ECOGRAPHY 27: 767–775, 2004

Estimation of the spatial autocorrelation function: consequences of sampling dynamic populations in space and time

Patrick C. Tobin

Tobin, P. C. 2004. Estimation of the spatial autocorrelation function: consequences of sampling dynamic populations in space and time. – Ecography 27: 767–775.

The estimation of spatial autocorrelation in spatially- and temporally-referenced data is fundamental to understanding an organism's population biology. I used four sets of census field data, and developed an idealized space-time dynamic system, to study the behavior of spatial autocorrelation estimates when a practical method of sampling is employed. Estimates were made using both a classical geostatistical approach and a recently developed non-parametric approach. In field data, the estimate of the local spatial autocorrelation (i.e. autocorrelation as the distance between pairs of sampling points approaches 0), was greatly affected by sample size, while the range of spatial dependence (i.e. the distance at which the autocorrelation becomes negligible) was fairly stable. Similar patterns were seen in the theoretical system, as well as greater variability in local spatial autocorrelation during the invasion stage of colonization. When sampling for the purposes of quantifying spatial patterns, improved estimates of spatial autocorrelation may be obtained by increasing the number of pairs of points that are close in space at the expense of attempting to cover the entire region of interest with equidistant sampling points. Also, results from the theoretical space-time system suggested that greater resolution in sampling may be required in newly establishing populations relative to those already established.

P. C. Tobin (ptobin@fs.fed.us), USDA Forest Service, Northeastern Research Station, 180 Canfield St., Morgantown, WV 26505-3180, USA.

The sampling of animal and plant populations and the associated spatial and temporal uncertainty is a fundamental issue for population biology. The ongoing development of computer and satellite technologies have continued to facilitate complex space-time analyses of populations (Rossi et al. 1992, Legendre 1993, Liebhold et al. 1993, Bascompte and Solé 1998, Bjørnstad et al. 1999, Liebhold and Gurevitch 2002, Legendre et al. 2002, Perry et al. 2002), and more sophisticated and elegant spatial statistical techniques have virtually replaced the mean-to-variance relationship approach (Southwood 1978) that once dominated the analysis of spatial distributions. Consequently, scientists across disciplines are quantifying spatial and temporal structure in increasingly more explicit detail.

However, when estimating spatial autocorrelation functions, a widely used statistical tool, there can be considerable uncertainty regarding their expected form and function when applied to dynamic space-time ecological processes (Bjørnstad and Falck 2001). After all, geostatistics, the theoretically- and empirically-robust approach and perhaps the most common spatial statistical tool in ecology (Rossi et al. 1992, Liebhold et al. 1993), was largely based on the spatial distribution of minerals and ores, which, compared to biological populations, is quite stable over time (Matheron 1963, Isaaks and Srivastava 1989). In some ways, the increase

Accepted 12 July 2004

ISSN 0906-7590

in the availability of user-friendly spatial statistics software has been both a benefit and bane to ecologists, as they permit spatial estimation for almost any set of data, including those for which the sample sizes may be inadequate to appropriately estimate spatial autocorrelation. This problem could be furthermore compounded when data consisting of small sampling sizes are used in the construction of spatially interpolated maps, in which these small sample sizes compound the problem of errors in predictions from smoothed surfaces.

For example, the statistical properties of the covariance function was studied by Stein (1988), who observed that if the sampling size was sufficient enough to estimate a covariance function that was compatible to the actual one, then there were negligible effects on predictions from kriging, a common spatial interpolation method (cf. Isaaks and Srivastava 1989). On the other hand, Stein and Handcock (1989) observed non-negligible effects on kriging predictors at distances near the range of the spherical covariance function when the covariance function was misspecified. Meuli et al. (1998) observed that the kriging variances of soil pollutants underestimated the actual variance of the spatial process, resulting in differences between the interpolated maps and actual values. Fortin et al. (1989) examined the effect of sample design and size on estimating the spatial structure in sugar maple *Acer saccharum* and observed that design was more influential than size in estimating spatial functions and using kriging in map generation.

Of course, it is difficult to assess the errors associated with and the potential instability of spatial autocorrelation estimates in populations without the benefit of robust benchmarking studies. One approach for optimizing sample size is to conduct an exhaustive spatial survey (Isaaks and Srivastava 1989). However, the harsh reality is that this is costly and hence a road not well, if at all, taken in field research. Even if such an endeavour was realized to the point where one felt comfortable about quantifying spatial variation in one population at one point in time, one would still have to address how spatial uncertainty manifests itself over temporal scales. Many ecological processes, such as those involving animal populations, are extremely dynamic in both space and time, and as a result, an exhaustive spatial survey at one point in time may be of little extended value.

For example, the spatial interpolations of insect abundance are prerequisite tools in site-specific pest management, whether the target populations are located at small spatial scales, such as nearby agricultural fields (Weisz et al. 1996, Blom and Fleischer 2001), or across large regions (Sharov et al. 1996). In site-specific management, insect populations are managed differentially at both spatial and temporal scales. There is generally wide acceptance of the benefits of this concept, whose benefits may include reductions in pesticide applications and consequent fewer resistance pressures on the targeted population (Fleischer et al. 1999) However, the implementation of site-specific plans still requires considerable sampling effort and thus a need for the optimization of sampling plan. In the past, the National Research Council (Anon. 1997), for example, has suggested that interpolated maps of mobile organisms, such as insects, may not be practical in site-specific management because of the temporal instability of spatial dynamics.

