



April 2014

Creating Spatially-Explicit Lawn Maps Without Classifying Remotely-Sensed Imagery: The case of suburban Boston, Massachusetts, USA

Nicholas M. Giner
Clark University, nmg5007@gmail.com

Colin Polsky
Clark University, cpolsky@clarku.edu

Robert G. Pontius Jr.
Clark University, rpontius@clarku.edu

Daniel M. Runfola
University of Colorado - Boulder, dan@danrunfola.com

Samuel J. Ratick
Clark University, sratick@clarku.edu

Follow this and additional works at: <https://digitalcommons.lmu.edu/cate>

Recommended Citation

Giner, Nicholas M.; Polsky, Colin; Pontius, Robert G. Jr.; Runfola, Daniel M.; and Ratick, Samuel J. (2014) "Creating Spatially-Explicit Lawn Maps Without Classifying Remotely-Sensed Imagery: The case of suburban Boston, Massachusetts, USA," *Cities and the Environment (CATE)*: Vol. 7: Iss. 1, Article 10. Available at: <https://digitalcommons.lmu.edu/cate/vol7/iss1/10>

This Article is brought to you for free and open access by the Center for Urban Resilience at Digital Commons @ Loyola Marymount University and Loyola Law School. It has been accepted for inclusion in Cities and the Environment (CATE) by an authorized administrator of Digital Commons at Loyola Marymount University and Loyola Law School. For more information, please contact digitalcommons@lmu.edu.

Creating Spatially-Explicit Lawn Maps Without Classifying Remotely-Sensed Imagery: The case of suburban Boston, Massachusetts, USA

Residential lawns are a dominant and growing feature of US residential landscapes, and the resource-intensive management of this landscape feature presents major potential risks to both humans and the environment. In recent years, scientists and policymakers have been increasingly calling for large-extent measures of lawns and other similar landscape features. Unfortunately, the production of such datasets using traditional, remotely sensed measurement approaches can be prohibitively expensive and time consuming. This study uses two statistical prediction methods to extrapolate the quantity and spatial distribution of residential lawns from a sample of mapped lawns in a large study area in suburban Boston, Massachusetts. The goal is to find an inexpensive, broad-coverage dataset that will provide useable estimates of landscape features in places where we do not have direct measurements of those landscape features. The first estimation method uses OLS regression in conjunction with the sample of mapped lawns and freely available US Census data representing theoretically informed social driver variables. The second, simpler, and less computationally intensive estimation method allocates the mean of the sample of mapped lawns uniformly across the study area. Both estimation methods are performed 1000 times in a Monte Carlo framework where the sample is drawn randomly each realization, to assess the sensitivity of the prediction results to the selection of CBGs in each simple random sample. The outputs of each estimation method are then compared to a reference map where the quantity and spatial allocation of lawns is known for each spatial unit of analysis. Results indicate that the OLS prediction method specified with the independent social driver variables performs better than a uniform prediction method when both are compared to the full-study area reference map.

Keywords

lawns, land-change science, Monte Carlo simulation, quantity and allocation disagreement, suburbanization, Boston

Acknowledgements

The United States' National Science Foundation (NSF) supported this work via the following programs: Long-Term Ecological Research via grants OCE-0423565 & OCE-1026859 for Plum Island Ecosystems, Coupled Natural-Human Systems via grant BCS-0709685, Research Experiences for Undergraduates via grant SES-0849985, Urban Long-Term Ecological Research Areas via grant BCS-0948984, and MACRO-BIO via grant EF-1065741. The Massachusetts Geographic Information Systems office (MassGIS) provided much of the geospatial data free of charge. Major mapping assistance was provided by J. Morgan Grove (US Forest Service) and Jarlath O'Neil-Dunne (University of Vermont) through the affiliation with the NSF LTER Baltimore Ecosystem Study (BES) project (<http://beslter.org>). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the funders. The authors would also like to thank Albert Decatur (Graduate School of Geography, Clark University) for data preparation and Professor Rinku Roy Chowdhury (Department of Geography, Indiana University-Bloomington) for her thoughtful comments on previous versions of this manuscript.

INTRODUCTION

Residential lawns are one major and growing feature in many US suburban landscapes, and the resource-intensive management of this landscape feature presents potential risks to society and environment, with important implications for local water quality, quantity, and human health (Alig et al. 2004; Groffman et al. 2004; Law et al. 2004; Steinberg 2006; Robbins 2007; Larson et al. 2009; Cook et al. 2011; Roy Chowdhury et al. 2011). Consequently, there has been an increased interest among scientists, policymakers, and planners in mapping both the quantity and location of US lawns, and associated resource-intensive lawn management practices (Robbins and Birkenholtz 2003; Milesi et al. 2005). However, the fine-scaled and spatially heterogeneous nature of the suburban landscape presents a number of fiscal, personnel, and computational challenges for comprehensively mapping lawns (Milesi et al. 2005; Mathieu et al. 2007). This article attempts to mitigate these challenges, examining to what degree the quantity and spatial distribution of residential lawns can be estimated using straightforward, transferable methods and readily-available socio-economic data.

In this paper, we use Monte Carlo simulation to calibrate two statistical prediction models from samples of mapped lawns in a suburban watershed (1143 km²) outside of Boston, MA, to predict the quantity and allocation of lawns in the unsampled areas of the study area. The performance of the two prediction models was assessed by comparing their respective outputs to a study area reference map where the measures of lawns are known with high confidence (Polsky et al. 2012). In Section 2, we discuss the importance of mapping the spatial extent and distribution of lawns and review previous attempts at mapping lawns and other turf grasses. Sections 3 and 4 cover the data, methods, and results of our modeling effort. We conclude the article in Sections 5 and 6 with a discussion of our findings and a summary of major conclusions.

LITERATURE REVIEW

Why map lawns?

The recent interest among geographers, urban planners, environmental scientists, and decision-makers in mapping the spatial extent and distribution of lawns is based on three main factors. First, lawns are an increasingly common feature of the American suburban landscape. Suburbanization is one of the greatest processes driving land-use and land-cover change in the US today (Jackson 1985; Otto et al. 2002; Hanlon et al. 2010). Based on recent growth projections for urban/suburban areas in the US (Alig et al. 2004), we can assume that increased growth in suburban areas will mean increased growth in the extent of the lawn landscape.

Second, lawns are features that are largely created, maintained, and managed by humans. Although short herbaceous turf grasses occur naturally in some areas, the golf-course style, mono-species residential lawn requires extensive human inputs in the form chemicals and water, especially in areas where local soils do not provide essential nutrients for growth, where local climate conditions provide insufficient precipitation for growth, or where undesirable weeds, insects, fungi, or bacteria threaten the mono-species lawn aesthetic (Robbins and Birkenholtz 2003; Robbins and Sharp 2003). There is also a non-trivial carbon input from the polluting two- and four-stroke engine equipment required for the lawn cutting (Steinberg 2006). Based on the

high-input characteristics of lawns in terms of both chemicals and water, the lawn landscape is associated with a number of ecological implications, both positive and negative. On the positive side, lawns can provide critical urban/suburban ecosystem functions such as carbon sequestration (Townsend-Small and Czimczik 2010), increased water infiltration in areas dominated by impervious surfaces (Brabec et al. 2002), and mitigation of urban heat island effects (Sproken-Smith et al. 2000; Gober et al. 2009). On the negative side, the chemical fertilizers used to facilitate lawn growth and greenness are one of the main contributors to overall nitrogen pollution in urban and suburban watersheds, and can lead to the reduction of vital soil nutrients, loss of biological diversity, and acidification of soils and local surface water (Carrico et al. 2012). Additionally, increased nitrogen loads contained in storm water runoff can potentially affect surface water biochemistry, increasing the risks of estuarine eutrophication and loss of marine biodiversity (Robbins and Birkenholtz 2003). The chemical pesticides used to eliminate insects and weeds may also be harmful to the biological health of local waterways, threatening fish, macroinvertebrates, and other aquatic life (Robbins et al. 2001).

In addition to the environmental implications of lawn chemical use, there are a number of potential negative human implications, including: human exposure to chemicals tracked into homes via clothing or shoes, and the subsequent accumulation of these chemicals into house dust on surfaces and carpets, where children spend much of their time. Recent research has suggested that chronic exposure of children to the neurotoxins and chlorpyrifos contained in some lawn chemicals is more dangerous than previously accepted (Zartarian et al. 2000; Robbins and Sharp 2003). As such, even though lawns provide some benefits to people and society, their associated ecological and human health risks suggest the need to explore them and their implications across large geographic areas.

The implications listed above are responsible for the third main reason why mapping the lawn landscape is important: mapping the lawn landscape represents the first step in quantifying and understanding both the social processes driving the transformation of this landscape (ct. Robbins and Birkenholtz 2003; Grove et al. 2006a; Troy et al. 2007; Zhou et al. 2009; Boone et al. 2010; Giner et al. 2013) and the socio-ecological implications resulting from it (ct. Groffman et al. 2004; Law et al. 2004; Milesi et al. 2005; Larson et al. 2009). Maps that describe the spatial extent and distribution of lawns will help researchers to understand the social processes responsible for the creation of these landscapes, and to develop reliable estimates of their socio-ecological implications, which in turn will allow planners and policymakers to respond to and mitigate these implications (Robbins and Birkenholtz 2003; Milesi et al. 2005).

Previous attempts at mapping the lawn landscape

There are only a small number of published studies that characterize the spatial extent and distribution of lawns or other turf grasses. The methods used in those studies fall into two broad categories: *direct mapping methods* and *inferential mapping methods*. Direct mapping methods involve the detailed, full-study area mapping of lawns through photogrammetric, remote sensing, field-based, or survey methods. In these methods, researchers map each individual lawn in the given study area such that the population of lawns is known. Inferential mapping methods also produce a full-study area map of lawns; however, the lawn measurements in an inferential map are inferred from a sample of mapped lawns within the study area. In the inferential case, the

researchers map only a sample of lawns in the study area, and then use this sample to calibrate a predictive model that estimates lawns in locations where lawns have not been mapped. The following two sub-sections highlight recent research using both direct and inferential mapping methods. For the purposes of this literature review, we include methods that map any form of turf grass, which may include residential lawns, grass on golf courses or other athletic fields, highway medians, public rights-of-way, corporate office complexes, parks, and cemeteries. For our particular study, we focus solely on residential lawns, which we define as any turf grass on household parcels located within residential land uses.

