Forum: Ideas

Spatial autocorrelation: an overlooked concept in behavioral ecology

Mihai Valcu and Bart Kempenaers
Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner Street, D-82319 Starnberg (Seewiesen), Germany

Spatial autocorrelation (SAC) is the dependence of a given variable's values on the values of the same variable recorded at neighboring locations (Cliff and Ord 1973; Fortin and Dale 2005). When high values are associated with relatively high values at neighboring locations, SAC is said to be positive and, conversely, where high values correspond to relatively low values at neighboring locations, SAC is negative. SAC can be a property of the variable itself (inherent or intrinsic SAC) or it can arise due to the dependence of the variable of interest on another spatially autocorrelated variable (induced SAC) (Legendre et al. 2002; Fortin and Dale 2005). Because it lies at the core of most spatial models, SAC is a fundamental concept of spatial analysis (Getis 2008).

During the last 2 decades, mostly after Pierre Legendre published his seminal paper "Spatial autocorrelation: trouble or new paradigm" (Legendre 1993), SAC received considerable attention from ecologists—in particular biogeographers investigating macroecological patterns of species distributions (Kissling and Carl 2008)—and from population geneticists investigating small-scale spatial genetic structure of populations (Guillot et al. 2009). These circumstances prompted Arthur Getis, in a review on the evolution of the SAC concept, to conclude that "Nearly all the major journals that concern themselves with the ecological aspects of their subjects print articles having a spatial autocorrelation foundation" (Getis 2008). In stark contrast, the issue of SAC has hitherto been largely ignored in behavioral ecology, despite the fact that many studies in this field deal with a spatial component. Thus, despite its recognized importance in adjacent fields of ecological research, even very basic topics in behavioral ecology remain unexplored with respect to SAC, and we could only find a handful of studies (van der Jeugd and McCleery 2002; Laiolo and Tella 2006; Duraes et al. 2007; Aarts et al. 2008; Giesselmann et al. 2008; Holdo et al. 2009), which included SAC in their research paradigm.

The general aim of this paper is to draw the attention of behavioral ecologists to the phenomenon of SAC. Specifically, we aim 1) to provide examples of spatially autocorrelated variables, indicating that SAC is widespread in variables commonly used in behavioral ecology studies, 2) to show why it is important to take SAC into account, and 3) to point to some tools to explore and model it.

CAN SAC BE DETECTED IN BEHAVIORAL ECOLOGY DATA SETS?

To illustrate the nature of SAC, let us consider territory size (Figure 1). The size of an animal's territory is usually the outcome of a well-understood behavioral process, namely the competition among neighboring individuals. This competition is reflected on the one hand, in the spatial distribution of individuals, whereby increased competition results in an increased spatial regularity (assuming—for simplicity—a uniform distribution of resources; Campbell 1992), and on the other hand, in the strength and number of interactions at the territory boundaries, whereby the degree of exclusion of the neighbors from the focal territory determines the amount of overlap between territories (Maher and Lott 1995). Because the size of an individual's territory is a result of interindividual competition, it can be predicted that territory size is intrinsically positively spatially autocorrelated (Valcu and Kempenaers 2010). Using both a simulation approach and a meta-analysis, we showed that all widely used measures of territory size are bound to be spatially autocorrelated due to the nature of territory formation (Valcu and Kempenaers 2010). SAC of territory size can be further increased if territory size is also a function of the amount of available resources (e.g., mating partners or food) and if those resources are themselves spatially autocorrelated at a scale larger than the scale of territory size (e.g., resources distributed along a gradient or in large patches across the study area).

The rationale previously applied to territory size can be straightforwardly generalized. We can thus argue that variables measuring processes such as competition, song matching, or extrapair paternity, which all reflect inter-individual interactions, are probably spatially autocorrelated (Table 1). Similarly, variables measuring processes driven by environmental (extrinsic) factors, such as clutch size decisions or brood sex ratio adjustment, can be spatially autocorrelated due to their dependency on an already spatially autocorrelated factor (Table 2). We therefore postulate that: 1) every measure of a process that results from interactions of (spatially distributed) individuals is potentially intrinsically spatially autocorrelated (i.e., inherent SAC) and 2) every measure of a process that is linked with a spatially distributed resource is potentially extrinsically spatially autocorrelated (i.e., induced SAC).

