.org/downloads/pdf/urban_tree_bib.pdf> accessed 15 Jul 2013.

Foster, A., and Dunham, I. M. (2015) 'Volunteered Geographic Information, Urban Forests, & Environmental Justice', *Computers, Environment and Urban Systems*, **53**: 65–75.

Fotheringham, A. S., Charlton, M. E., and Brunsdon, C. (1998) 'Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis', *Environment and Planning A*, **30**: 1905–27.

——, Brunsdon, C., and Charlton, M. (2003) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons.

Fox, J., and Weisberg, S. (2011) *An {R} Companion to Applied Regression*, 2nd edn. Thousand Oaks, CA: Sage. < http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.

Greenworks - City of Philadelphia. (2009) <http://www.phila.gov/green/greenworks/pdf/Greenworks_OnlinePDF_FINAL.pdf> accessed 3 Feb 2013.

Giner, N. M., and Rogan, J. (2012) 'A comparison of Landsat ETM+ and high-resolution aerial orthophotos to map urban/suburban forest cover in Massachusetts, USA', *Remote sensing letters*, **3**: 667–76.

—— (2013) 'Understanding the Social Determinants of Lawn Landscapes: A Fine-Resolution Spatial Statistical Analysis in Suburban Boston, Massachusetts, USA', *Landscape and Urban Planning*, **111**: 25–33.

—— (2014) 'Creating Spatially-Explicit Lawn Maps Without Classifying Remotely-Sensed Imagery: The case of suburban Boston, Massachusetts, USA', *Cities and the Environment (CATE)*, **7**: 4–25.

Grove, J. M. et al. (2006a) 'Data and Methods Comparing Social Structure and Vegetation Structure of Urban Neighborhoods in Baltimore, Maryland', *Society and Natural Resources*, **30**: 117–36.

—— et al. (2006b) 'Characterization of Households and its Implications for the Vegetation of Urban Ecosystems. Ecosystems', **9**: 578–97.

—— et al. (2006c) A report on New York City's present and possible urban tree canopy: Prepared for Fiona Watt, Chief of the Division of Forestry and Horticulture. New York Department of Parks and Recreation, USDA Forest Service, Northern Research Station. 28 pp. <http://nrs.fs.fed.us/nyc/local-resources/downloads/Grove_UTC_NYC_FINAL.pdf> accessed 5 Nov 2016.

—— et al. (2013) 'Building an Urban LTSER: The Case of the Baltimore Ecosystem Study and the D.C./B.C. ULTRA-Ex Project', in Singh S. J., Haberl H., Chertow M., Mirtl M., and Schmid M. (eds). *Long Term Socio-Ecological Research*, pp. 369–408. Dordrecht: Springer Netherlands.

——, Locke, D. H., and O'Neil-Dunne, J. P. M. (2014) 'An Ecology of Prestige in New York City: Examining the Relationships Among Population Density, Socio-Economic Status, Group Identity, and Residential Canopy Cover', *Environmental Management*, **54**: 402–19.

——, Pickett, S. T. A., Burch, W. R., and Machlis, G. (2015) *The Baltimore School of Urban Ecology: Space, Scale, and Time for the Study of Cities*. New Haven and London: Yale University Press.

Harrell, F. E., and Charles, D. (2015) Hmisc: Harrell Miscellaneous. R package version 3.16-0. <http://CRAN.R-project.org/package=Hmisc>.

Hamil, K. A. D. et al. (2016) 'Cross-scale Contradictions in Ecological Relationships', *Landscape Ecology*, **31**: 7–18.

Heckert, M., and Mennis, J. (2012) 'The Economic Impact of Greening Urban Vacant Land: A Spatial Difference-In-Differences Analysis', *Environment and Planning A*, **44**: 3010–27.

Heynen, N., Perkins, H. A., and Roy, P. (2006) 'The Political Ecology of Uneven Urban Green Space the Impact of Political Economy on Race and Ethnicity in Producing Environmental Inequality in Milwaukee', *Urban Affairs Review*, **42**: 3–25.