The issue of uncertainty in the analysis of ecological spatial data has motivated a number of studies, including those that address the effect of sample size on spatial autocorrelation estimation (Fortin et al. 1989, Bellehumeur and Legendre 1997, Bellehumeur et al. 1997, Webster and Oliver 2001, Dungan et al. 2002). In this study, I studied the effect of sample size on spatial estimation by using four sets of spatial census data as motivating examples. Each was collected at one point in time during a season, and because each was a census of a biological population, they provided an opportunity to provide a benchmarking measurement of spatial autocorrelation. I estimated the spatial autocorrelation function using the census data and then followed this analysis with estimate of the spatial autocorrelation function for a series of randomly-selected subsets of the census data. Furthermore, I used both a classical geostatistical approach (Isaaks and Srivastava 1989) and a recently developed non-parametric approach (Bjørnstad and Falck 2001). The latter method does not assume any functional form a priori, nor bin the data into distance classes, and thus may provide new insights on the effects of sample size in spatial data analyses. Lastly, I used a simplified idealized space-time insect system and used the non-parametric approach to examine the temporal properties of the spatial autocorrelation function within a theoretical context.

Methods

Field data collection

The first set of data was collected from a 50×100 m plot in Delaware City, New Castle Co., DE, USA during the winter of 1987–1988. The plot was characterized by a sandy dredging spoil habitat, and was adjacent to the Cheakespeake and Delaware Canal. The plot was evenly and completely divided into 2×2 m sampling quadrats. Within each quadrat ($N = 1250$), a census of the number of grass clumps, which are used by several species of coccinellids as overwintering habitats, was recorded, and for each clump, the diameter at the base was also measured. The total number of sevenspotted lady bird beetles (C7), *Coccinella septempunctata*,

overwintering in each clump was counted, from which standardized counts based on the m^2 of available overwintering habitat within each 2×2 m quadrat was calculated. Each quadrat was spatially-referenced, and these data are thus a spatial representation of the pattern of colonization by grasses and the adult coccinellids that use the grass habitat for overwintering.

The next set of data was collected from a 32×216 m plot within a Concord grape variety, *Vitis labruscana*, vineyard in the town of North East, PA, USA. A census of the number of oviposition scars on each grape vine ($N = 1092$) of the grape cane gallmaker (GCG), *Ampeloglypter sesostris*, was conducted in the winter of both 1985 and 1986 (Saunders and Tobin 2000). These data thus represent the spatial patterns of oviposition for two years in a perennial crop that was planted in a systematic pattern for agricultural purposes.

For each census, data were transformed using a $\log_{10}(Z+1)$ transformation and used to estimate spatial autocorrelation. Using the census data, I used two approaches to estimate spatial autocovariance: the conventional geostatistical spatial autocorrelogram (cf. Isaaks and Srivastava 1989, Deutsch and Journel 1998) and a novel non-parametric spatial covariance function (Bjørnstad and Falck 2001). In the former method, I used an exponential function (Deutsch and Journel 1998) to model the correlogram values at each lag distance class, ρ_h,

$$\rho_h = C\, e^{\left(\frac{-3h}{A}\right)}, \tag{1}$$

where C is an estimate of the correlation between samples as the distance between samples approaches 0, and A is an estimate of the range of spatial continuity (e.g., the lag distance, h, at which the correlogram is 95% of the asymptote) (PROC NLIN, Anon. 1999).

The non-parametric approach (Bjørnstad and Falck 2001) uses a smoothing spline to measure the correlation between the density of pairs of samples over a continuous function of the distance separating samples, without assuming any functional form a priori and without binning the data into distance classes. Let δ_{ij} be the Euclidean distance between spatial locations i and j, and $\rho_{ij,t}$ be the spatial correlation in abundance, z_i and z_j, at the two locations at time t,

$$\rho_{ij,t} = \frac{\left(z_{i,t} - \bar{z}_t\right)\left(z_{j,t} - \bar{z}_t\right)}{\frac{1}{M}\sum_{a=1}^{M}\left(z_{a,t} - \bar{z}_t\right)^2}, \tag{2}$$

where M is the total number of local populations and $\bar{z}_t$ represents the field-wide mean abundance at time t, so that the denominator represents the spatial variance of the population. Letting $\rho_t(\delta)$ be the expected spatial correlation between abundances as a function of distance, δ, the spatial covariance function can be estimated from non-parametric regression of $\rho_{ij,t}$ against δ_{ij} according to

$$\rho_t(\delta) = \frac{\sum_{i=1}^{M}\sum_{j=i+1}^{M} K\left(\frac{\delta_{ij}}{h}\right)\rho_{ij,t}}{\sum_{i=1}^{M}\sum_{j=i+1}^{M} K\left(\frac{\delta_{ij}}{h}\right)}, \tag{3}$$

where K is a kernel function with bandwidth, h, that defines curve smoothness (Härdle 1990, Hall and Patil 1994). The spline function was used as an equivalent kernel, and degrees of freedom based on the square root of M for the spline estimation. This approach also incorporates a bootstrap method to generate CI around the estimated non-parametric covariance function, and the CI was taken as the 2.5 and 97.5 percentiles of the bootstrap distribution based on 1000 replications (Efron and Tibshirani 1993, Bjørnstad and Falck 2001).

The spatial autocorrelation function was expected to change as sample size decreased. To address this issue, I used both the geostatistical and non-parametric approaches to estimate spatial autocorrelation when 75, 50, 25, 10, and 5% of the total number of sampling points in each census data set were used (Table 1). For each percentage, subsampled sets of data were randomly chosen and used to estimate autocorrelation. This was repeated 1000 times to derived Monte Carlo distributions, which were then used to summarize patterns of spatial covariance.

