

Identifying the Spatial Structure in Error Terms with Spatial Covariance Models: A Case Study on Urbanization Influence in Chaparral Bird Species

Diana Stralberg* and Shuming Bao†

* School of Natural Resources and Environment, the University of Michigan, Ann Arbor, MI 48109 USA.

† China Data Center, the University of Michigan, 1080 S. University, Ann Arbor, MI 48109-1106 USA.

Abstract

This study uses various spatial statistical methods to examine and model large- and small-scale spatial structure in bird abundance and urbanization. A set of chaparral-vegetated points across an urbanizing landscape in the Santa Monica Mountains of southern California was surveyed for birds in 1997 and mapped in a GIS. For each sample location, GIS landuse data were used to calculate surrounding urbanization proportion.

We first used semivariograms and correlograms to detect large-scale trends, and Moran's I statistic to test for small-scale spatial autocorrelation in bird abundance. Relationships between bird abundance and surrounding urbanization levels were then analyzed using ordinary least-squares (OLS) regression. Upon detection of spatial autocorrelation in model residuals, spatial covariance models were constructed to incorporate this small-scale spatial dependence statistically and obtain non-biased estimates of urbanization influence. As a conservative comparison, we also removed the large-scale spatial trends from bird abundance and regressed the detrended model residuals on surrounding urbanization proportion to examine the effects of local variations in urbanization on bird abundance.

Results of spatial covariance models, as well as detrended non-spatial models, indicate that chaparral bird populations are affected by changes in surrounding landscape composition, regardless of spatial location. Semivariogram and correlogram analysis provided further insight into the spatial structures of the bird populations examined.

I. INTRODUCTION

Increasingly, many large-scale ecological studies involve an explicit spatial component. With the growing availability and sophistication of Geographic Information Systems (GIS), as well as inexpensive Geographic Positioning Systems (GPS), ecologists can now georeference their sampling locations relatively easily and accurately. Oftentimes, however, this extra information is not used to its full potential. In ecology, GIS is now widely used as a map presentation and spatial analysis tool, facilitating the visualization and quantitative characterization of spatial variables and relationships among them. Commonly, however, spatial data are analyzed using simple statistical methods that ignore the underlying spatial structure and often improperly assume independence among sampling locations. This tendency is likely due to the statistical complexity and computation-intensiveness of many spatial statistical techniques, as well as their lack of widespread availability. But the omission of spatial dependence in statistical analyses can be problematic for the interpretation of results, and may result in biased model parameter estimates (Robertson 1987, Anselin 1989) and false detection of significant relationships (Legendre 1993).

Spatial dependence can occur at different scales, which

may be indicative of the process(es) driving the observed spatial patterns. Generally speaking, however, spatial autocorrelation in a data set indicates the presence of small-scale spatial dependence due to the interdependence of neighboring sites (Cressie 1993, MathSoft 1996). Small-scale spatial dependence is generally stochastic in origin and therefore difficult to predict. Conversely, large-scale or "structural" spatial dependence (Rossi et al. 1992) refers to a predictable pattern or spatial trend (Cressie 1993), which is often related to underlying physiographic features.

With respect to animal populations, large-scale spatial dependence in a species' distribution and abundance may reflect "continuous biogeographical dispersion routes" (Carroll and Pearson 1998), as well as underlying trends in the abiotic environment (e.g., topography, substrate, climate) that are generated slowly over geologic time (Brown 1984). Small-scale dependence tends instead to reflect patchiness in the biotic environment, stemming from stochastic disturbance events or complex species interactions that occur on an ecological time population dynamics, and density-dependent behaviors (e.g., territoriality, flocking) (Brown et al. 1995). These small-scale biological interactions may introduce temporal and/or spatial

1082-4006/99/0502-106\$5.00

©1999 The Association of Chinese Professionals in Geographic Information Systems (Abroad)

lags, decoupling a system from underlying geophysical factors (Wiens 1989) and making spatial variability difficult to predict.

As a result of these and other sources of spatial dependence, animal populations are likely to be non-uniformly distributed across a landscape, and sampling locations may not be equally independent of one another. Rather, sites that are closer in spatial proximity are likely to be more similar in species abundance (Brown 1984, Rossi et al. 1992). Thus, a fundamental statistical assumption is violated when populations exhibit significant spatial dependence. This leads to an over-specification of degrees of freedom, and an increased probability of rejecting a valid null hypothesis (type I or a error), meaning that results may too often be declared significant (Legendre 1993).

Ordinary least-squares (OLS) regression is typically used to model the relationships between ecological variables. If spatial autocorrelation is present in the error term, however, the estimated results will be biased and inconsistent (Anselin 1989). One must therefore control for spatial location when analyzing relationships between spatially distributed ecological variables. While the detection of spatial autocorrelation in regression model residuals may indicate the presence of an additional explanatory variable, this variable may not be easily measurable. Thus the spatial autocorrelation structure may instead be explicitly modeled, so that independence assumptions are met and significance tests are valid. Large-scale trends, which are easily misinterpreted as small-scale autocorrelation (Legendre and Fortin 1989), should first be investigated using semivariograms, correlograms, or other "structure functions" (Yaglom 1957).

This study attempts to address the above-named shortcomings in the analysis of spatial ecological data. Our primary goal was to assess the responses of selected songbird species to increasing levels of urbanization (residential development), taking into consideration the influence of small- and large-scale spatial dependence. Bird abundance data were sampled from the Santa Monica Mountains of southern California, which are facing urban encroachment on three sides by metropolitan Los Angeles. The study area contains a mosaic of interspersed suburban development and native chaparral vegetation and presents a unique opportunity to study the effects of urban encroachment at its intermediate stage, before the native habitat is severely fragmented.

Using non-spatial generalized linear regression models, the abundances of several chaparral- and woodland-associated bird species were previously found to

be negatively associated with levels of surrounding urbanization, while two urban-associated species exhibited positive associations with surrounding urbanization (Stralberg 1999). The validity of these results was called into question, however, when the underlying spatial patterns were examined.

Because the bird census locations used in this study were distributed over a large (~17,000 ha) geographic area, they were unlikely to constitute independent sampling points. Spatial distributions of both species appeared to follow an east-west urbanization gradient (Figures 1 and 2), suggesting the presence of large-scale spatial trends in their abundance patterns. Furthermore, small-scale interactions among sampling locations may have resulted in spatially autocorrelated abundances, which would invalidate the statistical tests for traditional non-spatial models.

In this paper, we introduce various methods for identifying spatial structure and estimating spatial models and apply them to the analysis of urbanization influence on two chaparral bird species. The analysis included four major steps: (1) identifying spatial trends and large- and small-scale spatial dependence in bird abundance and urbanization proportion, using semivariograms and correlograms; (2) estimating the impacts of urbanization levels on bird abundance using (non-spatial) ordinary least squares (OLS) regression, with diagnostics of the spatial dependence in model residuals; (3) re-estimating regression model of bird abundance on surrounding urbanization levels by removing large-scale spatial trends in bird abundance (regression of bird abundance on spatial coordinates), and investigating spatial structure in detrended model residuals; and (4) estimating spatial covariance models for bird abundance by incorporating small-scale spatial dependence, using spatial covariance structures suggested by semivariogram and correlogram analyses. The original conclusions of this study are modified in light of the new analyses, and potential mechanisms for the underlying spatial structure of the system are discussed.

II. METHODOLOGY AND EMPIRICAL EVIDENCE

Data Processing

Of the thirteen species exhibiting significant responses to some measure of surrounding urbanization (Stralberg 1999), two were selected for further analysis in this study, due to the relative robustness of their models, and the small departures from normality in their model residuals. The Wrentit (*Chamaea fasciata*) was negatively associated with surrounding urban-

ization proportion within a 500m radius, while the Northern Mockingbird (*Mimus polyglottos*) was positively associated with urbanization proportion within 2000m (Stralberg 1999).

Maximum per point abundance was estimated for several bird species at 107 locations (see Figure 1) by taking the maximum of two 100m fixed-radius point counts (Ralph et al. 1993) conducted during the summer of 1997. For each sample point, several measures of surrounding urbanization, including the proportion of the area within a circle of a specified radius contained in urban land uses, were calculated from GIS land use data in Arc/Info 7.1.1. Land use data were obtained from a 1:24000 Arc/Info polygon coverage, constructed from 1990 aerial photos by the Southern California Association of Governments. Buffer radii used to measure urbanization proportion ranged from 250m to 4000m. (See Stralberg 1999 for detailed field and GIS methods.) X-Y Coordinates and urbanization calculations for each sample location were exported to a database file and combined with bird abundance data for statistical analysis externally from the GIS.

S-Plus 4.5 (MathSoft 1997), S+SpatialStats (MathSoft 1996), and S-Plus for ArcView GIS (MathSoft 1998) were used for estimating semivariogram and correlogram. The Mixed Models Procedure in SAS 6.12 (SAS 1997, Littell et al. 1996) was used for estimating spatial covariance models.