Direct mapping methods

Within the last ten years, researchers in the Baltimore Ecosystems Study (BES) Long-Term Ecological Research (LTER) Site have used direct mapping methods to produce a series of full study area, fine-resolution land-cover datasets for use in socio-ecological and remote sensing research (see Grove et al. 2006a; Grove et al. 2006b; Troy et al. 2007; Zhou et al. 2008b; Zhou et al. 2009). The first of these datasets came from the Strategic Urban Forests Assessment (SUFA) for Baltimore City, Maryland, a 240 km² study area. This dataset contained four land-cover categories—developed, grass, forest, and water—which were extracted from pan-sharpened, 1m multispectral IKONOS imagery from 2001 using texture-based, remote sensing methods (Irani and Galvin 2003). Zhou and Troy (2008) produced a similar, full-study area land-cover dataset for the 171.5 km² Gwynns Falls Watershed, which contains portions of Baltimore City and Baltimore County, MD. This land-cover dataset was extracted using object-based image analysis (OBIA) methods from 0.6m aerial imagery acquired in 1999, and contains five land-cover categories—buildings, pavement, bare soil, coarse textured vegetation such as trees and shrubs, and fine textured vegetation such as grass. Their results indicated that the total quantity of grass vegetation was 48 km², which is roughly 30% of the study area. Mathieu et al. (2007) used OBIA methods to create a full-study area map of urban gardens in the 36 km² study area of Dunedin City, New Zealand, from 4m IKONOS imagery. Their land-cover dataset contained three garden categories, one of which was defined as containing greater than 70% lawn. The overall quantity of lawn of this category was 7.5 km², which is just over 20% of the study area. In another study using direct mapping methods, Farag et al. (2011) used a pixel-based, maximum likelihood methodology to classify fine-resolution (<1m) aerial photographs into eight categories—grass, trees and shrubs, concrete, asphalt, bare soil, shadow, water, and meadow—in three separate Salt Lake City, Utah suburbs. Urban vegetation (i.e. grass, trees and shrubs) was then identified from these eight categories, and combined with local evapotranspiration data to estimate household-specific irrigation demands. These estimated irrigation demands were then compared to the actual household-level water use data, such that systematic water over-users can be identified. Although the authors do not make the explicit distinction between lawns and all other turfgrass in the Farag et al. (2011) study, they do present an innovative approach to quantifying turfgrass under tree canopy using multi-temporal, leaf-off / leaf-on aerial imagery.

The most recent example of mapping lawns using direct mapping methods is by a research team at Clark University, who mapped 1143 km² of land cover in the Plum Islands Ecosystem (PIE) LTER site, Massachusetts (Polsky et al. 2012). This research team used OBIA methods to derive a seven-category land-cover map from 0.5m aerial photographs acquired over the study area in 2005. The results of their study indicate that there are 79 km² of lawn in PIE, accounting

for about 7% of the study area. Of all the direct mapping methods reviewed above, the Polsky et al. (2012) research was the only one to make the explicit distinction between lawns and other turf grasses. To make this distinction, turfgrasses within residential-use tax assessor's parcels (i.e. lawns) were isolated from all other turfgrasses such as golf courses, athletic fields, cemeteries, parks, and public rights-of-way.

The direct mapping methods are appealing to researchers because they produce a census of detailed observations, but these methods can be costly, resource intensive, time-consuming (Milesi et al. 2009), and often impractical for medium to large size cities, particularly because urban turfgrasses and lawns are fine-scaled and occur on spatially heterogeneous landscapes (Mathieu et al. 2007). Imagery of sufficient resolution to identify lawns—typically less than 4m spatial resolution (Jensen and Cowen 1999; Zhou et al. 2008a)—can cost up to \$17 US per km² (Giner and Rogan 2012), and processing it using either photogrammetric or remote sensing methods can require intensive computational power. As an example, it cost over \$20,000 US for the satellite imagery, GIS data, and computer hardware to complete the SUFA project in Baltimore City (Irani and Galvin 2003). Additionally, direct mapping methods can also require hundreds of personnel hours, especially in large study areas. Mathieu et al. (2007) found that their urban garden mapping research in Dunedin City, New Zealand, took roughly two months to map 36 km² using 4m spatial resolution pixels. For a large study area such as the PIE LTER, Mathieu et al. (2007)'s time estimate would translate to nearly 5.5 years.

Inferential mapping methods

Vinlove and Torla (1995) present the first peer-reviewed methods to map lawn cover. They extrapolated state-level, survey-based estimates of lawn acreage from ten states (Illinois, Montana, New Jersey, Maryland, North Carolina, Oklahoma, Kentucky, Michigan, Ohio, Pennsylvania) to the remaining 38 lower-United States through the use of ancillary data from the US Census Bureau and the Federal Housing Authority (FHA) for the period of 1977-1989. These state lawn acreage surveys were conducted at different years per-state from this period of time. For example, the Illinois survey was conducted in 1977, the Montana survey in 1983, and the Ohio survey in 1989. Their method consisted of deriving two household-scale lawn size estimates from the state surveys, then multiplying these lawn size estimates by the total number of single-family housing units in the US. Their results suggest that the total lawn area in the US ranges from 56,600-105,000 km², with a conservative estimate of about 73,000 km². It is important to note that Vinlove and Torla (1995) is the only indirect mapping study reviewed in this section that does not rely on remotely sensed data.

More recently, Robbins and Birkenholtz (2003) estimated, through aerial photo-interpretation and photogrammetric methods, that lawns cover about 23% of the total land cover in the ~1400 km² study area of Franklin County, Ohio. The authors photo-interpreted 1:600 scale black-and-white aerial photography and hand-digitized lawn cover on a sample of 63 spatially stratified tax assessor's parcels to develop a lawn coefficient value, which was multiplied by the non-built area of the unsampled assessor's parcels to predict the lawn area in each of the 79,894 unsampled parcels in the study area. This measure was then summarized to the US census tract level, for further analysis using socio-economic data.

Milesi et al. (2005) employed a remote sensing approach to estimate total turf grass in the lower-48 States as of the year 2000. Turf grass and impervious surface proportions were hand-digitized on a set of fine-spatial resolution aerial photographs covering a sample of 80, 1 km² grid cells across 13 major US urban centers (Atlanta, Boston, Chicago, Denver, Houston, Las Vegas, Miami, Minneapolis, New York, Phoenix, Portland, Sacramento, Seattle). The proportion of impervious surface derived from the aerial photographs was regressed against a national-scale, National Oceanic and Atmospheric Administration (NOAA)-derived impervious surface proportion estimate calculated at the 1 km pixel level from a combination of year 2001 radiance calibrated nighttime lights, a 1 km TIGER road density grid, and Landsat-derived urban land-cover categories (Elvidge et al. 2004). This regression equation was then applied to the lower-48 States to produce a nationwide impervious surface estimate at the 1 km pixel level. This impervious estimate was then regressed against the turf grass proportion value derived from the sample of aerial photographs to produce a nationwide estimate of 163,800 km² of turf grass. In 2009, the authors revised their previous national-scale turf grass estimate using the aerial photograph-derived turf grass data in conjunction with a new impervious surface proportion data set from the United States Geological Survey (USGS). The new estimate of lower-48 turf grass was 111,700 km², which is more than 50,000 km² less than the 2005 estimate. In an attempt to isolate lawns from other turf grasses, Milesi et al. (2009) built upon Vinlove and Torla (1995)'s methods using FHA and US Census data to develop national-scale lawn estimates, which range from 73,000 to 99,500 km² (Milesi et al. 2009).

The most recent research using inferential mapping methods is by Runfola (2012), who compared five methods for predicting turf grass in the Plum Island Ecosystems LTER site in northeastern Massachusetts. As a first step, the author aggregated the Polsky et al. (2012) 0.5m turf grass data to 120m cells, such that the proportion of turf grass in each 120m cell is known. Next, a simple random sample of 2.5% of the 120m cells was drawn and used to calibrate four interpolation models—polygonal, inverse distance weighting (IDW), ordinary kriging (OK), and mean allocation—to predict the proportion of turf grass in each unsampled 120m cell. The predicted turf grass values generated by each interpolative method were compared to the known turf grass values from the Polsky et al. (2012) PIE map to assess prediction error. This process was repeated 499 more times using Monte Carlo simulation, such that a total of 500 random samples were taken, generating 500 prediction maps for each inferential method, and thus 500 error measurements. The author also conducted a sensitivity analysis to assess how sample sizes of 2.5%, 5%, 7.5%, and 10% of the study area influence the error of the inferential model predictions. Results indicated that IDW performed the best in the study area for all sample sizes except for 10% of the study area, in which polygonal interpolation performed best. The overall estimate of PIE turf grass from the best performing model was 146 km², roughly 13% of the study area.

Although these inferential approaches may avoid some of the disadvantages of using direct mapping methods in terms of time, cost, and computational requirements, they too exhibit a few limitations. First, in both the Vinlove and Torla (1995) and Milesi et al. (2005; 2009) studies, the turf grass estimates rely on coarse-resolution data. Therefore, direct statements about household-level or neighborhood-level lawn cover are not possible because of the resolution mismatch between the fundamental pixel size (i.e. 1km² or 250 acres in the Milesi et al. studies) and the typical household parcel size (< 1 acre), of which lawns are only a portion. Additionally,

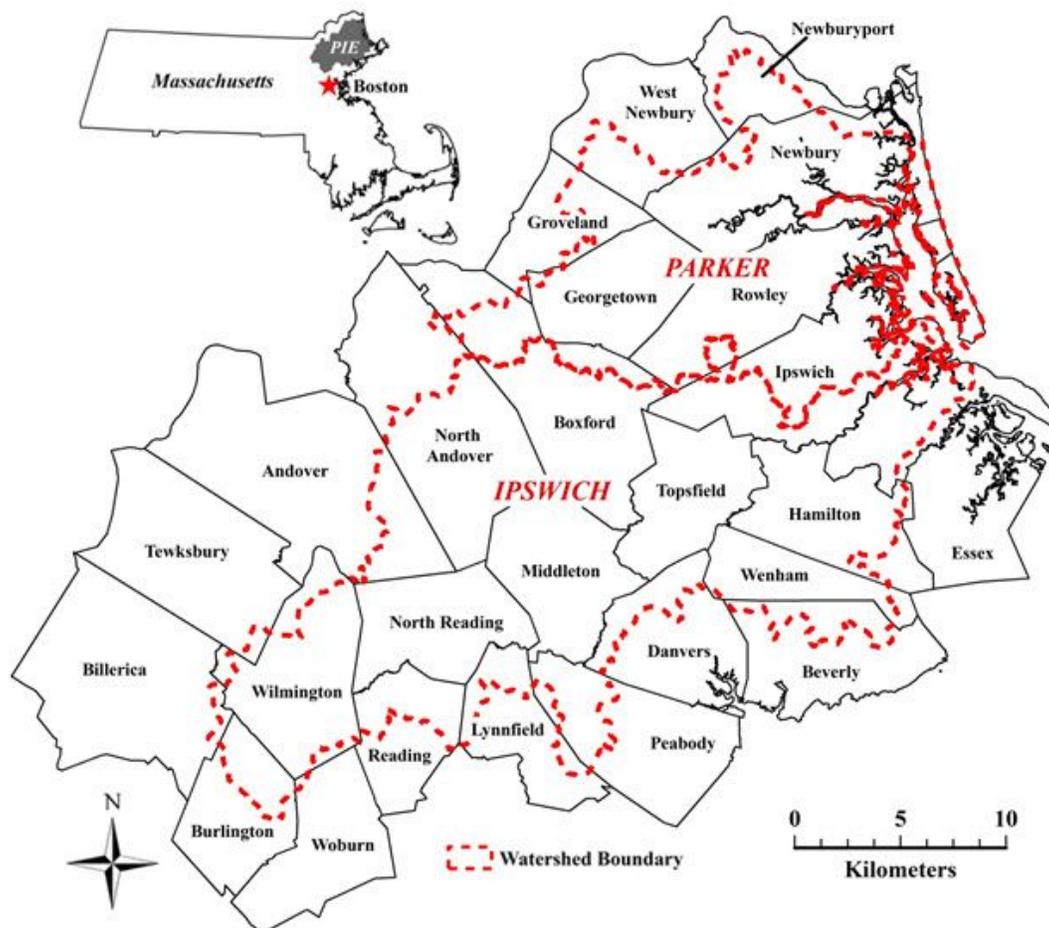
the coarse-resolution, nationwide lawn estimates in Vinlove and Torla (1995) did not rely on spatially explicit data, which means that their estimation models were not informed by any geographic characteristics of existing lawns such as distance, adjacency, or proximity, etc. Such a limitation can preclude many types of analyses – for example, the previously mentioned Farag et al. (2011) study hinged on having reliable estimates of land-cover on each residential property in their study area to identify systematic water over-users.

