Spatial extent and habitat context influence the nature and strength of relationships between urbanization measures

Brian G. Tavernia*, J. Michael Reed

Biology Department, Tufts University, 163 Packard Avenue, Medford, MA 02155, USA

1. Introduction

Urbanization can be described as an increase in human population density coupled with increased energy use and extensive alteration of the landscape (McDonnell and Pickett, 1990). The variety of measures used to quantify urban environments can be divided into three categories: demographic (e.g. population density), physical (e.g. road density), and landscape (e.g. mean patch size) (Hahs and McDonnell, 2006). Studies quantifying the effects of urbanization on species and ecological processes differ in urbanization measures selected (e.g. McIntyre, 2000; Marzluff et al., 2001; Theobald, 2004). If urbanization metrics are poorly or inconsistently correlated multiple measures would be required to sufficiently explore the relationships between urbanization and biodiversity, and integrating results across studies of ecological impacts of urbanization could be difficult. Landscape ecologists have used multivariate statistics to group landscape metrics into statistically independent subsets (e.g. Riitters et al., 1995; Cain et al., 1997); researchers have used these groups to describe the impact of urbanization on landscape structure and composition and have tried to link these patterns to socioeconomic and ecological processes (Luck and Wu, 2002; Seto and Fragiakas, 2005; Yu and Ng, 2007). Recently, Hahs and McDonnell (2006) used principal components analysis (PCA) to identify independent groups of urban measures. From each group, they selected the single variable that had the highest principal component loading, considering each to be representative of their respective group, and used these measures to describe the urban environment of Melbourne, Australia. However, they cautioned that the applicability of these results to other landscapes may be limited and that investigations of factors influencing the relationships between urbanization measures were needed.

Different methods of aggregating spatial data are known to influence the output and correlations between landscape metrics (Turner et al., 1989; Jelinski and Wu, 1996; Cain et al., 1997; Wu et al., 2002). A particular set of potential problems arising from these observations is known as the Ecological Fallacy Problem (EFP) (Robinson, 1950; Openshaw, 1984; Cao and Lam, 1997). The EFP encompass three issues that can lead to an inability to draw valid inferences across scales and study regions: individualistic, ecological, and cross-level (hereafter, cross-region) fallacies. Individualistic fallacies arise from the extrapolation of patterns from smaller to larger spatial scales, while ecological fallacies are the inverse, interpolating to smaller spatial scales from larger ones. An example of an interpolation error is assuming that human population density,
which can correlate positively with endangered species richness at large spatial scales (Dobson et al., 2001), will show the same pattern at smaller spatial scales. This is a fallacy because anthropogenic drivers, including urbanization, are causing species loss at small spatial scales (McKinney, 2002). The cross-region fallacy states that relationships developed from one study region may not necessarily apply to different regions. For example, Cain et al. (1997) found that correlations between landscape metrics can differ between geographic regions.

Our overall aim was to examine the effect of changing spatial extent and habitat context on correlations between urbanization measures. Studies of the influence of landscape features on ecological processes frequently assess landscape features in a buffer around a point or site (e.g. Hagan and Meehan, 2002) or in a single cell of a predefined grid (e.g. Naugle et al., 2001); the size of these buffers or cells is the spatial extent of individual landscapes (Turner et al., 1989). Efforts to assess the impacts of urbanization often range greatly in the extent considered (e.g. White and Greer, 2006; Clark et al., 2007). If correlations between urbanization measures change with scale as suggested by the previously outlined fallacies, studies conducted at different extents may need to use different suites of metrics to characterize the urban environment. Alberti et al. (2001) examined bivariate correlations between land-cover and land-use pattern metrics at two extents and found stronger average correlations at the larger extent (5 km² vs. 1 km² cells). We extend their work by looking at a larger range of spatial extents and using multivariate statistics to identify independent subsets of predictor variables at each extent.

A second potential cause of inconsistency in correlations between urbanization measures is the effect of habitat context on urban development. Studies of the ecological implications of urbanization have been conducted in multiple habitat types: deserts (Hostetler and McIntyre, 2001), grasslands (Collinge et al., 2003), and salt marshes (DeLuca et al., 2004). Following from the cross-region fallacy and given that previous studies have found covariation between metrics depends on landscape context (Cain et al., 1997), researchers working in different habitat contexts may need different subsets of urbanization metrics to characterize landscapes.