Investigating Spatial Patterns in Bird Abundance

One class of tools for examining spatial structure detecting scales of influence in ecological data comes from the field of geostatistics and the generalized regional variable model (Matheron 1963). Despite a relatively long kriging tradition in the mining and geology fields (David 1977, Journel and Huijbretts 1978), semivariograms (Matheron 1963) have not been widely applied to ecological studies (but see Grieg-Smith 1983 and Taylor 1984). Recently, however, semivariograms have been used to describe spatial structure and interpolate the values of ecological variables over space (e.g., Villard and Maurer 1996, Robertson et al. 1997, Meisel and Turner 1998).

The standard semivariance function $g(h)$ of geostatistics theory (Matheron 1963) is estimated by one-half the average squared difference between points separated by a distance h , defined as:

$$\gamma(h) = \frac{\sum_{i=1}^{N(h)} \sum_{j=1}^{N(h)} (z_i - z_j)^2}{2 |N(h)|}, \quad (1)$$

where $N(h)$ is the set of all pairwise Euclidean dis-

tances $i - j = h$, $|N(h)|$ is the number of distinct pairs in $N(h)$, and z_i and z_j are data values at spatial locations i and j , respectively (MathSoft 1996). Directional semivariograms, where g is a function of the direction as well as magnitude of h , may also be examined if the spatial autocorrelation of a variable is not "isotropic" (Cressie 1993).

Inversely related to the semivariogram is the covariogram, which may be used to describe the spatial difference between sampling points at various "lag distances" or distance classes, and to detect spatial trends. The empirical covariance function, $C(h)$, represents the portion of the population variance that is explained by spatial autocorrelation at lag distance h . It is defined by:

$$C(h) = \text{cov}(z_i, z_j) = \frac{\sum_{i=1}^{|N(h)|} \sum_{j=1}^{|N(h)|} (z_i - \bar{z})(z_j - \bar{z})}{|N(h)|} \quad (2)$$

The term "correlogram", $r(h)$, is often defined as the standardized covariogram ($C(h)/s^2$), but may also be used to refer to any plot of autocorrelation versus spatial lag distance (or temporal lags, in the case of time-series analysis). A commonly used autocorrelation index, related to $r(h)$, is Moran's I statistic (Moran 1950), which generally ranges from -1 to 1. For a given distance class h , Moran's statistic, $I(h)$, is defined as:

$$I(h) = \frac{\left[n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z}) \right]}{\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (z_i - \bar{z})^2 \right]}, \quad (3)$$

where w_{ij} takes on the value 1 when the pair (i,j) pertains to distance class h , and 0 otherwise; n is the number of sampled observations (Legendre and Fortin 1989). Because they are standardized, correlograms may easily be compared. A Moran's I correlogram can also be tested for statistical significance, making it less subjective than the semivariogram for describing spatial structure and detecting autocorrelation. Several recent ecological studies have used correlograms and spatial autocorrelation statistics to examine large- and small-scale dependence in species distribution patterns (e.g., Carroll and Pearson 1998, Koenig 1998, Koenig and Knops 1998). In terms of large-scale patterns, positive autocorrelation at small distances coupled with negative autocorrelation at large distances indicates the presence of a gradient, or large-scale trend. With respect to small-scale patterns, the evaluation of autocorrelation statistics at various lag distances in a correlogram may be useful in determining the maximum scale of significant spatial autocorrelation. This may indicate the size of the "ecological neighborhood" for the species or phenomenon being investigated (Addicott et al. 1987), or the

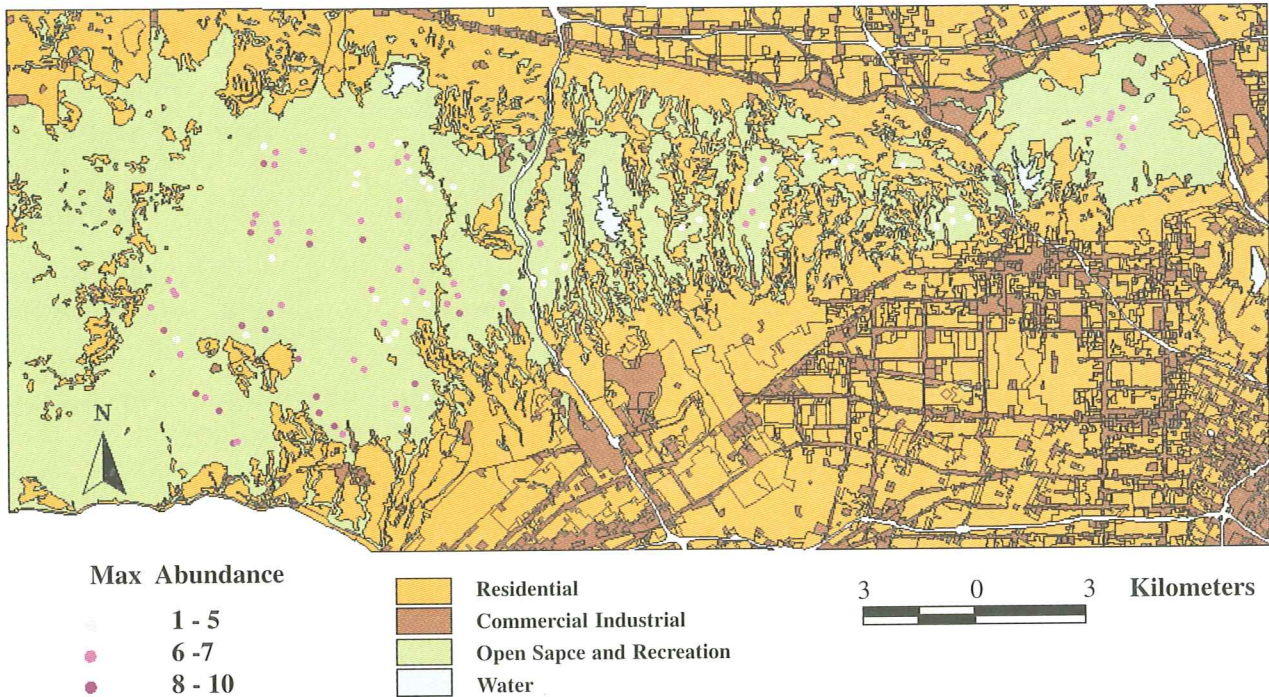


Figure 1a. Wrentit abundance. Maximum of two counts (1997). 107 sampling locations.

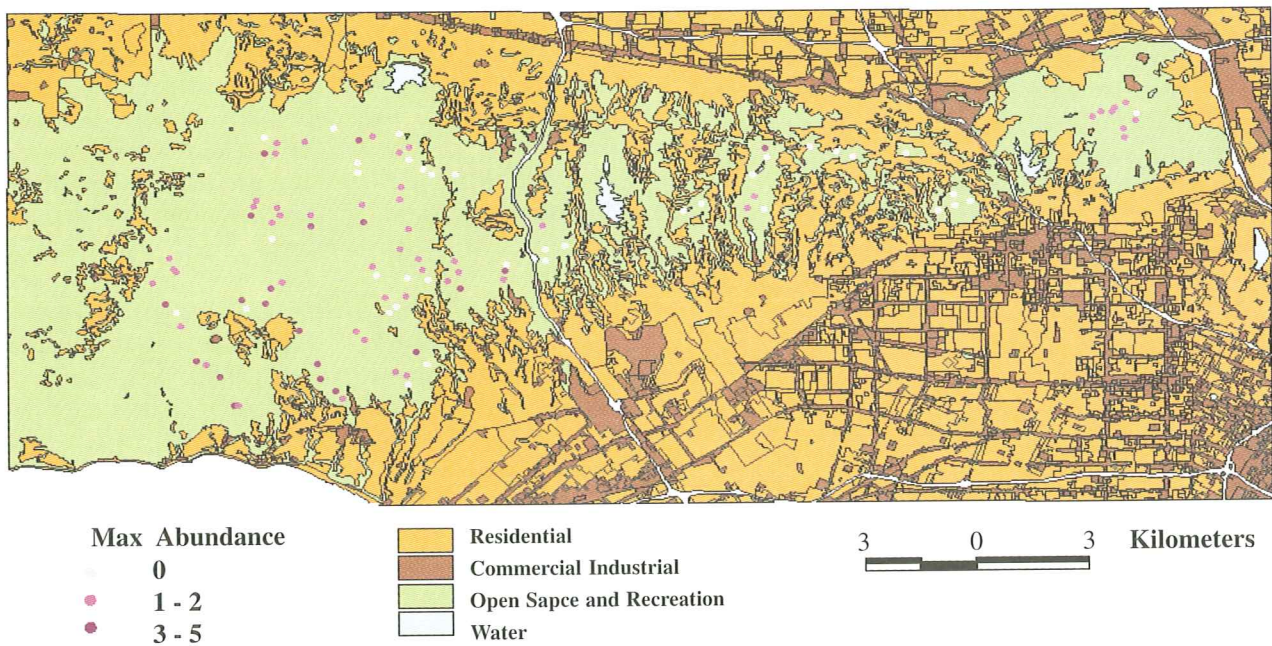


Figure 1b. Northern Mockingbird abundance. Maximum of two counts (1997). 107 sampling locations.

range of interaction between sites (Cressie 1993). Upon visual inspection of distribution maps, the Wrentit seemed to decrease in abundance from west to east, while the Northern Mockingbird increased (Figure 1a,b). The Wrentit also appeared to exhibit a north-south trend, with higher abundances in the south. To examine these apparent large-scale trends in bird abundance, as well as small-scale autocorrelation, the spatial covariance structure of the sampling locations was examined with respect to bird abundance, as well as urbanization proportion, using

two related structure functions, the semivariogram and the correlogram. For each variable, we examined large-scale trends, and determined the effective range of spatial autocorrelation (if any).

