

Spatio-temporal analysis

Introduction

This chapter expands the discussion of the analysis of spatial structure to include the temporal dimension of ecological processes resulting spatial patterns. The intimate relationship between spatial structure and temporal change in ecological systems was eloquently described by Watt (1947) in his famous discourse on pattern and process in plant communities. His theme was that a plant community could be viewed as a working mechanism with the dynamic behaviours of development, degradation, and regeneration. In many plant communities, the various phases of the dynamic process coexist, and may have an identifiable spatial relationship to each other (Figure 11.1). In communities and populations of animals, the relationship between spatial locations and dynamic processes can be even more obvious, as animals move through the spatial structure of their habitat to find resources or mates, and to avoid predators or unfavourable conditions. At the population level, we need to recognize that a population of a given density is not often homogeneously distributed, and that the dynamics of different subpopulations' densities may be very different depending on location and the conditions found there. At the level of the individual organism and its immediate environment, we need to include the fact that an individual is mostly affected by very local, and less to global, conditions (Fortin *et al.* 2012a), and that these may change significantly over relatively small distances and over relatively short time periods. In almost any system, our concepts of spatial structure and its

importance will include implicitly, if not explicitly, a temporal component.

The expansion to include time is not exactly the same as the elaboration from one spatial dimension to two, or from two to three; time is not 'just another dimension' because of the direction of causality. Even when cyclic phenomena are under study, in which a building phase and a degradation phase may seem just to be mirror images of each other (Figure 11.1), we will often find that different processes are responsible for the two temporal 'directions', even if the resulting sequence of patterns appears similar and merely reversed in time.

It has been made clear more than once that quantifying the changes in spatial statistics calculated at several points in time (described in Section 11.1) is usually not the best approach to what should be multi-dimensional analysis (e.g. Parrott *et al.* 2008). The methods we include in this chapter will be somewhat more technical than merely summary, as will become apparent. We have, however, limited the discussion almost entirely to analysis, avoiding extensive discussions of modelling because the topic of spatially explicit dynamic models, for example, is both outside the domain of this book and sufficiently rich and complex to deserve a book of its own (see Dieckmann *et al.* 2000; Dutilleul 2011; among others). On the other hand, we have been broader in our treatment of spatio-temporal analysis than just the statistics (or the Statistics) for application to such data. Cressie & Wikle (2011) provided excellent coverage of both statistics and models for that environment. We will not duplicate that material here.

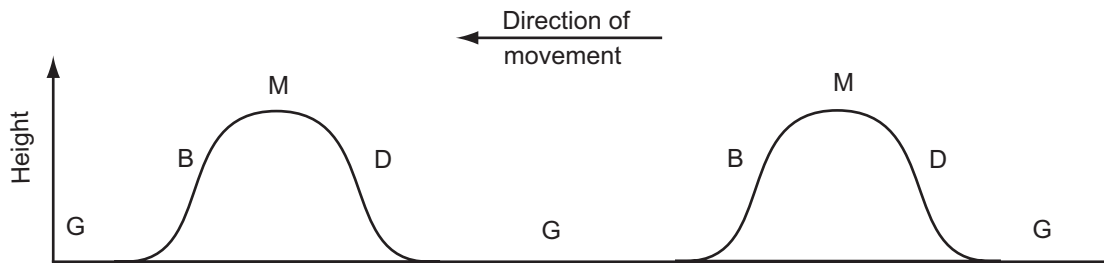


Figure 11.1 The relationship between pattern and process: the alternation between the mature hummock phase (M) and the hollow or gap phase (G) (e.g. in *Festuca ovina*-dominated sites) proceeds through building phase (B) from hollow to mature and through the degenerate phase (D) from mature to hollow. In this abstract example, building occurs on one side and degeneration on the other causing directional movement of the system. The same general scheme may apply in *Sphagnum*-dominated wetland hummock-hollow systems, brousse tigrée vegetation stripes, or wave-regenerating forest systems (based on Watt 1947).

The data used for spatio-temporal analysis can be classified in a number of different ways, but an important criterion is continuous versus discontinuous (discrete) data either in time or in space. For example, if we are monitoring the environmental conditions in a nature reserve by an array of hygrothermographs placed throughout the reserve, the data are continuous in time but discontinuous in space. On the other hand, if we are studying the movements of animals using global positioning system (GPS) radio collars, which report on position once every few hours or portions thereof, the data will be discrete both in space and time. In both examples, we may make assumptions based on interpolation: what the conditions are like between sites that have hygrothermographs, or where the animals would be found between reported positions. In cases such as permanent sample plots, in which tree stems are mapped and re-mapped at intervals, no interpolation may be necessary: stem no. 23 was alive in 1970, standing dead in 1978, and a downed log in 1985. In such cases, the observation and analysis of spatio-temporal pattern brings us very close to observing the processes that contribute to the pattern, because there is sufficient data to recover most of the important transitions.

To complete the classification of data types, it is possible to have data that are more-or-less continuous in both space and time, limited only by the resolution of the recording technology, for example the flight path of a butterfly, although the observer might want to divide the movement into temporal units for analysis.

Similarly, the tracks of animals in the snow are also of this kind. On the other hand, mark-recapture data are records of animals caught, labelled or tagged somehow for identification, released into the same environment, and then subsequently caught again; these are discontinuous both in space by trap locations, and in time by trapping session or date.

A concept that is basic to our discussions in much of this chapter is that of spatio-temporal autocorrelation. This refers to the lack of independence between objects, events, observations or measurements due to their positions in space and in time. The simplest kind is the case of short-range positive spatio-temporal autocorrelation where samples are more similar when they are closer together in space or in time (as discussed for space alone in the previous chapters). For example, Setzer (1985) used a Mantel test (see Chapter 7) on spatial and temporal distances between aphid galls on cottonwoods (*Populus deltoides*), and found that galls close in space were likely to suffer mortality close in time. In comparison, more complexity is likely to be found in cases involving cyclic behaviour such as diurnal migration, e.g. the vertical migration of zooplankton, where autocorrelation will be positive at short space and time lags, becoming negative over short space and longer time lags, and then positive again over even longer time lags (cf. Ohman 1990).

Just as there is a variety of measures for spatial autocorrelation (e.g. Moran's I , Geary's c), several different indices of spatio-temporal autocorrelation

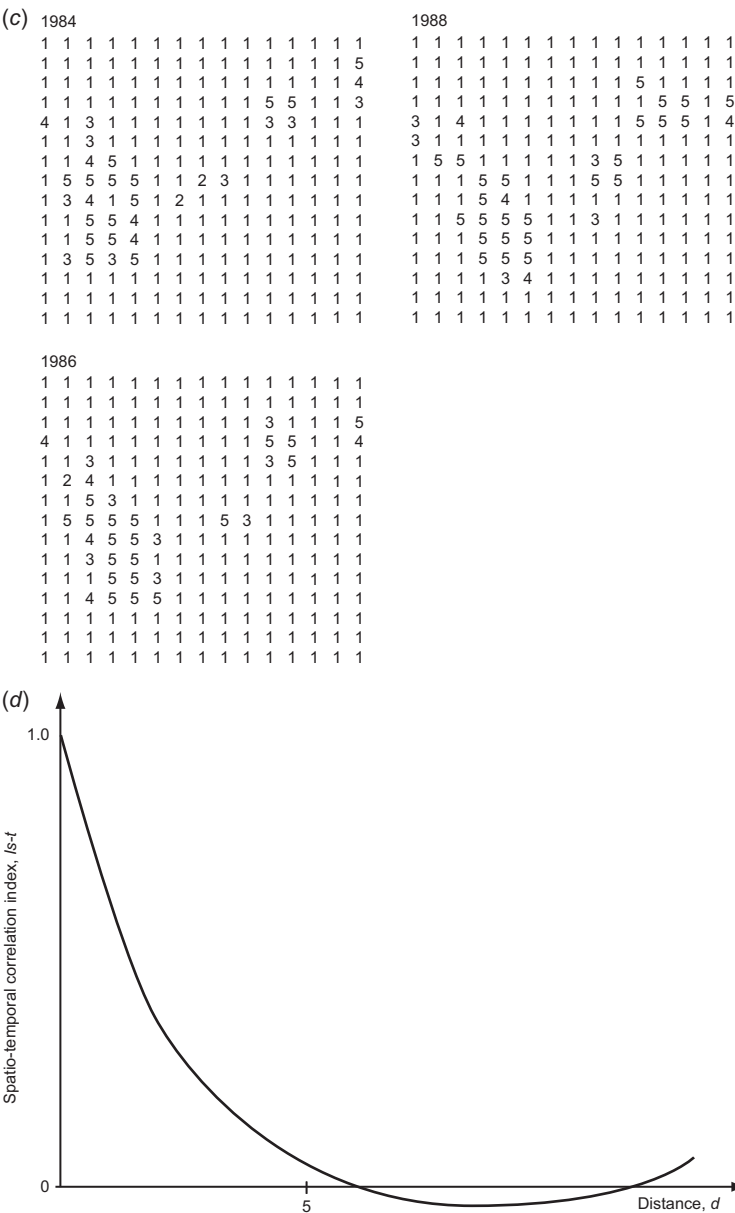


Figure 11.2 (cont.) (b) The analysis of the artificial data using Griffith's space-time correlation index. (c) Field data (*Nardus stricta*) from Law *et al.* (1997); higher values represent higher density. Each side of the 15×15 grid measures 50 cm. (d) The analysis of the field data using Griffith's space-time correlation index.

at change in some of the spatial statistics already described in this book. We then proceed to talk about kinds of truly spatio-temporal analysis based on join counts, cluster change detection, polygon change, and movement through space. The section following the group of methods attempts to tie together process and pattern by looking at situations in which we can analyse some kind of record created by the processes themselves, such as tree establishment, growth and mortality, or clonal plant mobility. We also provide a discussion of the concept and some applications of spatio-temporal graphs, analogous to the spatial graphs of Chapter 3, in which the nodes (and possibly edges) have locations both in time and in space.

11.1 Change in spatial statistics

One of the simplest kinds of spatio-temporal analysis is to examine changes in almost any of the spatial statistics described so far, as a function of time. This may not always be the best approach, but it is often the most readily available, and relatively easy to interpret. Small changes in a spatial measure can be accounted for by variation in the same underlying process, but large changes may suggest a change in the process itself. There are many examples of this approach in the literature. For example, Wu *et al.* (2000) studied the 'progress' of fragmentation in tiger bush (brousse tigrée) landscapes by comparing the curves of lacunarity as a function of scale (see Chapter 4), observed in different years. Brousse tigrée is a phenomenon of arid regions in which there are bands of woody vegetation, running across the direction of water flow, alternating with bare ground. It was described in more detail in Chapter 1 in relation to the concept of anisotropy. In their study of these landscapes in southwest Niger, Wu *et al.* (2000) found that lacunarity increased between 1960 and 1992, indicating the continuing fragmentation of the woody vegetation in the landscape (Figure 11.3).

Dale & Zbigniewicz (1997) studied a somewhat similar phenomenon at a smaller scale. They used 100 m transects of 10×10 cm contiguous quadrats in shrub-dominated communities near Kluane Lake in the Yukon

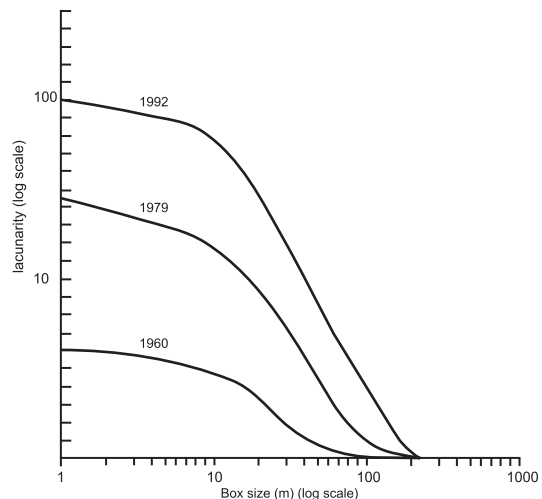


Figure 11.3 Changes in lacunarity curves for brousse tigrée sites. The lacunarity maximum increases markedly from 1960 to 1992, indicating more open ground, but the basic shape does not change noticeably, indicating little change in the other basic characteristics of the spatial pattern.

to examine the effect of a peak in snowshoe hare density on two of its winter food plants, *Salix* and *Betula*. The snowshoe hare (*Lepus americanus*) populations have cyclic fluctuations in density, with an amplitude of a factor of 5 to 25, and a period of about 10 years (cf. Krebs *et al.* 2001). They sampled the shrub vegetation in 1988, just before the peak in density, and in 1993, just after it. They used 3TLQV and Galiano's (1983) NQV (see Chapter 5) to detect the scale and patch size of the shrubs' patterns, expecting that the herbivore population peak would lead to fragmentation of the patches, leading to reduced scales of pattern and smaller patches. What they found, however, was that despite the intensive and extensive browsing of the twigs of these plants, the basic characteristics of the pattern recovered very quickly. This suggests that the basic characteristics of the spatial pattern of these shrub-dominated areas may persist for decades.

Vacek & Lepš (1996) studied the effect of neighbouring trees on tree vitality and mortality in five permanent plots in the mountains of the Czech Republic, monitored for 18 years. They used Ripley's *K*-function analysis (Chapter 4) to examine the spatial pattern of

Picea abies trees in 50×50 m plots. They found that although most of the plots started with a slight tendency to aggregation in 1976, by 1993, the pattern had changed to one of overdispersion at scales of 2–5 m, attributable to the higher mortality rates of trees with large and close neighbours. Kenkel *et al.* (1997) reported a similar trend for *Pinus banksiana* in Manitoba: a clumped pattern in the early stages of stand development gave way to a more-or-less random pattern at intermediate ages (30–40 years), with a tendency toward overdispersion at later stages. We will return to a more detailed analysis of tree establishment and mortality in Section 11.7.1, but this sequence of clumped to random to overdispersed is relatively common in many plant systems.

As a last example of this sort, we cite Nestel & Klein (1995) who used Moran's *I* (see Chapter 6) to study the spatio-temporal patterns of dispersion of adult leaf-hoppers (Homoptera: Cicadellidae) through the growing season of fruit orchards in Israel. The spatial pattern of one species, *Asymmetrasca decedens* (Paoli), began as aggregated and changed from aggregated to random with each successive generation. The pattern of a second species, *Edwardsiana rosae* L., remained aggregated throughout the entire season. The authors suggested that knowledge of the early season aggregation of both species could be used to develop an integrated management strategy for these insect pests.

This general approach to spatio-temporal analysis of using basic spatial statistics at two or more times is straightforward and clearly it can be very informative and useful. It will become more important as an analytical approach as more long-term studies continue or are established. One main disadvantage is that any summary statistics and changes in them may miss some of the important details of the actual changes to individual units or to particular subregions in the spatial structure. The other problem with this approach is that by averaging over space, as a summary statistic does, before examining changes in time in detail, this approach is prone to 'change of support' problems (see Jelinski & Wu 1996; Rahbek 2005). Parrott *et al.* (2008) therefore recommend approaches that do not 'collapse' the spatial dimensions before considering the temporal aspects of the data. As an

analogy, it is something like following the changes in plant community structure during succession by summarizing the complex information on species composition in a single measure, say species richness, and following the changes of that measure through time; it tells us something important about the community dynamics, but certainly not the whole story. Chapter 10 contains more detailed comments on the evaluation of measures of community diversity as they change through time.

One of many themes running through this book is the thought that methods that merely detect departures from randomness, such as spatial clustering, may not be telling us enough; for spatio-temporal analysis, for example, it may be important to know the size, spacing, positions and spatial evolution of the clusters (especially in pest management applications) in addition to the general degree and scale of aggregation as a function of time.

11.2 Spatio-temporal join count

In plant ecology, the concept of spatio-temporal pattern goes back at least to Watt (1947) who described how certain vegetation types tended to occur close together in space and time. One factor that contributes to this phenomenon is the clonal nature of some plants. Clonal growth forms are often described based on the spatial relationships among their growth units, known as 'ramets', which are related to patterns of establishment in space and time. The 'phalanx' form is characterized by a compact spatial structure of the ramets growing close together, maximizing intraspecific contacts, and the 'guerrilla' form is characterized by much more loosely arranged ramets, maximizing interspecific contacts (Lovett Doust 1981). These two forms are end-points of a continuum, and a general growth form can be described on the basis of the spatial arrangement of ramets, and the predictability of ramet establishment in space and in time.

Methods used in elucidating spatio-temporal pattern have been developed for single factor autocorrelation analysis to consider factors separated by two 'lags', that is intervals along axes of space and time.

One such method, Griffith's space-time index, given above (Griffith 1981; cf. Henebry 1995), is related to Moran's I , but examines the values of a variable at different locations and times. This index is appropriate to the analysis of continuous variables such as plant density, but if a binary variable such as presence : absence of individuals is considered, then join count statistics can be used. When the data are discrete count values then a Poisson space-time joint index should be used (Griffith & Paelinck 2009).

Join count statistics are a way of measuring association in nominal data distributed on a lattice or grid (Cliff & Ord 1981; see Chapter 6). Here we consider the case of binary data where each cell of a lattice can take only one of two values such as black or white. A join is defined as a connection of a particular lag between pairs of defined cell types, for example black to black. Figure 11.4a shows a simple example where joins of contiguous black cells are counted in a 7×7 square. In the spatio-temporal approach, a two-dimensional lattice represents a single spatial dimension, such as a transect of n quadrats and m intervals of time (Little & Dale 1999). Suppose the black cells represent either plant establishment or the presence of at least one individual of a particular species. If it represents establishment, then the occurrence of plants of different ages in the same quadrat can then be distinguished by placing several black cells in that column of the lattice. Join lengths are specified by combination of the intervals along the axes: $d = 2$, $t = 3$ signifies joins of length two on spatial axis and three in time. Spatio-temporal association can be determined by comparing the observed number in a class to the number expected from randomness (Jacquez 1996). The simple null hypothesis (H_{01}) is that the observed number can be accounted for by random occurrence. If more joins of a particular class (d , t) occur than are expected, this indicates a tendency for stems to be separated by a distance of d in space and t in time.

As a further refinement to this approach, Little & Dale (1999) were interested in the first stem that established in each quadrat; indicated by black cells in a lattice with only one black cell per column. Therefore, two further null models can be used, in which black

cells are randomly arranged but with no more than one in a column. In the uniform row model (H_{02}), the probability that each of the n columns contains a black cell is r and the probability of a black cell in any column is distributed uniformly among its rows. In the 'top black' model (H_{03}), black cells may occur in any cell of the lattice with probability p , but only those in the highest row in each column (representing the earliest to establish) are retained.