Hope, D. et al (2003) 'Socioeconomics drive urban plant diversity', *Proceedings of the National Academy of Sciences*, **100**(15): 8788–92.

—— et al (2003) 'Drivers of Spatial Variation in Plant Diversity Across the Central Arizona-Phoenix Ecosystem', *Society & Natural Resources*, **19**(2): 101–116.

Hox, J. (1998) 'Multilevel Modeling: When and Why', in Balderjahn I., Mathar R. and Schader M. (eds.) *Classification, Data Analysis, and Data Highways*, pp. 147–154. New York: Springer Berlin Heidelberg.

Hyman, H. H. (1942) The psychology of status. Archives of Psychology (Columbia University).

Keene, K. M. (2014) Personal communication Kevin M. Keene, Director of Mass Appraisal and Analysis City of Philadelphia Office of Property Assessment, 601 Walnut Street - 3rd floor Philadelphia, PA 19106.

Land Cover Philadelphia, PA. (2008, 2011) Raster Dataset. Spatial Analysis Laboratory at the University of Vermont, Burlington, VT. <http://gis.w3.uvm.edu/utc/Landcover/Philadelphia.zip>

Landry, S. M., and Chakraborty, J. (2009) 'Street Trees and Equity: Evaluating the Spatial Distribution of an Urban Amenity', *Environment and Planning A*, **41**: 2651–70.

——, and Pu, R. (2010) 'The Impact of Land Development Regulation on Residential Tree Cover: An Empirical Evaluation Using High-Resolution IKONOS Imagery', *Landscape and Urban Planning*, **94**: 94–104.

Locke, D. H. (2016) Data and R code supplement to "What's scale got to do with it? Models for urban tree canopy". Figshare. <https://dx.doi.org/10.6084/m9.figshare.3814665.v1> retrieved: 17 Sep 2016 (GMT).

—— and Grove, J. M. (2016) 'Doing the Hard Work Where It's Easiest? Examining the Relationships Between Urban Greening Programs and Social and Ecological Characteristics', *Applied Spatial Analysis and Policy*, **9**: 77–96.

—— and Baine, G. (2014) The good, the bad, and the interested: how historical demographics explain present-day tree canopy, vacant lot and tree request spatial variability in New Haven, CT. Urban Ecosystems, 1–19.

Logan, J. R., and Molotch, H. L. (1987) Urban fortunes: The political economy of place. University of California Pr.

Luck, G. W., Smallbone, L. T., and O'Brien, R. (2009) 'Socio-Economics and Vegetation Change in Urban Ecosystems: Patterns in Space and Time', *Ecosystems*, **12**: 604–20.

MacFaden, S. W. et al. (2012) 'High-Resolution Tree Canopy Mapping for New York City using LIDAR and Object-Based Image Analysis', *Journal of Applied Remote Sensing*, **6**: 063567–1.

Marco, A. et al. (2008) 'Gardens in Urbanizing Rural Areas Reveal An Unexpected Floral Diversity Related to Housing Density', *Comptes Rendus Biologies*, **331**: 452–65.

Martin, C. A., Warren, P. S., and Kinzig, A. P. (2004) 'Neighborhood socioeconomic status is a useful predictor of perennial landscape vegetation in residential neighborhoods and embedded small parks of Phoenix', *AZ. Landscape and Urban Planning*, **69**(4): 355–68.

McPhearson, T., Kremer, P., and Hamstead, Z. A. (2013) 'Mapping Ecosystem Services in New York City: Applying A Social–Ecological Approach in Urban Vacant Land', *Ecosystem Services*, **5**: 11–26.

Mennis, J. (2006) 'Mapping the Results of Geographically Weighted Regression', *The Cartographic Journal*, **43**: 171–9.

Merse, C. L., Buckley, G. L., and Boone, C. G. (2009) 'Street Trees and Urban Renewal: A Baltimore Case Study', *The Geographical Bulletin*, **50**: 65.