The specific goals of our study were to evaluate (1) how the nature and strength of correlations between urbanization measures changed with spatial scale and (2) the effects of habitat context on the nature and strength of these relationships. We distinguish “nature” (multivariate collinearity) and “strength” (bivariate correlation coefficient) (cf. Openshaw, 1984) of correlations to evaluate two related questions: (1) What subsets of interrelated variables are identifiable at different extents and in different habitats (nature)? and (2) Are variables within these subsets correlated to a sufficient degree to be used interchangeably (strength)? We conducted this research in Massachusetts (U.S.A.), a state whose eastern portion is dominated by the Boston metropolitan area. Urbanization is an ongoing process in this region, as evidenced by a 21.7% increase in the area developed from 1982 to 1992 (Noss and Peters, 1995; see also Porter and Hill, 1998).

2. Materials and methods

2.1. Data collection

We generated a regular grid of points across Massachusetts, extending from 41°21' to 42°51' N and from 70°1' to 73°28' W. Grid points were spaced every 4 km using Hawth’s tools, a downloadable extension for GIS ArcMap 9.2 (www.spatial ecology.com). For each point, we created circular buffers at five extents: 100 m, 250 m, 500 m, 1000 m, and 2000 m-radius; because of the grid spacing, buffers from adjacent points did not overlap. We kept for analysis only those points for which all buffers were completely contained within Massachusetts (n = 1105). Within each buffer, we determined the value of eight urbanization measures: (1) population density (number of people/ha), (2) agriculture cover (ha); area of cropland and pasture land), (3) forest cover (ha), (4) wetland cover (ha; area of freshwater wetlands), (5) dense residential cover (ha; total area of high density residences, including multifamily and <0.2 ha plots), (6) impervious surface cover (ha; area of paved roads plus commercial and industrial cover), (7) road length (km; total length of secondary, primary, and highway roads), and (8) greenspace cover (ha; combined areas of open land, cropland, urban open land, pasture, forest, and woody perennial). These variables cover the spectrum of urbanization measures outlined by Hahs and McDonnell (2006): demographic (population density), physical (road density), and landscape metrics (coverage of various land use types).

We determined the value for each metric at each extent using GIS data layers provided by the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA) (http://www.mass.gov/mgis/). The census information used to calculate population density was derived by the EOEEA from TIGER data sets made available by the U.S. Census Bureau. This information is available at a coarse scale in census blocks. To determine the population density of a buffer, we first calculated the population density of each census block and the area of the block that intersected the buffer. We then multiplied the intersected area of each census block by the population density to determine the number of people contributed by the census block to the buffer. Finally, the number of people was summed across census blocks and divided by buffer area. A land use layer provided by the EOEEA was used to determine the area of the land use types; this layer is based on photo-interpretation of 1:25,000 aerial photographs. Road length was determined using a GIS layer representing the major and minor roads present in Massachusetts, which is maintained jointly by the EOEEA and the Executive Office of Transportation. Road surface area, used in calculating impervious surface area, was determined by multiplying road length by average width: local and secondary roads 6.7 m wide, primary roads 11.6 m, and highway 23.8 m (Massachusetts Highway Design Manual, 1997).

The influence of habitat context on correlations between urbanization metrics was determined by isolating polygons of salt marsh, forest, and freshwater wetland areas from the land use layer. Using Hawth’s tools, we randomly generated 100 points within each habitat type, separated by at least 2 km. For the forest and freshwater wetland categories, we retained only those points greater than 1 km from the state border, generated 1 km-radius buffers, and calculated the urban measures listed above. Given the natural proximity to coastline, this procedure was not possible for salt marsh points. For this habitat context, we generated a 1 km-radius buffer around salt marsh points and determined urbanization measures on a per land area basis (e.g. percent greenspace, road density).

2.2. Data analysis

We analyzed data using SAS v 9.1.3 (SAS Institute, 2004). First, we evaluated the multivariate nature of the relationships among urbanization measures using PCA (Princomp procedure). This was done to identify subsets of related urbanization metrics with each subset describing a different aspect of the surrounding landscape. For this analysis we were interested in the number of significant principal components (eigenvector > 1.0), and in the variables that contributed significantly to each principal component (eigenvector > 0.3) (Hahs and McDonnell, 2006). A significant component revealed groups of measures that were correlated in
each treatment. If the nature of these correlations were stable, then urbanization measures should be associated with the same factors regardless of spatial extent or habitat context.