For each join class, 10 000 'data' lattices were generated for each of three random models, which were used to calculate a reference distribution and expectation. The association statistic of each join class was assessed based on the deviation of observed from expected join frequencies. The results are displayed as two factor 'correlograms', using circles, the sizes of which represent the probability and the positions of which represent the join classes. Figure 11.4b presents the *Populus balsamifera* establishment data from a transect at Ministik, Alberta, and Figure 11.4c shows the results of this analysis using three different null models. It shows that the fully randomized model has high values at (0.5, 1), (0.5, 3), (2.0, 4) and (2.5, 4). The other two models have high values at (1.0, 2) and (2.0, 4), the latter probably being a resonance peak. Overall, the results suggest that it is common for stems to be separated by two years in age and by 1 m along the transect, indicating that the clone advances in biennial pulses.

This technique is essentially a temporal adaptation of a lichen mosaic analysis (the 'random paired quadrat frequency' approach) described by Dale (1995; 1999, see chapter 5) but the important difference in this case is the use of several null models, rather than just one. As we have already commented, one way in which our spatial analyses can become more sophisticated, and thus more useful, is to use a range of null models rather than just simple 'randomness' for comparison with the outcomes we observe.

11.3 Spatio-temporal analysis of clusters and contagion

The approaches described in this section originated within the context of the study of clustering and spread

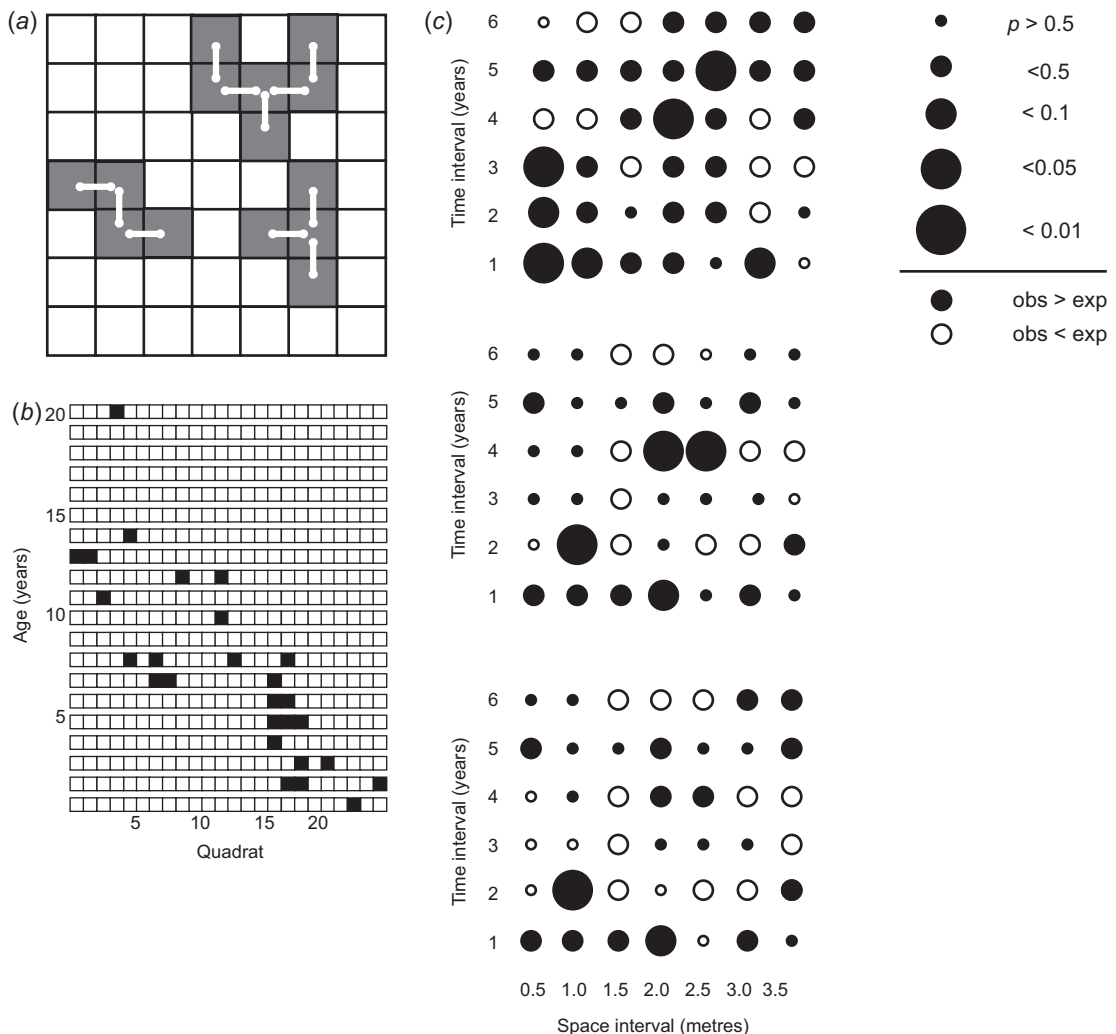


Figure 11.4 (a) Join counts of black-to-black contiguities in a 7×7 square; there are 11, as indicated by the dumb-bell 'simple join' symbols. (b) Space and time data for the establishment of *Populus balsamifera* at Ministik, Alberta. (c) Spatio-temporal join count analysis of the *Populus balsamifera* data. At each space and time lag combination, the size of the symbol represents the significance level and the colour codes (○, ●) represent the sign of the difference between observed and expected values. The three null hypotheses are fully randomized (upper), uniform row (middle) and 'top black' (lower).

of incidences of human disease, and some of that vocabulary persists in our discussion. The methods, however, translate well into many ecological areas, not just the incidence of disease, pathogens or parasites, but also the locations of rare epiphytes, nitrogen-fixing symbionts, and so on. The non-temporal version

of the basic approach is called 'cluster detection', which should not be confused with the multivariate technique known as 'cluster analysis'. Methods for detecting disease clusters for which the 'at risk' population is unknown are essentially versions of univariate point pattern analysis. The epidemiological literature

tends to emphasize approaches in which the locations (or at least the number in a given area) of individuals at risk are known, as well as affected individuals. The question is then whether the diseased cases are more clustered than can be explained by local variations in the at risk population (Jacquez 1996). This approach is a version of bivariate point pattern analysis, and strongly resembles other ecological questions of a similar nature, such as 'Are the *Solidago* plants with galls more aggregated than can be explained by the overall patchiness of the plants?' (Dale & Powell 1994).

Wakefield *et al.* (2000) provided a comprehensive review of the methods for cluster detection in the general area of spatial epidemiology (see also Fotheringham & Zhan 1996; Jacquez 1996) and Tango (2010) provided a recent and technical treatment of this topic. The methods fall into several categories. 'Traditional' methods include a simple comparison of the numbers of cases observed in different areal units (townships, counties) with the expected number (based on population and global disease rate), using a goodness-of-fit test. 'Distance: adjacency' methods include Moran's I for rates in contiguous areal units and Diggle & Chetwynd's (1991) variant of Ripley's bivariate K -function analysis for point data. 'Locally specific' methods include the moving window approach and risk surface estimation (see Wakefield *et al.* (2000) for details).

The first set of methods is best appreciated by looking at the four parts of Figure 11.5, which shows the progression of a disease (black dots) through a population (all dots) at four time periods. Clearly, the disease is spreading out from one corner of the figure, and the black colouring indicates a case that is new in that time period, rather than being a cumulative record.

A simple approach is to define a threshold value for 'near' versus 'far' in time and a threshold value for 'near' versus 'far' in space. Pairs of disease incidences are then categorized as 'near' or 'far' in time and 'near' or 'far' in space to produce a 2×2 contingency table. The table can then be tested with the usual statistics to determine whether incidences close in time also tend to be close in space (Knox 1964). For example, from Figure 11.5, using thresholds of two time steps and of

40% of the side of the sample area, we get the following counts.

The goodness-of-fit statistics are highly significant indicating spatio-temporal association among the dis-

	Time	
	Near	Far
Space		
Near	114	2
Far	125	110

ease incidences, in spite of the large number of pairs that are near in time but separated in space (chiefly in time period 4, Figure 11.5d).

In the introduction to this chapter, we mentioned the use of the Mantel test to determine spatio-temporal clustering in the example of aphid galls on cottonwood trees. The Mantel test compares two 'distance' matrices (see Chapter 7), and those distances can be of the spatial and temporal separations of events. In this context the Mantel test can be seen as an extension of Knox's approach, using the measured distances, rather than just 'near' and 'far' (Bailey & Gatrell 1995). Where d_{ij} is the separation of two events in space and s_{ij} is the separation in time, the basic Mantel statistic is

$$Z_M = \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}s_{ij}. \quad (11.3)$$

Evaluation of the test statistic is most easily accomplished by a randomization test. In the example shown in Figure 11.5, the observed Mantel statistic is $Z_M = 1219.14$ which was found to be highly significant by a randomization test, once again indicating an association between temporal and spatial proximity.

The approaches described so far do not use any of the information of the 'at risk', but disease-free, population. In fact, the next method we describe does not use that information either. That method is the spatio-temporal version of Ripley's K -function analysis, using only focus events (e.g. disease incidences) and counting

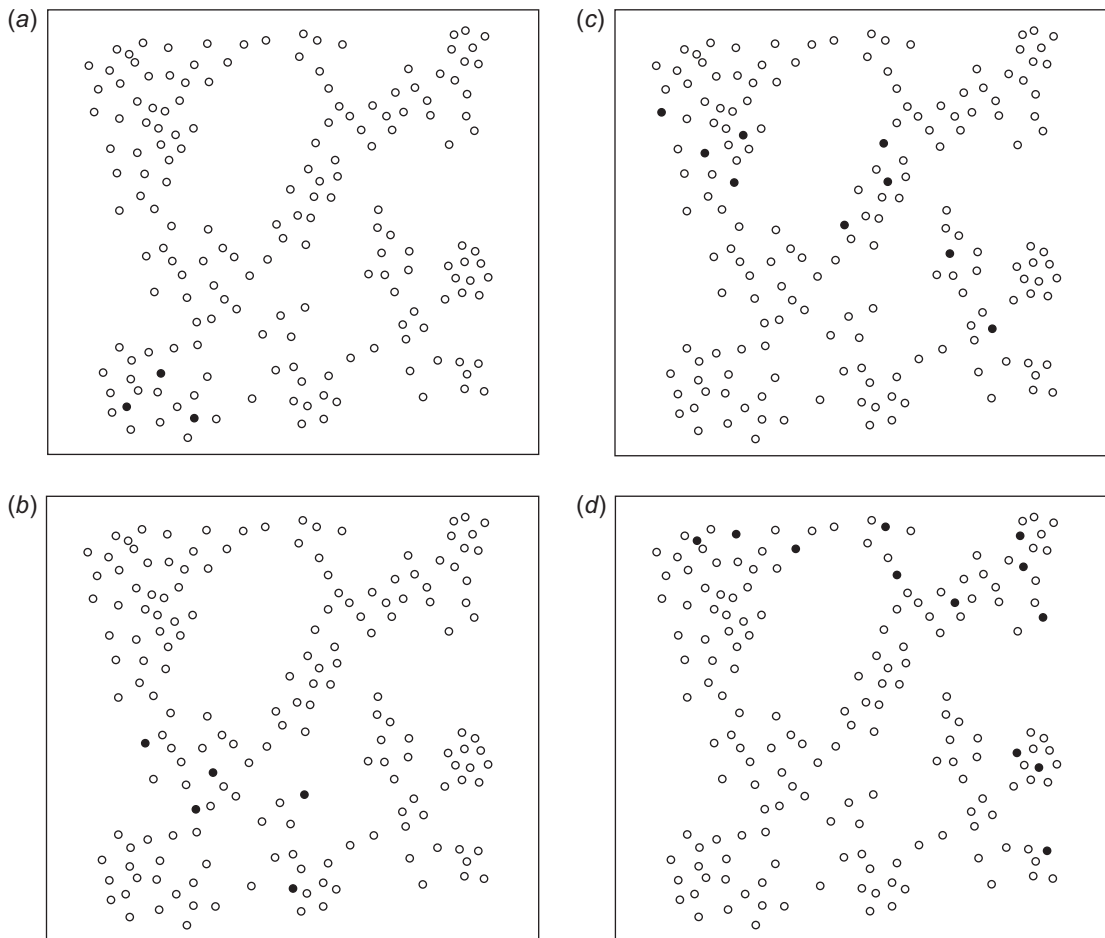


Figure 11.5 (a)–(d) Maps of the incidence of newly diseased organisms (filled circles) in a population at four different times. The spread of disease is obvious.

the number of focus events within spatial distance s and time distance t of each event (Diggle *et al.* 1995):

$$\hat{K}(s, t) = \frac{|A|T}{n(n-1)} \sum_i^n \sum_{j \neq i}^n w_{ij} v_{ij} I(d_{ij}^S \leq s) I(d_{ij}^T \leq t), \quad (11.4)$$

where w_{ij} and v_{ij} are weights for edge correction and the I s are the usual indicator functions. A and T are

the area and total time, their product giving the total spatio-temporal ‘volume’ being studied.

Observed and expected values are compared in the usual way and plotted as a function of t and τ . There are many potential problems with this approach, the first and most important being the possible incommensurability of time and space units as we have already discussed. The second is that unless there is a very long time series of observations, temporal edge effects can be an important factor. The third question is whether, because time is directional, a one-sided search template should be used rather

than the two-sided 't-bar' template depicted in Figure 4.13.

It seems that in the literature, a truly bivariate spatio-temporal analysis based on Ripley's K -function is missing. This is an area where further developments are needed and should be expected, but it may depend on a solution or good 'work-around' for the problem of space and time commensurability.

11.4 Spatio-temporal scan statistics

In Section 6.6 we described the concept of spatial scan statistics that can be used to detect and assess clusters of events, based on probability calculations of what is observed, usually using a moving window or template of some kind. We provided an example of the scan technique in a single dimension, which could be either spatial or temporal, the calculations being the same in either case. An obvious extension to the scan in one or more spatial dimensions is to include a temporal dimension, particularly when one of the motivations for the development of this field has been for the detection of clusters of disease, which clearly can have a temporal as well as a spatial component (see Kulldorff *et al.* 1998). As with spatial scan statistics, the spatio-temporal versions involve the calculation of a probability or a likelihood ratio in order to evaluate the result. The risk in interpreting these results associated with the potential for non-stationarity is clearly exacerbated by the possibility of temporal non-stationarity in addition to the spatial form discussed above. On the other hand, while we discussed potential complications for spatio-temporal analysis that might arise from the incommensurability of temporal and spatial dimensions, we may ask what does one unit in a spatial dimension compare to in the temporal dimension? For scan statistics, commensurability should not be an issue, just as Duczmal *et al.* (2006) described the usefulness of an elliptical template for spatial scan statistics, showing that there is no necessity for the spatial template to be isodiametric. Similarly, Iyengar (2005) portrayed the trade-offs in template shapes for spatio-temporal analysis using a cylinder (the spatial projection being a circle) and a square pyramid (spatially square, tapering in

time). Lastly, Takehashi *et al.* (2008) introduced a flexible approach to the shapes for spatio-temporal scan statistics templates, which will undoubtedly prove to be very useful in ecological contexts where irregular shapes occur naturally and inevitably.

As with the spatial version of scan statistics, the spatio-temporal version has great potential for ecological studies although it has not been much used to date, despite closely related applications for the detection of early disease outbreaks (Kulldorff *et al.* 2005; among many). One ecological example is by Marj *et al.* (2006) who used scan statistics to detect forest fire clusters in time and space using Florida daily fire records from the 2005 fire season. The use of these scan statistics to detect and evaluate spatio-temporal clusters of forest fires has received a general recommendation from Tuia *et al.* (2008).

11.5 Polygon change analysis

In Chapter 4, describing some basic methods of spatial analysis, there was an extensive discussion on the analysis of points or events in the plane, and considerably less on the analysis of irregular polygons in the plane; this reflects the emphasis in the literature. Similarly, published studies on spatio-temporal analysis also emphasize point events rather than polygons, quite naturally when the subject is the dynamics of tree stems, for example, or the locations of disease foci. The analysis of polygons offers considerably greater complexity than the analysis of point patterns, because of the greater complexity of the data themselves. The difficulties of dealing with polygons can sometimes be avoided by representing the centre of the polygon by a point with associated size and shape characteristics, and with the adjacency of polygons represented by lines between the points. These networks can then be analysed for autocorrelation in the neighbour network using statistics such as Moran's I or Geary's c (see Chapter 6). For example, the polygons in Figure 11.6a can be reduced to a graph of their connections to neighbours by contiguity and their areas (Figure 11.6b). Autocorrelation analysis shows significant similarity among neighbours, based on their

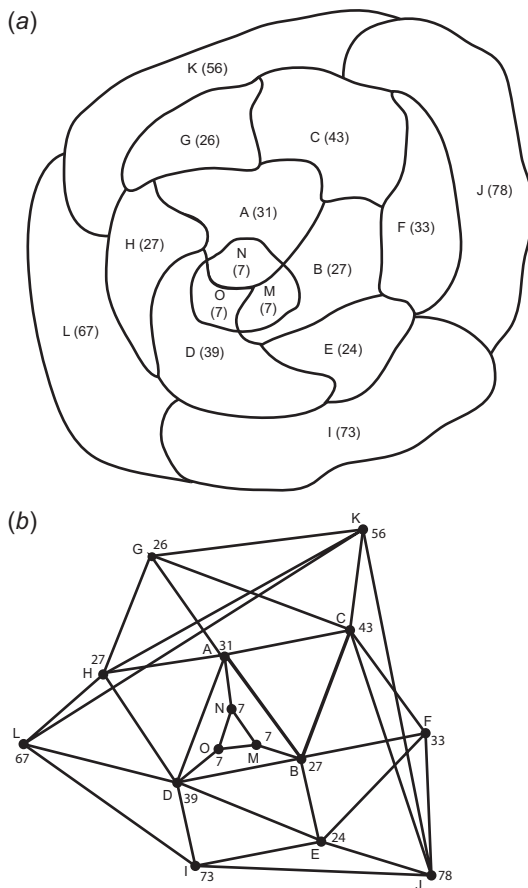


Figure 11.6 A set of polygons with their relative areas (a) portrayed as a graph (b) with the connections in the graph representing the contiguity (shared edge) of polygons. The nodes of the graph retain the areas of the polygons for further analysis.

areas, but information about other characteristics of the polygons is lost.