The primary objective of this paper is to examine the properties of spatial autocorrelation when sample size decreases, as opposed to compare explicitly the geostatistical and non-parametric approaches. Thus, for the geostatisical approach, I summarized patterns of local spatial autocorrelation for each subset using the estimates at the first lag distance class. For the non-parametric approach, which does not rely on binning, I used the estimate of $\rho_t(0)$ (cf. eq. 3) as an indicator of the local spatial autocorrelation, while the range was the distance at which the spline function estimated a value of 0.05 [Note, in this case, "local" spatial autocorrelation does not refer to local measures or indicators of spatial autocorrelation but to the estimate of spatial autocorrelation as the distance between pairs of sampled values approaches 0.]. In both cases, the Monte Carlo distibution of 1000 simulations was used to generate mean and 95% CI (using the 2.5 and 97.5 percentiles of the distribution) for each subset. Because of solid work by others demonstrating relative performance of different sample arrangements on spatial autocorrelation estimation (i.e. square versus hexagonal grids) (Olea 1984, Yfantis et al. 1987, Schotzko and O'Keefe 1990, Williams et al. 1992, Wollenhaupt et al. 1997), the focus of this paper was strictly on the changes in estimation and variability of the spatial autocorrelation function as sample size decreases relative to a census.

Table 1. Sampling characteristics and details for field data.

Percent of samples	Grass clumps/C7 adults		Grape cane gallmaker	
	Sample size	# pairs within 2.0–2.5 m Bin[a]	Sample size	# pairs within 2.0–2.5 m Bin[a]
100	1250	4777	1092	2081
75	938	2686	819	1170
50	625	1194	546	520
25	312	299	272	130
10	125	45	109	21
5	62	12	55	5

[a]Number of pairs are the mean from 1000 subsets of randomly selected points corresponding to the percent of samples.

Simulations

To explore the properties of the spatial autocorrelation function within a theoretical context, I used a stochastic Ricker function to simulate the spatial and temporal dynamics of an idealized population. Local pre-dispersal dynamics, N', at time t and spatial location i were

$$N'_{i,t+1} = \{(N_{i,t}) \exp\{r(1 - N_{i,t}) - (U_{i,t})\}, \quad (4)$$

where r is growth rate (Murray 1993), and U is a sequence of independent zero-mean Gaussian random variables with variance $=0.5$. The growth rate was fixed at 0.25 to represent biologically meaningful population growth during linear dynamics. The stochastic component was assumed to be spatially correlated but temporally independent, and represented a range of effects (e.g., random variability in host quality, predation, and abiotic stress). To simulate a population during the colonization phase of its dynamics, which is particularly relevant to transient systems (Hastings and Higgins 1994, Hastings 2001, 2004, Tobin and Bjørnstad 2003), 30 generations of spatiotemporal dynamics were simulated over a 30×30 coupled map lattice with absorbing boundaries. Abundance was thus represented as 900×1 matrices, $\underline{N}$. Following local dynamics, 25 or 75% of the individuals were assumed to disperse to the four adjacent cells according to

$$\underline{N}_{t+1} = D \times \underline{N}'_t, \quad (5)$$

where D is the 900×900 dispersal matrix with 0.25 (or 0.75) along the diagonal and 0.1875 (or 0.0625) at entries linking neighbors, and $\times$ denotes matrix multiplication. The two rates of dispersal were used as starting points for investigating the effect of dispersal on spatial and temporal stability of autocorrelation estimates, and were also chosen to represent different periods of transience.

Initial populations in each cell were randomly distributed at low abundance according to a uniform distribution from [0,1], and initial abundance was furthermore random over $\sim$5% of the cells and zero elsewhere to simulate the colonization phase. An example of the dynamics of this idealized system is shown in Fig. 1. Using the non-parametric approach (eqs 2–3, Bjørnstad and Falck 2001), the spatial autocorrelation at each generation and for each rate of dispersal was estimated for the entire 30×30 coupled map lattice, as well as for a random selection of 75, 50, 25, 10, and 5% of the total number of samples. The simulation was run 1000 times to derive Monte Carlo distributions for use in mean estimation and generating CI.

Results

The spatial patterns of the census field data sets are shown in Fig. 2. The colonization of grasses into a sandy dredging spoil habitat in Delaware shows distinct

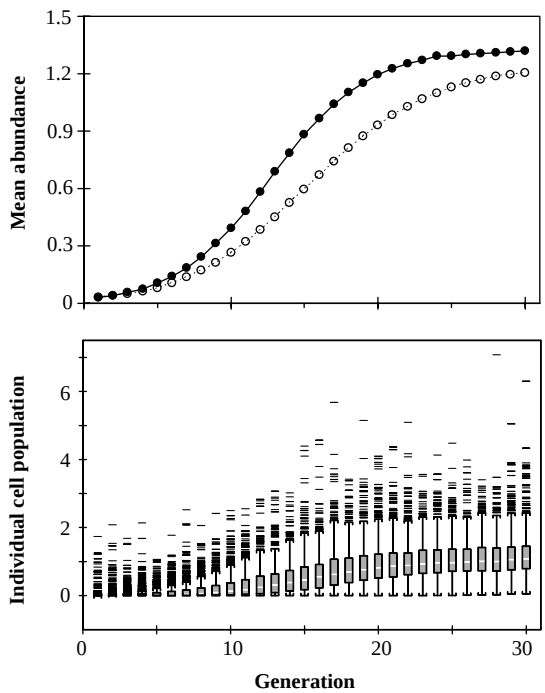

Fig. 1. Stochastic Ricker dynamics, showing mean post-dispersal abundance over the 30×30 coupled map lattice when 75% (solid circles and line) and 25% (open circles and dashed line) of individuals disperse (top graph). The bottom graph shows exemplified box-plots of the post-dispersal abundance of individualized cells when 25% disperse.