For maximum 1997 abundance of each species and urbanization proportion within a 2000m radius, an omnidirectional semivariogram, as well as directional semivariograms in the east-west (90° azimuth) and north-south (0° azimuth) directions (using a 45° tolerance), were constructed based on the 107 sampled

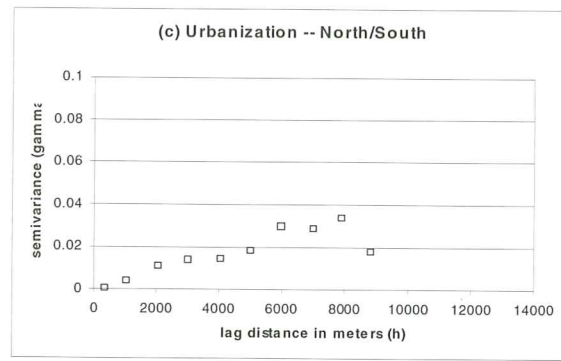
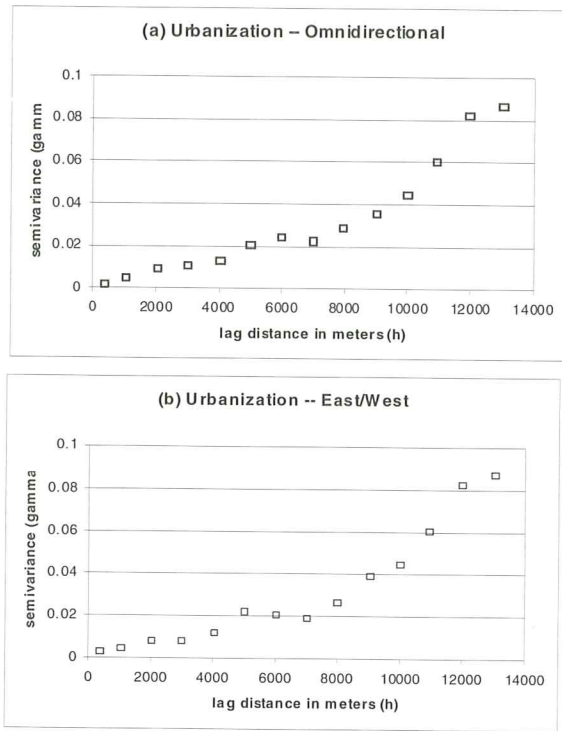


Figure 2. Empirical Semivariograms, $\gamma(h)$, for Urbanization Proportion (2000m radius).

East/West = 90° azimuth, North/South = 0° azimuth; 45° tolerance. Lag distance values represent averages distances between pairs of points within each 1000m distance class.*

locations. Fourteen lag distances, at 1000m intervals, were used for the omnidirectional and east/west semivariograms; nine 1000m lag distances were examined in the north/south direction. Semivariograms were generated using the variogram function in S+SpatialStats (MathSoft 1996).

Omnidirectional correlograms were also constructed in S+SpatialStats (MathSoft 1996) for each of the above variables, using Moran’s I autocorrelation statistics individually calculated for each 1000m lag distance (e.g., 0-1000m, 1000-2000m). In calculating the values of Moran’s I, point pairs within the given interval were considered to be spatial neighbors, and assigned a spatial weight (w_{ij}) of 1; all other point pairs were assigned a spatial weight of 0. We tested for overall significance of each correlogram using a Bonferroni-adjusted significance level ($\alpha = 0.01$) to account for multiple comparisons, as well as for significance of individual correlation values ($\alpha = 0.05$), as suggested by Legendre and Fortin (1989).

The overall trend apparent in the omnidirectional semivariogram for urbanization (Figure 2a) was driven primarily by east-west variation, as supported by the 90° directional semivariogram (2b), in which

the variation among points increases almost monotonically with distance, particularly at lag distances of 7000m or larger. The north-south (0°) semivariogram (2c) exhibited a similar trend between 0-8000m.

Both species exhibited distinct large-scale spatial trends (Figure 3) that appear to follow the urbanization trend. These trends occurred primarily in the east-west direction (3b,e), but also in the north-south direction (3c,f). The semivariograms did not increase monotonically, however, suggesting the presence of additional spatial structure not related to the east-west or north-south gradients.

In general, the omnidirectional semivariograms for both species resembled the east-west semivariograms, indicating that large-scale spatial trends were driven primarily by the east-west gradient (perhaps resulting from urbanization). The north-south semivariograms were most likely constrained by the configuration of the study area (long and narrow); thus they were based on fewer point pairs within each lag interval, and may be less reliable than the other semivariograms. The east-west semivariograms are most suspect at the shorter lag distances (<1000m) due to fewer point pairs, while both directional and omnidirectional semivariograms should be interpreted with caution at the larger lag distances (>8000m).

The omnidirectional Moran’s I correlogram for urbanization proportion displayed a nearly linear decrease

* Sample sizes for each distance class in omnidirectional semivariograms:

Avg.dist.	361	1044	2038	3010	4024	5011	5980	7013	7950	8987	9989	10939	11972	13041
# Pairs	72	282	357	567	580	582	504	388	287	158	158	151	110	137
East/West:														
Avg.dist.	379	1060	2033	3002	4024	5010	5994	7020	7973	9021	9989	10939	11972	13041
# Pairs	31	164	166	297	322	332	292	256	199	130	158	151	110	137
North/South:														
Avg.dist.	348	1021	2043	3018	4025	5012	5960	6999	7898	8832				
# Pairs	41	118	191	270	258	250	212	132	88	28				