Second, some of these studies rely on extrapolation of turf grass and lawn from small sample surveys, i.e. ten states in Vinlove and Torla, (1995), a transect sample of aerial photographs from each of 13 metropolitan areas in Milesi et al. (2005), or a sample of less than 1% of study area households in Robbins and Birkenholtz (2003). The use of such a small number of samples for calibration might lead to imprecise estimates due to within-city heterogeneity, or to regional variation across the country.

Third, these estimates, with the exception of Robbins and Birkenholtz (2003), include all turf grass, such as commercial and institutional grass cover, golf courses, athletic fields, parks, and lawns (Milesi et al. 2005). As such, using these turf grass datasets to make statements about only residential lawn landscapes is not possible.

In this paper, we use a combined remote sensing and inferential approach to estimate the quantity and spatial allocation of lawns in the PIE LTER site, Massachusetts, to answer two research questions: 1) how well do social driver variables predict lawns at the Census block group (CBG)-level when compared to a lawn map derived using direct mapping methods, and 2) how well does a simpler, less computationally intensive baseline uniform prediction method predict lawns at the CBG-level when compared to a lawn map derived using direct mapping methods? To our knowledge, this is the first large-study area, fine-resolution inferential lawn mapping exercise that uses theoretically-informed independent variables to proxy the social processes driving lawn landscapes.

Figure 1. Study area: The 26 towns of the Plum Island Ecosystems LTER that intersect the Ipswich and Parker River watersheds, Massachusetts. The total area of the 26 towns is 1143 km² (Polsky et al. 2012).



METHODS

Study area

The study area for this research is the US National Science Foundation-funded Plum Island Ecosystems (PIE) LTER site, located about 20 miles northeast of Boston, Massachusetts (see <http://ecosystems.mbl.edu/pie/>). This 1143 km² contiguous study area includes 26 towns that intersect the Ipswich and Parker River watersheds (Figure 1). In 2000, PIE housed approximately 460,000 people with an average of three people per household (US Census Bureau 2000). The PIE study area ranges from coastal wetland areas in the eastern portion of the study area to relatively densely populated, suburban residential, commercial, and industrial areas in the southern and western portions. The dominant land uses in PIE are forests and forested wetlands (47% of the landscape) followed by medium- and low-density residential areas (17%) (MassGIS 2011).

PIE is a compelling study area for this research for three main reasons. First, the Ipswich and Parker River watersheds drain into a large salt marsh—the largest wetland-dominated

estuary in New England. This estuary is highly biologically productive, which makes it of considerable ecological, commercial, and social value. Second, PIE is located within the Greater Boston metropolitan area, one of the most rapidly sprawling US regions in recent years, despite modest population growth (Otto et al. 2002). Third, our paper adds to a growing body of research literature focused on understanding the interactions between humans and the environment in PIE (see Harris et al. 2012a; Harris et al. 2012b; Giner et al. 2013; Runfola et al. 2013a; Runfola et al. 2013b).

Lawn mapping data and methodology

The full-study area PIE land-cover map (Polsky et al. 2012) was derived from 4-band (R,G,B,NIR), 0.5m spatial resolution aerial orthophotos acquired between 9 and 17 April 2005. The land-cover mapping methodology uses object-based image analysis (OBIA) methods (cf. Zhou and Troy 2008; Zhou et al. 2008a) to classify the aerial imagery into seven land-cover categories: bare soil, coniferous forest, deciduous forest, fine green turf grass, impervious surfaces, water, and wetlands. Lawns were isolated from other turf grasses using GIS methods and data concerning tax assessor's parcels and 2005 Massachusetts land-use. Additional ancillary data included GIS layers of impervious surfaces, protected lands, and a Massachusetts Department of Environmental Protection (DEP) wetlands layer. Data on the independent variables were provided by the US Census Bureau. All imagery, ancillary GIS data, and US Census data were obtained from the Massachusetts Office of Geographic and Environmental Information (MassGIS 2011). The unit of analysis for our regressions is the Census Block Group (CBG) (N=314).

The dependent variable – Percent lawn

The dependent variable for this study is *Percent Lawn Cover*, which we define as the ratio of lawn area in a given CBG to the total land area in that CBG. Our lawn measure contains only residential grass, and does not include grass on athletic fields, golf courses, public rights-of-way, highway medians, office complexes, parks, or cemeteries.

The independent variables and expected relationships with percent lawn

We select our independent variables for the regression model calibration following the nested, theoretical approach of Troy et al. (2007). This approach has been applied in research examining the social drivers of urban vegetation patterns in Baltimore (Grove et al. 2006a; Grove et al. 2006b; Troy et al. 2007; Zhou et al. 2008b; Zhou et al. 2009; Boone et al. 2010) and more recently in the Boston suburbs (Giner et al. 2013). This approach argues that there are three major socio-economic processes that influence the biophysical features of residential landscapes. *Population Density* is specified to reflect the likely landscape processes associated with increasing numbers of people per unit area. Recent research has shown that in the PIE study area, increasing population density leads to increasing lawn cover. It is important to note that this relationship is reflective of the largely low-density, suburbanized nature of the PIE study area (Giner et al. 2013). Previous research has revealed that in high-density urbanized areas such as Baltimore (Troy et al. 2007) and Denver (Mennis, 2006), the opposite relationship holds true—that higher population density correlates with less lawn cover. *Social Stratification* refers

to a process that highlights the importance of factors such as income, level of education, ethnicity, and housing age as strong correlates with preferences for residential landscape features. Social stratification theory hypothesizes that the level of socio-economic status of a neighborhood reflects that neighborhood's ability to invest in both public and private greening initiatives, such as lawn maintenance and tree planting. This theory also suggests that residents of higher socio-economic status may be drawn to neighborhoods where these greening initiatives are already in place (Troy et al. 2007). *Lifestyle Behavior* refers to a process whereby household consumption patterns, such as those associated with residential landscape management, may be motivated by neighborhood group identity and by homeowners' perceptions of social status and prestige that are associated with various lifestyles (Grove et al. 2006a; Zhou et al. 2009; Boone et al. 2010). In this paper, we specify twelve independent variables: one variable for *Population Density* (number of residents per square kilometer), six variables for *Social Stratification* (median household income, median home value, percent housing vacancy, percent high-school graduates, median house age, percent African-American), and five variables for *Lifestyle Behavior* (percent single-family detached homes, average household size, percent owner-occupied housing, percent protected land, and percent married). Recent research in the PIE study area shows that population density, percent single-family detached homes, average household size, and percent protected land are significant ($p < 0.05$) and explain 55% percent of the Census block group-level variation in lawn cover in PIE (Giner et al. 2013).

The geospatial database

The geospatial database used in the spatial lawn predictions contains information on the dependent variable and each of the twelve independent variables at the CBG-level (N=314). The Polsky et al. (2012) 0.5m PIE lawn map was aggregated to the CBG-level such that each CBG contains a percent lawn value. For the remainder of the text, this will be referred to as the "reference map", which is the best available information concerning the density of lawns at the CBG-resolution. The independent variables representing the three-theory approach were appended to the CBG-level PIE lawn database. Table 1 lists the definition and summary statistics for the dependent variable and each of the twelve independent variables.

Table 1. Dependent and independent variable definitions and summary statistics where the observational unit is the US Census block group (CBG). The number of CBGs is 314.

| | Theory | Variable | Definition | Mean | StDev | Median | Min. | Max. | Source | |
|-----------------------|---|---|---|---|---------|--------|---------|--------|---------|---|
| Dependent variable | | P_LAWN | Proportion of land area in the CBG that is lawn | 11.13 | 6.33 | 9.94 | 1.06 | 37.89 | 1 | |
| Independent variables | <i>Population density theory (one variable)</i> | POPD_SQKM | Population density (people per km ²) | 1152 | 1154 | 781 | 56 | 7169 | 2 | |
| | | <i>Social stratification theory (six variables)</i> | INC_MED_HS | Median household income (2000\$) | 68,293 | 20,271 | 66,478 | 13,790 | 146,981 | 2 |
| | | | VL_MED_OWN | Median value of all owner-occupied housing units (2000\$) | 243,493 | 65,852 | 229,800 | 0 | 575,400 | 2 |
| | <i>Lifestyle theory (five variables)</i> | PCT_VACANT | Percent of housing vacant | 2.17 | 2.9 | 1.81 | 0 | 28.69 | 2 | |
| | | PCT_HSGRAD | Percent high-school graduates | 19.17 | 6.58 | 19.46 | 2.08 | 35.13 | 2 | |
| | | HOUS_AGE | Median house age | 46.03 | 13.32 | 45 | 11 | 66 | 2 | |
| | | PCT_AFAM | Percent of population that is African-American | 0.80 | 1.50 | 0 | 0 | 8.90 | 2 | |
| | | PCT_SFDT | Percent of housing that is single-family detached homes | 71.78 | 26.3 | 80.18 | 0 | 100 | 2 | |
| | | HSHLD_SIZE | Average population of all housing units (number of people/housing unit) | 2.68 | 0.489 | 3 | 1 | 4 | 2 | |
| | | PCT_OWNOCC | Percent of housing that is owner-occupied | 77 | 20.95 | 84.82 | 0 | 100 | 2 | |
| | | PCT_PROT | Percent of land that is protected open space | 10.62 | 13.44 | 5.52 | 0 | 63.69 | 3 | |
| | | PCT_MARRIED | Percent of households with a married couple | 62.36 | 15.17 | 64.25 | 15.79 | 95.23 | 2 | |