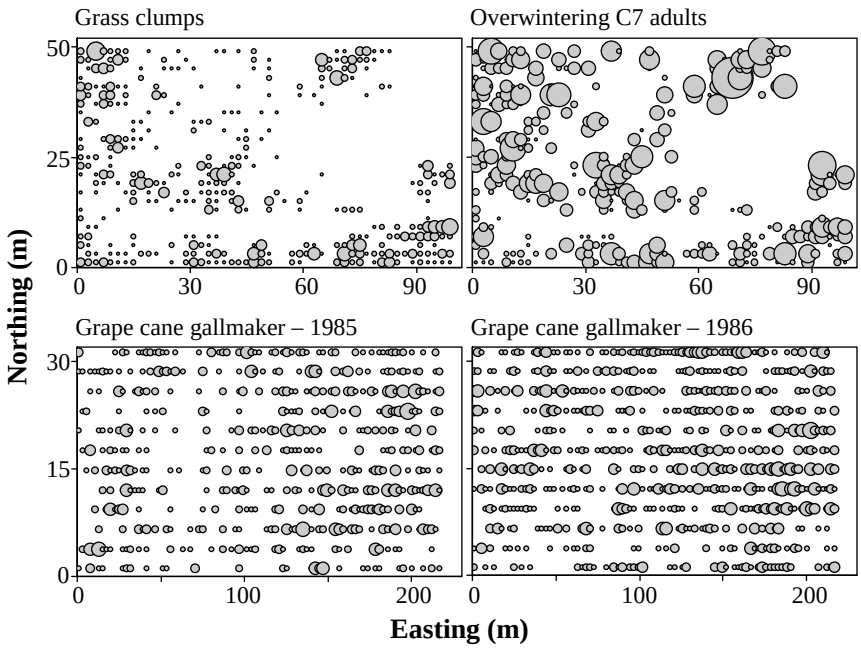

Fig. 2. Proportional spatial bubble plots of the numbers of grass clumps and overwintering C7 adults (top row), and grape cane gallmaker ovipositional scars (bottom row). Data are proportional for plots within each row. Ranges: grass clumps, 0–20; C7 adults, 0–96; GCG 1985, 0–9; GCG 1986, 0–10.

aggregations, as is the case with the *C. semptempuctata* adults who use the grasses as overwintering habitat. Grape cane gallmaker spatial patterns were far less distinct, though aggregations still existed. Estimates of the local spatial autocorrelation (i.e. correlation at lag distance =0) and the range of spatial dependence (i.e. lag distance at which correlation =0) in the census data from both the geostatistical correlogram and non-parametric covariance function is presented in Table 2, and can be considered benchmarking values when subsets of the census are used in spatial autocorrelation estimation.

Using the geostatistical approach, mean and CI values of the spatial autocorrelation estimate at the first distance class bin are shown for each subset in Fig. 3, as well as the corresponding value when using the census data. Also, the consequences within this first distance class bin when sample sizes are reduced are listed in Table 1. [Since the error in estimation as sample size decreased was most prominent near the origin, and for simplicity in reporting, the number of pairs within the 2.0–2.5 m distance class are shown for both the grass clump/*C. septempunctata* and grape cane gallmaker data.] For each set of data, the mean estimate from the Monte Carlo distributions is fairly stable though there is a marked increase in the variability as sample size decreases. This increase in variability in subset data was also observed when estimating local autocorrelation using the non-parametric approach (Fig. 4, inserts). In

Table 2. Benchmarking spatial statistical details of the local spatial autocorrelation correlation function and range of spatial dependence from census field data using the geostatistical and non-parametric approaches.

Census data set	Local correlation (CI)	Spatial range (CI)
Method = Geostatistical correlogram (Isaaks and Srivastava 1989)		
Grass clumps	0.53 (0.42, 0.64)	15.5 (12.3, 18.8) m
C7 Adults	0.30 (0.25, 0.36)	17.5 (13.9, 21.2) m
GCG 1985	0.13 (0.08, 0.17)	42.4 (21.3, 63.6) m
GCG 1986	0.25 (0.15, 0.36)	15.0 (7.1, 22.9) m
Method = Non-parametric Covariance Function (Bjørnstad and Falck 2001)		
Grass clumps	0.52 (0.35, 0.71)	22.6 (17.1, 26.5) m
C7 Adults	0.33 (0.21, 0.46)	23.2 (16.3, 36.4) m
GCG 1985	0.10 (0.05, 0.17)	66.7 (33.2, 63.8) m
GCG 1986	0.16 (0.10, 0.22)	70.6 (18.6, 66.1) m

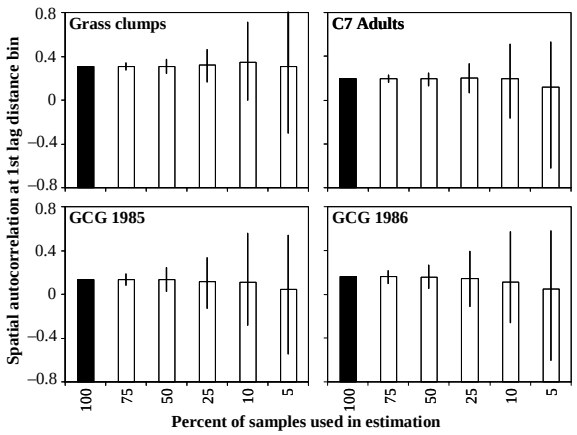

Fig. 3. Plots of the estimate of the geostatistical correlogram at the first distance class bin using the census data (black bars), and estimates from 75, 50, 25, 10, and 5% subsets of samples. For the subsets, the graphs depict the mean and CI from the Monte Carlo distribution of 1000 replications.