When we now consider analysing a dynamic system of polygons, the situation becomes even more complex, because not only do the characteristics of the polygons change (position, size, shape), as do their connections to neighbours, but also old polygons may disappear and new ones develop. One approach would be to calculate summary statistics for each of several observation times and then examine changes

in those summaries; for example, Peralta & Mather (2000), in a study of deforestation in Amazonia, used indices of lacunarity, patchiness and area-perimeter fractal exponent to summarize the changes. There is an advantage in using more than one summary statistic, obviously, but even so, a lot of detail on the characteristics of the polygons will be missed, entailing the loss of potentially important information.

Sadahiro & Umemura (2002) have developed a sophisticated approach to the analysis of changing polygons. The conditions of their scheme are that the polygons are immobile, so that individual animals, herds or flocks, are not eligible, and that change occurs in a discontinuous fashion, so that objects like temporary pools that shrink and expand in a continuous way are also excluded. With those exclusions aside, their treatment seems flexible and useful. They divide the stepwise behaviour of an individual polygon into six primitive events, illustrated in Figure 11.7:

- (1) generation, i.e. the appearance of a new polygon;
- (2) loss, i.e. the disappearance of a polygon;
- (3) expansion, i.e. the increase in area occupied;
- (4) shrinking, i.e. the loss of area;
- (5) union, i.e. two polygons merging; and
- (6) division, i.e. a polygon splitting into two.

The changes observed between two different times can then be described by combinations of these primitive events. This approach follows that of Claramunt & Thériault (1997) who proposed a more complicated scheme of sixteen primitive events, in part to be able to deal with possibly mobile polygons. An additional assumption is that there are no problems in tracing the identity of polygons; the time interval is assumed to be sufficiently short that polygons that overlap are two temporal versions of the same polygon.

The two sets of polygons from the two observation times, Γ_1 and Γ_2 , are overlaid to create a new set of polygons, Γ_u (Figure 11.8). These new polygons can then be classified into three groups: Ω_0 which existed at both observation times, Ω_1 which existed at time 1 but not time 2 and Ω_2 which existed at time 2 but not time 1. The arcs in the diagrams are classified into twelve groups based on the combinations of four possible states in which they existed at times 1 and 2: boundary, partition, internal to polygon, and absent.

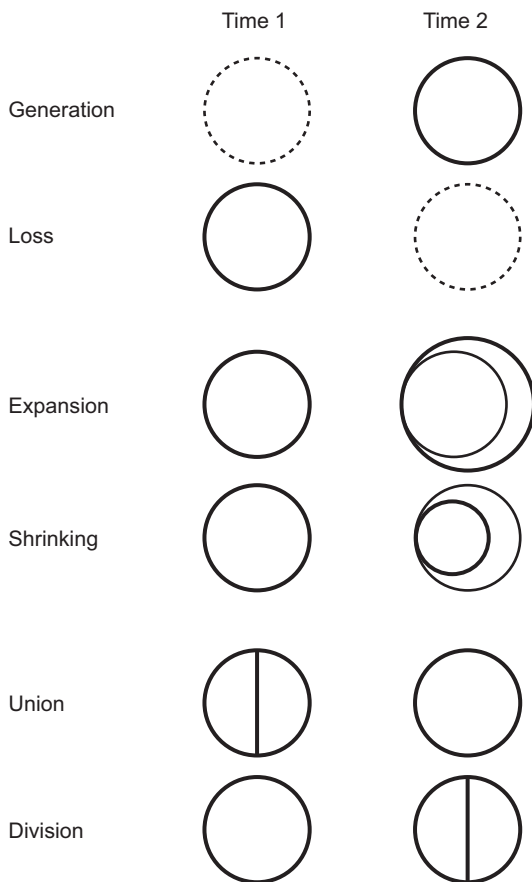


Figure 11.7 The primitive events for polygon change analysis: generation, loss, expansion, shrinking, union and division.

Figure 11.9 illustrates these classes. The two classifications are then used to deduce possible sequences of primitive events that gave rise to the observed changes. The changes between two observation times are decomposed into the smallest number of primitive events possible, and one useful statistic is then the number of such primitive changes, M_e , which can be standardized to the total number of polygons:

$$m_e = \frac{M_e}{|\Gamma_1| + |\Gamma_2|}. \quad (11.5)$$

Claramunt & Thériault (1997) suggested that, as a refinement, instead of raw event counts, the events

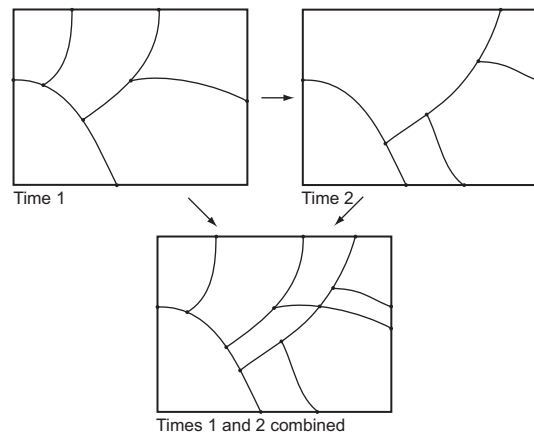


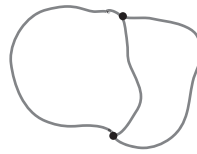
Figure 11.8 The two sets of polygons from time 1 and time 2 (Γ_1 and Γ_2 , upper diagrams) are overlaid to create a combined set (Γ_w , lower diagram). There are five polygons in each of the first two sets and nine in the combined set.

could be weighted by some function specific to the kind of event, f_x . For example, generation and expansion events could be weighted by the area gained, and shrinking and disappearance events by area lost, with partition and union events having weight 0. In that case, the area index would be

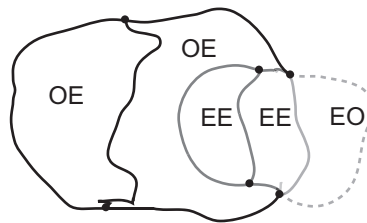
$$m_A = \frac{A(\Omega_1) + A(\Omega_2)}{A(\Gamma_1) + A(\Gamma_2)}. \quad (11.6)$$

Since its introduction, the analysis of polygon change has been extended to include the possibility of polygons that move, with the development of 'STAMP' (Spatial-Temporal Analysis of Moving Polygons) methods within a GIS environment (Robertson *et al.* 2007). The extension augments the earlier basic scheme with five movement events: displacement, convergence, divergence, fragmentation and concentration. Its usefulness was illustrated with reference to the spread of a wildfire in northwest Montana (USA), for which the method produced a good quantitative summary of changes in size and direction of movement. Grillet *et al.* (2010b) investigated the spatial dynamic of malaria incidences in Venezuela. Also, Vepakomma *et al.* (2012) applied this polygon change analysis to quantify gap spatial dynamics in the boreal forest in Québec.

Time 1



Time 2



Times 1 and 2 combined

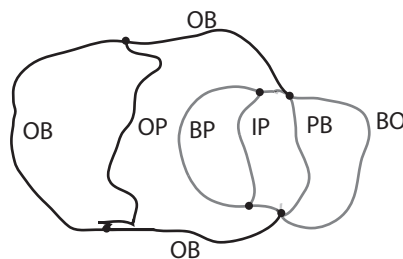


Figure 11.9 In this example, there are two polygons in the first set, four in the second and five in the combined set. The polygons can be classified as 'EO', extant at time 1, but not at time 2; 'OE', extant at time 2, but not at time 1; and 'EE', extant at both times. The edges that define the polygons can be classified as 'BO', boundary at time 1, but extinct at time 2; 'PB', partition then boundary; 'IP', interior then partition; 'BP', boundary then partition; 'OP', partition new at time 2; and 'OB', new boundary.

This approach is relatively new and has few examples of application, but it clearly presents a great deal of promise for future analysis. While it does deal with the dynamics of a set of polygons, it does not include characteristics such as their shape in the analysis. In addition, in this treatment, if the polygons

being studied form a complete mosaic filling the study region, the kinds of transitions that can occur become much more limited. The analysis of full mosaics (in which the polygons form a complete tiling of the plane) is another related area that deserves further work and exploration.

11.6 Analysis of movement

There are many circumstances in ecological studies in which the movement of individual organisms is of crucial interest: the spread of disease organisms or vectors in epidemiology, the identification of home ranges in wildlife ecology, or the spread of a clonal plant into new habitat, to suggest a few examples. The quantitative analysis of such movements will depend in part on whether the movement itself (or its record) is more or less continuous, like a beetle wandering through a grassland, or it occurs as discrete units, such as a butterfly's foraging journey stopping at individual inflorescences or the identifiable ramets of a growing clone. If the movement is continuous, it is often divided up into units, using some criterion (possibly arbitrary) such as the positions observed every five minutes. The concept of units of movement will be discussed further, later in this section.

When the movement occurs in units, or can be divided up into units, analysis usually proceeds based on the lengths of the units of movement and the angles between them (see Figure 11.10). Crucial to an understanding of the correct way to analyse this kind of data is an understanding of the calculation of the average of a set of angles. Cain (1989) pointed out that many authors in ecology have made mistakes in this process. We cannot just take a simple average the way we do for a linear scale like distance. For example, if measured on a scale from 0° to 360° , the angles 90° and 270° have an average of 180° ; measured on a scale of -180° to $+180^\circ$, the same angles, now 90° and -90° , have an average of 0° (see Upton & Fingleton 1989, chapter 9).

For any set of angles, whether the absolute angles of the steps, labelled α_i in Figure 11.10, or the 'turning angles', δ_i in the same diagram, the angles are represented by vectors of unit length and coordinates (x_i, y_i) , as in Figure 11.11. The coordinates of the mean vector are:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \cos(\alpha_i) \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n \sin(\alpha_i). \quad (11.7)$$

In polar coordinates, it is

$$(\bar{x}, \bar{y}) = (r_a \cos \varphi, r_a \sin \varphi), \quad (11.8)$$

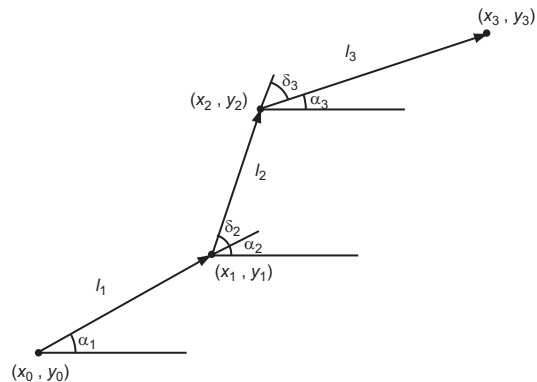


Figure 11.10 Movement portrayed as a series of straight-line units, l_1 to l_3 , with known start-points and end-points. The α s are the absolute angles and the δ s are the changes in angle.

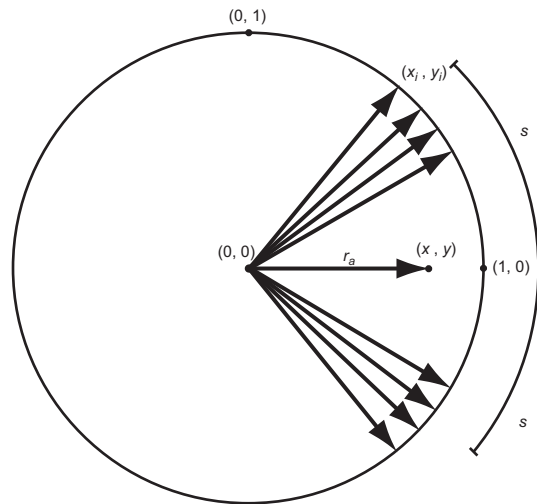


Figure 11.11 Calculation of the average of a set of eight unit vectors, (x_i, y_i) . The angular concordance is r_a and the angular deviation is s .

where $\varphi = \tan^{-1}(\bar{y}/\bar{x})$ and

$$\begin{aligned} r_a &= \sqrt{\bar{x}^2 + \bar{y}^2} = \frac{1}{n} \sqrt{n^2 \bar{x}^2 + n^2 \bar{y}^2} \\ &= \frac{1}{n} \sqrt{\left(\sum_{i=1}^n \cos \alpha_i \right)^2 + \left(\sum_{i=1}^n \sin \alpha_i \right)^2}. \end{aligned} \quad (11.9)$$

The measure r_a should not be referred to as a measure of angular correlation because it cannot take negative values; it is more appropriately referred to as a measure of angular concordance or angular concentration (Zar 1984). It takes value 1.0 when all the angles are the same and value 0 when the vectors cancel each other out (Upton & Fingleton 1989). For example, in Figure 11.11 there is a good agreement in the direction of the angles and r_a is 0.75. The circular equivalent of the standard deviation for linear data is s , the angular deviation (Batschelet 1981):

$$s = \sqrt{2(1 - r_a)}. \quad (11.10)$$

In our example, $s = 0.71$, which can be converted to degrees by multiplying by $180^\circ/\pi$ (Batschelet 1981), here 41° (see Figure 11.11), showing again that there is a relatively small variance in the set of angles.

In considering a path such as that depicted in Figure 11.10, and seeking a measure of the angular autocorrelation, for lag 1, the average of the cosines of the turning angles is a good candidate (Batschelet 1981). Figure 11.12 illustrates this concept. When the two steps of the path are aligned in the same direction, correlation is 1.0; when they are at right angles, the correlation is 0; and when they are directly opposite, it is -1.0 . Turchin (1998) suggested that this statistic should not be used

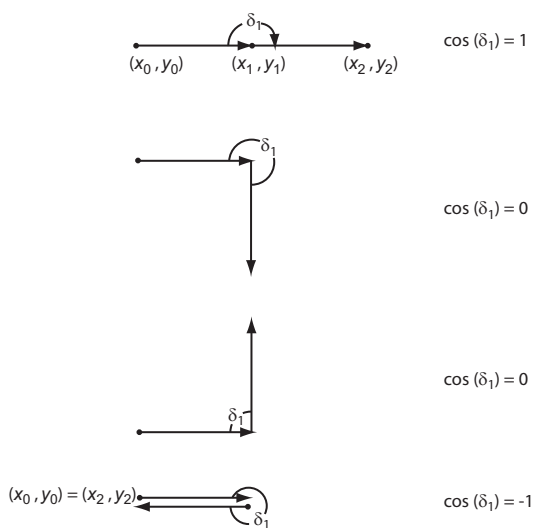


Figure 11.12 The cosine of the turning angle as a measure of angular autocorrelation.

for turning angles because 'turning angles are typically concentrated around zero (so that) the difference between two successive turning angles is likely to be near zero, even if there is no autocorrelation. This would result in a significantly positive, but spurious, angular autocorrelation'. We disagree and would argue that turning angles concentrated near zero, indicating a tendency for motion to continue in the same direction, is real autocorrelation as a characteristic of the data.

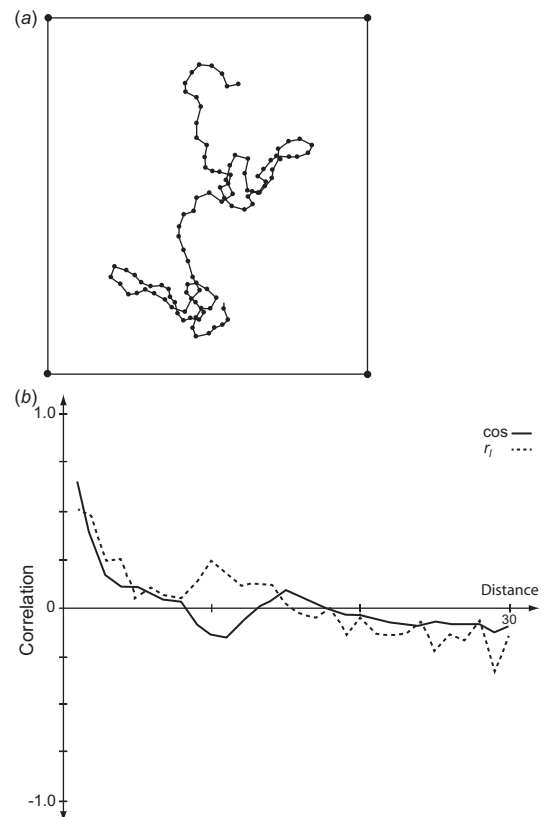


Figure 11.13 Pairs of paths and the autocorrelations of the units' lengths and angles as a function of distance or lag: (a), (b) short-range positive autocorrelation in both length and angle; (c), (d) angles and lengths both independent; (e), (f) directional bias, but lengths independent; (g), (h) directional cycles, lengths independent; and (i), (j) cyclic behaviour in lengths, angles independent. The dots in the illustrations of paths are the corners of the study area.

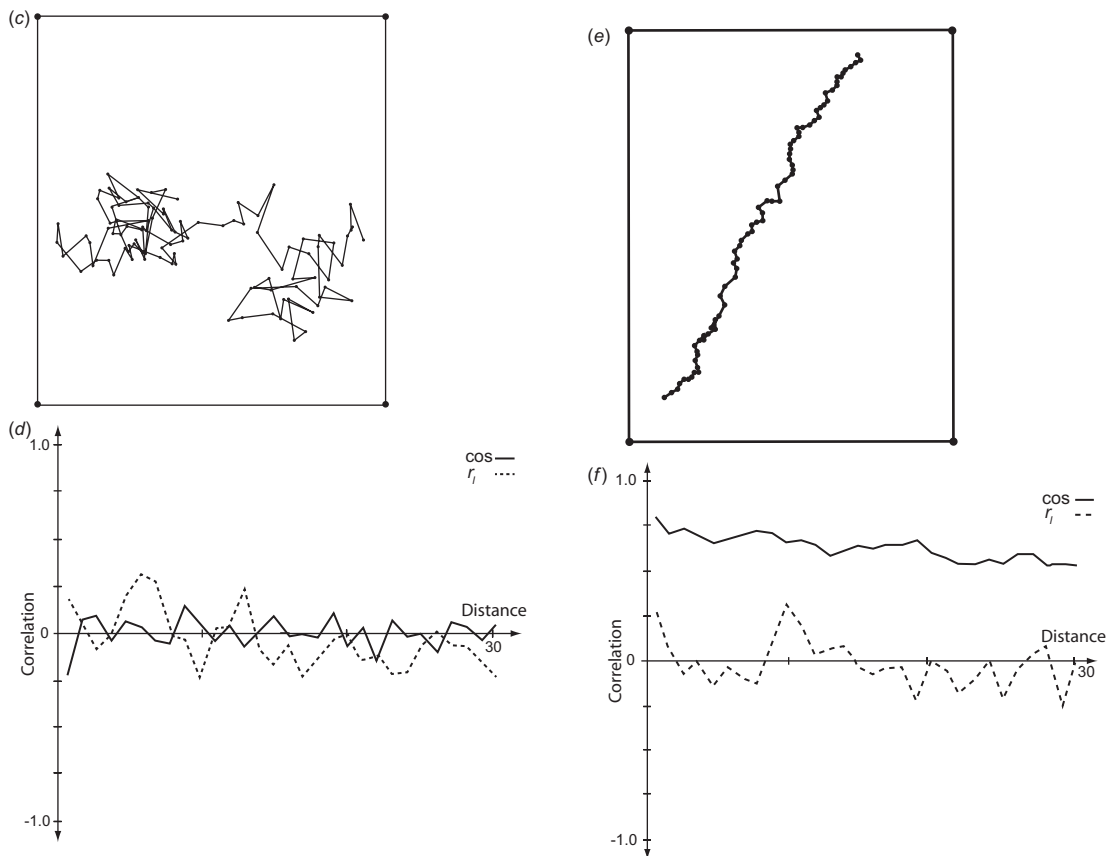


Figure 11.13 (cont.)