1 = Polsky et al. (2012)

2 = US Census Bureau (2000)

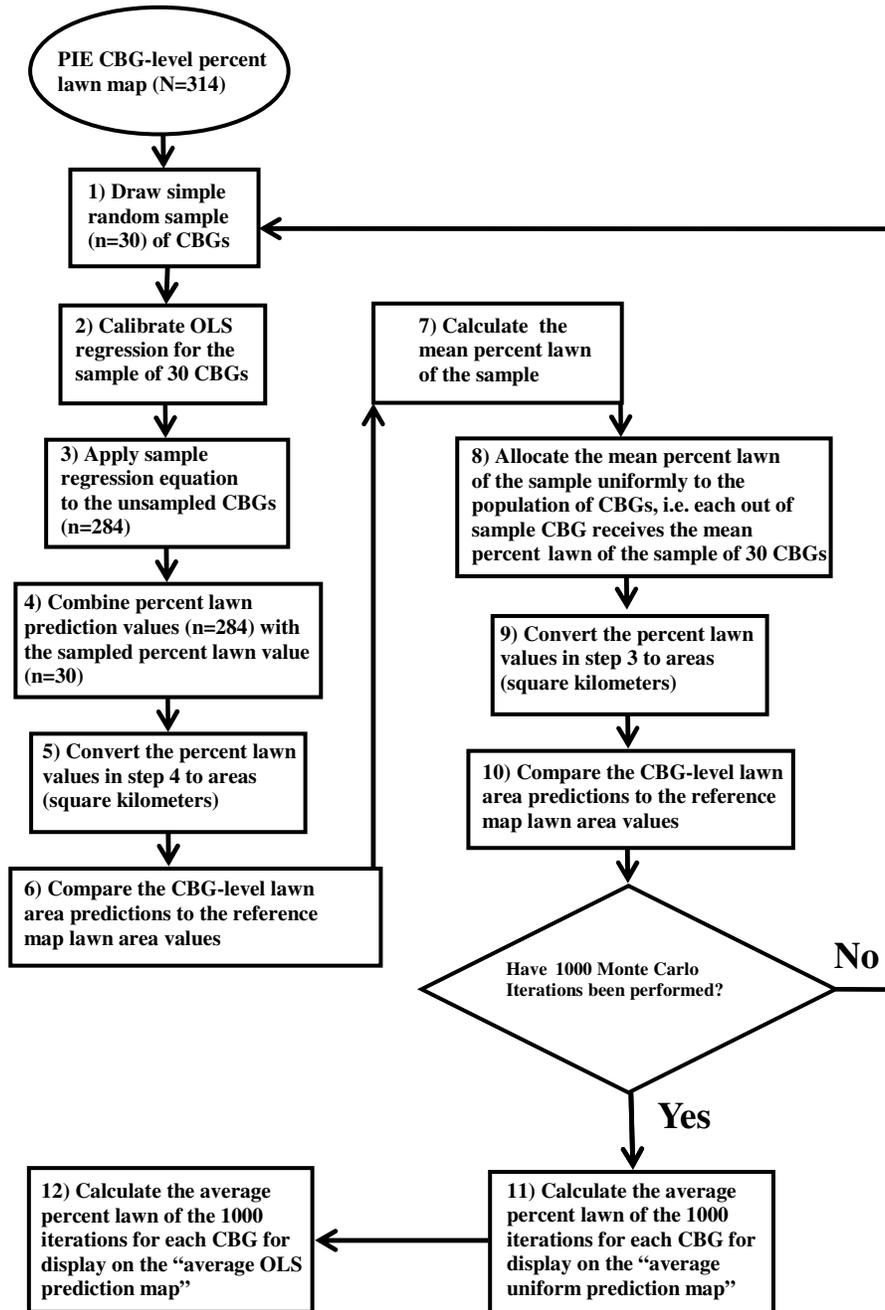
3 = MassGIS (2011)

Methodology for spatial lawn prediction using social driver variables

Steps 1-6 in the flowchart in Figure 2 show the methodology for creating the CBG-level percent lawn predictions using the 12 independent variables listed in Table 1. After aggregating the 0.5m lawn data to the CBG-level, step 1 is to draw a simple random sample of 30 CBGs, which is roughly 10% of the population of CBGs. This step is designed to simulate a limited investment in turf grass mapping. In step 2, we calibrate an ordinary least squares (OLS) regression model using the sample of 30 CBGs, where the dependent variable is percent lawn and all independent variables are entered together. The equation generated from this OLS regression is then applied to the 284 unsampled CBGs, creating a percent lawn prediction value for each unsampled CBG (step 3). Next, the predicted percent lawn values are combined with the observed percent lawn values drawn in the simple random sample, to create a full-study area,

combination sampled and predicted population map ($N=314$) (step 4). Step 5 is to convert the 314 percent lawn values to areas for error assessment. Step 6 is to perform the error assessment by comparing the OLS prediction to the reference map such that the predicted lawn area value for each CBG is compared to the known lawn area value for the corresponding CBG in the reference map. The above steps are repeated 999 more times using Monte Carlo simulation, such that a total of 1000 spatial lawn predictions are created. The Monte Carlo method was chosen for this analysis to assess the sensitivity of the results of each prediction to the effect of simple random sampling – i.e. to simulate the likely outcome of a practitioner selecting 10% of the PIE study area at random to map (Ratick and Schwarz 2009).

Figure 2. Flowchart depicting the steps to create the OLS prediction and uniform prediction maps.



Methodology for the uniform lawn prediction

Steps 7-10 in the flowchart in Figure 2 show the methodology for creating the uniform allocation spatial lawn predictions. After aggregating the 0.5m lawn data to the CBG-level and drawing the simple random sample of 30 CBGs (step 1 in the flowchart), step 7 is to calculate the mean percent lawn of the simple random sample. Step 8 is to allocate this mean percent lawn value to

the population of 314 CBGs. For example, if the mean percent lawn of the sample of 30 CBGs is 10% lawn, then each of the 314 CBGs is assigned a value of 10% lawn. Step 9 is to convert the 314 percent lawn values to areas for error assessment. Step 10 is to perform the error assessment by comparing the uniform prediction to the reference map such that the predicted lawn area value for each CBG is compared to the known lawn area value for the corresponding CBG. The above steps are repeated 999 more times using Monte Carlo simulation, such that a total of 1000 uniform lawn predictions are created. The final steps are to calculate the average percent lawn value of the 1000 OLS simulations for each CBG (Step 11) and the average percent lawn value of the 1000 uniform simulations for each CBG (Step 12) for display on maps. We refer to these maps from here forward as the “average OLS prediction map” and the “average uniform prediction map”, respectively.

Model assessment

We conduct two map comparisons to answer our two research questions:

- 1) OLS prediction map versus Polsky et al. (2012) reference map, and
- 2) uniform prediction map versus Polsky et al. (2012) reference map.

The first map comparison addresses our first research question and tests how the inferential OLS prediction model specified with the independent social driver variables performs in relation to the direct mapping method created using remotely sensed data and object-based image analysis methods. The second map comparison addresses our second research question and tests how the computationally simpler, baseline inferential method of uniform prediction performs in relation to the direct mapping method. By comparing the results of these two map comparisons, we will gain an understanding of the contribution of the independent variables to the accuracy of the CBG-level prediction of lawns. We assess the performance of both inferential methods in relation to the direct mapping method in terms of the disagreement budget of lawn area. The disagreement budget is split into two types of disagreement: quantity disagreement and allocation disagreement, which when summed equals the total disagreement between two maps (Pontius et al. 2008; Pontius and Millones 2011). Quantity disagreement measures how much less than perfect a model assigns the quantity of lawn in the entire study area, while allocation disagreement measures how much less than perfect the model allocates this quantity of lawn spatially among the CBGs. The total absolute disagreement between two maps is given as:

$$D_t = \sum_{b=1}^B |X_b - Y_b| \quad \text{Equation 1}$$

where D_t is the total disagreement, B is the total number of census block groups, b is a unique identifier for each census block group, X_b is the observed square kilometers of lawn within census block group b in the reference map, and Y_b is the predicted square kilometers of lawn within census block group b in either the OLS-generated map or the uniform prediction map. The quantity disagreement D_q is calculated by equation 2:

$$D_q = |\sum_{b=1}^B (X_b - Y_b)| \quad \text{Equation 2}$$

Allocation disagreement D_a is calculated by equation 3:

$$D_a = D_t - D_q \quad \text{Equation 3}$$

where D_a is allocation disagreement, D_t is total disagreement, and D_q is quantity disagreement in square kilometers between the Polsky et al. (2012) reference map and the two inferential methods. Because the disagreement statistics are calculated in units of area, all percent lawn maps are converted to square kilometers of lawn. We generate the disagreement budget statistics for each of the 1000 Monte Carlo simulations for each inferential method. Additionally, we compare the prediction results for each of the 1000 pairs of Monte Carlo simulations, where one pair consists of a single initial calibration sample that is then used to produce both an OLS prediction and uniform prediction. This allows us to understand which of the two inferential methods predicted lawn areas more accurately on a paired simulation-by-simulation basis. For example, the OLS prediction lawn area error for each CBG in simulation 1 is compared to the uniform prediction lawn area error for each CBG in simulation 1, followed by simulation 2, 3, and so on to the 1000th simulation. To perform this comparison, we first calculate the difference between the observed value and the predicted value (for both inferential methods) for each CBG in each simulation, which gives the total error in units of area between the observed value and the predicted value. We then subtract the total error for the OLS prediction from the total error for the uniform prediction to determine the area deviations between each simulation pair. A positive deviation indicates that the OLS prediction method produced less error, while a negative deviation indicates that the uniform prediction method produced less error.