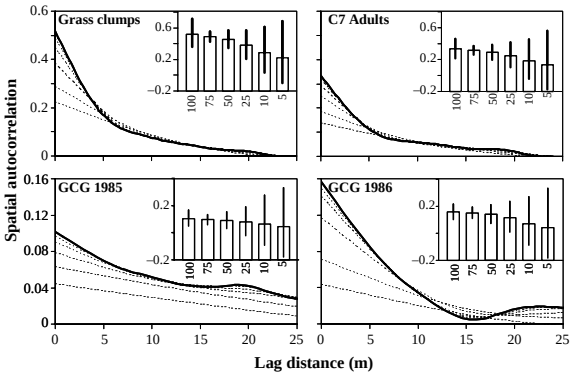

Fig. 4. Plots of the non-parametric spatial autocorrelation function for each set of field data. The thick line represents the estimation from census data, while the dashed lines represent the estimates from 75, 50, 25, 10, and 5% subsets of samples in decreasing order from the census estimate. The insert graphs depict the mean and CI for the local spatial autocorrelation (i.e. correlation at lag distance =0) for the census (100) and each subset.

addition, the mean estimate of the local autocorrelation decreased in concert with decreasing sample size though the mean estimate of the spatial range was more stable across subsets (Fig. 4).

Because of similar results between the geostatistical and non-parametric approaches in estimating the local spatial autocorrelation in census data (Table 2), and in the qualitative dynamics (i.e. variability) of these values in subset data, the non-parametric approach was used in the theoretical space-time system in investigating spatial covariance properties. Also, because this approach does not bin the data, nor rely on a priori functional forms in estimating spatial patterns, it is more tractable in estimating and summarizing the data from 1000 replicated simulations in which spatial locations are iteratively and randomly selected.

Expectations of the local spatial autocorrelation through time realized from simulations in which 25% of individuals disperse at each time step are shown in Fig. 5. For this stochastic Ricker model, the behavior of the spatial function revealed an increase in local structure in conjunction with initial colonization, and then a gradual decrease in structure over the course of the invasive phase of population growth. Similar patterns were seen when 75% of individuals disperse (Fig. 6).

Estimates of local correlation from both dispersal fractions showed that some subsets tended to overestimate mean spatial structure (75, 50, and 25%), while subsets of smaller sample sizes (10, 5%) underestimated local correlation. Moreover, the variance around the local spatial correlation function differed greatly among subsets (Figs 5 and 6), and for the subset data, the variances were less stable – to varying degrees – during initial colonization. In contrast, the estimates and variability of spatial range among the subsets were more likely to be stable, particularly during the invasive phase of dynamics. Once exception was the use of 5% of the lattice data – particularly when the dispersal fraction was 0.75 – that tended to be visibly more variable than the other subsets (Fig. 6). The consequences of reduced sample sizes can be furthermore examined by assessing the number of distinct pairs in lag distance classes near the origin. For the field data, this was previously reported in Table 1. As a companion to this table, these consequences are shown in Fig. 7, using the 30×30 lattice as an example in which the total number of quadrats is 900 and quadrats are separated by 1 unit of distance.

Discussion

Degradation in the level of spatial aggregation in field data occurred when subsets of samples were used in lieu of a census (Figs 3 and 4). The most common feature is that for all data sets, the degradation was largely in the local spatial autocorrelation, with the level of degradation inversely related to the number of samples, while the spatial ranges tended to be more stable across sample size. Moreover, when using 10% of the census data, for example, sample sizes were >100 for all sets, while the average minimum distance between pairs of sampling locations was ≈ 4.5 m, which is not unreasonable given the labor costs of spatial sampling in field studies.

In addition to the general underestimation of local autocorrelation in subset data, the variability around the estimate increased with decreased samples (Figs 3 and 4). The consequences of this variability can be severe. For example, in all cases, using 5% of the data resulted in an underestimation such that the local correlation was insignificant (i.e. the CI from the Monte Carlo distributions included 0). This is not trivial, as when estimates of the local autocorrelation are not significant, the variable cannot be considered regionalized (Matheron 1963); thus, the process is essentially spatially random and should not be consequently mapped using spatial interpolation, such as kriging. Moreover, due to the variability, it is also possible to overestimate local spatial structure.