As a preliminary analysis of movement data, we suggest calculating the radial (distance) and angular correlation for a range of lags. Figure 11.13 shows the results of this kind of analysis for five sets of artificial data with different kinds of autocorrelation. (We could also look at the autocorrelation of net displacement as a function of lag.) Turchin (1998) discussed the possibility of problems associated with 'oversampled' data sets; that is if an animal's location is recorded every second, there is so much autocorrelation in the data that we may seem to have too much data for the information we can get out of it. The exploratory analysis recommended here, however, will allow a simple evaluation of the 'resolution' of the data.

Similarly, Turchin (1998) also discussed the aggregation of individual steps (the finest scale of resolution of the movement) into 'moves', with more recognizable breaks or turns between them (Figure 11.14a). This may not be necessary or desirable; the evaluation of radial and angular correlation as a function of lag should provide the information needed. In addition, the angular correlation graph will allow an assessment of whether there tends to be an alternation of right and left turns, without the need for a separate runs test (cf. Sokal & Rohlf 1995). If that question is of particular interest, that approach can also be used to provide an actual statistical test rather than just an exploratory indication of tendency.

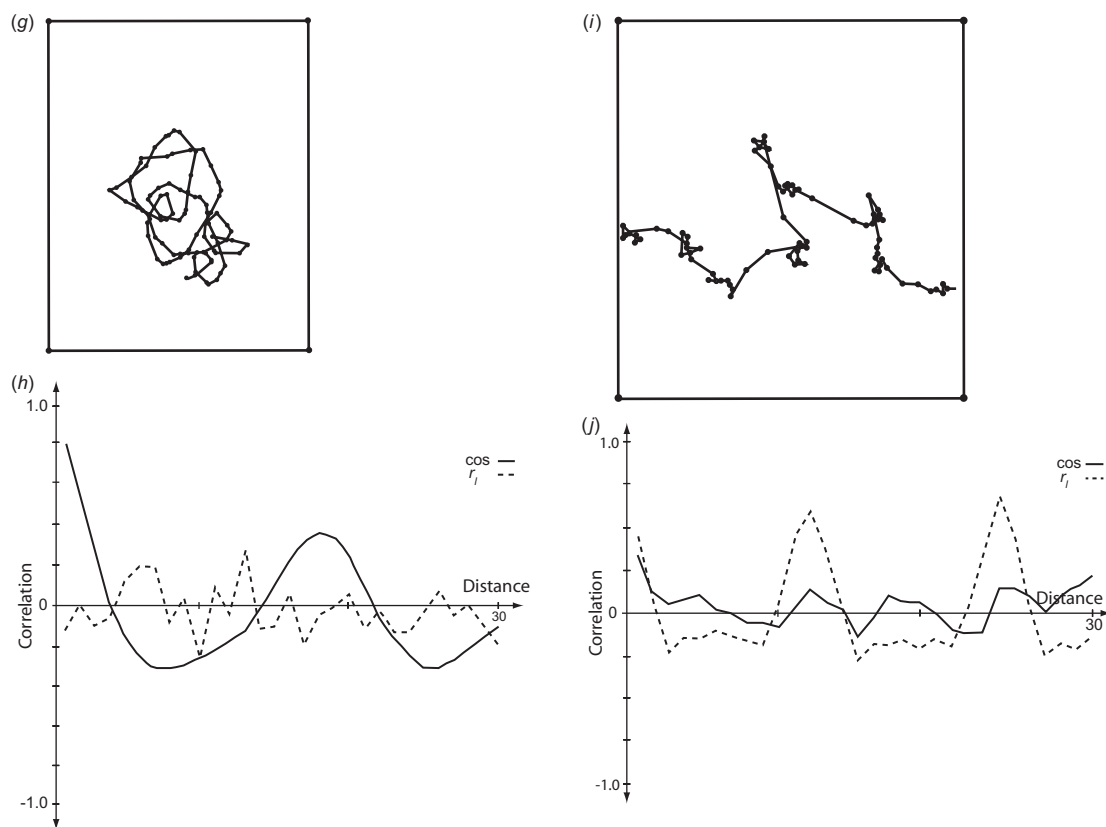


Figure 11.13 (cont.)

The most difficult problem to deal with in real field data may lie not in the resolution of the data, or in over- or undersampling; it may lie in the lack of stationarity of the process. As an example, a colleague has provided us with about 3000 locations of an individual elk taken every two hours with a GPS radio collar (Merrill, unpublished). Even one-tenth of the data represents about a month of time, and it is easy to imagine how an elk's behaviour might change from the first half of a month, such as July, to the second half, as the montane vegetation changes rapidly. This possible departure from the underlying assumptions must be taken into account both in analysis and in interpretation.

As a simple null hypothesis, with which to compare the observed characteristics, we might consider the well-known 'random walk' (cf. Turchin 1998). The null model of the random walk is that the direction of each unit of movement is randomly chosen from the full circle of possible directions and that the length of the unit is drawn from a distribution of such lengths, which can be estimated from the data. It is easy to calculate the expected displacement of the individual (the expected net displacement is zero) and other characteristics, such as the distributions of angles, resulting from a random walk; however, in general the paths of clonal plants and mobile animals do not match those well. Angles near 180° are less common

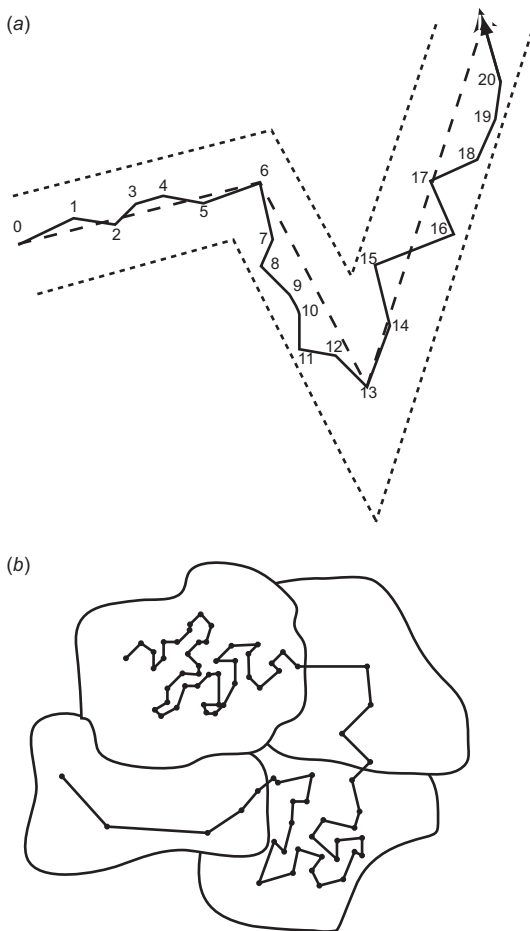


Figure 11.14 (a) The proposed aggregation of unit steps into 'moves'. (b) Movement has more turns and shorter lengths in preferred habitat. (The curved boundaries delineate different habitat types.)

and displacement tends to be greater than in this model (Turchin 1998).

As an alternative to the random walk model, comparisons with a 'correlated random walk' (CRW) model are popular (Kareiva & Shigesada 1983). In this model, the expected net displacement is still zero as in the simple random walk, but its other properties are different because of autocorrelation of the units that make up the path. The model usually looks at the first-order autocorrelation of length and direction of the units of

movement and compares the observed and expected net squared displacement and path 'tortuosity' (Wiens *et al.* 1993) or 'sinuosity' (Williams 1992; Sanuy & Bovet 1997). Kareiva & Shigesada (1983) examined this model and provided details on the derivation of the expected values. Another approach to the evaluation of the appropriateness of the CRW model is to compare the observed net displacement after n steps (or the net squared displacement) with that predicted from the model. Many studies have used one or other of these approaches to evaluate the movement of caribou, clover, golden rod, toads and caterpillars (Cain 1990; Cain *et al.* 1995; Sanuy & Bovet 1997; Bergman *et al.* 2000; Doak 2000). Kareiva & Shigesada (1983) recommend that we should go beyond the simpler first-order models and look at more complicated examples using higher-order Markov models. That is, the length and direction of a movement unit can depend not just on the one immediately preceding it, but perhaps on the characteristics of the last two, three, or more units of movement. For example, Figure 11.13f illustrates behaviour that has strict cycles, although a more realistic version of the same cyclic behaviour but with a random component could be modelled by a high-order Markov model.

Another alternative is to examine 'biased correlated random walks' in which a general directional tendency, either absolute or relative to some habitat element, is included (Turchin 1998), so that the expected net displacement is not zero. This approach is similar to adding 'drift' to the model, by including a set absolute directional component to each unit of movement (Wiens *et al.* 1993). For example, Schultz & Crone (2001) looked at the tendency of the 'Fender's blue' butterfly (*Icaria icarioides* ssp. *fenderi*) to move toward its host plant (*Lupinus sulphureus* ssp. *kincaidii*) by including a bias component of movement toward a nearby lupin patch in a correlated random walk model, whatever its absolute direction. The bias component was significant for a range of distances from a patch. Whichever kind of model is used, randomization techniques based on re-ordering the individual units of the movement or Monte Carlo procedures using parameters determined from the observed characteristics of the path can be used

to evaluate the results (cf. Manly 1997), but we will describe some general concerns about randomization tests in Chapter 8.

In characterizing the spatial complexity of movement, a number of authors have advocated measuring the fractal dimension of the path, either using the 'dividers' method (see Chapter 6) or based on re-normalization procedures (Wiens *et al.* 1993). Turchin (1996) advised that this is a risky procedure unless the path is truly self-similar so that the fractal dimension remains constant over a range of spatial scales. This seems to be good advice, and this method should be used only with an understanding of that risk of misinterpretation. In addition, Schultz & Crone (2001) pointed out that descriptive measures of fractal dimension (or tortuosity) are difficult (impossible?) to translate into useful movement parameters that can then be used to predict the distributions and dynamics of populations. This is an important deficiency because for the wildlife biologists who use these methods, prediction is what it is all about.

Another important theme in the analysis of this kind of data is the search to associate the locations and the characteristics of movement paths with the types of habitat through which the animal moves and thus to evaluate differential habitat preference and habitat use (see Figure 11.14*b*). A difficult and controversial aspect of this analysis is how to carry out statistical tests relating animal movement to habitat type, and Manly *et al.* (2002) provided a useful review. Before we proceed to discuss this topic we suggest that you . . . STOP.

Stop reading this book for a few minutes and check your file of reprints (or 'e-prints' more likely!) and re-read Hurlbert's (1984) discussion of pseudo-replication and the design of ecological field experiments. We need to make sure that our understanding of 'units' and 'replication' is clear and thorough. An important concept that is brought out in Hurlbert's discussion begins with a description of an experiment to compare the rates of decomposition at 1 and 10 m depth in water. Eight netting bags filled with leaf material are placed at 1 m depth and eight at 10 m depth. The important concept here is that, in this case, the leaf bags are not the experimental units; the locations at which they are placed are, and the bags are really just measuring

devices. If the eight bags for each depth are placed at only one location for each depth, we cannot really test for differences between depths. We can only test for differences between two locations, one of which happens to be at 1 m and the other at 10 m; that is we have simple pseudo-replication. In order to examine the effects of depth, the leaf bags must be dispersed to allow us to make inferences more general than the results of just two locations. Sacrificial pseudo-replication occurs when there are true replicates, but the data for replicates are pooled before analysis or where two or more samples taken from each experimental unit are treated as independent samples. Hurlbert's discussion reminds us that we need to consider the identity of the experimental unit and the identity of a replicate in designing statistical tests for locational data. It also reminds us that we need to approach the pooling of data with great caution . . .

NOW, . . . to return to the topic of testing the relationship between animal location and habitat type . . .

The simplest approach is to look only at the locations of the animals, without considering the properties of the path of movement. Neu *et al.* (1974) advocated a simple goodness-of-fit test to compare the frequency of animals' presences in particular habitat types with the availability of those habitat types in the landscape. For example, with only two habitat types, A and B, known to be present in a ratio of 2:1 in the landscape, the observed data of 175 animal locations in type A and 125 in type B would be compared with the expected values of $\frac{2}{3} \times 300 = 200$ for A and $\frac{1}{3} \times 300 = 100$ for B by calculating

$$\chi^2 = \frac{(175 - 200)^2}{200} + \frac{(125 - 100)^2}{100} = 9.375. \quad (11.11)$$

The test statistic χ^2 is compared with the χ^2 distribution with one degree of freedom and in this example the result is highly significant. The conclusion is that the animal is preferentially using habitat B. It is worth reminding ourselves that this goodness-of-fit test is sensitive to the overall sample size. Given the same proportions but only 60 observations in total, 35 and 25 in the two habitat types, the result would not be significant. With 10 000 observations, the result of this test is almost certain to be significant if there is any

Table 11.1 A comparison of habitat use with habitat availability

Habitat type	Observed count	Expected count	Freeman–Tukey ^a <i>z</i>
A	180	200	−1.43
B	110	100	1.00
C	30	60	−4.48
D	50	30	3.21
E	30	10	4.64
Total	400	400	$\Sigma z^2 = 54.9$

^a The Freeman–Tukey statistic is a standardized residual that can be compared with $N(0, 1)$ for assessment.

tendency at all to depart from non-preferential habitat use. If more than two habitat types are being considered, following the overall test, a Freeman–Tukey standardized residual could be calculated, comparing the observed and expected value for each habitat type, to determine which contribute most to the overall significance. Values of magnitude greater than 2.0 indicate important contributions to the overall significance, but we cannot ascribe a particular significance value to individual cells (cf. Bishop *et al.* 1975). Table 11.1 provides an example.

The table shows an apparently significant departure of the observed from the expected ($\chi^2 = 71.3$ on 4 d.f.; $G = 58.5$; compare with $\Sigma z^2 = 54.9$) with the avoidance of habitat type C and the preferential use of D and E contributing most to the overall significance.

The same table is based on the situation in which the proportions of the five habitat types can be treated known, perhaps from air photo interpretation or GIS analysis. If, on the other hand, the values in the third column represent the frequencies of the habitat types in 400 random samples, the analysis is different. The null hypothesis now is that columns 2 and 3 in the table are both estimates of a common frequency distribution and so expected values are calculated for both. For example, the expected value for both columns and type A is 190. Now, $\chi^2 = 26.5$ on 4 d.f. ($G = 27.2$) and only habitat type E has a Freeman–Tukey standardized residual with an absolute value greater than 2. The fact that the evaluation of habitat

availability is based on a sample has greatly lowered the value of the test statistic and has changed the interpretation to one that would concentrate on the disproportionate overuse of habitat type E.

This contingency table approach, as proposed by Neu *et al.* (1974), has some advantages, including the fact that it can be used for one or for several animals (with suitable caution in how the results for several animals are combined; White & Garrott 1990). Remembering that Hurlbert (1984) pointed out that this kind of goodness-of-fit test is the most misapplied statistical procedure, we might have some concerns about what is being implicitly defined as the experimental unit and as the replicate if we include data from several animals. Thomas & Taylor (1990) shared the concerns about the possible misuse of the goodness-of-fit test and point out a number of other problems including tests that do not control the ‘experiment-wise’ error and the sensitivity of tests to the subjective inclusion or exclusion of resources. Even leaving those concerns aside, this approach cannot be used as described because it has one truly fatal flaw.

The problem is, of course, that in using the standard goodness-of-fit test, autocorrelation in the data has not been accounted for. Positive autocorrelation makes statistical tests too liberal, giving more apparently significant results than the data actually justify. We discussed this problem at length in Chapter 8, in the context of the effect of spatial autocorrelation on statistical testing. The same sorts of considerations will apply, however, to the spatio-temporal autocorrelation inherent in animal movement data. It is a bit strange that in recommending Neu’s approach, White & Garrott (1990) mentioned the problem of autocorrelation without any suggestion of how it can be addressed.

The optimistic view is that if you thin out the data so that the observations used in the analysis are further and further apart in time and space, at some point they will become independent of each other. This is the concept of ‘time to independence’, which has received some attention in the behavioural ecology literature (Swihart & Slade 1985, 1986; Solow 1989; Salvatori *et al.* 1999). It seems unlikely to apply well in the case of studying a single animal, the behaviour of which may be more consistent than comparisons among

animals, just because it is the same animal with its own idiosyncrasies, memories, and so on. Differences in the approaches used depend on whether the focus is on a single animal or on a population (Manly *et al.* 2002). Millsaugh *et al.* (1998) emphasized the importance of considering the biological characteristics of the organisms in attempting to assess autocorrelation and the independence of observations. Second, behaviour may exhibit cyclic patterns, which may introduce negative autocorrelation at some lags (temporal or spatial), and does not provide independent samples or observations either.