In most empirical data sets, intrinsic and extrinsic factors are likely to interact. Thus, SAC of a given variable can be caused by both extrinsic and intrinsic factors. Moreover, an extrinsic factor can modulate the intensity of an intrinsic spatially autocorrelated variable (e.g., SAC of territory size can increase under limited resource availability due to increased competition). Conversely, the SAC of an extrinsically spatially autocorrelated variable can be masked by environmental variables covarying at the same spatial scale.

WHY SHOULD BEHAVIORAL ECOLOGISTS BE CONCERNED WITH SAC?

From a statistical analysis perspective, SAC can lead to several types of spurious results. 1) Increased type I error rate. In common bivariate tests, such as the Pearson correlation coefficient, the risk of type I error increases when SAC is present in both variables, even at small levels (Lennon 2000; Legendre et al. 2002; Legendre et al. 2004). Likewise, one of the important assumptions of general and generalized linear models (GLMs) and of their extensions, the independence of residual errors (e.g., Hill 2007), will be violated when SAC is present in the residuals of the fitted model (Haining 1990). SAC can
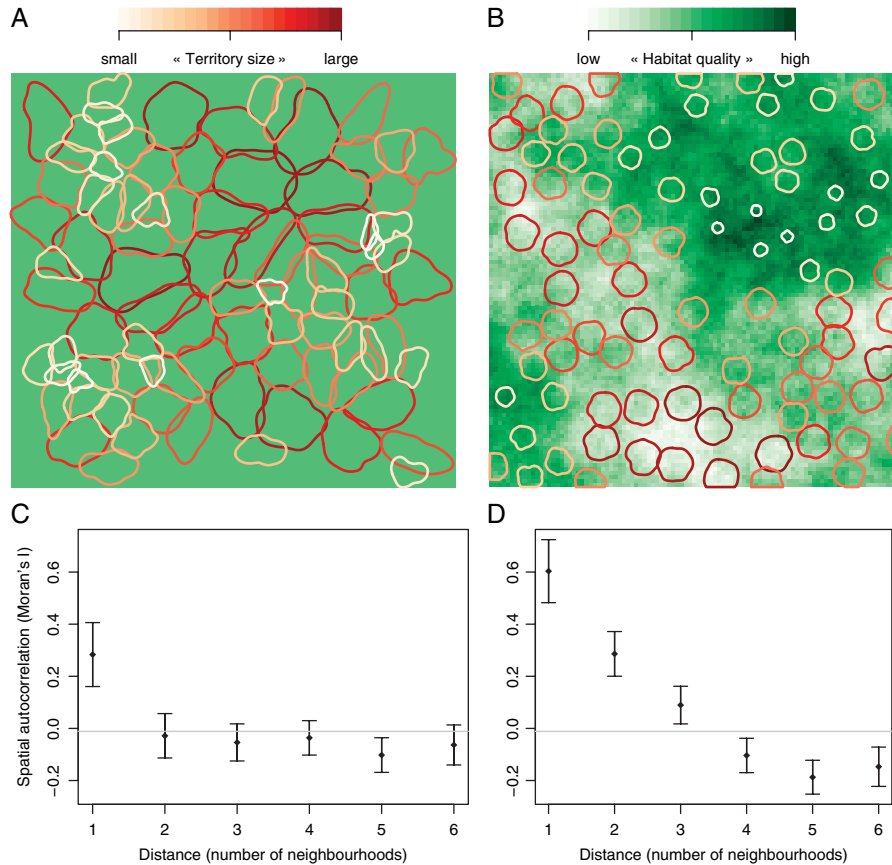
Figure 1
Examples of inherent and induced SAC of territory size at different spatial scales. Inherent SAC results solely from interindividual interactions and space partitioning without any influence of the habitat (A). The corresponding correlogram (C) shows positive and significant[*] SAC only at the scale of the nearest neighbors. Induced SAC appears as a result of the negative correlation between territory size and habitat quality (B). The corresponding correlogram (D) shows significant[*] SAC at all spatial scales. In both scenarios, the same number of territories (bivariate normal kernels) were generated; for simulation procedures see Valcu and Kempenaers 2010). The correlograms show Moran's I coefficient $\pm$ standard deviation at 6 distance classes (distance 1 refers to closest neighbors; distance 2 refers to second-order neighbors, and so on). Moran's I coefficient is comparable with a Pearson's correlation coefficient; it takes values between 0 and 1 in the case of positive SAC and between -1 and 0 in the case of negative SAC (Fortin and Dale 2005). The horizontal line is the expected value of Moran's I under the null hypothesis of no autocorrelation. ([*]after Bonferroni correction).