Finally, we determined bivariate correlations between urbanization measures within each spatial scale and within each habitat context for all survey points. Results from the PCA suggested that a single group of landscape measures characterized urban environments (see Section 3), so we focused on these variables in assessing the strength of correlations between urbanization measures. We arbitrarily selected a Pearson’s correlation coefficient of 0.7 ($r^2 = 0.49$) as the minimum relationship to consider a correlation to be strong, i.e., the two urban measures might be considered as proxies. This is a conservative approach as previous studies have considered a correlation coefficient of 0.8 as an indicator that landscape measures are interchangeable (Hargis et al., 1998). Levene’s test indicated that habitat context, but not extent, groups displayed unequal variances. Consequently, we used a one-way ANOVA and Welch’s ANOVA to determine if spatial extent and habitat context, respectively, were significantly associated with the strength of bivariate correlations (GLM procedure with a post hoc Tukey studentized range test).

3. Results

3.1. Extent

PCA results strongly suggested that the relationships between urbanization measures did not change with extent (Table 1). At each extent, the gradient of urban conditions was captured by principal component (PC) 1. Specifically, PC 1 identified a continuum from “undeveloped” sites dominated by forest and greenspace to “developed” regions characterized by urban features (i.e., impervious surface area, road length, and dense residential area). PC 2 and PC 3 were not directly related to urbanization, but they provide indications of other anthropogenic impacts on landscape composition. PC 2 was associated with conditions ranging from sites with high agriculture and wetland habitat cover to those with high forest cover. Finally, PC 3 represented a gradient from agricultural sites to those with wetland habitat. The first 3 principal components explained 72–89% of the variance, and the amount of variance explained by each principal component rose monotonically with spatial scale (Table 1). Over a third of the correlations between urbanization measures loading onto PC 1 were strong (Table 2; 26/75 = 34.7% with $|r| ≥ 0.70$) with only a few being considered weak (5/75 = 6.7% with $|r| ≤ 0.30$). The proportion of strong correlations increased at larger spatial scales, with only 1 of 15 (6.7%) strong correlations at the 100 m spatial scale, but 13 of 15 (86.7%) at the 2000 m scale (Table 2). The average strength of these correlations increased significantly at larger spatial scales (Fig. 1; $F = 14.36$, d.f. = 4, 70, $p < 0.01$; post hoc significance at $\alpha = 0.05$).

3.2. Habitat context

Although spatial scale (extent) did not have an effect on the nature of the correlations between urbanization measures, PCA results suggest that habitat context did influence the nature of correlations (Table 3). As with the spatial scale analysis, for all habitat contexts, principal component (PC) 1 was associated with a gradient from “undeveloped” regions dominated by forest and greenspace to “developed” regions characterized by urban features (i.e., impervious surface area, road length, and dense residential area). However, within the salt marsh context, PC 2 indicated that a subset of these urbanization measures shift from being negatively to being positively correlated. In particular, salt marsh PC 2 did not distinguish between some measures of urbanization and

Table 1
Results of a principal component analysis between urbanization measures at five landscape extents in Massachusetts, USA. Urbanization measures were calculated for 1105 non-overlapping areas with mid-points set on a 4-km grid throughout the state. Principal components with an eigenvalue ≥ 1.0 were retained for comparisons. Variables with eigenvector values ≥ 0.30 or ≤ −0.30 (bolded) were considered strongly associated with a principal component.

<table>
<thead>
<tr>
<th>Extent</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Wetland</th>
<th>Dense residential</th>
<th>Impervious surface</th>
<th>Road length</th>
<th>Greenspace</th>
<th>Cumulative variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 km</td>
<td>0.36</td>
<td>−0.09</td>
<td>−0.40</td>
<td>0.00</td>
<td>0.41</td>
<td>0.32</td>
<td>0.46</td>
<td>0.47</td>
</tr>
<tr>
<td>2 km</td>
<td>0.37</td>
<td>−0.10</td>
<td>−0.39</td>
<td>0.02</td>
<td>0.41</td>
<td>0.40</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>500 m</td>
<td>0.35</td>
<td>−0.07</td>
<td>−0.41</td>
<td>0.02</td>
<td>0.41</td>
<td>0.36</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>250 m</td>
<td>0.33</td>
<td>−0.07</td>
<td>−0.42</td>
<td>0.02</td>
<td>0.42</td>
<td>0.34</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>100 m</td>
<td>0.33</td>
<td>−0.04</td>
<td>−0.43</td>
<td>0.02</td>
<td>0.41</td>
<td>0.32</td>
<td>0.46</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Fig. 1. Relationship between the strength of bivariate correlations of urbanization measures and landscape extent in Massachusetts, USA. Urbanization measures were calculated at 1105 points that were set on a 4-km grid throughout the state. Each bar represents the mean of 15 bivariate correlations within each extent. Error bars represent one standard error of the mean. Extents labeled with different letters were significantly different ($p < 0.01$).
rural or undeveloped landscape metrics: forest area, greenspace, impervious surface cover, and road length loaded positively, as did agriculture. This axis appears to identify areas with no systematic pattern of human land use, mixing industrial and agriculture areas fragmented by roads. In forest and freshwater wetland contexts, in contrast, PC 2 is a continuum from agricultural areas to forest. While not including urbanization measures, the third principal component (PC 3) offered additional evidence that characteristics of the surrounding landscapes can differ between habitat contexts. Within a forest context, PC 3 represented a continuum from landscapes high in wetland cover to landscapes without wetlands. Within freshwater wetland and salt marsh contexts, in contrast, PC 3 represented a continuum from landscapes high in agriculture cover to areas with wetlands in the surrounding landscape. The first 3 principal components explained 76–89% of the variance in these habitat contexts. With regards to strength, a greater number of bivariate correlations were strong (|r| > 0.70) in forest and freshwater wetland contexts (10 of 15 correlations; 66.6%, for both) than in a salt marsh context (3 of 15; 20.0%). Urbanization measures in forest and wetland habitat matrices showed significantly higher average correlations than did measures in salt marsh habitat matrices, but they did not differ significantly from one another (Fig. 2; F = 11.65, d.f. = 2, 26.2, p < 0.01; post hoc significance at α = 0.05).