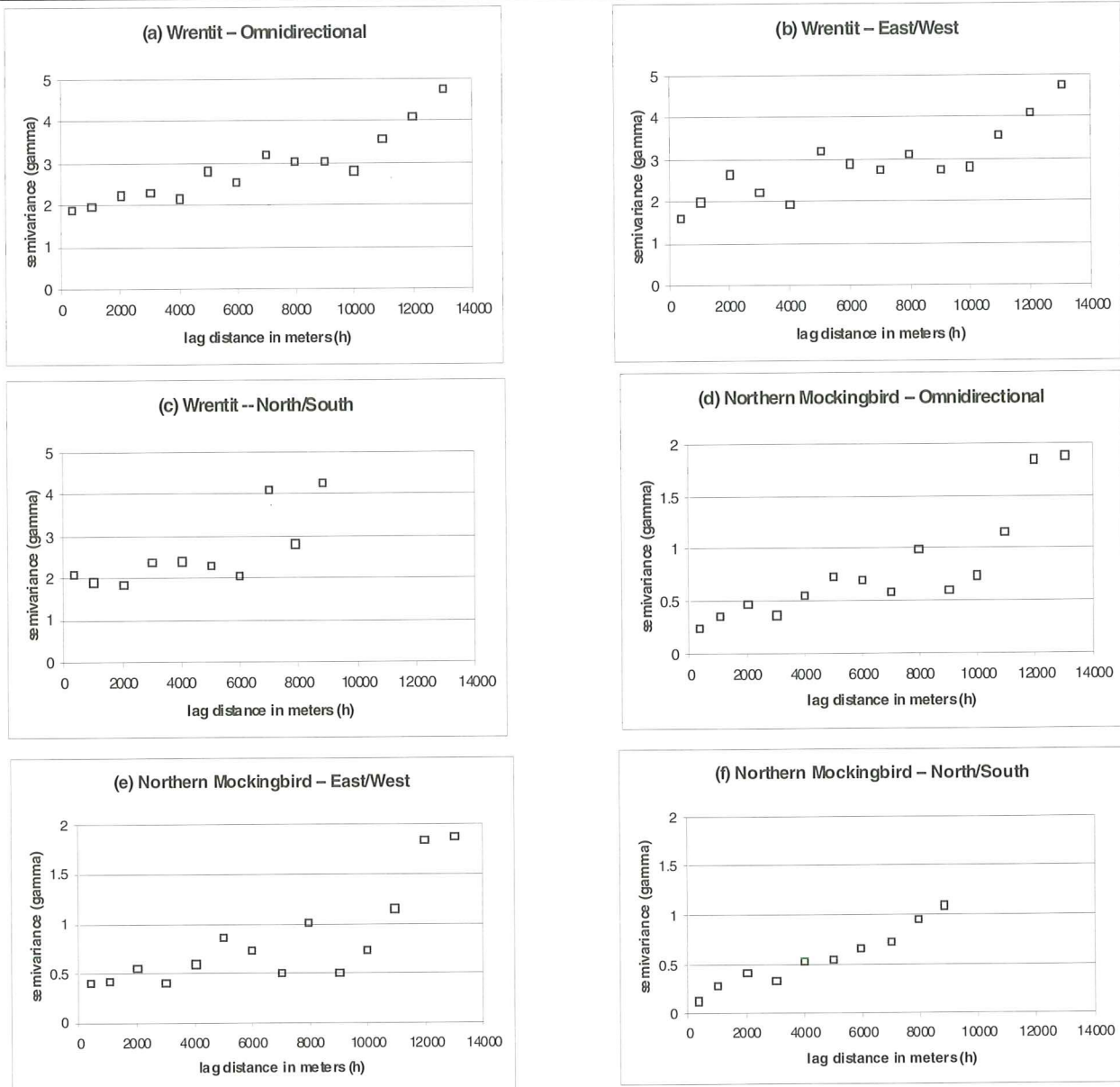


Figure 3. Empirical Semivariograms, $\gamma(h)$, for Maximum 1997 Bird Abundance.

East/West = 90° azimuth, North/South = 0° azimuth; 45° tolerance. Lag distance values represent averages distances between pairs of points within each 1000m distance class (see Figure 2 note).

in autocorrelation (covariance) with increasing lag distance (Figure 4a). It demonstrated overall significance ($p < 0.01$), as well as significant autocorrelation values for almost all distance classes. This indicates a clear, large-scale gradient, supporting the conclusions of the semivariogram analysis.

Correlograms also confirmed that both species were distributed along a spatial gradient, based on the steady (if not monotonic) decrease from positive to negative autocorrelation and overall significance ($p < 0.01$) of each correlogram (Figure 4b,c). Both species exhibited significant positive autocorrelation ($p < 0.05$) in more than one of the closer distance classes, while only the mockingbird also showed significant negative autocorrelation ($p < 0.05$) at greater distance

classes. Autocorrelation ranges suggested by the correlograms were approximately 7000m for the Wrenit (4b) and 5000m for the mockingbird (4c).

Examining Spatial Dependence in Non-Spatial Model Error Terms

Results from the previous section indicated large-scale trends in the abundances of both species, which could invalidate tests for spatial autocorrelation (Legendre 1993). That is, the similarity between sites that are close in spatial proximity may be due to their position on a gradient, rather than some intrinsic heterogeneity-producing spatial process. To obtain a reliable picture of small-scale autocorrelation, large-scale trends must first be removed.

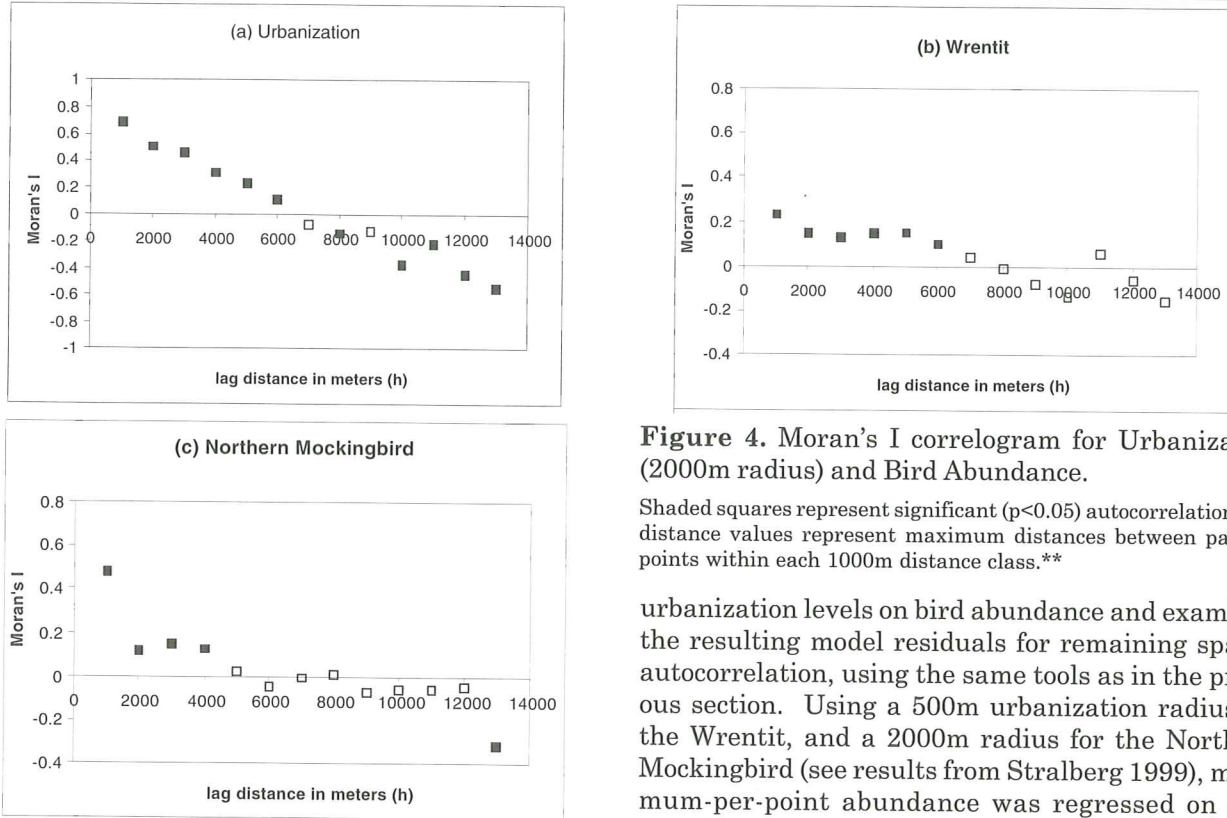


Figure 4. Moran's I correlogram for Urbanization (2000m radius) and Bird Abundance.

Shaded squares represent significant ($p < 0.05$) autocorrelation. Lag distance values represent maximum distances between pairs of points within each 1000m distance class.**

urbanization levels on bird abundance and examined the resulting model residuals for remaining spatial autocorrelation, using the same tools as in the previous section. Using a 500m urbanization radius for the Wrentit, and a 2000m radius for the Northern Mockingbird (see results from Stralberg 1999), maximum-per-point abundance was regressed on surrounding urbanization proportion, using ordinary least squares (OLS) regression.

Because similar trends were also observed in landscape urbanization proportion (within a 2000m radius of each point), we suspected that these large-scale trends were actually products of the urbanization gradient. Thus we modeled the effects of surrounding

urbanization as a significant predictor of abundance for both species (Table 1). Due to the high variation in the system, R^2 values were low, although regres-

Table 1. Non-spatial and detrended models for bird abundance.

	OLS model: bird abundance ~ urbanization measures	Trend model residuals ~ urbanization measures	Trend model: bird abundance / urbanization measures ~ x,y coordinates (x = east/west, y =north/south)	
<i>Wrentit</i>				
$\beta \pm$ S.E. (p-value)	-7.83 ± 1.41 (<0.0001)	-3.82 ± 1.24 (0.003)	Significant effects ($p < 0.05$)	x, y, x*y
F-statistic (p-value)	30.61 (<0.0001)	9.43 (0.003)	F-statistic (p-value)	18.36 (<0.0001)
R^2	0.23	0.08	R^2	0.35
<i>Northern Mockingbird</i>				
$\beta \pm$ S.E. (p-value)	3.42 ± 0.40 (<0.0001)	1.65 ± 0.43 (<0.0001)	Significant effects ($p < 0.05$)	x, y, x*y
F-statistic (p-value)	72.05 (<0.0001)	14.89 (<0.0001)	F-statistic (p-value)	10.67 (<0.0001)
R^2	0.41	0.12	R^2	0.24

β is the regression coefficient for the effect of urbanization proportion within a surrounding radius (500m for Wrentit; 2000m for Northern Mockingbird).