thus bias model selection because spatially autocorrelated variables will get narrower confidence intervals and consequently be picked up as having a significant contribution to the fitted model more often than by the desired significance level (Lennon 2000). Hence, SAC can be seen as a form of pseudoreplication (Hurlbert 1984), whereby the effective sample size is smaller than the observed sample size (Dutilleul 1993). It has further been suggested that the presence of SAC can also reduce the power of a test statistic (Legendre et al. 2002, 2004). 2) Bias in parameter estimates. Neglecting SAC

Table 1

Examples of possible scenarios leading to inherent SAC

Behavioral process	Variables	Possible reasons for SAC
Competition for breeding space	Territory size	Competition at territory boundaries and the spatial distribution of individuals.
Song as a signal	Song type, frequency	Song matching between close neighbors.
Extrapair mate choice	Number of extrapair young	Spatial distribution of potential extrapair mates leads to local "hot spots."
Conspecific/heterospecific attraction	Breeding synchrony (e.g., in lay date), breeding density	Conspecific/heterospecific attraction leads to differences in breeding onset and/or breeding density across the landscape.
Aggregation of territories into hidden leks	Extrapair paternity rates, breeding density	Extrapair mating behavior leads to aggregations of territorial males with particular characteristics.

Table 2
Examples of possible scenarios leading to induced SAC

Behavioral process	Variables	Possible reasons for SAC
Nonrandom settlement via breeding/natal dispersal	Life-history trait measures	High-quality individuals select high-quality breeding sites across a heterogeneous habitat.
Clutch size decisions	Clutch size, number of hatchlings and fledglings	Clutch size is adjusted to habitat quality, mate quality, and/or breeding density across a heterogeneous habitat.
Sex ratio adjustment	Sex ratio of litter or brood	Litter or brood sex ratio is adjusted to habitat quality and/or to the social environment across a heterogeneous habitat.
Lombard effect	Song amplitude	Spatial heterogeneity of environmental noise.
Mate vs. territory choice (polygyny threshold)	Number of mates	Spatial heterogeneity of habitat quality at a scale larger than territory size.
Coping style/personality	Aggressiveness, activity, and sociability	Individuals with certain copying styles are associated with a particular habitat type or quality.

can lead to a large upward bias in parameter estimates as shown in a recent meta-analysis of species distribution studies (Dormann 2007). Because SAC is expected to occur at all spatial scales, behavioral ecologists should be aware that some of the large highly significant effect sizes could be generated by SAC instead of reflecting a causal relationship or a treatment effect.

However, SAC need not be seen as a nuisance; it can be a useful method for data analysis both during the descriptive stage and during hypotheses testing. Understanding and modeling SAC may lead to a deeper biological understanding of the investigated variables. For example, a visual inspection of a correlogram (a graph where SAC values are displayed on the y axis and e.g., neighborhood relations, distance classes, or nearest neighbors are shown on the x axis; see Figure 1C,D) (Legendre and Fortin 1989; Bivand et al. 2008, p. 267) will allow to explore SAC of a given variable at multiple spatial scales (Figure 1).