4. Discussion

Given the myriad of available landscape measures, urban ecologists face the important task of identifying a subset of statistically independent, biologically relevant metrics that will allow quantification of the urban environment (e.g., Hahs and McDonnell, 2006). This task, however, is not trivial, and there might not be a small number of variables that can capture urbanization across studies. In our study area, which covered ~27,000 km² and included 1 major and 2 lesser urban centers, we found that several factors that commonly differ among studies, landscape extent and habitat context, influence the nature and strength of correlations between urbanization measures. As a consequence, the appropriate set of metrics for quantifying the urban environment is likely to differ between studies, reinforcing the cautionary words of Hahs and McDonnell (2006). These results suggest (1) that integrating results from previous studies will be difficult, as unaccounted-for differences in the quantification of the urban environment may be driving documented patterns, and (2) that future studies of the relationship between urbanization and species and ecological processes need to fully characterize the landscapes surrounding their sites.

4.1. Extent

Following from the EFP (Cao and Lam, 1997), potential scale constraints on extrapolation and interpolation suggest that the nature
and strength of relationships between urbanization measures may not translate across extents. In our study area, we found no evidence that the nature of correlations between measures changed across the range of extents we considered. In contrast, Cain et al. (1997) investigated the correlation of landscape metrics at two extents (18,000 km² rectangles, sub-watersheds) in the Chesapeake Bay watershed, and found that the nature of correlations depended on the extent considered. This study incorporated a wide variety of structural landscape metrics, and not all of these metrics respond the same way to changing extents (Wu et al., 2002; Wu, 2004). Most of the metrics we considered were coverage metrics and consequently they might be expected to respond in the same fashion to changing extent size (Wu, 2004). Our selection of urbanization measures may have contributed to the consistent nature of the relationships between urbanization measures in our study.

Within our study region, the strength of correlations between urbanization measures differed greatly ($r = 0.17-0.93$), a finding consistent with earlier studies (Medley et al., 1995; Alberti et al., 2001). In contrast to the nature of the relationships among urbanization measures, the strength of correlations between measures increased significantly with spatial extent, as did the proportion of strong correlations. This pattern is consistent with the findings of Alberti et al. (2001) in the Seattle metropolitan area. To some degree, the increasing strength of correlation may be a function of the varied resolutions (or grain size) of the input data sets. For example, census data were available at relatively coarse resolution (i.e., census block) as compared to other urbanization measures. Consequently, measures of population density at smaller extents are more inaccurate, and this could contribute to lower correlations between population density and other urbanization measures. In some studies, once independent subsets of variables have been identified via PCA, researchers have suggested that a single variable may represent each group (principal component) (e.g. Riitters et al., 1995; Hahs and McDonnell, 2006). Based on the strength of correlations observed, this approach may only be feasible at larger extents as correlations become strong enough for variables to serve as proxies for one another. For example, at 2000 m in our study area, dense residential areas might serve as a proxy for increasing road length and impervious surface, while simultaneously representing decreased forest and greenspace cover ($|r| > 0.80$). However, it should be noted that the ability to serve as a proxy in a statistical sense does not suggest that these urban features will have similar biological effects. At smaller extents, indicators of urbanization appear to operate somewhat independently in the Massachusetts' landscape. Therefore, as has been recognized in other studies (Luck and Wu, 2002; Seto and Fragkias, 2005; Yu and Ng, 2007), it might be impossible to capture urbanization using one or a few metrics. Studies that rely on single metric estimates of the degree of urbanization could be missing important facets of urbanization that might have significant effects on biodiversity or ecological processes. For example, Marzluff (2001) reviewed studies examining the influence of urbanization on bird communities and found a wide range of results, with studies reporting decreasing, stable, and increasing species richness in urbanized areas. This mix of results might be due to a lack of consistent measures of urbanization across the various studies.