** Sample sizes for each distance class in Moran's I Correlograms:

Maximum distance	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	12000	13000
Number of Pairs	198	317	480	556	583	577	403	384	211	148	162	120	120

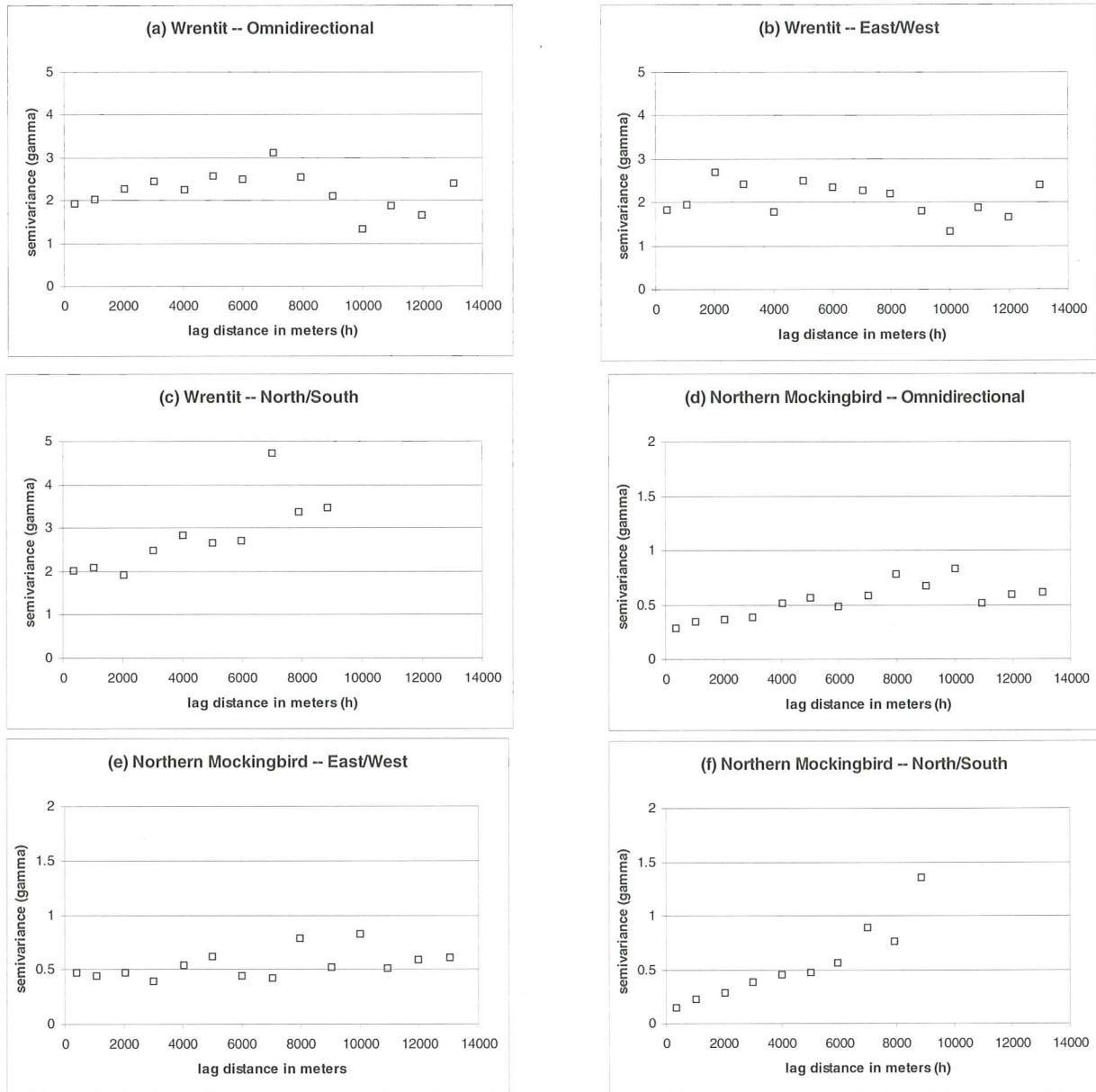


Figure 5. Empirical Semivariograms, $\gamma(h)$, for Non-Spatial Urbanization Model Residuals.

East/West = 90° azimuth, North/South = 0° azimuth; 45° tolerance. Lag distance values represent averages distances between pairs of points within each 1000m distance class (see Figure 2 note).

sion coefficients (b) were highly significant. The mockingbird exhibited a positive response to surrounding urbanization proportion, while the Wrentit was negatively influenced.

For both species, omnidirectional semivariograms for non-spatial model residuals increased initially, but eventually reached a maximum, indicating the possible presence of spatial autocorrelation in model residuals up to the distance at which the sill was reached (range) (Figure 5a,d). Although the sill and range differed, semivariograms for both species exhibited several small peaks and troughs, which may indicate that several processes, operating at different spatial

scales, govern bird abundance. They may also reflect patchiness in the sampling locations, however (Legendre and Fortin 1989, Meisel and Turner 1998). The omnidirectional semivariograms for Wrentit and Northern Mockingbird residuals reached their respective sills at ranges of 7000m (Figure 5a) and 10,000m (5d). Thus these species still appeared to exhibit a slight large-scale trend in their abundance patterns, even after the effect of urbanization was removed. These trends appeared to be driven by remaining north-south dependence (5c,f), which may have been a product of topography, in that study area consists of many north-south running canyons. Within these canyons, bird abundance may be more similar than

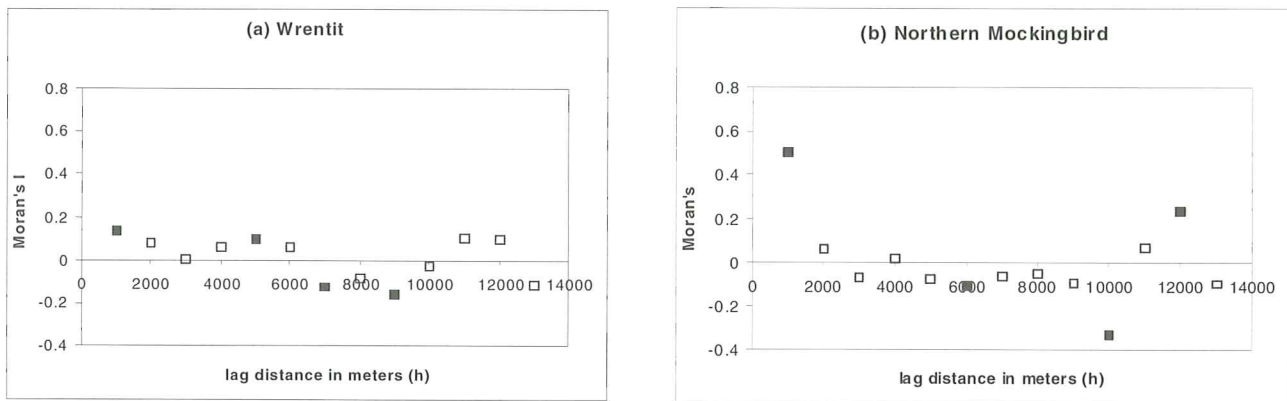


Figure 6. Moran's I correlograms for Non-Spatial Urbanization Model Residuals.

Shaded squares represent significant ($p < 0.05$) autocorrelation. Lag distance values represent maximum distances between pairs of points within each 1000m distance class (see Figure 4 note).

between sites in different canyons that are closer in a Euclidean sense. Both species may also vary with elevation, which generally increases from south to north in the study area. The comparatively flat east-west semivariograms (5b,e) suggested that autocorrelation in these species may be anisotropic (direction-dependent).

The presence of a sill in each semivariogram indicated that the residuals were stationary over the study area, which in turn suggested that large-scale trends in bird abundance were largely explained by variations in urbanization. However, the large autocorrelation ranges revealed by the semivariograms (7000-10,000m) indicated that model residuals still contained large-scale spatial dependence, perhaps due to the influence of other large-scale patterns (smaller than the size of the study area), such as patterns of variation in vegetation or elevation.

Examination of omnidirectional Moran's I correlograms provided a more complete picture of spatial structure in urbanization model residuals. Correlograms were overall significant ($p < 0.01$) for both species, with spatial autocorrelation generally declining with distance and oscillating around zero (Figure 6a,b). What the correlograms suggested, beyond the semivariograms, was a maximum range of "small-scale" autocorrelation (when the correlogram first reaches zero). Both species exhibited significant small-scale spatial autocorrelation ($p < 0.05$), with maximum autocorrelation ranges apparently near 2000m. Additional autocorrelation beyond these ranges may represent remaining large-scale trends and/or patchiness in the sampling sites.

Again, both species exhibited somewhat periodic dips and crests in the correlograms for their model residuals. For the Wrentit, significant positive spatial autocorrelation occurred at 1000m and 5000m; negative autocorrelation was present at the 7000m and 9000m lag distances (Figure 6a). The mockingbird

was positively autocorrelated at 1000m and 12,000m, and negatively autocorrelated at 6000m and 10,000m (6b). Positive autocorrelation at large distances may represent similarities at extreme ends of the study area, caused by large-scale processes, but more plausible is the influence of outliers, given the high variability and large gaps in the data.