Merton, R. K., and Kitt, A. S. (1950) 'Contributions to the Theory of Reference Group Behavior', *Continuities in Social Research: Studies in the scope and method of "The American Soldier*, 40–105.

Mooney, S. J., Richards, C. A., and Rundle, A. G. (2014) 'There Goes the Neighborhood Effect: Bias Owing to Nondifferential Measurement Error in the Construction of Neighborhood Contextual Measures', *Epidemiology*, **25**: 528–35.

Nowak, D. J. (2012) 'Contrasting Natural Regeneration and Tree Planting in Fourteen North American Cities', *Urban Forestry and Urban Greening*, **11**: 374–82.

O'Brien, R. M. (2007) 'A Caution Regarding Rules of Thumb for Variance Inflation Factors', *Quality and Quantity*, **41**: 673–90.

O'Neil-Dunne, J. P. M. (2011) A report on the City of Philadelphia's Existing and Possible Tree Canopy. The Spatial Analysis Lab at the University of Vermont's Rubenstein School of the Environment and Natural Resources. <http://gis.w3.uvm.edu/utc/Reports/TreeCanopy_Report_Philadelphia.pdf> accessed 15 Nov 2014.

—— et al. (2012) 'An Object-Based System for LiDAR Data Fusion and Feature Extraction', *Geocarto International*, 1–16 (June 2012) doi:10.1080/10106049.2012.689015.

——, MacFaden, S., and Royar, A. (2014) 'A Versatile, Production-Oriented Approach to High-Resolution Tree-Canopy Mapping in Urban and Suburban Landscapes Using GEOBIA and Data Fusion', *Remote Sensing*, **6**: 12837–65 doi:10.3390/rs61212837.

OPA, Office of Property Assessment. (2014) "City of Philadelphia: Property Assessment Data." <http://www.phila.gov/OPA/Assessments/Pages/AssessmentData.aspx http://www.phila.gov/OPA/Documents/TaxYear2014DataSet.zip> accessed 28 July 2014.

Pearsall, H., and Christman, Z. (2012) 'Tree-Lined Lanes or Vacant Lots? Evaluating Non-Stationarity Between Urban Greenness and Socio-Economic Conditions in Philadelphia, Pennsylvania, USA at Multiple Scales', *Applied Geography*, **35**: 257–64.

Perkins, H. A., Heynen, N., and Wilson, J. (2004) 'Inequitable Access to Urban Reforestation: The Impact of Urban Political Economy on Housing Tenure and Urban Forests', *Cities*, **21**: 291–9.

Pickett, S. et al. (2011) 'Urban Ecological Systems: Scientific Foundations and A Decade of Progress', *Journal of Environmental Management*, **92**: 331–62.

Pham, T. -T. -H. et al. (2012a) 'Predictors of the Distribution of Street and Backyard Vegetation in Montreal, Canada', *Urban Forestry and Urban Greening*, doi:10.1016/j.ufug.2012.09.002.

—— et al. (2012b) 'Spatial Distribution of Vegetation in Montreal: An Uneven Distribution or Environmental Inequity?', *Landscape and Urban Planning*, **107**: 214–24.

Plant One Million. (2013) Plant One Million: 3 States. 13 Counties. 1 Tree at a time. <http://www.plantonemillion.org/Accessed> accessed 6 Apr 2013.

R Core Team (2015a) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing <https://www.R-project.org/>.

—— (2015b) Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, …. R package version 0.8-65. <http://CRAN.R-project.org/package=foreign>.

Raciti, S. M., Hutyra, L. R., and Newell, J. D. (2014) 'Mapping Carbon Storage in Urban Trees with Multi-Source Remote Sensing Data: Relationships Between Biomass, Land Use, and Demographics in Boston Neighborhoods', *The Science of the Total Environment*, **500–501C**: 72–83.

Robbins, P., and Sharp, J. (2003) 'Producing and Consuming Chemicals: The Moral Economy of the American Lawn', *Economic Geography*, **79**: 425–51.