4.2. Habitat context

Understanding the influence of habitat context on the correlations between urbanization measures is important as studies examining the ecological impacts of urbanization cover a wide range of habitat types (Hostetler and McIntyre, 2001; Collinge et al., 2003; DeLuca et al., 2004). Even if studies in different habitats focus on the same extent, the potential for cross-region fallacies (Cao and Lam, 1997) suggests that one should look for differences in the nature and strength of the relationships between urbanization measures. Our results show that the habitat context for urbanization (from forest, to freshwater wetland, to salt marsh) impacts the nature and strength of correlations between urbanization measures. With regards to nature, the same urbanization gradient was identified in all habitat types (PC 1, Table 3). However, in the salt marsh habitat context, some urbanization measures were involved in characterizing a separate aspect of the surrounding landscape (PC 2, Table 3) in a manner that fundamentally changed their relationship to one another. Specifically, greenspace and forest were negatively correlated with dense residential cover and road density in urban areas but were positively related in other areas. Therefore, relative to forest and fresh water marsh contexts, the pool of measures that are exclusive to the urban gradient is reduced in the salt marsh context. This does not suggest that changes in these variables are not having ecological impacts across the urban gradient, but rather it underscores the importance of multivariate statistics in evaluating the ability of a given metric to represent urbanization. Even if studies in different habitat contexts use the same urbanization measures, our results indicate that comparing results across studies may be hindered by the shifting nature of the relationships between measures. Our findings reinforce those of earlier studies that found that the values of and correlation between landscape metrics depend upon landscape context (Cain et al., 1997; Hargis et al., 1998; Wu et al., 2002).

Habitat context also influenced the strength of correlations between urbanization measures in our study. Specifically, urbanization measures in forest and freshwater wetland landscapes were...
more strongly correlated than were urbanization measures in a salt marsh context. Simulations have suggested that unique histories of landscape disturbance may alter the strength of correlations between landscape metrics (Hargis et al., 1998); it could be that in our study region urbanization followed different trajectories in the habitat contexts we examined, leading to differences in the strength of correlations between urbanization measures. Regardless, the effectiveness of a given urbanization measure to serve as a proxy for other urban features was greater in forest and freshwater wetland contexts than in a salt marsh context. As previously discussed for small extents, this suggests that in a salt marsh context more metrics (and studies) may be necessary to characterize the full impact that urbanization may have upon ecological processes.

5. Conclusions

The complexities associated with changing relationships among urbanization metrics across extents and habitat contexts will likely increase the difficulty of distilling consistent patterns across studies. This will make it difficult to generate rules of thumb to aid the conservation of species and important ecological processes in an urbanizing world. To advance understanding of the effects of urbanization on biodiversity, we join other authors in suggesting that future studies: (1) identify specific measures of urbanization with hypothesized mechanisms linking those metrics to the species and ecological processes of interest, a step widely recognized as critical when considering relationships in ecology (Wiens et al., 1993; Huston, 2002; Li and Wu, 2004), but apparently not done in many studies (McIntyre, 2000; Marzluff et al., 2001); (2) use a suite of statistical techniques to fully characterize the nature and strength of the relationships among selected urban measures (Riitters et al., 1995; Cain et al., 1997; Hahs and McDonnell, 2006); (3) evaluate relationships at a variety of spatial scales based on the target species or ecological processes (Hostetler, 2001; Marzluff et al., 2001); (4) finally, explicitly report summary statistics (at least means and ranges of all variables) for the urban metrics, because these will be needed for valid comparisons and synthesis among studies.

Acknowledgements

We thank D. Des Rochers, B. Parmenter, P. Florance, and three anonymous reviewers whose comments improved an earlier version of this manuscript. Ideas for this paper were developed while doing related research supported by the Massachusetts Environmental Trust, Massachusetts Water Resources Research Center, and the Charles Blake Fund through the Nuttall Ornithological Club.

References