The empirical and theoretical components of this study provided overlapping observations. Estimates of the spatial range in both field and theoretical data generally tended to be stable as sample size decreased (Figs 4–6), as was, for the most part, the variability in spatial range estimates over time from the theoretical system (Figs 5 and 6). However, estimates and the variability associated with the estimates of local spatial autocorrelation were greatly affected by sample size (Figs 3–6). These results suggest important consequences in spatially- and temporally-referenced sampling

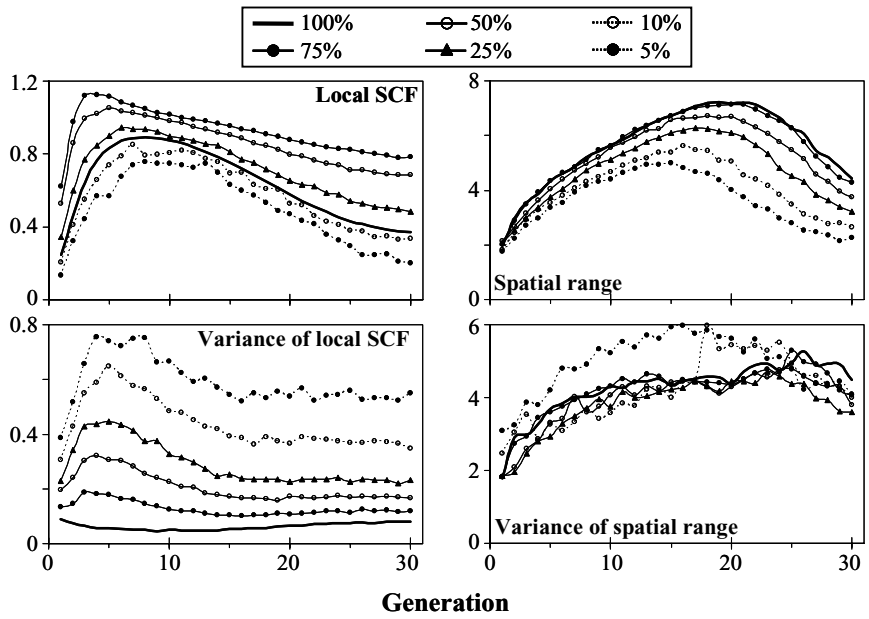

Fig. 5. Behavior of the local non-parametric spatial autocorrelation function and range of spatial dependence (top row) and their corresponding variance (bottom row) using the entire coupled map lattice and subsets, while assuming that 25% of individuals disperse to adjacent cells.

efforts of dynamic populations. The most critical aspect is the error in estimation of the local spatial autocorrelation estimate. Thus, in the optimization of spatial sampling protocols, it may be more practical to increase the number of pairs of points that are close in space at the expense of attempting to cover the entire region of interest with equidistant sampling points.

Such a sampling approach could be accomplished by incorporating sampling subregions, in which sampling points within the subregions are closer in space than the subregions are to each other (Pettitt and McBratney 1993, Blom and Fleischer 2001). In turn, the subregions can be arranged in a manner so that the entire region of interest is reasonably represented. Thus, the local spatial structure is captured without compromising the ability to detect spatial trends (cf. Rossi et al. 1992) across the region. There are several examples of this type of sample design. Blom and Fleischer (2001) used a systematic cluster design to study of the spatial dynamics of Colorado potato beetle *Leptinotarsa decemlineata*. Fortin et al. (1989) used a similar approach to quantify spatial patterns in *A. saccharum*, in which a

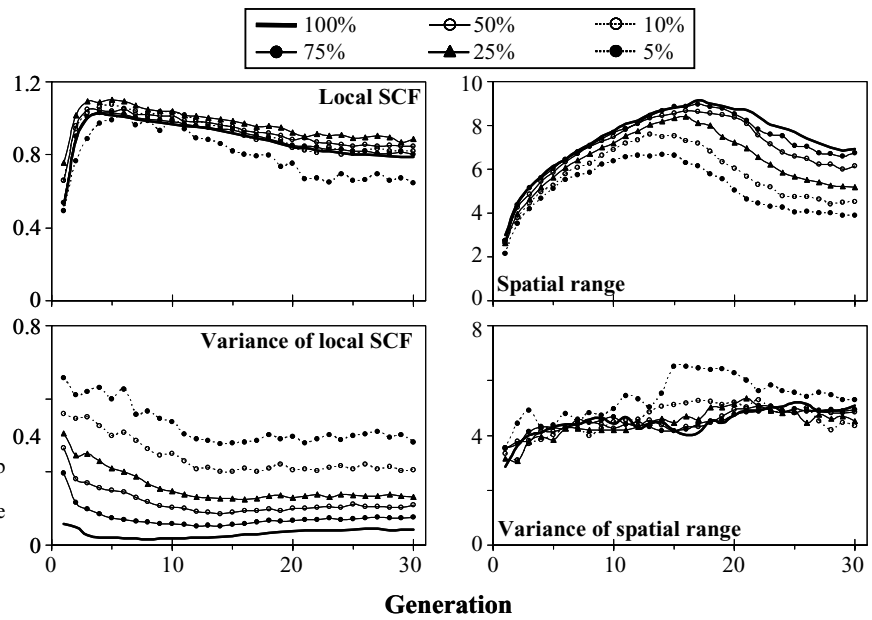

Fig. 6. Behavior of the local non-parametric spatial autocorrelation function and range of spatial dependence (top row) and their corresponding variance (bottom row) using the entire coupled map lattice and subsets, while assuming that 75% of individuals disperse to adjacent cells.