Romolini, M., Grove, J. M., and Locke, D. H. (2013) 'Assessing and Comparing Relationships Between Urban Environmental Stewardship Networks and Land Cover in Baltimore and Seattle', *Landscape and Urban Planning*, **120**: 190–207.

Roy Chowdhury, R. et al. (2011) 'A Multi-Scalar Approach to Theorizing Socio-Ecological Dynamics of Urban Residential Landscapes', *Cities and the Environment (CATE)*, **4**: 1–19.

Runfola, D. M. et al. (2013a) 'A Growing Concern? Examining the Influence of Lawn Size on Residential Water Use in suburban Boston, MA, USA', *Landscape and Urban Planning*, **119**: 113–23.

—— et al. (2013b) 'Future Suburban Development and the Environmental Implications of Lawns: A Case Study in New England, USA', in Czamanski D., Benenson I., and Malkinson D. (eds.) *Modeling of Land-Use and Ecological Dynamics: Cities and Nature*, pp. 119–141. New York: Springer.

—— et al. (2014) 'Using Fine Resolution Orthoimagery and Spatial Interpolation to Rapidly Map Turf Grass in Suburban Massachusetts Using Fine Resolution Orthoimagery and Spatial Interpolation to Rapidly', *International Journal of Geospatial and Environmental Research*, **1**: 1–17.

——, and Hughes, S. (2014) 'What Makes Green Cities Unique? Examining the Economic and Political Characteristics of Grey and Green Cities', *Land*, **3**: 131–47.

Smith, R. M. et al. (2005) 'Urban Domestic Gardens (V): Relationships Between Landcover Composition, Housing and Landscape', *Landscape Ecology*, **20**: 235–53.

Smith, M. L. et al. (2010) 'Evaluation of the National Land Cover Database for Hydrologic Applications in Urban and Suburban Baltimore, Maryland', *JAWRA Journal of the American Water Resources Association*, **46**: 429–42.

Snijders, T. A. B., and Bosker, R. J. (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edn. New York: Sage.

Stone, B. (2004) 'Paving Over Paradise: How Land Use Regulations Promote Residential Imperviousness', *Landscape and Urban Planning*, **69**: 101–13.

Subramanian, S. V., Duncan, C., and Jones, K. (2001) 'Multilevel Perspectives on Modeling Census Data', *Environment and Planning A*, **33**: 399–417.

—— (2004) 'The Relevance of Multilevel Statistical Methods for Identifying Causal Neighborhood Effects', *Social Science and Medicine*, **58**: 1961–7.

Troy, A. R. et al. (2007) 'Predicting Opportunities for Greening and Patterns of Vegetation on Private Urban Lands', *Environmental Management*, **40**: 394–412.

——, Morgan Grove, J., O'Neil-Dunne, J. (2012) 'The relationship between tree canopy and crime rates across an urbanrural gradient in the greater Baltimore region', *Landscape and Urban Planning*, **106**(3): 262–270. http://doi.org/10.1016/j.landurbplan.2012.03.010.

United States Census Bureau. (2011) "Table 1. Annual Estimates of the Resident Population for Incorporated Places Over 50,000, Ranked by July 1, 2011 Population: April 1, 2010 to July, 2011" (CSV). 2011 Population Estimates. United States Census Bureau, Population Division. <http://www.census.gov/popest/data/cities/totals/2011/tables/SUB-EST2011-01.csv> accessed 4 Aug 2012.

Wu, J., and David, J. L. (2002) 'A Spatially Explicit Hierarchical Approach to Modeling Complex Ecological Systems: Theory and Applications', *Ecological Modelling*, **153**: 7–26.

Zhou, W. et al. (2009) 'Can Money Buy Green? Demographic and Socioeconomic Predictors of Lawn-Care Expenditures and Lawn Greenness in Urban Residential Areas', *Society and Natural Resources*, **22**: 744–60.