Modeling Large-scale Spatial Trends

Because the east-west trend in urbanization was so prominent, separating the effects of urbanization and spatial location on bird abundance was difficult. Results from the previous section suggested that large-scale trends in bird abundance are largely a reflection of variations in urbanization, but that additional large-scale spatial dependencies may have been present, primarily in the north-south direction. The question remained as to whether these trends indicated non-stationarity over the study area (a "true gradient"), or intrinsic autocorrelation among sampling locations separated by large distances (a "false gradient") (Legendre 1993). If the former is true, then bird abundance should vary as a function of spatial location, and residuals should be uncorrelated once the trend is removed (Legendre 1993).

To statistically remove the large-scale trend, a trend model was estimated for the abundance of each species on the x - y coordinates of the sampling locations and the interaction between them ($z \sim x + y + x*y$). The trend model residuals were then regressed on urbanization proportion for each species. Finally, semivariograms and correlograms were used for investigating the remaining spatial trends and dependence in the residuals of the de-trended urbanization models.

For the Wrentit, spatial location explained a greater portion of the variation in abundance than urbaniza-

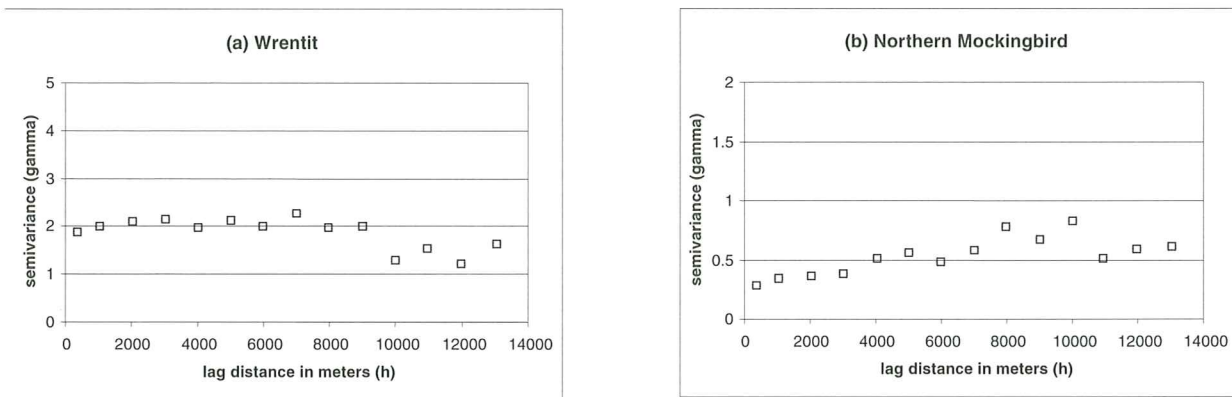


Figure 7. Empirical Omnidirectional Semivariograms, $\gamma(h)$, for Detrended Urbanization Model Residuals. Lag distance values represent averages distances between pairs of points within each 1000m distance class(see Figure 4 note).

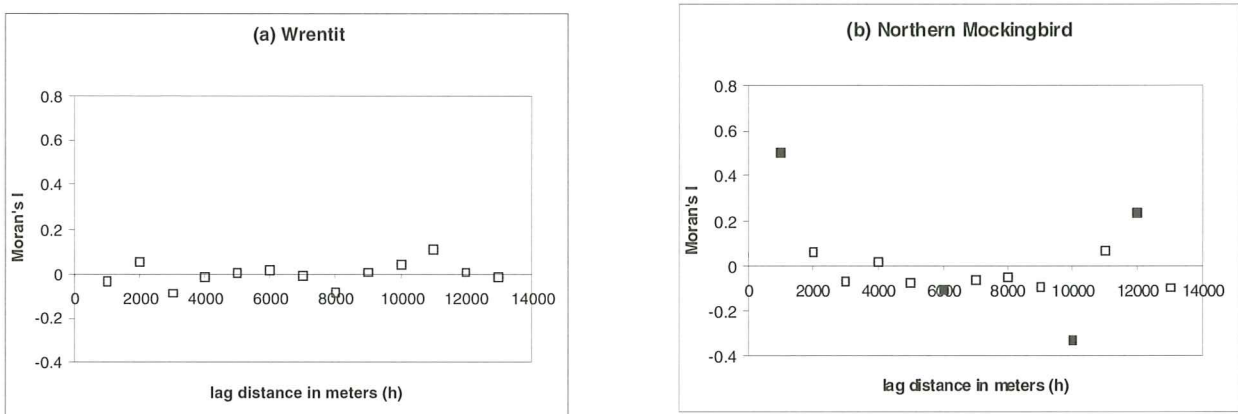


Figure 8. Moran's I correlograms for Detrended Urbanization Model Residuals.

Shaded squares represent significant ($p < 0.05$) autocorrelation. Lag distance values represent maximum distances between pairs of points within each 1000m distance class (see Figure 4 note).

tion proportion—35% vs. 23% (Table 1). For the Northern Mockingbird, however, surrounding urbanization levels explained more of the variance than spatial location (41% vs. 24%). Over 50% of the variation in urbanization (at the 2000m radius) was explained by spatial location, which means that the two effects are difficult to separate.

After removing the large-scale spatial trend, urbanization proportion was still a significant predictor of abundance for both species, although the explanatory power of these models was generally low, compared with the initial OLS models (i.e., lower R^2 and p-values) (Table 1). These results indicate that urbanization, in and of itself, did explain a significant portion of the variation in abundance of these species, even under the most conservative assumptions.

Semivariogram and correlogram analyses suggested that Wrentit abundance may have followed a “true” gradient, in that detrended model residuals appeared to be uncorrelated (Figures 7a,8a), while the Northern Mockingbird, whose residuals were still correlated, may exhibit a “false gradient” (Figures 7b,8b). The

extent to which these large-scale trends in abundance were a product of urbanization remains unclear, however, especially since surrounding urbanization proportion at a particular scale may not describe all of the variation in urbanization.

Estimating Spatial Error Term Models

Most GIS and statistical software packages provide functions for estimating traditional regression models such as Ordinary Least Squares regression (OLS), which are commonly used to estimate structural relationships among variables. If residuals are not independent, however, parameter estimates from OLS models are biased (Robertson 1987, Anselin 1989), and confidence intervals may be too small (Legendre 1993). Upon detection of spatial dependence in regression model residuals (e.g., using Moran's I statistic), several types of error models can be specified to incorporate spatial dependence, assuming that residuals are stationary (i.e., no trend).

One method is to explicitly model the spatial covariance structure of the error terms, using

semivariogram/covariogram models to define the nature of spatial dependence in the residuals and estimate the appropriate spatial parameters. This type of covariance modeling can be integrated with a mixed model framework (Zimmerman and Harville 1991, Littell et al. 1996) and takes a geostatistical approach, which is best suited for data like bird abundance, which varies (more or less) continuously over space (Cressie 1993). The general mixed model with spatial variability is of the form $y = X\beta + Z\gamma + e$, where y is a vector of observed data, β is an unknown vector of fixed-effects parameters with known design matrix X , γ is an unknown vector of random-effects parameters with known design matrix Z , and e is an unknown random error vector (SAS 1997). The spatial covariance is modeled through $R = \text{var}(e)$, using restricted maximum likelihood estimation (REML) to simultaneously estimate regression coefficients and covariance parameters. If y_i occurs at location s_i , then the covariance structure is defined by assuming that the covariance of two locations, s_i and s_j , is a function of d_{ij} , the distance between them. The covariance model then takes on the general form:

$$\text{Var}(e_i) = \sigma^2 = \sigma_1^2 + \sigma_2^2, \quad (4)$$

$$\text{with } \text{Cov}(e_i, e_j) = \sigma^2 [f(d_{ij})]$$

where d is the range, σ^2 is the sill, σ_1^2 is the nugget, and σ_2^2 is the remaining variance, or partial sill. Since the covariance and semivariance are directly related (under the assumption of stationarity), these parameters can be expressed in terms of semivariogram parameters. The function $f(d_{ij})$ depends on the shape of the semivariogram, and is generally specified according to the results of the empirical variogram.

Spatial dependence detected in the residuals from non-spatial regression models meant that the models were improperly specified (Anselin 1989). If this autocorrelation represents positive, small-scale autocorrelation, rather than large-scale trends, then spatial covariance models can be used to adjust for the spatial dependence. Thus, for each species, we fit a spatial covariance model (Littell et al. 1996) corresponding to the non-spatial model. The models were fit using mixed model regression in SAS 6.12 with REML (Restricted Maximum Likelihood) estimation of model parameters (SAS 1997) and an exponential covariance function:

$$f(d_{ij}) = e^{\left(\frac{-d_{ij}}{\rho}\right)} \quad (5)$$

The range parameter (ρ) was fixed at a value consistent with the empirical semivariogram for non-spatial model residuals, and the partial sill (σ_2^2) and nugget (σ_1^2) values were estimated simultaneously with estimation of regression coefficients (β). Using the maximum values of the empirical semivariograms, the

approximate ranges estimated were 7000m for the Wrentit and 9000m for the Northern Mockingbird. These range estimates are rather large, however, and the spatial covariance model is only intended for positive, small-scale dependence (Littell et al. 1996). Thus we used as a second range estimate the approximate distance at which the correlogram for non-spatial model residuals first reached zero: 2000m for both species. For the Wrentit, this smaller range also corresponded with the range suggested by the east-west directional semivariogram. The use of two different ranges (small vs. large) also provided a sensitivity analysis for the regression parameters. Based on the apparent convergence of the semivariograms to a sill, we assumed stationarity in both cases, despite irregularities and large autocorrelation ranges in the semivariograms and correlograms. We also assumed isotropic behavior, despite differences in east/west and north/south semivariograms.