Without repeating too much of the material of the earlier chapter (Chapter 8 of this book), it is sufficient to say that there is no easy solution to this problem and no simple 'time to independence' trick that will make it disappear (see Rooney *et al.* 1998). The fact that autocorrelation is not significantly different from zero (cf. Swihart & Slade 1985) does not mean that it has been shown to be zero and can therefore be treated as such. In Chapter 8, we gave a good example of how 'insignificant' autocorrelations can have a large cumulative effect. Minta (1992) concluded that time to independence may be 'practically unachievable' for many species, and Rooney *et al.* (1998) recommended the use of short temporal sampling intervals to produce a rich data set. Otis & White (1999) pointed out that there seems to be disagreement in the literature concerning the importance of autocorrelation in the analysis of habitat selection. (Surprisingly, Otis's 1997 paper on habitat selection with multiple patches does not discuss problems arising from autocorrelation.) They recommended using tests based on the variation among individual animals for which the number of degrees of freedom is not affected by the number of locations for each animal. For example, if the habitat use by each of ten animals can be used to create ten rankings of habitat preference, the rankings can be compared using a non-parametric rank test (Friedman 1937; Conover 1980). What would be the experimental unit and what is a replicate here? Of course, if some of the animals used are part of the same family or the same herd, problems with pseudo-replication can still arise because of lack of independence among the units treated as replicates (cf. Weber *et al.* 2001). Manly *et al.*

(2002) provided detailed advice on designing studies depending on whether an individual or a population is the subject of interest. They also described the use of a resource selection function, which is based on the ratios of observed to expected sample counts in the different habitat categories.

The question remains, however, whether there is a way, using the characteristics of the path of a single animal, to look at preferential habitat use. We will now describe one approach. It is based on the fact that in favoured habitat, the movement will have greater tortuosity with more frequent and tighter turns; this leads to less net displacement and greater residence time in those patches (Turchin 1998). Any set of k steps in a movement path can be assigned an index of tortuosity, or as we express it 'compactness'. A number of different measures could be used for this purpose, but we will describe one that we believe is new to the literature. The index of compactness can be based on the 'convex hull' of that portion of the path, which is the smallest convex polygon that contains it. Where m is the diameter of the convex hull, the largest node-to-node distance in the convex hull, and L is the total path length within it, then a simple measure of path tortuosity is L/m (Claussen *et al.* 1997). The same authors also discussed a measure based on turning angles but those can be used only when the total path is available for analysis, not just a sample of it, such as spaced observations. Figure 11.15 shows two examples of this measure; a number of other measures are also possible for this purpose.

For each point i in the path and integer k , we can calculate the compactness for a sub-path of length k centred on i . Then, those scores can be compared for different habitat or vegetation types in which the centre points of the sub-paths occur, either by averaging or in some spatially explicit way (e.g. contour maps) for a range of sub-path lengths. Figure 11.16 gives an example. Statistical testing can be carried out by superimposing the path itself, after random translation, rotation and reflection on the habitat map, say 1000 times, and recalculating the scores. The scores from the original position of the path can then be compared with the values from the randomizations and thus evaluated for significance. As always, the

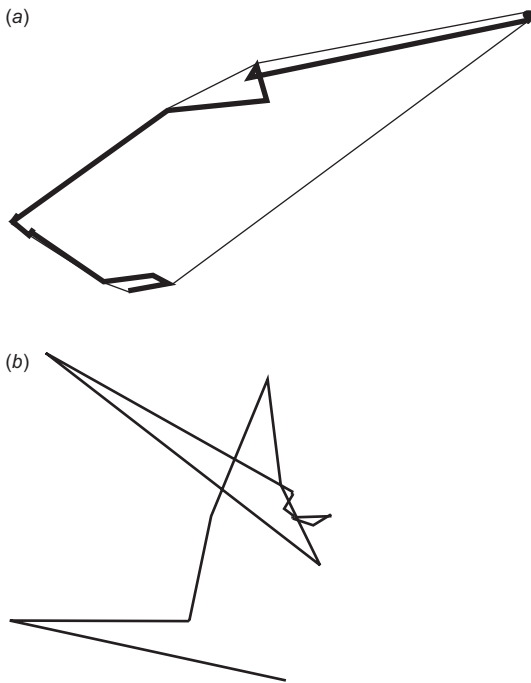


Figure 11.15 Measures of path compactness or tortuosity. (a) A sample of elk movement of low tortuosity: $L/m = 1.56$. The bold lines are the path and the fine lines complete the convex hull. (b) A sample of elk movement with higher tortuosity: $L/m = 4.12$.

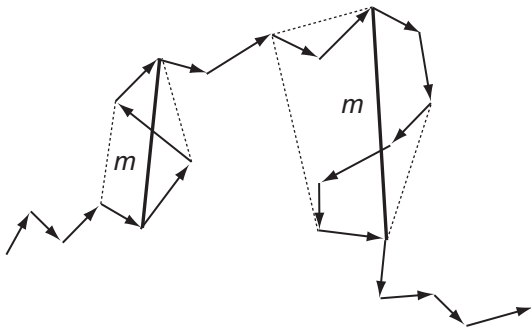


Figure 11.16 Measures of path compactness or tortuosity can be localized for given lengths of sub-path, here four or eight units; the lines are the path and the fine dotted lines complete the convex hulls in the two locations. The bold lines are the diameters, m . The measure of compactness is 1.64 for the first sub-path and 2.31 for the second.

assumptions that seem reasonable and the questions being asked will affect the randomization procedure chosen. The usefulness of compactness as a measure will depend on the pattern of behaviour.

As with the discussion of using position data to evaluate habitat use, there is much discussion in the literature of the proper evaluation of an animal's home range or of an animal's territory (Moorcroft & Lewis 2006). Again, the presence of autocorrelation in the data is an important feature that must be considered. We are not going to comment on the technical aspects of home-range evaluation, but we will point out that spatial (Valcu & Kempenaers 2010) and temporal autocorrelation is implicit in the concept of an animal's home range (Börger *et al.* 2006). The presence of these spatial and temporal autocorrelations is also due to animal behaviour resulting in site fidelity. Beyer *et al.* (2013) used a Bayesian state-space movement model to identify behaviour states from moose movement data.

There are other topics related to the analysis of the movement of individuals that we have not addressed here. For example, we have not discussed the mark-recapture methods associated with trapping session on a grid of traps. Usually the aim is an estimate of density and other characteristics related to the temporal dynamics of the population being studied, but it would be easy and straightforward to extract some spatial information from such data. We agree with the overall recommendations of de Solla *et al.* (1999) and Rooney *et al.* (1998) that more data will provide clearer answers, but that autocorrelation needs to be evaluated in determining those answers. As a general comment, however, it is clear that the topic of the spatio-temporal analysis of the movement of individuals is both important and worth further research on methods and their application.

11.7 Process and pattern

11.7.1 Tree regeneration, growth and mortality

In both population ecology and forestry, there has been an abiding interest in the fates of trees after the

regeneration of a forest following a major disturbance such as a fire or harvesting. From the point of view of population ecology, the emphasis is on the density of stems and the 'self-thinning' process, as density declines with ongoing tree mortality (Kenkel *et al.* 1997). The question for naturally regenerating stands is to what extent the size and proximity of neighbours determine the probability of mortality. From the point of view of forest managers, the emphasis is on the appropriate densities for planting and the size and age at which thinning should be carried out to maximize economic yield. In mixed forest (e.g. deciduous and conifer) there may also be questions about the timing and intensities of interventions to adjust the ratios of different species. For either kind of question, there is value in the long-term study of permanent plots in which the establishment, growth and mortality of identified individual stems are recorded at regular intervals over many years. Data on tree growth increments are typically analysed by multiple regression procedures, with the spatial factors being included as the distances to or densities of neighbouring trees, together with their heights and canopy volumes. Mortality can be analysed in a similar way, but using logistic regression to evaluate the probability of mortality as a function of those kinds of variables (Woolfons 1998).

In the absence of established permanent plots, it may sometimes be possible to obtain some of the same kinds of information by reconstructing the history of a forest stand based on tree rings, the sizes and positions of living trees, and the sizes and positions of recent and not-so-recent mortalities. For example, Carrer & Urbinati (2001) analysed structural and tree-ring variables of a timberline forest (Italian Alps) at fine spatial and temporal scales, and found that the positive autocorrelations they observed could be attributed to microsite differences, but that the spatial structure of radial growth was sensitive to extreme meteorological events. They concluded that succession seemed to be moving the forest toward a system governed by gap regeneration dynamics that would maintain the coexistence of the two main tree species, *Larix decidua* and *Pinus cembra*. Brodie *et al.* (1995) reconstructed the history of a clone of poplar (*Populus balsamifera*) in

northern Québec from the age, diameter and positions of all living and dead stems. Partial Mantel tests suggested that the clone developed in three phases: post-fire colonization, consolidation, and then directional expansion. The numerous small-diameter dead stems were aggregated and the mortality seemed to be density dependent, probably due to intracohort competition. These two examples (of the many available) illustrate the wealth of information that can be obtained from such retrospective spatial analyses of trees and the insight they can provide in the ecological processes that they document.

11.7.2 Plant mobility

The interest in the spatial dynamics of forest stands is driven at least in part by their economic importance, and the persistence of tree rings, as records of past processes, provides a useful insight to the past. An obvious disadvantage of trees as study organisms is that the processes take such a long time. Many authors have looked at the spatio-temporal structure of plant populations and communities by using herbaceous plants for which the processes are considerably more rapid. Here the rhizomes or runners, rather than annual growth rings, provide the record of past processes, such as clonal expansion.

For example, Evans & Cain (1995) studied the 'foraging behaviour' of a clonal plant, *Hydrocotyle bonariensis* (Apiaceae), in response to patches of grass. Instead of calculating a 'bias' component from growth angles, as described for Schultz & Crone (2001) above (Section 11.6), they recorded rhizome growth behaviour in the vicinity of grass patches in three categories: veers toward patch, veers away from patch, and no change in direction. The numbers in the three categories were then subjected to contingency table analysis which showed that where the grass was patchy, the rhizomes tended to veer away from it. Where the grass was uniform or absent, no significant veering behaviour was observed. We might be tempted to use this approach to study animal movements, rather than using the biased random walk approach (Section 11.6), but we would have to be very concerned in the statistical analysis about the large amount of positive

spatial autocorrelation in the data (from using one or only a few animals), as we have already discussed.

Cain & Damman (1997) studied the patterns of reproduction and clonal growth in the woodland herb, *Asarum canadense*, comparing the characteristics in an early successional forest with those in a late successional forest. They used the Pearson correlation coefficient to examine the autocorrelation of rhizome length and branching angles over time. Rhizome lengths were autocorrelated (first-order, we assume, but it is not stated) in both ages of forest, but apical angles were not. Lateral branching angles, however, were negatively autocorrelated (first-order again), with a tendency to alternate branching to the left and to the right of the parent rhizome. This study provides a relatively rare ecological example of the spatial analysis of branching linear structures.

This topic again leads us back in the direction (!) of a discussion of spatial autocorrelation and its effect on our evaluations. For the purposes of the current topic, we can end by pointing out only that it is an important first step to assess and quantify the autocorrelation in the data. Even that first step may provide some insight into the characteristics of the processes underlying the observed spatio-temporal patterns.

11.7.3 Population synchrony

The basic concept of spatio-temporal autocorrelation analysis is that samples that are closer together in space or time probably are more similar than samples that are taken further apart. In a simple world that might be true but, in both space and time, it is also possible for more widely spaced samples to be similar. In space, a simple alternation of patches of high density with gaps of low density will produce cyclic behaviour for almost any measure of spatial autocorrelation as a function of distance. Temporal cycles will produce the same behaviour as a function of time and we are familiar with many examples in ecology, such as masting by trees and the famous population cycles of snowshoe hares and lynx, or small mammals such as lemmings and voles. Often these cyclic phenomena are more-or-less synchronized over large areas, and a number of studies have examined the relative effects of

the characteristics of population dynamics and dispersal and of external forces, such as climatic events, in causing large-scale synchrony. Synchronous cycles are not the only form of spatio-temporal organization observed in natural systems or in the models used to investigate their properties. Expanding circles, travelling waves, and spirals of high density are other possibilities, somewhat like the wave-regeneration in some forests or the development of banded vegetation in 'brousse tigrée' described above (Section 11.1). Much of the literature on spatio-temporal organization has concentrated on the obvious cycling of certain populations because they are widespread (in more ways than one!) and provide a challenge to our understanding of the factors that determine the abundances and distributions of organisms (cf. Krebs 2002).

Given a cyclic system that is found over a wide geographic area, it is of interest to ask about the relationship between the cycles at different locations, as a function of distance. Do the cycles exhibit the same periodicity or does the period change with location? If the cycle lengths are more or less the same, are the cycles in synchrony over the geographic range or does synchrony decline with distance?

There are several different, but closely related, ways of analysing data from a set of spatial locations in order to answer these kinds of questions. In general, the data for any particular location, i , will consist of a time series of population densities, $N_i = N_{i1}, \dots, N_{it}, \dots, N_{iT}$. Density data are often log-transformed before analysis:

$$X_t = \log(N_t + 1). \quad (11.12)$$

The method described by Hanski & Woiwod (1993) uses these log-transformed data, but removes the first-order temporal autocorrelation effects by fitting the equation

$$X_{t+1} = a + bX_t, \quad (11.13)$$

and then using the residuals for further analysis

$$R_t = X_t - (\hat{a} + \hat{b}X_{t-1}). \quad (11.14)$$

One general approach is the calculation of the cross-correlation coefficient between the two series with a

time lag of zero. Where there are two time series, x_1 and x_2 , both of length T , their cross-correlation is:

$$r_{12}(0) = \frac{\sum_{t=1}^T (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)}{\sqrt{\sum_{t=1}^T (x_{1t} - \bar{x}_1)^2 (x_{2t} - \bar{x}_2)^2}}. \quad (11.15)$$

Where the data consist of two time series for each of a number of locations, it is also possible to examine the effect of spatial distance on the cross-correlation coefficient of the two time series. Tobin & Bjørnstad (2003) provided an interesting example of the application of this approach to a study of the spatio-temporal relations between a prey species, the house fly (*Musca domestica*) which is a serious pest in commercial hen houses, and a predatory beetle, *Carcinops pumilio*. Given the cross-correlation of the two time series for each of a number of locations (108 in one large hen house and 162 in another), a kernel function can be used to give an estimate of the cross-correlation at any given distance. They found that during the exponential growth phase of the fly population, the beetles were strongly negatively cross-correlated with their prey at local spatial scales.

Hanski & Woiwod (1993) used a related approach to examine the spatial synchrony of populations of single species. Studying the densities of different kinds of insects (moths and aphids) in England, they used the cross-correlation between the residuals of conspecific population densities for each pair of sites, r_{ij} , and plotted them as a function of distance between sites, d_{ij} . The y intercept of the linear regression of correlation as a function of distance was used as a species-specific measure of synchrony to be compared with a measure of population variability through time. They found a positive relationship between synchrony and temporal variability for aphids and noctuid moths, but not for geometrid moths.

Another method is to make a direct comparison of the $n \times n$ matrix of correlation coefficients, r_{ij} , with the $n \times n$ matrix of inter-site distances, d_{ij} . The usual technique for comparing such a pair of matrices is the Mantel test (Chapter 7), which uses a

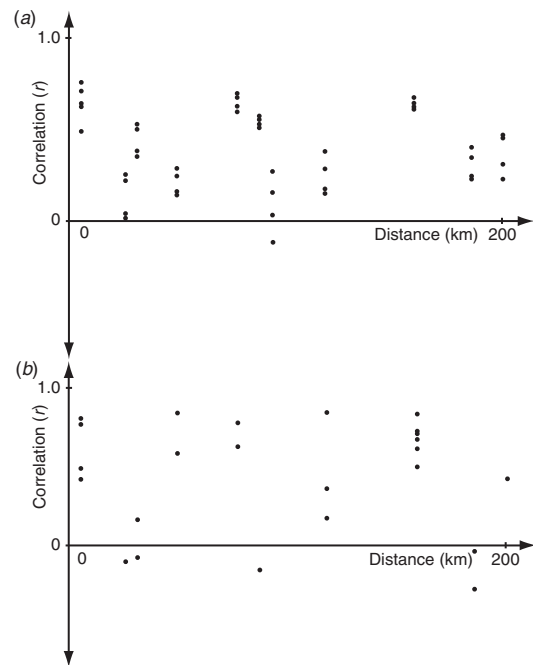


Figure 11.17 Correlation of *Picea glauca* tree-ring widths as a function of geographical distance in central Alberta. (a) Five sites, the same two trees per site. (b) The same five sites, two trees per site, but the trees not reused (sampling without replacement).

randomization procedure to test for the significance of the relationship between the two. Koenig & Knops (1998) cautioned that because there are more pairwise correlation coefficients than there are sites, 'the potential for pseudo-replication biasing the results of statistical tests cannot be ignored'. They suggested that, because autocorrelation is expected to decline with distance, the Mantel correlogram may not produce results of ecological interest, when it detects only a decline in autocorrelation as a function of distance. Figure 11.17a gives an example of the correlation of tree-ring widths in *Picea glauca* as a function of geographic distance in Alberta (Peters 2003). The results for two trees per site and five sites show little evidence of a systematic decline in correlation with distance, indicating that, in this case at least, the situation is not the simple one that Koenig & Knops (1998) described. In fact, the simple decline with distance

may not be as common as those authors suggest, as we described in Chapter 8.

The same authors (Koenig & Knops 1998) recommended the use of a 'modified correlogram' to display the results graphically, by plotting the mean correlation coefficient between the time series of randomly chosen pairs of sites within specified distance classes; to avoid pseudo-replication problems, each site is used only once. Ranta *et al.* (1997) used a randomization technique to compare the level of synchrony at any particular site with others, by choosing other sites at random, thus avoiding the same problem. Figure 11.17*b* shows a reanalysis of the same *Picea glauca* tree-ring data of Figure 11.17*a*, but using randomly chosen pairs of trees that are then not reused (sampling without replacement). The conclusions drawn would be the same. Where each site has two time series of data, for example acorn production and annual growth in oak trees, the cross-correlation coefficient between the two series can be used in the same way (Koenig & Knops 1998). It is not clear to us how great a problem this reuse of data really is; as described in Chapter 5, many exploratory analysis techniques are based on the repeated use of the same data; TTLQV being an extreme example (Section 5.2). There is the usual trade-off between independence and the effective use of the information available. This comment applies to a wide range of methods not just to the correlation techniques being discussed here.

There are many variations on these basic methods. Sutcliffe *et al.* (1996) used the non-parametric Spearman's correlation to evaluate the synchrony of butterfly population dynamics. This approach uses the ranks of the abundances rather than the values themselves and may be less affected by extreme values (cf. Conover 1980). As with other correlation statistics, it applies to pairs of series and for more than two series in a region, the measures for all pairs could be averaged to give a regional measure of synchrony. An alternative is to use Friedman's measure of concordance for several series (Conover 1980). Bjørnstad *et al.* (1999a, b) averaged the pairwise cross-correlations to get a regional measure of synchrony but based them on the year-to-year changes in density, not on the densities themselves.