Although the effects of SAC on most empirical data sets will be difficult to predict, by corroborating the information exposed by the correlogram with detailed knowledge of the studied system (including the distance over which intrinsic and extrinsic factors operate) one can make further predictions and design experiments at the correct spatial scale. For example, when SAC is only apparent on a small spatial scale (e.g., among close neighbors), as depicted by the correlogram (Figure 1C), it can be hypothesized that it is caused by interindividual interactions and the strength of the SAC will reflect the strength of these interactions (e.g., competition) (Figure 1A). Alternatively, when SAC is gradually decreasing with distance (Figure 1D), it can be hypothesized that the studied variable depends on an environmental variable distributed along a gradient and SAC reflects the habitat heterogeneity (Figure 1B). Modeling SAC can thus inform us or lead us to hypothesize about, for example, the scale at which habitat quality is heterogeneous, the distance (scale) over which males influence each other through vocal communication, or the distance over which females sample mates.

TOOLS FOR EXPLORING AND MODELING SAC

Due to the recent advances in spatial statistics and geographical information systems a wide range of tools are now available to model SAC (e.g., Haining 1990; Fortin and Dale 2005; Bivand et al. 2008). Describing a general framework for dealing with SAC is beyond the scope of this note; hence, we will just highlight a few points that may be of interest to behavioral ecologists.

The prerequisite of spatial data analysis and thus of SAC analysis is the existence of geographical coordinates (ideally transformed in a projected coordinate system like Universal Transverse Mercator to ensure a constant distance relationship throughout the map) associated with each variable. Once the data set is augmented with the geographical coordinates, the investigator can proceed to the first step of exploratory data analysis, which is mapping the target variables. To get further insight into the data, a graphical representation of SAC at increasing spatial scales, for example, a correlogram (Figure 1C,D), can be created. This step requires the identification of the spatial relationships between observations (i.e., neighbors). Among the most common criteria used here are distance bands (individuals are considered neighbors if they are not farther apart than a given distance), nearest neighbors (the first k nearest neighbors are considered), or graph-based neighbors (based on the relationship among geometrical constructs, e.g., Dirichlet polygons) (Bivand et al. 2008, p. 239). The choice of such criterion should be made based on the life-history traits under consideration. For example, distance classes can be used in the case of song or calls, based on knowledge of their range of action in a particular habitat, k nearest neighbors can be used in order to account for differences in densities across the habitat, or territory boundaries can be used for a straightforward delineation of neighbors in case of territorial species.

The identification of the spatial relationships among neighbors is also of great importance for the last step of data analysis; modeling and hypothesis testing. Simultaneous autoregressive (SAR) models are a useful class of spatial models dealing with SAC (e.g., Fortin and Dale 2005; Bivand et al. 2008), particularly because they are a straightforward extension of the GLM. SAR and other types of models make use of the spatial relationships among neighbors in order to construct a spatial weights matrix, which is further used to model nonindependent (i.e., autocorrelated) errors. In short, the SAR models are a particular case of GLM: $Y = X\beta + e$, where Y is the dependent variable, β are the coefficients, X is the matrix of predictors, and e is the error term. The way in which e is modeled determines the type of the SAR model. The SAR error model is defined as $Y = X\beta + \lambda Wu + e$ and the SAR lagged model as $Y = \rho WY + X\beta + e$ where λ and ρ are the spatial autoregression coefficients, u is the spatially dependent error term, and W is the matrix of spatial weights (Fortin and Dale 2005; Bivand et al. 2008). Thus, the SAR error model assumes that SAC is to be found in the error term because of either inherent or induced SAC, whereas the SAR lagged model assumes that SAC is a property of the response variable because of inherent

SAC. A study comparing SAR models (Kissling and Carl 2008) recommends the SAR error model as the most reliable model in terms of precision of parameter estimates, SAC reduction, and type I error control. Once the SAR model is fitted, the last step is checking the model assumptions. A specific model assumption check is to test whether the residuals of the SAR model are spatially autocorrelated. Because any spatial structure in the residuals can be indicative of some nonmodeled spatial structure in the data, careful examination of the residuals should also enable the detection of misspecified models.