RESULTS

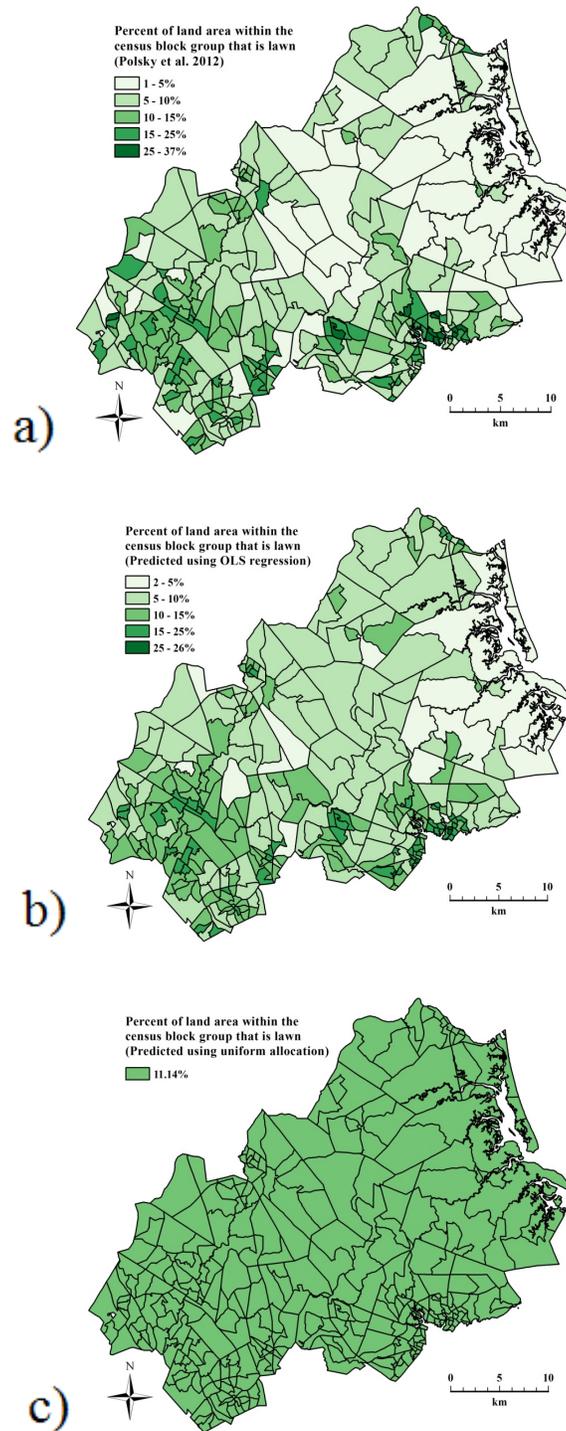
Direct and inferential mapping methods

Figures 3a-c and Table 2 show the spatial results and summary statistics for the reference map, the average OLS prediction map, and the average uniform prediction map, respectively. The Polsky et al. (2012) reference map produced by the direct mapping methods contains 78.41 km² of lawn, accounting for 7.13% of the study area (Figure 3a). The average OLS prediction map has 89.79 km² of lawn, which is 8.17% of the study area (Figure 3b), while the average uniform prediction map has 122.46 km² of lawn, which is 11.13% of the study area (Figure 3c). In both the reference map and the OLS prediction map, smaller CBGs in both the southern and southwestern portions of the study area have higher percentages of lawn than the larger CBGs in the eastern and north eastern portions. The uniform prediction map has the same percentage of lawn assigned to each CBG.

Table 2. Summary statistics for the three PIE lawn maps: 1) Polsky et al. (2012) reference map, 2) the average OLS prediction map, and 3) the average uniform prediction map. The units are percentage of the CBG.

| Map | Mean | Stdev | Median | Min | Max | N |
|------------------------------|-------|-------|--------|-------|-------|-----|
| Polsky et al. 2012 reference | 11.13 | 6.34 | 9.94 | 1.07 | 37.89 | 314 |
| OLS prediction | 11.38 | 4.72 | 10.98 | 2.87 | 26.32 | 314 |
| Uniform prediction | 11.13 | 0 | 11.13 | 11.13 | 11.13 | 314 |

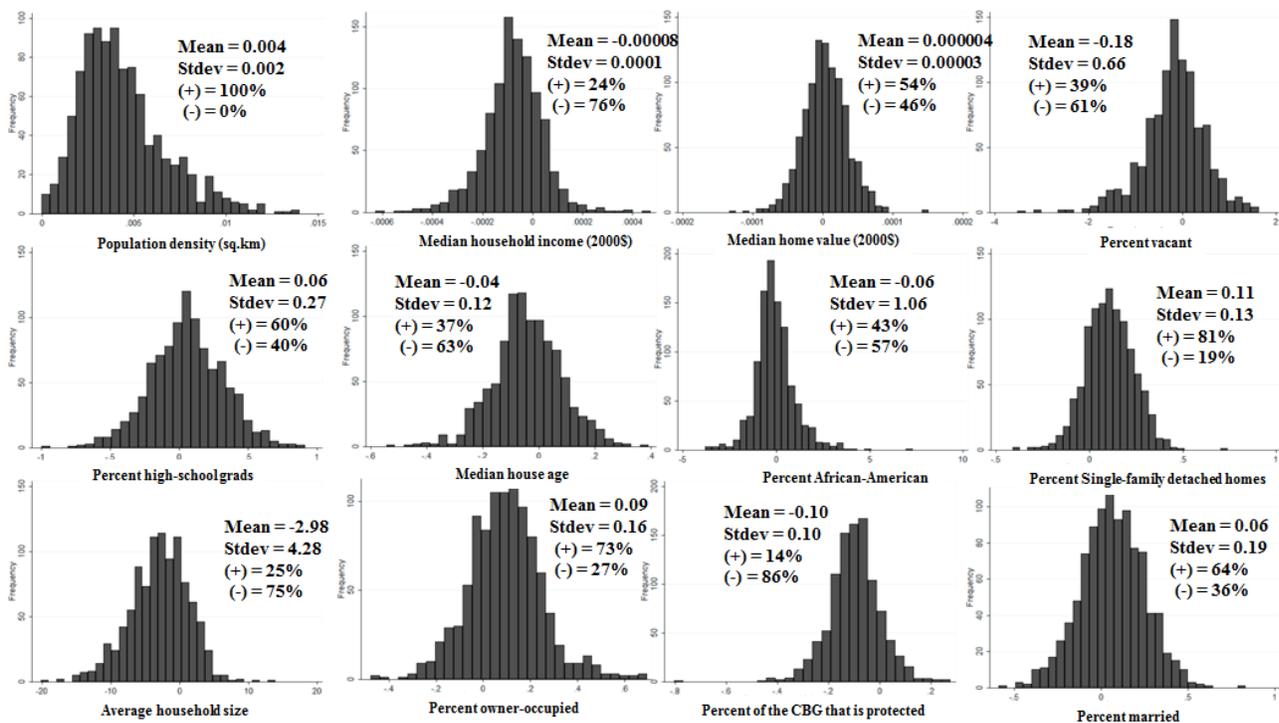
Figure 3a-c. The (a) Polsky et al. (2012) lawn map, (b) the average OLS prediction map (b), and (c) the average uniform prediction map. All areal units are measured as “percent lawn,” i.e., the area of grass within residential land in the CBG.



In addition to the 1000 lawn predictions produced by the 1000 Monte Carlo simulations of each respective inferential method, the Monte Carlo simulations of the OLS prediction method

produced 1000 coefficients for each independent variable. Figure 4 shows histograms of the coefficient estimates for the 1000 Monte Carlo simulations in the OLS prediction method. The figure also contains values for the percentage of the 1000 simulations in which the algebraic sign of the estimated coefficient is positive or negative. As Figure 4 illustrates, the only independent variable that consistently had a non-zero coefficient value (i.e. was statistically significant) was *population density*. In 100% of the 1000 OLS predictions, the algebraic sign of the coefficient for the population density variable is positive.

Figure 4. Histograms of the independent variable coefficient estimates for the 1000 Monte Carlo simulations.



Global model assessment results

The results of the global model assessment in terms of the disagreement budget statistics for the 1000 OLS prediction maps and the 1000 uniform prediction maps are presented in Figures 5a and 5b, respectively. In each figure, the y-axis represents square kilometers of total disagreement, split between quantity (black) and allocation (gray) disagreement. The x-axis shows the Monte Carlo simulation, ordered from smallest to largest total disagreement for both inferential methods. When the figures are compared qualitatively, the OLS prediction method produces less total disagreement and less quantity disagreement than the uniform prediction method. The average total disagreement for the 1000 OLS predictions is 39.67 km², with an average of 13.63 km² of quantity disagreement and 26.04 km² of allocation disagreement. The average total disagreement for the 1000 uniform predictions is 59.80 km², with an average of 44.05 km² of quantity disagreement and 15.75 km² of allocation disagreement. The lowest total disagreement (38.54 km²) from the 1000 uniform predictions is nearly two square kilometers

greater than the average total disagreement (36.67 km^2) for the entire set of 1000 OLS predictions.

Figure 5a-b. Distribution of the disagreement budgets in terms of total disagreement, quantity disagreement, and allocation disagreement across the 1000 Monte Carlo simulations, for (a) OLS prediction maps compared to the reference map, and (b) uniform prediction maps versus the reference map. The simulations are ordered by total disagreement from smallest to largest total disagreement.