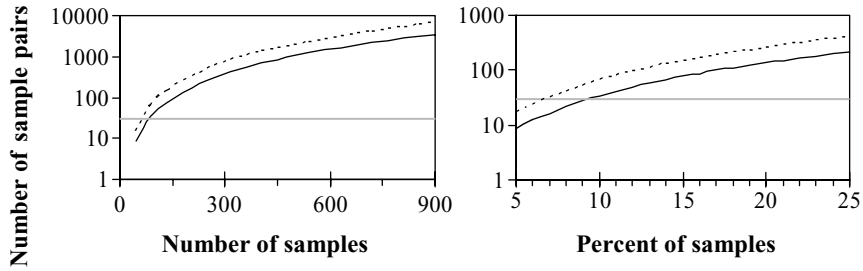

Fig. 7. Relationship between the number and percent of samples over a 30 × 30 lattice (N = 900), and the number of distinct sample pairs in the first (solid line) and second (dashed line) distance classes (1 and 2 units of distance, respectively). Number of pairs is the mean from 1000 subsets of randomly selected points corresponding to the number or percent of samples. The thin horizontal line represents 30, or the minimum number of distinct pairs of sample points in each distance class required to measure spatial autocorrelation (cf. Journel and Huijbregts 1978, Cressie 1993, Liebhold et al. 1993).

systematic cluster design seemed to better quantify spatial autocorrelation than a systematic approach that covered the region of interest with equidistant sampling points. Also, a recent study of the spatial estimation and mapping of aeolian sediment transport showed that a nested sampling approach, in which sampling locations are spaced along transects and nested according to an a priori understanding of the spatial scale of the study system, was superior to both grid and random designs in capturing the spatial variation of the population (Chappell et al. 2003). Indeed, Legendre et al. (2002) were motivated by the potential use of existing knowledge of spatial patterns in explanatory variables in the optimization of sampling plans. [Although the design employed in this study used a random selection of sampling points, the results were derived from Monte Carlo distribution of 1000 simulations for the purposes of quantifying spatial autocorrelation and the associated variability as opposed to a one-time random selection of sampling points.]

The importance of the local spatial structure should not be overlooked, as the spatial statistical literature suggests that the behavior of the spatial autocorrelation function near the origin (i.e. the local spatial estimate) is the most critical component in kriging and other spatial interpolation methods (Cressie 1993, Bailey and Gatrell 1995). The failure to capture the local structure can result in the misspecification of the covariance structure for which Stein and Handcock (1989) reported critical consequences in kriging estimation. Furthermore, theoretical results from the idealized system suggest that the local spatial structure during the invasive phase of establishing populations are subject to increased temporal instability (Figs 5–6), thus requiring more spatially- and temporally robust sampling regimes relative to established populations. This can be particularly problematic in, for example, pest management because many pest decisions target the colonization phase of the pest. In contrast, the estimates and variability of spatial range among the subsets were more likely to be stable, particularly during the invasive phase of pest dynamics. This concept can apply to many types of transient systems, such as agricultural systems, most of which are relatively short-lived and involve seasonal reestablishment of the community, and natural systems in which communities may continuously re-establish themselves due to disturbances or phenology.

For geostatistical applications, including the use of kriging, the optimal number of samples and the number of sample pairs within each lag distance class is often specific to the study design. Regardless of the target organism, however, past work in spatial statistics applications indicates that at least 30 distinct pairs of sample points should be in each distance class to measure spatial autocorrelation (Journel and Huijbregts 1978, Cressie 1993), and Liebhold et al. (1993) recommended 30–50 pairs when applying geostatistics to insect ecology. In field data, sample sizes of 55 or 62 (i.e. 5% of the available sampling points, Table 1) would not be sufficient in spatial autocorrelation estimate, with the former only allowing for 5 distinct pairs. For grape cane gallmaker, using 109 samples (10%) only provided 21 pairs, still less than the 30 recommended.

Spatial analyses of the abundance of populations through time can provide valuable insight into population ecology and consequently enhance our knowledge of a population's interaction with its environment and other organisms. Such analyses are also paramount in the development of management strategies. Although the incorporation of complex space-time analyses within the context of population studies has greatly improved our understanding, the limitations and challenges of these approaches must be recognized by the researcher during the development and implementation of spatial surveys. The gathering of spatial data that is in turn temporally-referenced is a formidable task that may not always yield desirable conclusions, but the optimization of such surveys, particularly in manner that focuses increased effort on the local spatial structure, may improve our ability to describe and explain the ecological consequences in dynamic systems.

Acknowledgements – I thank Paul Blom (USDA Agricultural Research Service, Prosser, WA, USA) and Andrew Liebhold (USDA Forest Service, Morgantown, WV, USA) for their critical comments in preparation of this manuscript, and Paul Schaefer (USDA Agricultural Research Service, Newark, DE, USA) for access to the *C. septempunctata* and grass clump data.

References

Anon. 1997. Precision agriculture in the 21st century. Geospatial and information technologies in crop management. – National Research Council, National Academy Press.

Anon. 1999. SAS/STAT user's guide, ver. 8. – SAS Inst.

Bailey, T. C. and Gatrell, A. C. 1995. Interactive spatial data analysis. – Longman Group.

Bascompte, J. and Solé, R. V. 1998. Spatiotemporal patterns in nature. – Trends Ecol. Evol. 13: 173–174.

Bellehumeur, C. and Legendre, P. 1997. Aggregation of sampling units: an analytical solution to predict variance. – Geogr. Anal. 29: 258–266.

Bellehumeur, C., Legendre, P. and Marcotte, D. 1997. Variance and spatial scales in a tropical rain forest: changing the size of sampling units. – Plant Ecol. 130: 89–98.

Bjørnstad, O., Ims, R. A. and Lambin, X. 1999. Spatial population dynamics: analyzing patterns and processes of population synchrony. – Trends Ecol. Evol. 11: 427–431.

Bjørnstad, O. N. and Falck, W. 2001. Non-parametric spatial covariance functions: estimating and testing. – Environ. Ecol. Stat. 8: 53–70.