CONCLUSION

There is no doubt that SAC is an important concept as has been widely acknowledged in several areas of ecological research in the last decades. Behavioral ecologists can benefit by assimilating the tools and the concepts developed in spatial ecology, among which SAC is of central importance. Data sets collected by behavioral ecologists should therefore be kept spatially explicit by recording the geographical coordinates associated to each observation. SAC is multidirectional and operates at multiple scales, so the effects it will have on an empirical data set are difficult, if not impossible, to predict. We suggest that testing for SAC, both as an exploratory exercise and during statistical modeling, should be a standard method to append to the current statistical toolset of field behavioral ecology.

Address correspondence to M. Valcu. E-mail: valcu@orn.mpg.de.

REFERENCES

Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J. 2008. Estimating space-use and habitat preference from wildlife telemetry data. Ecography. 31:140–160.

Bivand R, Pebesma EJ, Gómez-Rubio V. 2008. Applied spatial data analysis with R. New York: Springer.

Cliff AD, Ord JK. 1973. Spatial autocorrelation. London: Pion.

Campbell DJ. 1992. Nearest-Neighbour graphical analysis of spatial pattern and a test for competition in populations of singing crickets (*Teleogryllus commodus*). Oecologia. 92:548–551.

Dormann CF. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob Ecol Biogeogr. 16:129–138.

Duraes R, Loiselle BA, Blake JG. 2007. Intersexual spatial relationships in a lekking species: blue-crowned manakins and female hot spots. Behav Ecol. 18:1029–1039.

Dutilleul P. 1993. Modifying the t-test for assessing the correlation between 2 spatial processes. Biometrics. 49:305–314.

Fortin M-J, Dale MRT. 2005. Spatial analysis: a guide for ecologists. Cambridge: Cambridge University Press.

Getis A. 2008. A history of the concept of spatial autocorrelation: a geographer's perspective. Geogr Anal. 40:297–309.

Giesselmann UC, Wiegand T, Meyer J, Vogel M, Brandl R. 2008. Spatial distribution of communal nests in a colonial breeding bird: benefits without costs? Austral Ecol. 33:607–613.

Guillot G, Leblois R, Coulon A, Frantz AC. 2009. Statistical methods in spatial genetics. Mol Ecol. 18:4734–4756.

Haining R. 1990. Spatial data analysis in the social and environmental sciences. Cambridge: Cambridge University Press.

Hill AGaJ. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press.

Holdo RM, Holt RD, Fryxell JM. 2009. Opposing rainfall and plant nutritional gradients best explain the wildebeest migration in the Serengeti. Am Nat. 173:431–445.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecol Monogr. 54:187–211.

van der Jeugd HP, McCleery R. 2002. Effects of spatial autocorrelation, natal philopatry and phenotypic plasticity on the heritability of laying date. J Evol Biol. 15:380–387.

Kissling WD, Carl G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Glob Ecol Biogeogr. 17:59–71.

Laiolo P, Tella JL. 2006. Landscape bioacoustics allow detection of the effects of habitat patchiness on population structure. Ecology. 87:1203–1214.

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology. 74:1659–1673.

Legendre P, Dale MRT, Fortin MJ, Casgrain P, Gurevitch J. 2004. Effects of spatial structures on the results of field experiments. Ecology. 85:3202–3214.

Legendre P, Dale MRT, Fortin MJ, Gurevitch J, Hohn M, Myers D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography. 25:601–615.

Legendre P, Fortin MJ. 1989. Spatial pattern and ecological analysis. Vegetatio. 80:107–138.

Lennon JJ. 2000. Red-shifts and red herrings in geographical ecology. Ecography. 23:101–113.

Maher CR, Lott DF. 1995. Definitions of territoriality used in the study of variation in vertebrate spacing systems. Anim Behav. 49:1581–1597.

Valcu M, Kempenaers B. 2010. Is spatial autocorrelation an intrinsic property of territory size? Oecologia. 162:609–615.