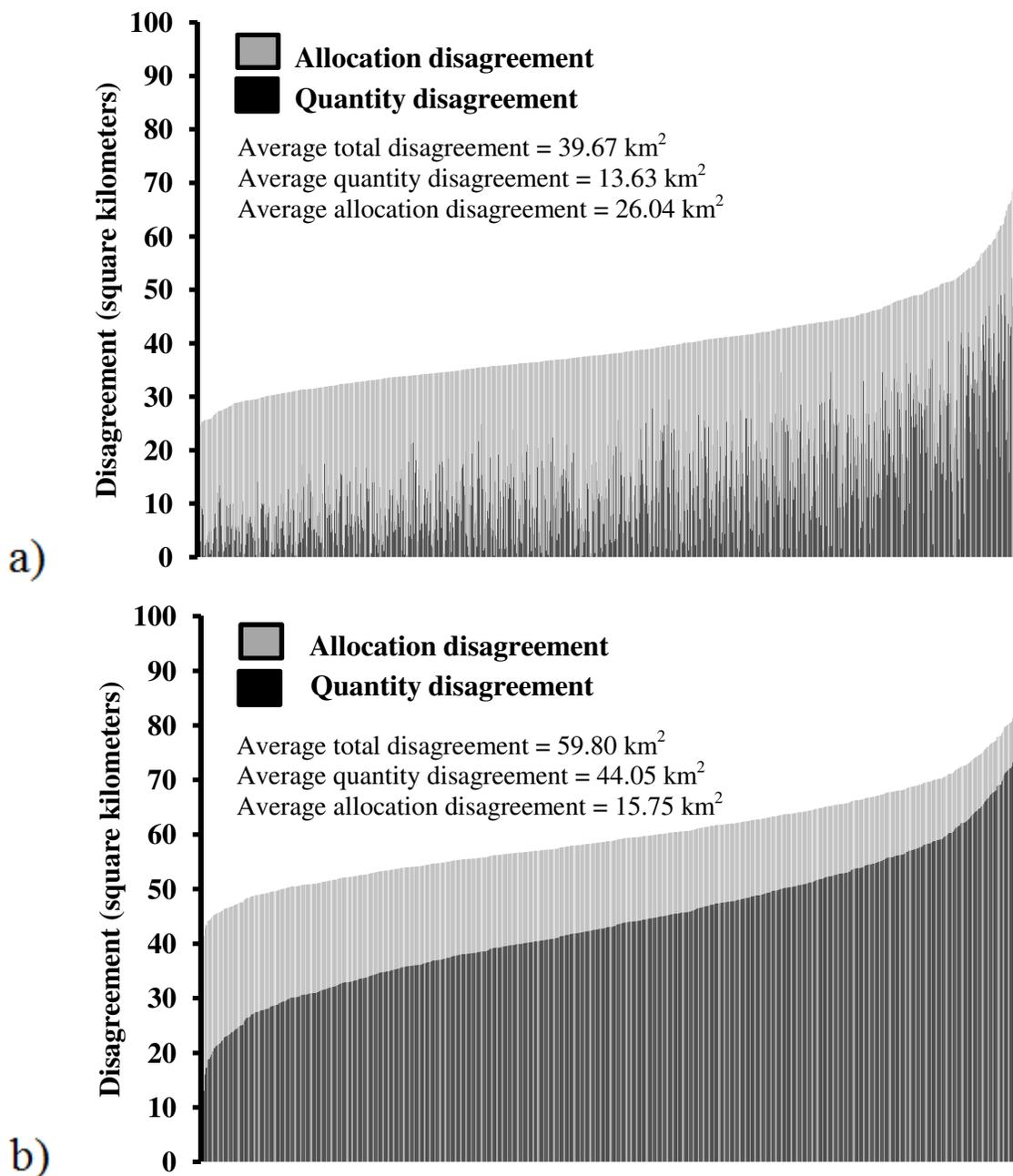
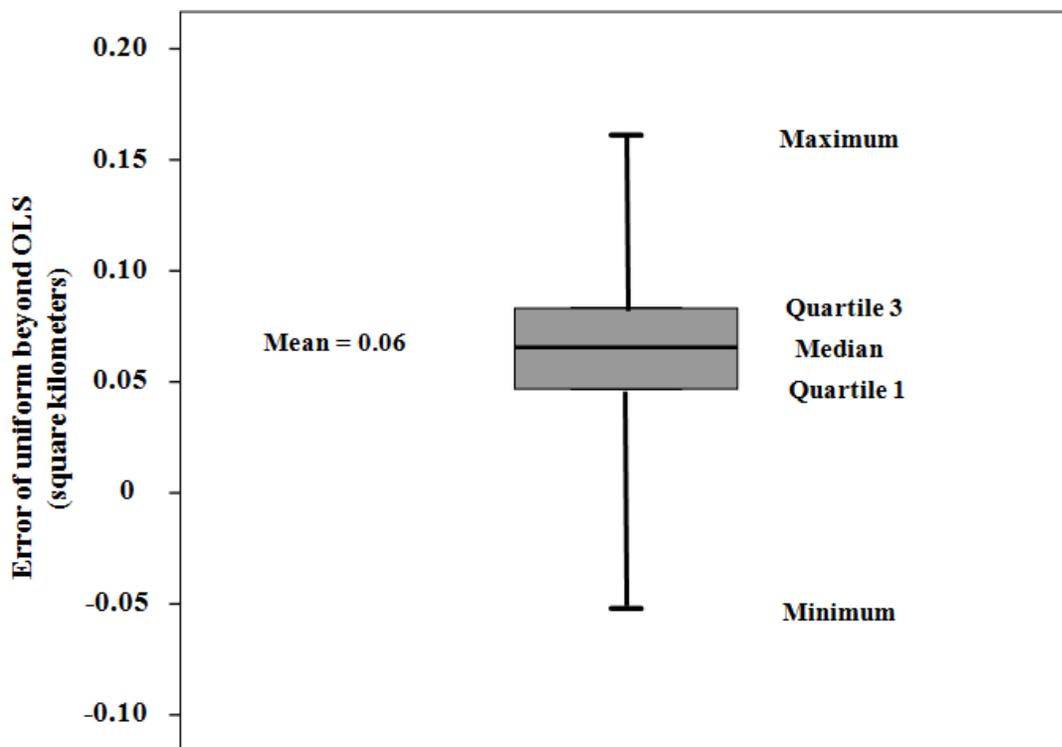


Figure 6 shows the results of pairwise differences between the uniform prediction error and the OLS prediction error for each of the 1000 Monte Carlo simulation pairs. The y-axis

shows the square kilometers of error produced by the uniform prediction method minus the error produced by the corresponding OLS prediction method that is based on the same sample. Averaged across the 1000 Monte Carlo simulations, the uniform prediction method produced 0.06 square kilometers more error than the OLS prediction method on a pairwise basis, which suggests that the errors in the uniform prediction method are systematically larger than errors in the OLS prediction. Additionally, the OLS prediction method produced less error than the uniform prediction method in 966 of the 1000 simulations.

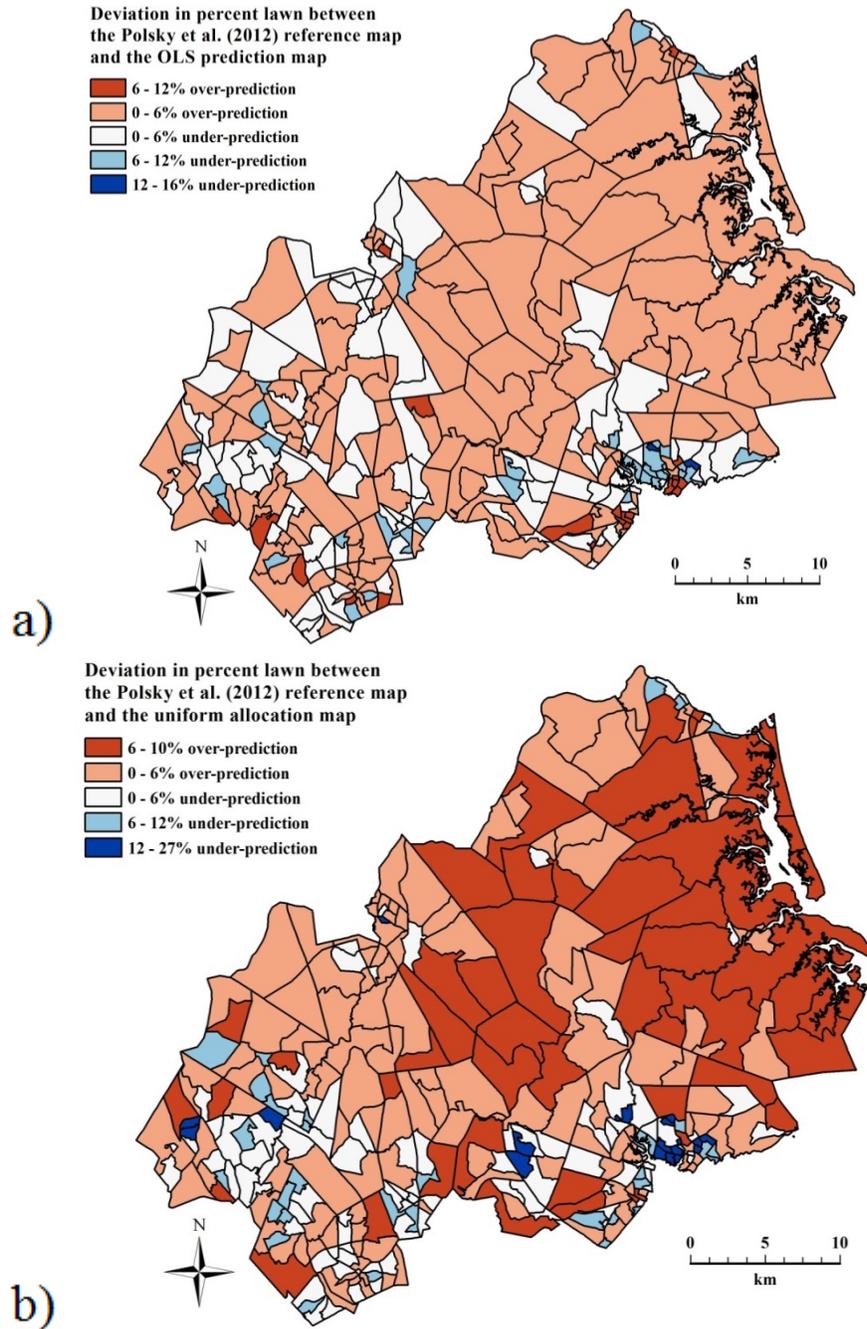
Figure 6. Distribution of the difference in total error between the uniform prediction and OLS prediction for the 1000 Monte Carlo samples. The box plot shows that the error in the uniform prediction is systematically larger than the error in the OLS prediction. The OLS prediction method produced less error than the uniform prediction method in 966 of the 1000 simulations.



Local model assessment results

Figures 7a-b show the average deviations in percent lawn between the reference map versus the average OLS prediction and average uniform prediction maps, respectively. Generally speaking, both inferential methods tend to over-predict percent lawn in the largest CBGs of the study area and under-predict percent lawn in the smallest CBGs, with the uniform prediction method over- and under-predicting more severely.

Figure 7a-b. Maps showing the percent lawn deviation between the Polsky et al. (2012) reference map compared to (a) the average OLS prediction map and (b) the average uniform prediction map. Both maps show over-prediction in larger CBGs and under-prediction in smaller CBGs; this bias is more severe in the uniform prediction.