Blom, P. E. and Fleischer, S. J. 2001. Dynamics in the spatial structure of *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). – Environ. Entomol. 30: 350–364.

Chappell, A. et al. 2003. Simulations to optimize sampling of aeolian sediment transport in space and time for mapping. – Earth Surf. Processes 28: 223–1241.

Cressie, N. A. C. 1993. Statistics for spatial data. – Wiley.

Deutsch, C. V. and Journel, A. G. 1998. GSLIB: geostatistical software library and user's guide, 2nd ed. – Oxford Univ. Press.

Dungan, J. L. et al. 2002. A balanaced view of scale in spatial statistical analysis. – Ecography 25: 626–640.

Efron, B. and Tibshirani, R. J. 1993. An introduction to the bootstrap. – Chapman and Hall.

Fleischer, S. J., Blom, P. E. and Weisz, R. 1999. Sampling in precision IPM: when the objective is a map. – Phytopathology 89: 1112–1118.

Fortin, M.-J., Drapeau, P. and Legendre, P. 1989. Spatial autocorrelation and sampling design in plant ecology. – Vegetatio 83: 209–222.

Hall, P. and Patil, P. 1994. Properties of nonparametric estimation of autocovariance for stationary random fields. – Probab. Theory Rel. 99: 399–424.

Härdle, W. 1990. Applied non-parametric regression. – Cambridge Univ. Press.

Hastings, A. 2001. Transient dynamics and persistence of ecological systems. – Ecol. Lett. 4: 215–220.

Hastings, A. 2004. Transients: the key to long-term ecological understanding? – Trends Ecol. Evol. 19: 39–45.

Hastings, A. and Higgins, K. 1994. Persistence of transients in spatially structured ecological models. – Science 263: 1133–1136.

Isaaks, E. H. and Srivastava, R. M. 1989. An introduction to applied geostatistics. – Oxford Univ. Press.

Journel, A. G. and Huijbregts, C. J. 1978. Mining geostatistics. – Academic Press.

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

Legendre, P. et al. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. – Ecography 25: 601–615.

Liebhold, A. M. and Gurevitch, J. 2002. Integrating the statistical analysis of spatial data in ecology. – Ecography 25: 553–557.

Liebhold, A. M., Rossi, R. E. and Kemp, W. P. 1993. Geostatistics and geographical information systems in applied insect ecology. – Annu. Rev. Entomol. 38: 303–327.

Matheron, G. 1963. Principles of geostatistics. – Econ. Geol. 58: 1246–1266.

Meuli, R., Schulin, R. and Webster, R. 1998. Experience with the replication of regional survey of soil pollution. – Environ. Pollut. 101: 311–320.

Murray, J. D. 1993. Mathematical biology, 2nd ed. – Springer.

Olea, R. A. 1984. Sampling design optimization for spatial functions. – Math. Geol. 16: 369–392.

Perry, J. N. et al. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. – Ecography 25: 578–600.

Pettit, A. N. and McBratney, A. B. 1993. Sampling designs for estimating spatial variance components. – Appl. Stat. 42: 185–209.

Rossi, R. E. et al. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. – Ecol. Monogr. 62: 277–314.

Saunders, M. C. and Tobin, P. C. 2000. Grape cane gallmaker (Coleoptera: Curculionidae), and its impact on cultivated grapes. – J. Econ. Entomol. 93: 795–799.

Schotzko, D. J. and O'Keefe, L. E. 1990. Effect of sample placement on the geostatistical analysis of the spatial distribution of *Lygus hesperus* (Heteroptera: Miridae) in lentils. – J. Econ. Entomol. 83: 1888–1900.

Sharov, A. A., Liebhold, A. M. and Roberts, E. A. 1996. Spatial variation among counts of gypsy moths (Lepidoptera: Lymantriidae) in pheromone-baited traps at expanding population fronts. – Environ. Entomol. 25: 1312–1320.

Southwood, T. R. E. 1978. Ecological methods with particular reference to the study of insect populations. – Chapman and Hall.

Stein, M. L. 1988. Asymptotically efficient spatial interpolation with a misspecified covariance function. – Ann. Stat. 16: 55–63.

Stein, M. L. and Handcock, M. S. 1989. Some asymptotic properties of kriging when the covariance function is misspecified. – Math. Geol. 21: 171–190.

Tobin, P. C. and Bjørnstad, O. N. 2003. Spatial structuring and cross-correlation in a transient predator-prey system. – J. Anim. Ecol. 73: 460–467.

Webster, R. and Oliver, M. A. 2001. Geostatistics for environmental scientists. – Wiley.

Weisz, R., Fleischer, S. J. and Smilowitz, Z. 1996. Site-specific integrated pest management for high-value crops: impact on potato pest management. – J. Econ. Entomol. 89: 501–509.

Williams, L. III, Schotzko, D. J. and McCaffrey, J. P. 1992. Geostatistical description of the spatial distribution of *Limonius californicus* (Coleoptera: Elateridae) wireworms in the northwestern United State, with comments on sampling. – Environ. Entomol. 21: 983–995.

Wollenhaupt, N. C., Mulla, D. J. and Gotway-Crawford, C. A. 1997. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. – In: Pierce, F. J. and Sadler, E. J. (eds), The state of site-specific management for agriculture. American Society of Agronomy, pp. 19–53.

Yfantis, E. A., Flatman, G. T. and Behar, J. V. 1987. Efficiency of kriging estimates for square, triangular, and hexagonal grids. – Math. Geol. 19: 183–205.