Bascompte & Solé (1998) described how the analysis of two data sets confirmed the predictions from spatially explicit dynamic population models of spontaneous self-organization in the form of spirals or travelling waves of population density. Since then, a number of studies have demonstrated the phenomenon of travelling waves of density, for example in field voles (*Microtus agrestis*; MacKinnon *et al.* 2001) and in red grouse (*Lagopus lagopus* ssp. *scoticus*; Moss *et al.* 2000). At a larger scale (the entire Canadian boreal forest, of the order of 5000 km across), Viljugrein *et al.* (2001) found that although there is broad-scale synchrony in the population cycles of mink (*Mustela vison*) and muskrat (*Ondatra zibethicus*), peaks and troughs in these cycles generally appear first in the Athabasca basin and spread from this epicentre. Similarly, while the basic oscillation of spruce budworm (*Choristoneura fumiferana* Clem.) is the same across all of Ontario (about 1000 km), large outbreaks appear first in the eastern zone, followed by the central zone, and then five or six years later in the western zone (Candau *et al.* 1998). It is not our purpose to review the mechanisms believed responsible for this travelling wave phenomenon, but see Sherratt *et al.* (2000) and Sherratt (2001) for interesting discussions. Travelling waves are detected in the spatio-temporal data by looking for anisotropy in the spatial covariance (Bjørnstad *et al.* 1999a, b; Lambin *et al.* 1998). If there is a travelling wave, cross-correlation declines markedly with distance in directions perpendicular to the wavefront, but does not decline with distance parallel to the wavefront. The plot of cross-correlation as a function of distance, described above, is merely divided into a few direction classes and examined for differences. Lambin *et al.* (1998) and colleagues have described a modelling method to help estimate the speed and direction of such a travelling wave and to determine its statistical significance.

There are two related questions of the consistency of spatial pattern through time and of the synchrony of temporal patterns in space. This section has attempted to provide a description of the various methods that can be used to answer these two questions. Answering these questions merely leads (of course) to more

questions, now concerning the ecological processes that lead to the spatio-temporal patterns we detect.

11.7.4 Spatio-temporal chaos

In the preceding section, we discussed some of the spatial aspects of population dynamics, or the temporal aspects of patterns of population density, but we avoided the topic of spatial (or spatio-temporal) chaos, because it is sufficiently interesting and important to merit a section of its own. Used in a technical sense, 'chaos' refers to a system's behaviour in time or space that is irregular and possibly very complex but that is strictly deterministic. Chaos is not the same as randomness or stochastic behaviour, although it may appear to be unpredictable and aperiodic. One distinctive feature of chaos is that the overall behaviour can be very sensitive to very small changes in conditions (the so-called butterfly effect; Schroeder 1991). We will begin our discussion of chaos as it relates to spatio-temporal analysis with a short review of an example that is probably very familiar to the reader.

Consider the following difference equation that describes the dynamics of the population density at a particular location and time interval, N_t , as a function of the density at the preceding time interval, the growth rate of the population, r , and the carrying capacity for the population, K :

$$N_t = rN_{t-1}(K - N_{t-1})/K, \quad (11.16)$$

which can be rewritten as

$$n_t = rn_{t-1}(1 - n_{t-1}). \quad (11.17)$$

This equation describes logistic population growth, which is almost exponential when the population is well below the carrying capacity, with the growth rate being slower at higher densities and decreasing to zero at the carrying capacity. The growth rate is negative when the carrying capacity is exceeded and so the population declines. The behaviour of the population that is derived from this equation depends on the intrinsic growth rate, r . When r takes the value 2.5, the equilibrium value of n_t is 0.6, and the population converges to this value, no matter what the starting density. If we plot a two-dimensional diagram of n_t

versus n_{t-1} , any trajectory will converge to the point (0.6, 0.6), which can be thought of as an 'attractor' under these conditions (Figure 11.18a). When r is 3.2,

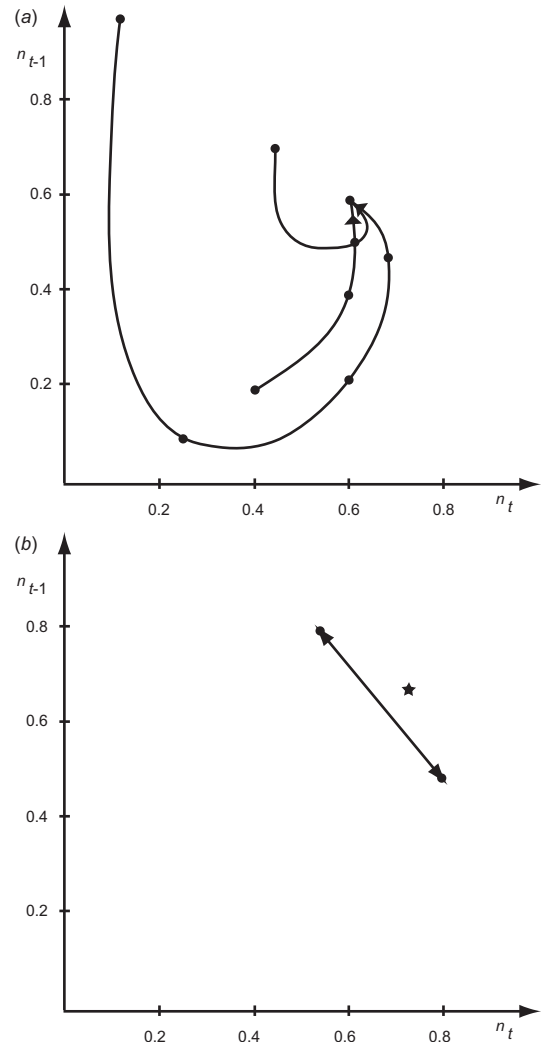


Figure 11.18 (a) When $r = 2.5$, the different trajectories in the phase space converge to the equilibrium point of (0.6, 0.6). (b) When $r = 3.2$, the trajectories do not converge to the unstable equilibrium value of 0.6875 (the star), but converge to an alternation between two other values. (c) When $r = 3.4$, the 'equilibrium' is a cycle among four different densities.

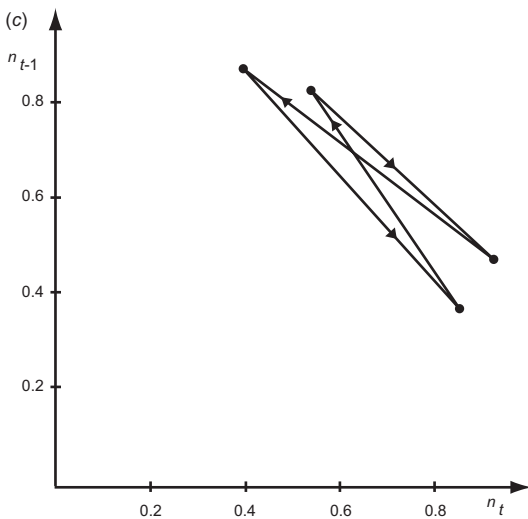


Figure 11.18 (cont.)

there is an equilibrium value of 0.6875, but it is almost never reached, because the equilibrium is unstable, and n_t alternates between two values: 0.513 and 0.7995 (Figure 11.18b). When r is increased to 3.4, n_t cycles among four values: approximately 0.875, 0.383, 0.827, and 0.501 (Figure 11.18c). Further increases cause doubling of the lengths of the cycles, but the behaviour soon becomes aperiodic (that's chaos!). What is most fascinating about this simple system is that as r continues to increase, the behaviour returns to simple cycles, then back to chaos and so on (usually illustrated with the 'bifurcation to chaos' figure that appears in many places, e.g. Schroeder 1991, figure 12.11). While as a simple model of population behaviour, this approach to chaos seems unrealistic because the values of r required to achieve chaos are unnaturally high, it provides an important lesson on the potential importance of nonlinear dynamics. The diagram plotting n_t against n_{t-1} , as in Figure 11.18, is called a phase space diagram, in which time is not represented as an axis, but is included explicitly in the drawing of the trajectories, through time, of the combinations of values. In complex situations, we may plot only the 'attractor' to which the trajectories converge (Figure 11.18a). The attractor may be a single point,

as mentioned above, a finite loop of 2, 4, 8, ... points, or in the case of chaos, a 'strange' attractor that is infinite but bounded and often (always?) fractal (of fractional dimension).

How does this relate to spatio-temporal analysis? The first question is how to detect chaos in time series from individual locations and to determine whether natural systems indeed can be chaotic. In 1995, Solé & Bascompte wrote: 'Detection of chaos in ecological data is one of the most challenging problems in contemporary ecology. It is not enough to certify that time series are complex, we need a comprehensive approach to that complexity. ... [T]he question of chaos in nature is still an open one'.

In that decade, much effort was devoted to trying to meet that challenge, but without complete success, some of which we shall describe below (cf. Stone & Ezrati 1996; Perry *et al.* 2000).

If temporal chaos does exist in ecological systems, a second challenge is to determine the relationships among chaotic data series at a number of locations, and to find whether those relationships have a spatial component. A related question is whether there is such a thing as spatial chaos. If there is, how do we detect it and what are the characteristics of its dynamics? We might also ask about the relationship between space and time in chaotic systems and how we might determine that relationship.

We will begin by considering the first question, concerning the detection of chaos in a single time series. At first glance, a chaotic time series seems indistinguishable from a noisy stochastic series, and it may seem to be a challenging problem to detect chaos, but actually it is not, at least in theory. The basis for detecting chaos goes back to a characteristic of chaos described above: that small differences become amplified. Let us consider the example of Eq. (11.17), using two different values of r to create two series both starting with $n = 20$ and $m = 21$. When we compare $\Delta i = |m_i - n_i|$ for the two series, with $r = 3.95$ these differences increase and chaos is the result; with $r = 3.25$ these differences tend to decline and the result is not chaotic.

Where there is chaos, the trajectories that started close together diverged in the phase space; where the

behaviour was not chaotic they converged. To put it in more mathematical terms, consider two trajectories in a phase space that are separated by a small amount, ε , at time t ; at time $t + \tau$, they are separated by $\varepsilon_\tau = \varepsilon e^{\lambda\tau}$, where λ is a constant that is characteristic of the system, known as the Lyapunov exponent. In fact, it is the Lyapunov exponent that will allow us to detect chaos: if it is greater than zero, there is chaos. If it is less than zero, there is convergence. An exponent of zero indicates cyclic behaviour such as the limit cycles predicted from simple predator–prey models, familiar from ecology textbooks (Figure 11.19). It seems that all that is required to detect chaos, then, is to determine the Lyapunov exponent from the data and then draw a conclusion based on its value.

Problem solved ...? Well, no, not exactly. The difficulty is that to get a reliable estimate of λ from real data, long data series may be needed, and those are seldom available in ecological studies. One feature of ecological data is that even if there is a chaotic deterministic foundation to the data, we would expect a stochastic component of exogenous factors to be added on somehow (Dennis *et al.* 2001), and some methods of analysis may be affected by this ‘noise’ in the data (cf. Ellner & Turchin 1995). Even if long data series are available, the characteristics of the population, such as the intrinsic growth rate, may change over time, so that the Lyapunov exponent we are trying to estimate is not constant. As Ellner & Turchin (1995) concluded, it may be more sensible to ask of any particular system ‘When and how often is it chaotic?’ rather than ‘Is it chaotic?’. This suggestion fits with one theme in this book, the distinction between global and local evaluation of the characteristics of the phenomena being studied.

In addition to the technique of estimating the Lyapunov exponent, a number of other methods to detect chaos have been proposed; see Stone & Ezrati (1996) and Perry *et al.* (2000). Using response surface methods, Turchin & Taylor (1992) found a number of different dynamics in natural populations of insects and vertebrates, but only one, the aphid *Phyllaphis fagi*, was thought to be chaotic. Perry *et al.* (1993) included further data from the same population of *Phyllaphis fagi*, and from other populations of the

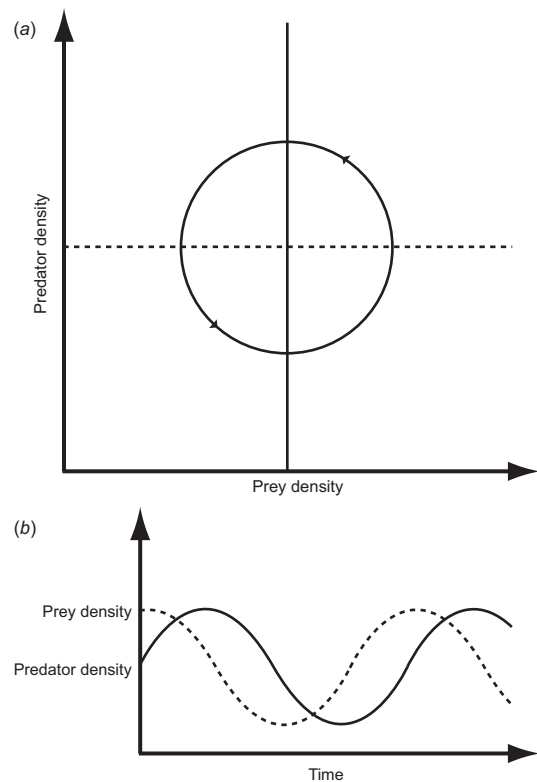


Figure 11.19 The cyclic behaviour predicted by simple predator–prey models, depicted in phase space (a) and as relative densities as a function of time (b). The predator density increases when prey are plentiful, but high predator densities drive prey density down. Predator density declines when prey are scarce.

same species, and decided that its dynamics were actually stable, not chaotic. Knowledge of the biology of the system being studied seems to be essential for interpreting the dynamics, whatever techniques are used to investigate them. From Perry *et al.* (2000), one might conclude that many systems are on the ‘edge of chaos’, with Lyapunov exponents around zero. While chaos may turn out to be rare in nature, it is important when studying dynamics and spatio-temporal systems to keep in mind that it is possible.

Given that chaos is possible in purely temporal systems, what is possible when a spatial structure is

included? The easy answer is that, in theory, almost anything can happen. A straightforward way to include a spatial element in simple dynamic models is to have two or more metapopulations, each governed by the same underlying model, but linked by dispersal between adjacent subpopulations. Gonzalez-Andujar & Perry (1993) investigated such linked populations and concluded that linking the populations reduced the occurrence of chaos. Ruxton (1993) responded pointing out that linked populations may still be chaotic. Doebeli & Ruxton (1998) extended this work on metapopulation dynamics, showing that long-range dispersal can stabilize otherwise complex dynamics and that short-range dispersal can destabilize otherwise stable dynamics.

To illustrate these effects, we will use a model of two populations, both governed by Eq. (11.17), and either linked by dispersal or isolated from each other. We will consider several different situations, all with $r = 3.58$. With starting densities of $n = 0.50$ and $m = 0.80$, the behaviour is chaotic, with the attractor as shown in Figure 11.20*a*. When the same starting values are used, but the populations are linked by an exchange (dispersal in each direction) of 5% of each population, the behaviour becomes a cycle of period 2, as shown in Figure 11.20*b*. Because the system is 'on the edge of chaos', the outcome can be changed merely by changing the starting densities, while the degree of linkage is unchanged. Starting with $n = 0.90$ and $m = 0.80$, the result is a cycle of period 4 (Figure 11.20*c*). Starting with densities of 0.80 and 0.85, the two subpopulations essentially act as one and chaos returns, and the situation reverts to that shown in Figure 11.20*a*. This example, based on a very simple system of two subpopulations, illustrates well the potential complexities of the behaviour of chaotic systems with spatial structure. Think of the possibilities with greater spatial complexity! This kind of interaction between subpopulations should be considered in interpreting spatial dynamics or the spatio-temporal patterns of density.

Spatial chaos? Why not? At least in theory. Any model that can generate chaos in time can generate chaos in space, but it is not clear how applicable such models would be in describing real systems. On the other hand, Petrovskii & Malchow (2001) described

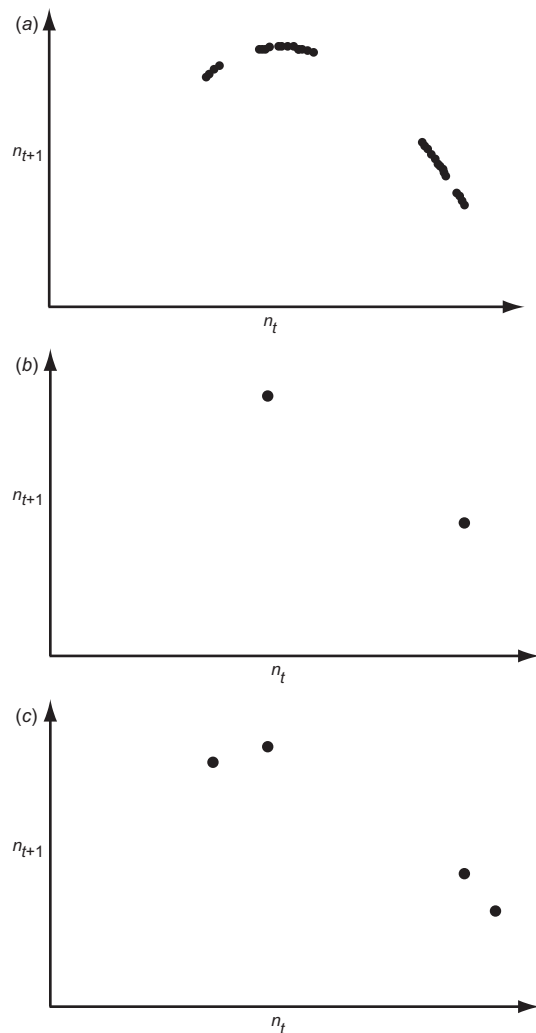


Figure 11.20 The effects of interchange between populations on the edge of chaos, with $r = 3.58$. (a) $n = 0.5$, $m = 0.8$: chaotic behaviour; (b) 5% interchange: chaos suppressed, a cycle of period 2. (c) $n = 0.9$, $m = 0.8$, cycle period 4 with 5% exchange; but with $n = 0.85$, $m = 0.8$, the chaotic behaviour returns as in (a).

a spatially explicit predator-prey model in which spatio-temporal chaos appears in a subdomain of the system and then spreads to take over the whole space. A kind of purely spatial chaos can be found in the

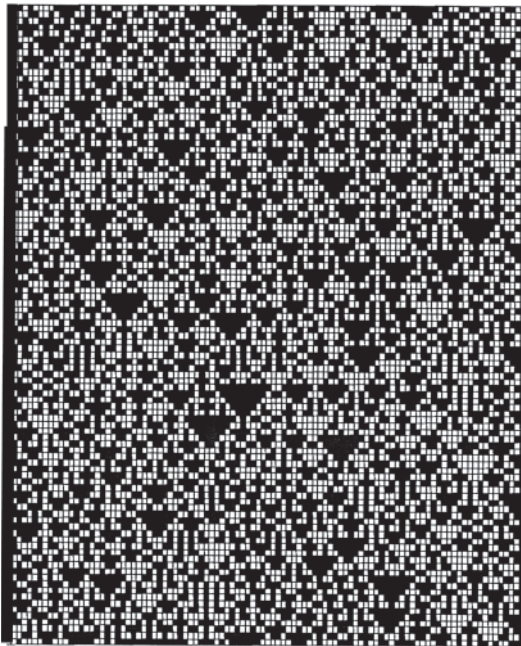


Figure 11.21 Something like spatial chaos in a cellular automata system, Wolfram's (2002) 'rule 150'.

development of some cellular automata models. For example, Wolfram (2002) displayed a number of examples which seem to have complex aperiodic behaviour, which he refers to as 'random', but which, because they result from deterministic rules, should more properly be referred to as chaotic (e.g. p. 227, 'rule number 150'; Figure 11.21). Some cellular automata resemble chaos in their sensitivity to initial conditions; Wolfram's 'rule 73' produces qualitatively different patterns for different starting conditions. These cellular automaton systems resemble temporal systems more than spatial chaos, in some ways, because there is a strict directionality in them for cause and effect (top to bottom in our figure).