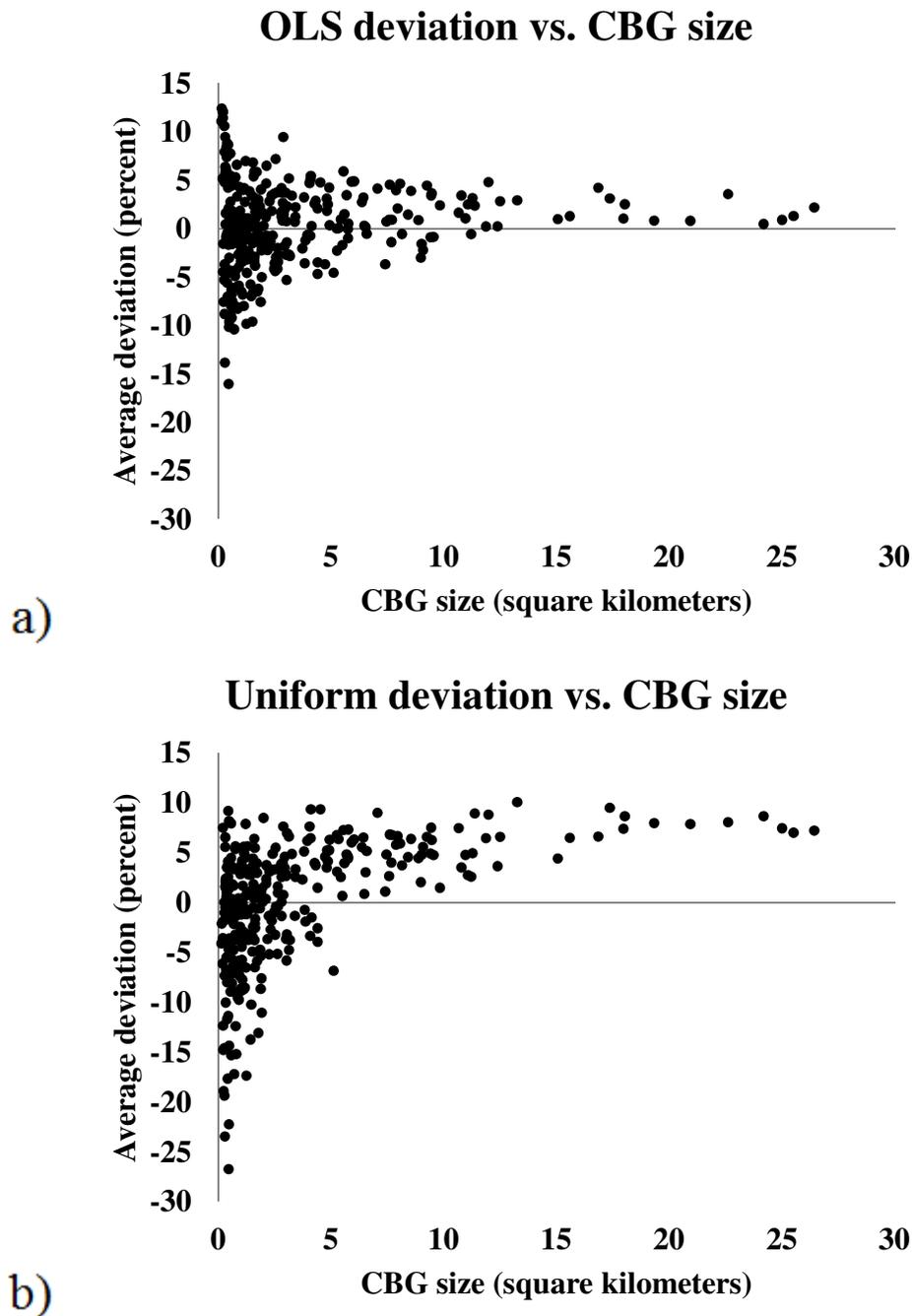
DISCUSSION

Assessment of methods and limitations

The results presented above illustrate the differences in the amount of lawn in PIE produced by the Polsky et al. (2012) direct mapping method and the two inferential methods—OLS prediction and uniform prediction. The results also provide answers to our two research questions: 1) how well do the social driver variables predict lawns at the Census block group (CBG)-level when compared to a lawn map derived using direct mapping methods, and 2) how well does a simpler, less computationally intensive baseline uniform prediction method predict lawns at the CBG-level when compared to a lawn map derived using direct mapping methods? When we compare the results of the two inferential methods to the reference map, we find that on average, the OLS prediction method performed better than the uniform prediction method in terms of both total disagreement and quantity disagreement across the 1000 Monte Carlo simulations.

Globally, both inferential methods tend to over-predict the quantity of lawn in the study area. In the case of the uniform prediction method, the systematic over-prediction of lawn quantity can be attributed to the logic with which Census block group boundaries are drawn, which is an example of the Modifiable Areal Unit Problem (MAUP) (Openshaw 1984). The US Census Bureau attempts to draw CBG boundaries to contain roughly 600 to 3,000 people, with an optimal size of 1,500 people (US Census Bureau 2000). Their ultimate goal is to draw the CBG boundaries with as close to equivalent populations as possible. As a result, dense clusters of people are collected into smaller CBGs, while sparse populations are drawn into larger CBGs, leading to high population densities in small CBGs and low population densities in large CBGs. The uniform prediction method assigns each CBG the same percent lawn value, so it over-predicts lawn in large CBGs and under-predicts lawn in small CBGs, but overall, the uniform method predicts more total lawn than the reference lawn because the uniform method applies the sample average taken from a random sample of CBGs to all other CGBs, many of which are large (Figure 8b). The OLS prediction method also globally over-predicts the quantity of lawn in the study area, but the OLS prediction method does this less severely than the uniform prediction method (Figure 8a).

Figure 8a-b. Scatterplots depicting the relationship between CBG area and percent lawn deviation between the Polsky et al. (2012) reference map compared to (a) an OLS prediction map and (b) a uniform prediction map. Each point is a CBG. Both plots show how the prediction is systematically larger than the reference information for CBGs that are larger than 10 square kilometers, while this bias is more severe for the uniform prediction.



Transferability of methods

Our results suggest that, as expected, a lawn area estimation that includes predictor variables leads to less total disagreement and less quantity disagreement, but greater allocation disagreement, than the uniform prediction method. Although we specified twelve independent variables, the only variable that exhibited statistical significance across the 1000 Monte Carlo simulations was population density. This is not to suggest that other variables should not be included in studies examining rapid mapping approaches for lawn grass—rather, this is reflective of our study area, which is generally characterized by low population and infrastructure densities. It is important for practitioners to consider how such relationships may differ (or remain the same) for their individual case studies. Further, practitioners seeking to map broader turf grass categories – i.e. other than lawns – may find different results, as certain correlations such as those between residential lawns and population may be weaker when vegetation found on surfaces other than residential properties is included.

Our research covered one study area in northeastern Massachusetts; it is a step in updating the national-scale lawn estimation of Vinlove and Torla (1995) and Milesi et al. (2005; 2009) at a finer spatial resolution and with explanatory variables that predict lawns with less total disagreement and less quantity disagreement than a baseline uniform prediction method. The independent variable data used in this study are freely available at the CBG-level for the entire United States, so it is possible to test our methods in other study areas in the US, with the goal of creating a CBG-level lawn map for the US. This would require some direct mapping approaches to map the sample of lawns used as the dependent variable in the regressions, but would not require the time, money, personnel, and computational resources for a comprehensive, full-study area mapping effort. That said, the decision of which CBGs to map for the sample is key, and would require an intelligent sampling design.

More generally, the transferability of this inferential mapping method will be predicated on the purpose of a given study, the size of the study area, and the researchers' access to data, time, and resources. For example, broad-scope studies examining nitrogen balance in urban environments may be able to leverage this approach, pending the accuracy and scalar requirements needed. Conversely, our results suggest that studies that require high levels of quantity and allocation accuracy will still need to rely on the more expensive, direct-mapping techniques.

Future directions

The findings in this study offer important opportunities for future research. In this study, we used a simple random design to choose the sample of CBGs that calibrated the predictive models. In the future, we may consider a different sample design such as weighted or stratified sampling, so we can account for the fact that CBG size is inversely related to lawn density. Regardless of sample design, it would also be interesting to conduct a sensitivity analysis to explore the impact of how adjusting the sample size influences the accuracy of the model predictions. Additionally, it would be interesting to examine whether the disagreement results change if CBG size is included as an independent variable in the OLS prediction models, or if CBG size is used as a replacement for the population density variable. Specifying the regression

models to include additional independent variables relating to land use policy (e.g. development standards in local zoning codes) or examining the interactions between variables could provide strong insights into the social dynamics behind lawn landscape management. Incorporating information on spatial dependence in the prediction of lawns to enable predictive approaches such as Geographically Weighted Regression, which models the regression relationships between dependent and independent variables locally (Brunsdon et al. 1998; Ogneva-Himmelberger et al. 2009; Gao and Li 2011; O’Loughlin and Witmer 2011; Tu 2011; Troy et al. 2012) or regression kriging, which performs kriging on the spatially autocorrelated residuals of an OLS model and combines the kriged residual surface with the OLS-predicted values (Hengl 2007; Hengl et al. 2007), may be fruitful avenues for future research. Once maps of lawns are produced for broader-scopes than are currently available, more in-depth understandings of the tradeoffs between the positive and negative effects of lawns can be produced, and how these effects vary over space and context can be better understood.

CONCLUSIONS

In this paper, we assess the performance of two inferential methods for predicting the quantity and spatial allocation of residential lawns at the CBG-level in relation to a reference map derived using OBIA mapping methods. The first inferential method uses OLS prediction calibrated with social driver variables to create the lawn predictions. The second inferential method predicts lawns uniformly across the study area, and does not contain information on the social driver variables. Our results indicate that the OLS prediction method produces less total disagreement and less quantity disagreement than the uniform prediction method. Our results provide support for the usefulness of one of the theoretically-informed social driver variables—population density—as a predictor of lawn in this study area, and suggest that the OLS prediction method is a reasonable alternative to the more costly, time-consuming, and resource intensive direct mapping approaches when it comes to mapping lawns. Our OLS prediction method also produces a finer-resolution lawn land-cover product than has been previously produced, and the freely available independent variable data encourages the application of this method to other locations in the United States.

LITERATURE CITED

- Alig, R.J., J.D. Kline and M. Lichtenstein. 2004. Urbanization on the US landscape: Looking ahead in the 21st century. *Landscape and Urban Planning* 69:219-234.
- Boone, C.G., M.L. Cadenasso, J.M. Grove, K. Schwarz and G.L. Buckley. 2010. Landscape, vegetation characteristics, and group identity in an urban and suburban watershed: Why the 60s matter. *Urban Ecosystems* 13:255-271.
- Brabec, E., S. Schulte and P. Richards. 2002. Impervious surfaces and water quality: A review of current literature and its implications for watershed planning. *Journal of Planning Literature* 16(4):499-514.