Likelihood ratio χ^2 tests (-2 times the log likelihood from the non-spatial model minus -2 times the log likelihood from the spatial model) were used to determine the significance of the spatial (vs. non-spatial) models. The "best" covariance model could not be determined statistically, but was evaluated by comparing log likelihood values and significance of individual spatial parameter estimates. Regression coefficients and standard errors were also compared between spatial and non-spatial models to determine whether modeling of spatial structure in the residuals changed the substantive interpretations of the regression analysis.

For both species, the spatial models exhibited significantly better fit than the corresponding non-spatial models (χ^2 likelihood ratio test, $p < 0.05$) (Table 2), indicating significance of the spatial covariance parameters. The Wrentit exhibited the largest variation in regression parameter estimates between the three models, with somewhat higher magnitude values estimated by the non-spatial model than in either spatial covariance model, although confidence intervals for all parameter estimates were overlapping (Table 2). Spatial parameter estimates (σ_1^2 and σ_2^2) were larger in the large range (7000m) model, but the nugget effect was only significantly different from zero ($p < 0.05$) in the small range (2000m) model. The log likelihood value for the large-scale model was slightly higher, indicating that the incorporation of large-scale spatial structure (up to 7000m) improved model fit for this species, but with a consequent reduction in the urbanization parameter estimate. This suggests that part of the urbanization association may indeed be attributable to spatial autocorrelation. The most appropriate model depends on the range of autocorrelation assumed.

Table 2. Non-spatial vs. spatial covariance models for Bird Abundance.

	Non-Spatial Model	Spatial Covariance Models	
<i>Wren tit</i>			
$\beta \pm \text{S.E.}$	-7.83 ± 1.41	-6.16 ± 1.78	-4.93 ± 1.82
(<i>p-value</i>)	(<0.0001)	(0.0008)	(0.0078)
-2 Log Likelihood	400.93	384.34	382.59
Range (ρ)		2000m	7000m
Nugget (σ_1^2) \pm S.E.		1.28 ± 0.51	1.77 ± 0.81
Partial Sill (σ_2^2) \pm S.E.		1.45 ± 0.29	1.65 ± 0.27
Model Likelihood Ratio Chi-Square (<i>p-value</i>)		16.59 (0.0009)	18.34 (0.0004)
<i>Northern Mockingbird</i>			
$\beta \pm \text{S.E.}$	3.43 ± 0.41	3.23 ± 0.74	2.96 ± 0.90
(<i>p-value</i>)	(<0.0001)	(<0.0001)	(0.0014)
-2 Log Likelihood	241.85	204.84	210.39
Range (ρ)		2000m	9000m
Nugget (σ_1^2) \pm S.E.		0.69 ± 0.22	2.30 ± 0.87
Partial Sill (σ_2^2) \pm S.E.		0.10 ± 0.06	0.14 ± 0.06
Model Likelihood Ratio Chi-Square (<i>p-value</i>)		37.01 (<0.0001)	31.45 (<0.0001)

β is the regression coefficient for the effect of urbanization proportion within a surrounding radius (500m for Wren tit; 2000m for Northern Mockingbird).

The Northern Mockingbird exhibited a slightly smaller relative reduction in regression parameter estimates with the inclusion of spatial parameters (Table 2). Spatial covariance parameter estimates were also significantly different from zero ($p < 0.05$) in both spatial models, but the nugget estimate was quite a bit larger in the large range model (9000m) than in the small range (2000m) model. For this species, the 2000m range distance appeared to generate the best-fitting model, based on the higher log likelihood value. Thus the incorporation of large-scale autocorrelation (above 2000m) was unnecessary and inappropriate for this species.

III. DISCUSSION

The two species examined in this study demonstrated significant population responses to urbanization at large spatial scales, suggesting that their abundances may be affected by changes in landscape composition and pattern. Using a variety of spatial statistics tools, we were able to consider the influence of spatial dependence in our study, in order to verify these results. Equally important, these techniques also allowed a better understanding of the spatial structures of the species under study, and the landscape mosaic that they inhabit. Different insights were gained for each species, while common spatial patterns were also illuminated.

Both species exhibited large-scale spatial trends in their abundance, primarily coinciding with the east-west urbanization gradient along which they were sampled (Figures 3,4). After accounting for urbanization effects, however, these trends were largely absent from model residuals (Figures 5,6), suggesting that urbanization explained most of the large-scale spatial structure in these species. Conversely, when the spatial trends were removed, both species still demonstrated significant responses to changes in urbanization (Table 1).

Abundance patterns for both species also exhibited additional spatial autocorrelation unrelated to urbanization (Figures 5,6). The use of spatial covariance models to incorporate this small-scale spatial dependence (where possible) allowed conclusions about urbanization effects to be more robust. In general, the observed relationships between urbanization and bird abundance held, despite small-scale spatial structure (autocorrelation) in the data. Furthermore, the spatial covariance models had better predictive capabilities, indicated by the significant improvement in log likelihood statistics. If desired, the spatial and regression parameter estimates obtained in these analyses could be used to interpolate bird abundance at points between those sampled.

Ranges of Spatial Autocorrelation

Ranges of small-scale spatial autocorrelation, as suggested by the correlogram analysis, can be interpreted as indications of the ecological neighborhoods for each species—that is, the “regions of activity or influence during periods of time appropriate to particular ecological processes” (Addicott et al. 1987, p.340). Depending on the process and time-scale, a mobile animal’s ecological neighborhood can be defined by factors like (in order of increasing spatial scale) breeding territory size, home range size, natal/adult dispersal distance, and spatial extent of a local population.

The large autocorrelation distances found in this study—as compared with average dispersal ranges and territory sizes (see, e.g., Erickson 1938, Derrickson and Breitwisch 1992, Baker et al. 1995)—may indicate the influence of population (vs. individual) dynamics, which is reasonable given the intermediate temporal and spatial scales of the urbanization processes being studied. Time lags may also result in a decoupling of ecological variables and responses in bird abundance (Wiens 1989). Given the coarse resolution of this analysis, however, autocorrelation at smaller scales, resulting from interactions among individuals, could not easily be detected.

Of the two species examined, the Northern Mockingbird demonstrated the strongest small-scale autocorrelation, even after large-scale trends were removed. This indicated a strong interdependence among sampling sites, independent of the location in the landscape, but may also have been due in part to the clustering of this species near urban developments. Similar results were obtained for the Wrentit, a species that was negatively associated with surrounding urbanization, although for this species, residuals were no longer correlated after large-scale trends were removed (Figure 8). The Wrentit also seemed to exhibit anisotropic variation in the shorter distance ranges (Figure 3), which likely explained the weaker overall autocorrelation values.

Large-scale Trends vs. Small-scale Autocorrelation

Spatial covariance modeling is most appropriate for small-scale spatial dependence resulting from the interactions among sites that are close in spatial proximity. Spatial autocorrelation can also occur at larger scales, when the influence of one site carries a long way through geographic space (Legendre 1993). But large-scale spatial dependence commonly reflects a spatial gradient, and if so, it should be removed before spatial autocorrelation is modeled, as it violates

the stationarity assumption (Legendre 1993). The distinction between the two is subtle, and also depends upon the temporal scale and window of analysis (Wiens 1981). For the purpose of geostatistical analyses, a true gradient should encompass and extend beyond the study area. Large-scale patterns that are smaller than this can be incorporated in a spatial covariance model.

In this study, large-scale spatial dependence in bird abundance was certainly present, but most of it was explained by the urbanization gradient. Remaining large-scale structure appeared to be operating on a smaller scale than that of the study area (as demonstrated by semivariograms in Figure 5) and may have followed patterns of variation in vegetation, topography, or other geophysical factors. It may also have been a product of aggregated sampling locations (Legendre and Fortin 1989), resulting from a combination of habitat fragmentation by urbanization, clustering of census sites along trails and fire roads, and gaps in census sites due to restricted access. Regardless of the source, this large-scale spatial dependence appeared to be adequately described in the spatial covariance models.