Given the interaction of deterministic effects, stochastic factors, and the underlying spatial structure, it will probably be even more difficult to demonstrate true spatial chaos in nature than it has been to demonstrate true temporal chaos. Diks *et al.* (1997) warned that if spatio-temporal chaos exists and we study it only through time or in space, we may be misled

because we need to look at both space and time together.

For the purposes of this book and of this chapter on spatio-temporal analysis, we have probably said enough on the fascinating topic of chaos. The main point is to be aware that in biological systems, non-linear dynamics should be considered as a possibility, and that chaos and the near approaches to chaos cannot be totally ruled out as reasonable explanations of the observed behaviour, but it may be very difficult to determine unambiguously.

11.8 Spatio-temporal graphs

In Chapter 3 (Section 3.6), we introduced spatial graphs to analyse patterns of movement in an ecological context, as one of several phenomena, and that introduction leads naturally to considering graphs as a general approach to spatio-temporal analysis. This is an extension of the use of now-familiar spatial graphs by adding the dimension of time to produce spatio-temporal graphs. Spatio-temporal graphs are those in which the nodes have explicit locations in time and in space, and so the edges have locations in space and time for their end-points at least, even if the edge does not trace a realized spatio-temporal trajectory.

In some versions, spatio-temporal graphs will have the same sets of nodes and edges throughout the period being studied, but with changes to the state or nature of the nodes (representing, for example, healthy versus diseased plants) or to the state and nature of the edges (preventing or facilitating the movement of disease vectors, for example), thus depicting the movement of a state or disease through a population, as in a study of percolation thresholds. In other versions, the set of nodes or the set of edges may themselves change through time, with the loss or gain of either kind of element. In that case, the spatio-temporal graph can depict how changes represented in the graph structural connections or 'topology' affect the ecological processes (e.g. dispersal of aquatic organisms through intermittent stream networks) or how the processes affect the graph structure (e.g. fire in a heterogeneous landscape removing stands of trees as nodes of a

network). These possibilities lead us to consider how the characteristics of spatio-temporal graphs might help in classifying them for the ecological context.

11.8.1 Characteristics and classification

Before introducing our own thoughts on classifying spatio-temporal graphs into ecologically meaningful categories, we can start with a summary of *dynamic graphs* by Harary & Gupta (1997). Dynamic graphs are not necessarily spatio-temporal graphs, because they are initially aspatial in context, but with a temporal setting. In addition to a graph's usual two sets V (the nodes) and E (the edges) already described in Chapter 3 of this book, Harary & Gupta (1997) also include functions f and g that assign weights to the nodes (f) and edges (g), greatly increasing the flexibility and range of applications for these graphs. The graphs of traditional graph theory are *static*, because they do not change over time or because they deal with events that can be aggregated in a summary graph that is static (Kostakos 2009; Tang *et al.* 2009). These can be thought of as *atemporal*, because time is not explicit in their development or interpretation, as well as being aspatial; *dynamic* graphs, on the other hand, change in some of their characteristics with time, as the name suggests (Harary & Gupta 1997). This consideration leads to the following classification of basic types of these dynamic graphs.

- Node dynamic: the set V changes with time, in that nodes can be gained or lost (and lost nodes take their edges with them when they go).
- Edge dynamic: the set E changes with time, in that edges can be gained or lost (but as usual, edges can only exist between two nodes of the graph).
- Node-weight dynamic: node weights (or states) change with time because f can vary.
- Edge-weight dynamic: edge weights change with time because g can vary.
- Fully dynamic weighted graph: both f and g can vary with time.

A further augmentation of the basic concept is to allow dynamic digraphs (the edges have directions) and dynamic signed graphs (the edges are positive or negative) in the same basic classes, and thus dynamic

networks should also be included. In fact, the most familiar ecological example of a dynamic graph is a dynamic network: a food web is a dynamic aspatial network with variable weights on the nodes representing species biomass and varying weights on the edges representing rates of consumption which fluctuate with prey availability and predator switching behaviour.

The Harary-Gupta description of dynamic graphs is not the only version of the basic concept, nor does it provide the only terminology in use. Kostakos (2009) referred to *temporal graphs* and Casteigts *et al.* (2011) preferred *time-varying graphs*, but also used the related term *dynamic networks* (Casteigts *et al.* 2011) where Kempe *et al.* (2002) used *temporal network*.

When we introduce the characteristics of dynamic graphs into an explicitly spatial graph, we arrive at the concept of a spatio-temporal graph, in which the nodes have locations in space and time, and edges exist in a spatio-temporal context. The edges have fixed end-points, the nodes, and their lengths represent displacement rather than actual distance measured with respect to the interaction being represented. The organisms of interest may be short-lived or long-lasting, and it is possible for the nodes to maintain their identities while moving to different locations through time.

Spatio-temporal graphs have the potential to be complex in structure, but that complexity permits their application in a wide variety of situations. For example, time and space can each be treated in either continuous or discrete form (steps for time or raster format for space). The time steps do not need to be of equal duration if only the order of events is important. For example, in a study of spatial variation in phenology, the order of records of first leaf, second leaf, first anthesis, and so on, may be preserved even if the dates of these indicators vary by weeks. A spatio-temporal graph could be a three- or four-dimensional structure with the nodes as points located in the two or three dimensions of space, but run out through the dimension of continuous time to form 'linear' objects; the edges would also have to be extended through time, forming planes. More usually, the structure is divided into discrete temporal units of planes (two spatial dimensions) or volumes (three spatial dimensions),

so that the nodes can remain as points with dimension zero, as in the more familiar examples of graphs. In general, the events of interest may have short durations through the period of record, and the temporal slices aggregate the events that occur. The relationships between nodes are shown by edges that are lines within each of the temporal units. If necessary, a special category of edge can be created to run between temporal units to keep track of node identities by joining up instances of the same node, or edges can be categorized according to whether they experience change between observations. For example, Figure 11.5 shows a map of a population of organisms (the nodes of the spatial graph are stationary organisms like plants) observed at four different times, giving four different graph panels, and we could add 'edges of identity' running between identical nodes in the different panels to show the continuity of individuals as nodes. Here, the organisms (the nodes) do not move between observation time periods and that addition is probably not necessary. Another possibility is for edges running between the panels, from one time period to the next, showing the most probable sources of new infections, the individuals that were infected in the preceding time period that are closest in space. These edges then link the most likely pairs of 'ancestors' and 'descendants' in chains of disease transmission. These same edges could be used in a similar way in a single graph that summarizes the four time periods, with temporal labels on the infected nodes to indicate the time of first infection.

While the multidimensional space format may seem ideal, visualization and analysis may be more practical when divided into temporal slices at equal (or unequal) time intervals. The dimension of time can also be reduced to a single time interval 'snapshot' or condensed to provide an integrative summary of the structures throughout a finite time course. There may be a choice, therefore, between having time represented as an explicit dimension and having the dimension of time, and the temporal locations within it, being given by numerical or categorical labels associated with the nodes or edges.

In a spatio-temporal graph, the nodes can be spatially mobile or they can be fixed at permanent

spatial locations for the duration of the period under analysis. In addition, the nodes can also be classified as persistent, when they exist throughout the time period, or instantaneous or transient when they do not. Intermediate possibilities for nodes through time include intermittent (a sequence of being 'on' or 'off' like 1s and 0s, with a weight of zero being equivalent to non-existence) such as a seasonal pond that alternates between being water-filled and dry, or a variable measure of effective existence (such as water depth in an intermittent pool). Where time is a discrete rather than continuous variable, it may be common for a node to last for only a single time interval (for radio-collar locations, the animal is usually at any one location for only one observation). The nodes can then represent the locations of objects, or the objects themselves, and their identity will determine their temporal characteristics such as longevity. As in spatial graphs, nodes in spatio-temporal graphs can obviously have different properties associated with them such as categorical, qualitative, or quantitative labels (e.g. species, health index, size, and so on) or mathematical functions describing processes (e.g. a sine function describing evaporation rate as it varies throughout a day or a year).

Similar to the nodes, the edges may be persistent or transient in some applications for the same reasons, or just time-indexed when they designate an event or relationship that lasts for only a short time. Again, an appropriate example is the set of radio-collar locations for tracking animal movement in which the edge joins nodes for times t and $t + 1$. Edges, like nodes, can also be intermittent or variable, such as a temporary stream that only flows after rainstorms, or a measure of its flow rate. The characteristics of the edges are not independent of the characteristics of the nodes because the existence of any edge depends on the existence of the nodes that define its end-points.

Where the interactions depicted by edges of a spatio-temporal graph are transient, a summary graph that includes edges from a set of time intervals (or just over a time range) may be misleading in giving the impression of which nodes can be reached from which others. As a simple case, if the overall connection among three nodes looks like $A - B - C$, and we are

interested in the transmission of information (or disease) from A to C, it cannot happen if the B — C edge disappears before the A — B edge forms, but it is possible if A — B precedes the edge B — C in time. Tang *et al.* (2009) made this point clearly in their discussion of distance metrics for analysing social networks, but the same is true for ecological examples, and Blonder & Dornhaus (2011) described the limitations on information flow in ant colonies by reference to *time-ordered networks*. The same argument about order may apply equally well to nodes as to edges. Where nodes represent ephemeral habitat patches used as stepping stones for dispersal, the order of appearance and disappearance of nodes may combine with limitations on dispersal distances to permit or to limit stepping-stone dispersal.

In addition to having temporal properties of their own, edges can indicate temporal relationships, such as contemporaneity in an integrative summary (showing which nodes exist at the same time) or identity (which nodes represent the same object in different time periods), or ‘ancestor and descendent’ relationships (which nodes are derived from which others through time, by whatever processes are appropriate to the study, see Del Mondo *et al.* (2010)). We have already indicated that edges in spatio-temporal graphs can have direction, as with other versions that create digraphs, and the edges can represent objects (e.g. dispersal corridors), processes (e.g. pollination), or relationships (e.g. competitive interaction). In spatio-temporal graphs, edges can have properties associated with them such as categorical, qualitative, or quantitative labels or mathematical functions describing processes, such as flow rates as a function of time. Table 11.2 provides a summary of criteria that can be used to classify spatio-temporal graphs for the ecological context.

For studies in ecology, the development of applications of spatio-temporal graphs will provide great analytical flexibility and power. The possibilities and combinations are seemingly endless, and we have had difficulty in providing a truly coherent and complete classification of those possibilities. However, we have provided a preliminary assessment of some of them, and will now describe a few of the many potential applications.

Table 11.2 Classification criteria for spatio-temporal graphs

Time: Discrete (equal steps or order) versus continuous (usually made discrete)
Full dimensional time-course versus snapshot versus integrative summary
Depiction by an explicit dimension versus temporal labels
Space: Discrete (raster, objects, x - y) versus continuous (may be made discrete)
Number of spatial dimensions: 1D, 2D, 3D, ...
Nodes:
Time: instant, transient, intermittent, variable, persistent (constant)
Temporal characteristics: snapshot (contemporaneous) versus integrative summary (node as 1D object in 4D, condensed to 0D point) versus time-course (‘ancestors’ and ‘descendents’)
Space: fixed, moving
Representing: locations, objects, events
Properties: categorical, qualitative, or quantitative labels; associated functions
Edges:
Time: instant, transient, intermittent, variable, persistent (constant)
Temporal characteristics: one time versus contemporaneity in summary graphs versus time-course (linking ‘ancestors & descendents’, or linking instances of the same object)
Space: directional, symmetric versus asymmetric
Representing: locations, objects, processes, relationships
Properties: categorical, qualitative, or quantitative labels (including signs); associated functions

11.8.2 Animal movement with spatio-temporal graphs

For the study of animal movement, there are at least two contexts for the choice of how to deal with a spatial graph with the added temporal dimension. The first has to do with territory and home range, encamped with daily movement, where an animal has a behavioural attachment to a defined area, within which most of its activities occur, but with occasional forays of exploratory movement outside the imaginary boundaries of that region (Beyer *et al.* 2013). There is no overall net change in the organism’s average position

expected through a long series of observations, but only spatial variance around some central mean location, as the organism consistently returns to its home base. The second has to do with the locations of an animal as it moves through a landscape of different patch types as it forages or migrates, so that there is net displacement between the beginning and the end of the observations and between successive observations. In both cases, one problem is that the locations are usually a sample of where the animal has been, and while the order of locations is known, the intermediate locations between observation i and observation $i + 1$ are not known. This means that both the actual path is not known and the rate of movement can be determined only as net displacement over time elapsed between the point data. As the time between locations increases, this lack of information becomes of greater concern, but even for closely spaced observations (e.g. 15 minutes versus 24 hours) the gaps can represent a significant unknown. For studies of animal territory or home range, while the order of observations is known, they are sometimes analysed unordered or unlabelled, with the result that the data can appear very disordered when mapped.

To present a simple example, we consider the case of $n = 6$ locations, which are the nodes of a spatial graph, and $(n - 1) = 5$ transitions, which are the edges of a graph (Figure 11.22). (The edges also have an order to them, following that of the nodes, and a direction following the arrow of time.) In comparing examples (a) and (b) in Figure 11.22, one of the most striking differences is the fact that the edges in (b) cross, whereas in (a) they do not. The number of crossings is therefore one statistic that can be a measure of path complexity, call the statistic X , which can be evaluated by considering the random re-labellings of the nodes and their associated edges. A second statistic of interest measures whether the crossings occur between edges of the path that are close to each other in time or whether they are temporally separate; a simple measure would be the pairwise correlation of the temporal labels of edges that cross. Again, its significance can be evaluated by comparison with random re-labellings of the nodes. This is one case

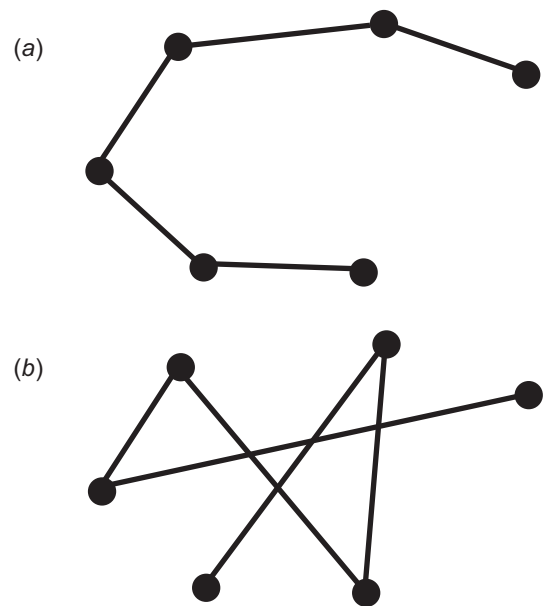


Figure 11.22 Two graphs of six locations with five edges, one with no edge crossings (a), and one with four (b).

where the actual time periods between observations are less important than the order in which they occur. Equal time intervals are not required, although equal spacing may help in clarifying the interpretation of the findings.

In complicated situations or for large data sets, this kind of analysis could be carried out for short sub-paths of the sequence, with the results recorded, mapped or plotted in order to detect changes in time or space.

For the second set of circumstances, where non-zero net displacement is expected as a result of foraging or migration, the question of interest is how to detect important behavioural patterns from the incomplete information provided by the periodic data acquisition. For example, in Figure 11.23, can we determine whether characteristics of the path indicate avoidance of the area marked as a shaded rectangle, representing habitat gap that might be avoided such as a clear-cut in a continuous forest, or unusually rapid movement across that patch.

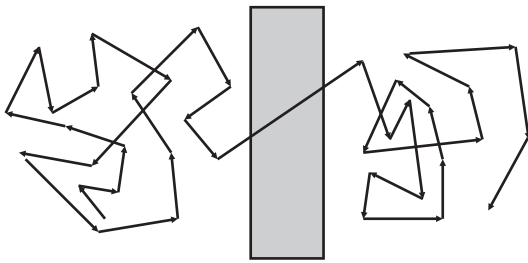


Figure 11.23 Locations of an animal tracked through a landscape at equal time intervals, with the order indicated by the edges. The rectangle represents a patch of habitat that is possibly avoided by the animal.

This thinking leads to a hierarchy of questions and a hierarchy of tests.

- (0) *Complete randomness.* The nodes as end-points of the edges are randomly and independently positioned in the entire area being considered.
- (1) *Complete avoidance* of the gap. There are no nodes of the spatial graph within a given area, here for example the shaded rectangle. There are 28 observations, and the shaded area represents 14% of the total area. With all the usual randomness and independence assumptions, the probability that the shaded area has no nodes is $(0.86)^{28} = 0.0146$.
- (2) *Few crossings* of the patch. Given the positions of the nodes, if they were randomly joined in pairs by edges of the graph, the frequency of realizations with 1 or fewer intersections of the patch can be shown to be significantly small. The concern might be whether completely random pairing of nodes constitutes a fair test, since it will tend to produce longer edges than observed; we could use a restricted randomization that limits edge lengths.
- (3) *Crossing unusually rapid.* Compared to other single transition path segments, it is by far the longest, indicating very rapid movement. Hence the individual is avoiding a specific landcover type. This can be approached using randomization test, although the number of edges may sometimes be too small. Comparison with the model of a correlated random walk (CRW; see Codling *et al.* 2008) should be useful here.