- Brunsdon, C., S. Fotheringham and M. Charlton. 1998. Geographically weighted regression—Modeling spatial non-stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician)* 47(3):431-443.
- Carrico, A.R., J. Fraser and J.T. Bazuin. 2012. Green with envy: Psychological and social predictors of lawn fertilizer application. *Environment and Behavior* DOI: 10.1177/0013916511434637.
- Cook, E.M., S.J. Hall and K.L. Larson. 2011. Residential landscapes as social-ecological systems: A synthesis of multi-scalar interactions between people and their home environment. *Urban Ecosystems* DOI: 10.1007/s11252-011-0197-0.
- Elvidge, C.D., C. Milesi, J.B. Dietz, B.T Tuttle, P.C Sutton, R.R. Nemani and J.E. Vogelmann. 2004. U.S. constructed area approaches the size of Ohio. *EOS, Transactions, American Geophysical Union* 85(24):233-240.
- Farag, F.A., C.M.U. Neale, R.K. Kjelgren, and J. Endter-Wada. 2011. Quantifying urban landscape water conservation potential using high resolution remote sensing and GIS. *Photogrammetric Engineering & Remote Sensing* 77(11):1113-1122.
- Gao, J. and S. Li. 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography* 31:292-302.
- Giner, N.M and J. Rogan. 2012. A comparison of Landsat ETM+ and high-resolution aerial orthophotos to map urban/suburban forest cover in Massachusetts, USA. *Remote Sensing Letters* 3(8):667-676.
- Giner, N.M., C. Polsky, R.G. Pontius, Jr. and D.M. Runfola. 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.
- Gober, P., A. Brazel, R. Quay, S. Myint, S. Grossman-Clarke, A. Miller and S. Rossi. 2009. Using watered landscapes to manipulate urban heat island effects: How much water will it take to cool Phoenix? *Journal of the American Planning Association* 76(1):109-121.
- Groffman, P. M., N.L. Law, K.T. Belt, L.E. Band and G.T. Fisher. 2004. Nitrogen fluxes and retention in urban watershed ecosystems. *Ecosystems* 7:393-403.
- Grove, J.M., A.R. Troy, J.P.M. O'Neil-Dunne, J.W.R. Burch, M.L. Cadenasso and S.T.A. Pickett. 2006a. Characterization of households and its implications for the vegetation of urban ecosystems. *Ecosystems* 9(4):578-597.

- Grove, J.M., M.L. Cadenasso, W.R. Burch, Jr., S.T.A. Pickett, K. Schwarz, J. O'Neil-Dunne and M. Wilson. 2006b. Data and methods comparing social structure and vegetation structure of urban neighborhoods in Baltimore, Maryland. *Society and Natural Resources* 19:117-136.
- Harris, E.M., D.G. Martin, C. Polsky, L. Denhardt and A. Nehring. 2012a. Beyond 'lawn people': The role of emotions in suburban yard management practices. *The Professional Geographer* DOI: 10.1080/00330124.2012.681586.
- Harris, E.M., C. Polsky, K.L. Larson, R. Garvoille, D.G. Martin, J. Brumand and L. Ogden. 2012b. Heterogeneity in residential yard care: Evidence from Boston, Miami, Phoenix. *Human Ecology* DOI: 10.1007/s10745-012-9514-3.
- Hanlon, B., J. Rennie Short and T.J. Vicino. 2010. *Cities and suburbs*. Oxford, UK: Routledge.
- Hengl, T. 2007. *A practical guide to geostatistical mapping of environmental variables*. JRC Scientific and Technical Reports. European Commission, Joint Research Center, Institute for Environment and Sustainability. Luxembourg: Office for Official Publications of the European Communities.
- Hengl, T., G.B.M. Heuvelink and D.G. Rossiter. 2007. About regression-kriging: From equations to case studies. *Computers & Geosciences* 33(10):1301-1315.
- Irani, F.W. and M.F. Galvin. 2003. Strategic urban forests assessment: Baltimore, Maryland. In: *Proceedings from the American Society of Photogrammetry and Remote Sensing 2003 Conference, American Society of Photogrammetry and Remote Sensing, 5 – 9 May 2003, Anchorage, Alaska, USA*.
- Jackson, K.T. 1985. *Crabgrass frontier: The suburbanization of the United States*. New York, NY: Oxford University Press, Inc.
- Jensen, J.R. and D.C. Cowen. 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering & Remote Sensing* 65(5):611-622.
- Larson, K.L., D. Casagrande, S.L. Harlan and S.T. Yabiku. 2009. Residents' yard choices and rationales in a desert city: Social priorities, ecological impacts, and decision tradeoffs. *Environmental Management* 44:921-937.
- Law, N.L., L.E. Band and J.M. Grove. 2004. Nitrogen input from residential lawn care practices in suburban watersheds in Baltimore County, MD. *Journal of Environmental Planning and Management* 47(5):737-755.

- MassGIS, 2011. Office of Geographic and Environmental Information (MassGIS), Commonwealth of Massachusetts, Information Technology Division. Available online at: <http://www.mass.gov/mgis/>. Accessed December 16, 2011.
- Mathieu, R., C. Freeman and J. Aryal. 2007. Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. *Landscape and Urban Planning* 81:179-192.
- Mennis, J. 2006. Socioeconomic-vegetation relationships in urban, residential land: The case of Denver, Colorado. *Photogrammetric Engineering & Remote Sensing* 72(8):911-921.
- Milesi, C., S.W. Running, C.D. Elvidge, J.B. Dietz, B.T. Tuttle and R.R. Nemani. 2005. Mapping and modeling the biogeochemical cycling of turf grasses in the United States. *Environmental Management* 36(3):426-438.
- Milesi, C., C.D. Elvidge and R.R. Nemani. 2009. Assessing the extent of urban irrigated areas in the United States. In *Global mapping of irrigated and rainfed cropland areas using remote sensing*. P.S. Thenkabail, J.G. Lyon, H. Turrall and C.M. Biradar (eds.), pp. 217-236. London, UK: Taylor & Francis.
- O'Loughlin, J. and F.D.W. Witmer. 2011. The localized geographies of violence in the North Caucasus of Russia, 1999-2007. *Annals of the Association of American Geographers* 101(1):178-201.
- Ogneva-Himmelberger, Y., H. Pearsall and R. Rakshit. 2009. Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts. *Applied Geography* 29:478-487.
- Openshaw, S. 1984. The modifiable areal unit problem. *Concepts and Techniques in Modern Geography (CATMOG)* 38:1-41.
- Otto, B., K. Ransel, J. Todd, D. Lovaas, H. Stutzman and J. Bailey. 2002. *Paving our way to water shortages: How sprawl aggravates the effects of drought*. American Rivers: The Natural Resources Defense Council and Smart Growth America. Online: <http://www.smartgrowthamerica.org/documents/DroughtSprawlReport09.pdf>.
- Polsky, C., R.G. Pontius, Jr., N.M. Giner, A. Decatur, D. Runfola and R. Rakshit. 2012. *HERO object-based lawn mapping exploration of suburbia: Rationale, methods and results for the NSF Plum Island Ecosystems Long-Term Ecological Research Site*. George Perkins Marsh Institute Working Paper No. 2012-24. Worcester, MA: Clark University.
- Pontius Jr., R.G., O. Thontteh and H. Chen. 2008. Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics* 15(2):111-142.

- Pontius, Jr., R.G. and M. Millones. 2011. Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32(15):4407-4429.
- Ratick, S. and G. Schwarz. 2009. Monte carlo simulation. In *The International Encyclopedia of Human Geography: Volume 1*. R. Kitchin and N. Thrift (eds.), pp. 175-184. Amsterdam: Elsevier, Ltd.
- Robbins, P. A. Polderman and T. Birkenholtz. 2001. Lawns and toxins: An ecology of the city. *Cities* 18(6):369-380.
- Robbins, P. and T. Birkenholtz. 2003. Turf grass revolution: Measuring the expansion of the American lawn. *Land Use Policy* 20:181-194.
- Robbins, P. and J. Sharp, 2003. The lawn-chemical economy and its discontents. *Antipode* 35(5):955-979.
- Robbins, P. 2007. *Lawn people: How grasses, weeds, and chemicals make us who we are*. Philadelphia, PA: Temple University Press.
- Roy Chowdhury, R., K. Larson, J.M. Grove, C. Polsky, E. Cook, J. Onsted and L. Ogden. 2011. A multi-scalar approach to theorizing socio-ecological dynamics of urban residential landscapes. *Cities and the Environment (CATE)* 4(1):1-19.
- Runfola, D.M. 2012. *Human-environment interactions across space and time: Examining lawns and land change in the United States*. Unpublished PhD Dissertation, Geography, Clark University, Worcester, Massachusetts.
- Runfola, D.M., C. Polsky, C. Nicolson, N.M. Giner, R.G. Pontius, Jr. and A. Decatur. 2013a. A growing concern? Examining the influence of lawn size on residential water use in suburban Boston, MA, USA. *Landscape and Urban Planning* 119: 113-123.
- Runfola, D.M., C. Polsky, N.M. Giner, R.G. Pontius, Jr. and C. Nicolson. 2013b. Future suburban development and the environmental implications of lawns: A case study in New England, USA. In *Cities and Nature: Modeling of land-use and ecological dynamics*. D. Malkinson, D. Czamanski, and I. Benenson (eds.), pp. 119-141. Heidelberg, Germany: Springer-Verlag.
- Spronken-Smith, R.A., T.R. Oke and W.P. Lowry. 2000. Advection and the surface energy balance across an irrigated urban park. *International Journal of Climatology* 20(9):1033-1047.
- Steinberg, T. 2006. *American green: The obsessive quest for the perfect lawn*. New York, NY: W.W. Norton & Company, Inc.

- Townsend-Small, A. and C.I. Czimczik. 2010. Carbon sequestration and greenhouse gas emissions in urban turf. *Geophysical Research Letters* 37(L06707): DOI: 10.1029/2009GL041675.
- Troy, A.R., J.M. Grove, J.P.M. O'Neil-Dunne, M. Cadenasso and S.T.A. Pickett. 2007. Predicting patterns of vegetation and opportunities for greening on private urban lands. *Environmental Management* 40(3):394-412.
- Troy, A.R., J.M. Grove and J. O'Neil-Dunne. 2012. The relationship between tree canopy and crime rates across an urban-rural gradient in the greater Baltimore region. *Landscape and Urban Planning* 106(3):262-270.
- Tu, J. 2011. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography* 31:376-392.
- US Census Bureau, 2000. US Census Bureau: Cartographic Boundary Files. <http://www.census.gov/geo/www/cob/bg2000.html>.
- Vinlove, F.K. and R.F. Torla. 1995. Comparative estimations of U.S. home lawn area. *Journal of Turfgrass Management* 1:83-97.
- Zartarian, V.G., H. Ozkaynak, J.M. Burke, M.J. Zufnall, M.L. Rigas and E.J. Furtaw, Jr. 2000. A modeling framework for estimating children's residential exposure and dose to chlorpyrifos via dermal residue contact and nondietary ingestion. *Environmental Health Perspectives* 108(6):505-514.
- Zhou, W. and A.R. Troy. 2008. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing* 29(11):3119-3135.
- Zhou, W. A.R. Troy and J.M. Grove. 2008a. Object-based land cover classification and change analysis in the Baltimore Metropolitan area using multitemporal high resolution remote sensing data. *Sensors* 8:1613-1636.
- Zhou, W. A.R. Troy and J.M. Grove. 2008b. Modeling residential lawn fertilization practices: Integrating high resolution remote sensing with socioeconomic data. *Environmental Management* 41:742-752.
- Zhou, W. A.R. Troy, J.M. Grove and J.C. Jenkins. 2009. Can money buy green? Demographic and socioeconomic predictors of lawn-care expenditures and lawn greenness in urban residential areas. *Society and Natural Resources* 22:744-760.