For the Wrentit, residuals from the detrended models exhibited no spatial autocorrelation (Figure 8), suggesting that spatial covariance models may not be necessary to describe its abundance patterns if large-scale trends are removed. Here emerges a possible trade-off between describing the distribution of abundance and detecting urbanization-related changes. If the primary goal is to model the abundance of this species in this particular landscape, then factoring in spatial location, in addition to the effect of urbanization influence, may be adequate. The remaining residuals were uncorrelated, so the model should be appropriately specified. Given the distinct urbanization gradient, however, such a model diminishes the potential for detection of urbanization influences. If autocorrelation is thought to occur over large distances, than this spatial dependence can instead be explicitly incorporated in a spatial covariance model.

The decision to de-trend a variable, in cases such as this when the trend is not obvious, depends on the assumptions made about the system. Here, because the birds under study are fairly widespread and common throughout the Santa Monica Mountains, and vegetation was similar at all sites sampled, we would expect urbanization and landscape pattern to cause more variation in abundance than spatial location *per se*.

Directional semivariograms indicated that trends in bird abundance were primarily driven by east-west

variation (Figure 3), but after factoring out urbanization, north-south autocorrelation was responsible for remaining large-scale dependence (Figure 5). These results suggest that east-west trends were primarily urbanization-related, while north-south trends were likely due to natural features (i.e., the north-south orientation of the canyons and trails along which sampling sites were located). Anisotropic covariance models, explicitly incorporating direction, as well as distance, may therefore be more appropriate for these species. Furthermore, the use of local spatial statistics (Getis and Ord 1992, Anselin 1995, Bao and Henry 1996) may help identify specific regions of autocorrelation.

IV. CONCLUSION

Fundamentally, spatial autocorrelation in nature is caused by unknown factors that could potentially be measured and modeled explicitly, as many have done, particularly in small-scale studies. Measuring these unknown variables is often infeasible or impractical, however, especially when they involve things like animal behavior, indirect effects of biotic interactions, and time lags. Furthermore, such "bottom-up" approaches may sacrifice generalizability (Wiens 1989, Root and Schneider 1995).

In this study, landscape-level urbanization influences on bird abundance were the primary focus of interest, while habitat type (i.e., vegetation) was kept constant. Since small-scale autocorrelation has a large stochastic component, it is difficult to predict. Thus, in the absence of the elusive "right" variables, the use of spatial statistics is a good way to ensure that conclusions are valid for the explanatory factors examined.

ACKNOWLEDGMENTS

The authors are deeply grateful for the insightful comments and suggestions of two anonymous reviewers. Funding was provided by the Southwest Parks and Monuments Association, as well as the Population-Environment Dynamics Project, Rackham Graduate School and School of Natural Resources and Environment at the University of Michigan.

REFERENCES

- [1] Addicott, J.F., J.M. Aho, M.F. Antolin, D.K. Padilla, J.S. Richardson and D. A. Soluk, 1987, Ecological neighborhoods: scaling environmental patterns, *Oikos*, 49:340-346.
- [2] Anselin, Luc, 1989, *Spatial Econometrics: Methods and Models*, Kluwer Academic, Dordrecht, Netherlands.
- [3] Anselin, Luc, 1995, Local indicators of spatial association – LISA, *Geographical Systems*, 27:93-115.
- [4] Baker, M., N. Nur and G.R. Geupel, 1995, Correcting biased estimates of dispersal and survival due to limited study area: theory and an application using Wrentits, *Condor*, 97:663-674.
- [5] Bao, S. and M. Henry, 1996, Heterogeneity Issues in Local Measurements of Spatial Association, *Geographical Systems*, 3:1-13.
- [6] Brown, J. H., 1984, On the relationship between abundance and distribution of species, *American Naturalist*, 124:255-279.
- [7] Brown, J. H., D. W. Mehlman, and G. C. Stevens, 1995, Spatial Variation in Abundance, *Ecology*, 76:2028-2043.
- [8] Carroll, S.S. and D. L. Pearson, 1998, Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as a bioindicator taxon, *Ecological Applications*, 8:531-543.
- [9] Cressie, N.A.C., 1993, *Statistics for Spatial Data*. Wiley, New York.
- [10] David, M., 1977, *Geostatistical ore reserve estimation*, Elsevier Scientific, New York.
- [11] Derrickson, K.C. and R. Breitwisch, 1992, Northern Mockingbird (*Mimus polyglottos*), *Birds of North America* 7, A. Poole, P. Stettenheim and F. Gill, Eds. Philadelphia, PA.
- [12] Erickson, M. M., 1938, Territory, annual cycle and numbers in a population of wren-tits (*Chamaea fasciata*), *University of California Publications in Zoology*, 42:247-334.
- [13] Getis, A. and K. Ord, 1992, The analysis of spatial association by the use of distance statistics, *Geographical Analysis*, 24:189-206.
- [14] Grieg-Smith, P., 1983, *Quantitative Plant Ecology*, University of California Press, Berkeley, CA.
- [15] Journel, A. G. and C. J. Huijbregts, 1978, *Mining Geostatistics*, Academic Press, London.
- [16] Koenig, W. D., 1998, Spatial Autocorrelation in California Land Birds, *Conservation Biology*, 12(3):612-620.
- [17] Koenig, W.D. and J.M.H. Knops, 1998, Testing for spatial autocorrelation in ecological studies, *Ecography*, 21:423-429.
- [18] Legendre, P., 1993, Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:1659-1673.
- [19] Legendre, P. and M.-J. Fortin, 1989, Spatial pattern and ecological analysis, *Vegetatio* 80:107-138.
- [20] Littell, R.C., G.A. Milliken, W.W. Stroup, and R. D. Wolfinger, 1996, *SAS System for Mixed Models*, Chapter 9: Spatial Variability, SAS Institute Inc., Cary, NC.
- [21] Matheron, G., 1963, Principles of geostatistics. *Economic Geology*, 58:1246-1266.
- [22] MathSoft, 1996, *S+SpatialStats User's Manual, Version 1.0*, MathSoft, Inc., Seattle, WA.
- [23] MathSoft, 1997, *S-Plus User's Guide*, MathSoft, Inc., Seattle, WA.
- [24] MathSoft, 1998, *S-Plus for ArcView User's Guide*. MathSoft, Inc., Seattle, WA.
- [25] Meisel, J.E. and M.G. Turner, 1998, Scale detection in real and artificial landscapes using semivariance analysis, *Landscape Ecology*, 13:347-362.
- [26] Moran, P. A. P., 1950, Notes on continuous stochastic

- phenomena, *Biometrika*, 37:17-23.
- [27] O'Neill, R.V., S.J. Turner, V.I. Cullinan, D.P. Coffin, T. Cook, W. Conley, J. Brunt, J.M. Thomas, M.R. Conley and J. Gosz, 1991, Multiple landscape scales: an intersite comparison, *Landscape Ecology*, 5:137-144.
- [28] Ralph, C. J., G. R. Geupel, P. Pyle, T. E. Martin, and D.F. DeSante, 1993, *Handbook of Field Methods for Monitoring Landbirds*, U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, Albany, CA. Gen. Tech. Rep. PSW-GTR-144.
- [29] Robertson, G. P., 1987, Geostatistics in ecology: interpolating with known variance, *Ecology*, 68:744-748.
- [30] Robertson, G.P., K.M. Klingensmith, M.J. Klug, E.A. Paul, J.C. Crum, and B.G. Ellis, 1997, Soil resources, microbial activity, and primary production across an agricultural ecosystem, *Ecological Applications*, 7: 158-170.
- [31] Root, T.L. and S.H. Schneider, 1995, Ecology and climate: research strategies and implications, *Science* 269:334-341.
- [32] Rossi, R. E., D. J. Mulla, A. G. Journel and E. H. Franz, 1992, Geostatistical tools for modeling and interpreting ecological spatial dependence, *Ecological Monographs*, 62:277-314.
- [33] SAS, 1997, SAS/STAT® Software: Changes and Enhancements through Release 6.12. SAS Institute Inc., Cary, NC.
- [34] Stralberg, D., 1999, *A Landscape-Level Analysis of Urbanization Influence and Spatial Structure in Chaparral Breeding Birds of the Santa Monica Mountains, CA. M.S. Thesis*, University of Michigan. 119 pp.
- [35] Taylor, L. R., 1984, Assessing and interpreting the spatial distributions of insect populations, *Annual Review of Entomology*, 20:321-357.
- [36] Villard, M-A. and B. A. Maurer, 1996, Geostatistics as a tool for examining hypothesized declines in migratory songbirds, *Ecology*, 77:59-68.
- [37] Wiens, J. A., 1981, Scale problems in avian censusing, *Studies in Avian Biology*, 6:513-521.
- [38] Wiens, J. A., 1989, Spatial scaling in ecology, *Functional Ecology*, 3:385-397.
- [39] Yaglom, A. ,M., 1957, Some classes of random fields in n-dimensional space, related to stationary random processes, *Theory of Probability and its Applications*, 2:273-320.
- [40] Zimmerman, D. L. and D. A. Harville, 1991, A random field approach to the analysis of field-plot experiments and other spatial experiments, *Biometrics*, 47:223-239.