Other questions that can be answered include the following.

- (4) Is the path self-avoiding? The number of self-crossings is only five. Random pairing of nodes would produce more, but is that a fair test? Again, a correlated random walk might be a more appropriate model for comparison.
- (5) Given information on habitat classification for the entire area, not just the shaded rectangle, can we use a resource selection function or other approaches (such as a step-function, or patch-focused CRW, and so on) to characterize the possibly nonrandom use of different habitat types?

These kinds of questions can be best approached by using a spatio-temporal graph to model and analyse the essential features of the system. Other interesting examples of spatio-temporal graphs for studying animal movement to be pursued include the analysis of a group of identifiable individual animals, such as a family group or herd, to study the dynamics of their relative positions at a series of time intervals as they move through the landscape.

11.8.3 Other applications

Because of the great flexibility and range of characteristics that can be accommodated within the general structure of spatio-temporal graphs, the number of applications in ecology is too large to describe, or even enumerate, but we will mention several here in order to provide some impression of that variety.

We can return to the example of the incidence of disease in a population of sessile organisms as illustrated in the multipart Figure 11.5, with newly infected individuals recorded in each of four time intervals. The nodes are persistent and fixed, although they change categories from 'uninfected' to 'newly infected', and perhaps then to 'resistant', 'previously infected' or 'diseased and infective', or 'deceased'. Similarly, any edge between nodes can belong to several categories, such as 'potential transmission', 'transmission occurred', and 'transmission no longer possible'. From the same set of nodes, we can create at least three different kinds of spatio-temporal graph.

- (1) The first approach is to create a simple planar subgraph, using any topological algorithm (MST, LDNT, etc.), for each subset of nodes that are newly infected in the same time interval. The first graph to consider is then just the union of these subgraphs (Figure 11.24a). When taken together as a single graph, the edges of each subgraph are given the labels of the appropriate time interval. [Aside: one index of spatial complexity would be to count the crossings of unlike edges, and compare with the counts produced by random labellings. A low number of crossings would be consistent with an orderly centrifugal progression of the disease through the population.]
- (2) A second approach is to create a series of graphs depicting the situation as it evolves, with the nodes changing categories from uninfected to infected, and so on. Several different sets of edges can be created, say MST and DT, with those that are likely routes of infection changing categories as time proceeds. Using more than one set of edges for potential transmission routes allows us to compare hypotheses about the spatio-temporal characteristics of transmission. (Not shown.)
- (3) The third approach to consider is a single graph in which each node that is newly infected in time t is connected to the nearest node that was newly infected in time $t - 1$ (or in any previous time interval). This graph is a spanning tree of the ever-infected individuals, and some of the edges have direction, from previous to current infection, thus producing a spatio-temporal digraph. Here the edges may represent likely routes of disease transmission, although the actual pathways of infection may not be known.
- (4) Create a single graph connecting all nodes ever infected (Figure 11.24b).

A more explicit but similar circumstance would be a map of forest trees or tree stumps as nodes with edges being the *Armillaria* rhizomorphs that are the infection mechanisms for the spread of the fungus through the forest stand (Lamour *et al.* 2007). Here also, the edges could be labelled with the year of their establishment, giving a spatio-temporal graph that records the dynamics of the infection process through

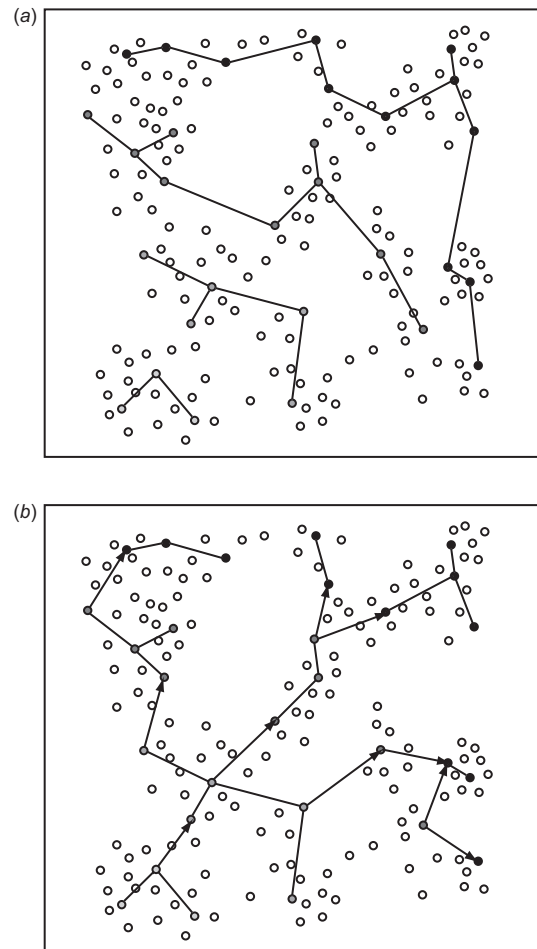


Figure 11.24 (a) A population with four time periods of records of new infections. Subgraphs for each time period. There are no edge crossings in this graph. (b). A population with four time periods of records of new infections. A MST for all nodes ever infected. We can count the number of edges between nodes of different times. Some edges (those ones) have directions. There are no edges between time periods more than one unit apart. See also Figure 11.5.

the years. The difference between the *Armillaria* example and other illustrations of disease spread is that the routes of transmission are both physical and persistent. In other examples, the route that infection travels is transient and leaves no physical evidence. That is especially true when we look at disease in

mobile organisms for which transmission as a result of short-term contact, physical or social, is hard to trace with any accuracy. This is one area of study where graphs that are both spatial and temporal will prove very useful through the comparison of graphs of contact and graphs of transmission, for example, building on the work of Craft *et al.* (2010) on the networks of the territorial populations of Serengeti lions.

Staying with applications related to disease in the ecological setting, one use of spatio-temporal graphs is for the detection and subsequent analysis of spatio-temporal clusters, a procedure similar to the use of scan statistics already described (Sections 6.6 and 11.4 of this book; for background see Knox 1964; Kulldorff *et al.* 1998; Iyengar 2005; Demattei & Cucala 2011). In the standard multivariate context, cluster detection usually proceeds by determining the distances between all pairs of events in k -dimensional space and then creating clusters of the closest groups of events based on any one of several rules (Legendre & Legendre 1998). For spatio-temporal data, the difficulty is that time and space are not commensurable in most circumstances, so that there is no easy comparison of proximity in time and proximity in space when it comes to creating clusters. Demattei & Cucala (2011) suggested a formula for creating a combined space-time distance by rescaling the temporal axis in proportion to the ratio of the observation area and length of the observation interval. They provided evaluations of this technique using simulations and an example of application to earthquake data. Applications of this approach to producing useful spatio-temporal graphs based on clustering algorithms are yet to be explored. Tang *et al.* (2009) suggested a temporal distance metric for social network analysis that is based on the speed with which information flows from node to node in the network.

Ecological data are patchy, or even cyclic, in time just as they tend to be patchy in space, and proximity in such temporal structures may be most relevant if related explicitly to those cycles. Spatial and temporal clusters of phenological events such as flower opening or spore release, may be obvious when proximity in time is evaluated not only by simple Euclidean distance, but also by similarity of time of day, stage of

lunar cycle, or number of days since last frost (or snow-melt or spring equinox). These considerations clearly add complications to the difficulties of commensurability of time and space, but they are important to the ecological phenomena that are the subjects of our analysis.

11.8.3.1 Graph evolution

In applications of graph theory, the development of any graph, spatial or not, through stages, perhaps indicated by the addition of nodes or edges, is referred to as the graph's *evolution* (Gross & Blasius 2008), which may be somewhat confusing in an ecological context. As any ecological community develops, the aspatial graph, with nodes for its species and edges for the interactions between species within it, evolves. As a new species becomes established, it adds a node to the graph and usually one or more edges, representing its direct interactions, but other edges may be added or lost as the less direct effects of the new species affect relationships previously established. These aspatial community graphs may also evolve through time without the loss or gain of species, either in response to environmental changes, or as the community structure changes due to the interactions developing within it.

Turning from aspatial to spatial graphs, and reflecting on the discussions in the last chapter, the spatio-temporal relationships among communities that are re-established from 'zero' each year or at other intervals would provide many insights on spatial aspects of community structure and assembly, as the spatial graphs of immigration and establishment events evolve with community development. One example that would be appropriate for this approach to studying the spatio-temporal patterns of metacommunity structure is the pitcher plant inquiline community (see Kneitel & Miller 2003; Harvey & Miller 1996; among many), as would those of some temporary ponds (Wilbur 1997) or the annual algal communities of intertidal zones where ice-scour cleans the surface every winter (Scrosati & Heaven 2007). In fact, spatio-temporal graphs seem to be an ideal method for the investigation of metacommunities and how spatial

structure contributes to their dynamics (see Mouquet & Loreau 2002; Leibold 2009).

In recent work, Chan *et al.* (2008) described an approach to determining the amounts of correlation in spatio-temporal change in different regions of graphs as they evolve. The approach applies to graphs with persistent nodes and intermittent edges between them, as might occur in studies of network management, webpage linkages, or bioinformatic analysis of large sets of genomic sequences. The analysis is based on comparing waveforms in the temporal dimension that summarize the presence or absence of any edge (1 or 0) rather like a square wavelet. The shapes of the waveforms are compared and pairs with high correlation indicate regions of the graph with similar dynamics. As an aside, the same approach should work for edges that have a variable flow rate or capacity between 0 and 1, just as it does when the edge has weight 0 or 1 indicating connection or not.

The evolution of a structural spatial graph has, of course, many more applications in ecology than the development of interactions within a community or changes in a metacommunity. For example, a 'natural' application of this approach is in the study of the structure of foraging trails created by ants. For example, Kost *et al.* (2005) provided some details on the foraging trails of leaf cutter ants (*Atta* sp.) in the New World (Brazil, French Guiana, and Panama); these ant trail systems evolve through time both in the short term, with ephemeral foraging trails being established and abandoned as branches of longer-lasting trunk trails, but with the characteristics of the systems changing between young colonies and mature colonies, the latter having larger and more complex trail systems that exploit a broader range of leaf sources. The authors used a combination of fractal analysis, circular and conventional statistics in their analysis, but graph theoretic measures such as graph diameter and branch order (among many) would prove useful for evaluating the dynamics of these trail networks which are *rooted trees* (the ant nest is the root from which the sub-trees emerge) in a graph theoretical classification.

In a somewhat more abstract application, we can consider the changes through time of the functional

relationships in an ecological system, whatever they are, and the extent to which these relationships can continue in the presence of stress or disturbance, and their ability to recover after the adverse conditions are removed. The continuity and resilience often observed in an ecological system can be thought of as the system's 'memory' (Chapter 2, Section 2.1): the extent to which past behaviour of the system affects current and future processes, thus persisting through a number of realizations or observations of the processes. Very simply, 'ecological memory' is the set of processes and the degree to which an ecological system is affected by its past (Peterson 2002). Ecological memory also derived from the interaction of plants and propagules, through biological legacies, such as seeds and rhizomes, remnant trees and vegetation patches in a disturbed landscape; mobile links such as seed dispersers; and more distant support areas which serve as refuges for the mobile links and from which they disperse can provide inputs to the disturbed area (see Garcia Molinas & Donohue 2011). Ecological memory also results from the network of species in a system, their interactions with each other and with the environment, and the set of structures that allow recovery after disturbance, thus being directly related to ecological resilience, which is the capacity of the system to survive disturbance (Bengtsson *et al.* 2003).

A landscape is a spatially explicit system, and like any other system, it can exhibit landscape memory or landscape legacy, which is the degree to which current processes are influenced by the history of past processes in the same landscape (Hendry & McGlade 1995; Peterson 2002; James *et al.* 2007). In order to evaluate this landscape memory effectively, the analysis needs to be spatially explicit, as well as temporally referenced, and that requires the use of spatio-temporal graphs. Earlier approaches treated landscape patches at different stages of development as elements in a first-order Markov chain, with transitions between stages in that chain occurring with certain probabilities. In this scheme all patches follow the same sequence of stages, although possibly at very different rates, determined by the probabilities of transitions. Real dynamics are often less predictable, and patch

transitions between stages are likely to depend on past conditions beyond the stage immediately preceding (longer memory or history) and on the states of spatial neighbours, either contiguous patches or at a greater distance (the spatial context). First-order Markov models, however, do not include longer-term patch history, neighbour effects or non-stationarity (whether temporal or spatial), whereas in real ecological systems, patch state transitions probably depend on multi-step history and on spatial context, perhaps beyond the immediate first-order neighbours (Mladenoff & Baker 1999). Spatio-temporal graphs can include these various factors in the development of this concept and its analysis. We can consider a hierarchy of effects that begins with the first-order Markov model of a patch at time t that depends only on its own state at time $t - 1$, then a spatio-temporal graph that includes a one-unit temporal memory of the patch itself and the contemporaneous states of first-order spatial neighbours, and lastly a more complicated version that uses two steps of time for the patch itself and the recent history as well as current state of its spatial neighbours.

11.8.3.2 Phenology

As a final example of the use of spatio-temporal graphs for ecological analysis, we turn to the rich and fascinating topic of spatial variation in phenological sequences. The general topic is, of course, of great interest in the current context of the effects of global climate change on natural processes and seasonal events of importance in conservation, but it also presents interesting challenges to the techniques of analysis for complex phenomena. The data consist of a number of recognizable stages or events (the nodes) that can be located accurately in time, and for individual plants (or perhaps for clones or populations) located also in space. The events may or may not have a prescribed order; for example, depending on the study organism, a second leaf emergence always occurs after a first leaf emergence, but the first flower anthesis may occur either before or after the emergence of the seventh leaf. In that case, each sequence of observations has a time associated with each event,

and a place, because of the location of the individual (s) observed. The timings of events, as well as their order treated as a simple ranking, can be analysed in conjunction with the locations, as the structure of a spatio-temporal graph.

In Section 3.4.1 of this book, we described the concept of a graph-of-graphs, in which the nodes of a higher level graph structure are themselves graphs of a lower level, with nodes and edges of their own. In the case of spatial variation of phenological sequences, the higher level graph could be a spatial graph in which the nodes are regions or sites and the edges indicate proximity or climatic similarity, each region having a spatio-temporal graph depicting the phenological sequences of individuals or populations together with their spatial relationships. The higher level graph could also be temporal, as well as spatial, in depicting year-to-year changes in seasonal phenology in different regions. Many combinations of space and time can be accommodated in the complexities of a graph-of-graphs structure. The application of spatio-temporal graphs lends itself well to an extended study of graphs-of-graphs, with different combinations of space or time at the two levels of graphs, but we will not pursue that topic in any detail here. We can leave that for another opportunity.

11.8.4 Final comment on spatio-temporal graphs

The potential applications of spatio-temporal graphs to summarize both spatial and temporal relationships and interactions among ecological units from individuals to communities seem limitless. There is such a wealth of spatio-temporal data now available, although much of it is in spatial lattice form (see examples in Cressie & Wikle 2011), which may not initially seem to produce 'interesting' spatio-temporal graphs, but the choices of node-pairs for edge construction will be a key issue for understanding the rich information in these data structures. Spatio-temporal graphs, as we describe them here, can help determine the weights in an adjacency or connectivity matrix in space and time that is the basis for calculating the spatio-temporal statistics described in Cressie & Wikle (2011). These

graphs have many applications in a broad range of areas. In fact, the possibilities are almost overwhelming, and careful thought will be required so as not to get lost in the technical details to ensure that the focus is on answering questions of ecological importance. Of particular importance is the reminder that configuring the ecological data as a spatio-temporal graph is only the first step. Many methods of spatio-temporal analysis implicitly use these graphs as the skeleton for their analytical procedure; for example the space-time *K*-function (Diggle *et al.* 1995) is essentially based on a threshold-based nearest-neighbour graph with space distance threshold, s , and time distance threshold, t . The real power of the graph theory approach lies in using the rich and flexible techniques of graph theoretical analysis, such as path lengths, clustering, modularity, connectivity, or node centrality; the works already cited (Tang *et al.* 2009; Del Mondo *et al.* 2010; and so on) provide some ideas as to how to proceed, but there is much more work to be done. As with our comments on the importance of spatial graphs to future approaches to purely spatial analysis in ecological studies, the development of applications of spatio-temporal graphs to ecological studies that include time series, seasonality, succession, movement, or system dynamics will be critical contributions to ecological analysis.

11.9 Concluding remarks

The area of spatio-temporal analysis, and the phenomena with which it deals, is definitely one of the most fascinating and rapidly developing in ecology today. It is critical to a mature ecological understanding, not just in the sense of spatial pattern and temporal process, but also in the sense of spatial dynamics and temporal pattern. For example, the work on spatial synchrony and asynchrony described in Section 11.7.3 is providing important insights into the basic drivers of population dynamics and community interactions. The analysis of animal movement through its

habitat, and how that is related to the habitat structure is an area of active research where we expect to see rapid developments. The analysis of polygon change is also one that deserves further work and effort.

11.9.1 Recommendations

Spatio-temporal analysis is a field which can handle, or perhaps requires, rich data sets; many of the techniques described here will be most rewarding with detailed spatial information and many times of observations.

In conducting statistical tests for this sort of study, we need to be constantly aware of spatial and temporal autocorrelation, their effects and the processes that give rise to them. In many instances, we also need to be concerned about the possible pitfalls of pseudo-replication. A constant theme throughout this book is the lack of independence between observations. While that lack of independence causes problems for statistical tests through autocorrelation, it is also the property that makes prediction possible, and prediction (of many kinds, including interpolation as well as extrapolation) is crucial to the scientific value of ecology. Predictions, in general, are more powerful and potentially more useful when they are quantitative, and so our advice is to have lots of high-quality numerical data. We have discussed the problems and interesting qualities of spatial and temporal autocorrelation, and it is an important consideration in analysing spatio-temporal data of all kinds. We do not agree with the concept of somehow thinning out data in order to get 'independent' observations because it is wasteful and it may not work, as discussed in Chapter 8. The concept of time-to-independence or distance-to-independence is mistaken. We need to learn to take advantage of that lack of independence in the data and use it for our own purposes. Therefore, it is much better to use all the information available and to evaluate the characteristics of autocorrelation in the data to be used in later analysis.