12

Closing comments and future directions

Introduction: myths, misunderstandings and challenges

It has become common for ecological studies to recognise the importance of the spatial aspects of ecological systems (e.g. Wagner & Fortin 2005; Schroeder & Seppelt 2006a; Hanski 2009; Leibold 2009; Fortin *et al.* 2012a; Gelfand 2012), and to include them in the design and analysis of these studies at least conceptually (Legendre *et al.* 2002, 2004). Spatial effects are various and many may act simultaneously or interdependently, including factors such as spatial scale, autocorrelation, locational and neighbour effects, functional connections, and the ubiquitous mutual interaction between spatial pattern and temporal processes. The set of topics in this constellation of ideas is complex and potentially confusing, especially when time is included as well as space. While the increasing awareness of the importance of spatial effects is helpful, the complexity has given rise to some misunderstandings and to a number of 'myths' about the issues and possible solutions. Although careful instruction and clear advice are available in the literature and in a number of textbooks (Cressie 1993; Legendre & Legendre 1998; Haining 2003; Fortin & Dale 2005; Schabenberger & Gotway 2005; Cressie & Wikle 2011), some challenges remain for researchers who face particular kinds of data and particular details of analysis. In this section we will make explicit a few of the prevalent myths about spatial effects, with an attempt to correct the related misunderstandings. Also we will acknowledge the real (not mythical at all!) challenges that including the spatial context presents for ecological studies.

When characterizing spatial pattern, several statistical and ecological concepts come into play and potential mismatches among them can occur in the appropriateness of their usage. Furthermore, a number of other considerations need to be acknowledged.

- Without spatial pattern there is no scale in the system. A totally homogeneous or totally random system provides no landmarks that determine spatial scale or from which scale can be detected; it really has no scale to be detected.
- The observed pattern may result from the action of several different processes (Figure 1.2 of this book). A central question in ecology is that of the relationships between observed pattern (here in the form of spatial structure) and the processes that generate it, or that arise from it. What is generally agreed, however, is that we cannot safely deduce process from pattern, in part because different processes can give rise to similar spatial signatures (as shown Figure 1.2). For example, there is ongoing discussion about the several possible mechanisms responsible for the fascinating phenomenon of the vegetation stripes observed in arid regions, with no consensus (cf. Barbier *et al.* 2006; Deblauwe *et al.* 2011; and many others).
- The processes causing pattern may intermingle as they act together, possibly with interactions reinforcing or interfering (and possibly with spatial or temporal lags in their action), or they may be fully confounded, so that they are impossible to disentangle.
- Most processes involve a variable's change in state and so we need to quantify those variables spatially,

in order to detect spatial dependence or autocorrelation (see Chapter 8).

- Finally, even ecological questions that do not have an explicit spatial component should account for spatial context if only to rule out possible spatial effects in the observed data pattern. For example, in studying the dynamics of predator and prey populations, if both populations are territorial, any potential for spatial effects may need to be removed before non-spatial questions about the dynamics can be answered.

It may be difficult to distinguish between inherent spatial autocorrelation and other forms of spatial dependence (see Chapter 8), and so an understanding of the ecological properties of the system (Chapter 2) may be needed to determine the cause(s) of any dependence detected by spatial statistics. While it may be difficult to determine the cause, there is a wide choice of methods available to detect and to evaluate spatial dependence as presented throughout this book (Chapters 7 and 8). The irony is that the lack of independence is what makes prediction and hypothesis generation possible, but its very existence causes problems for the testing of hypotheses, producing both positive and negative aspects while analysing ecological data.

- The positive aspect is that spatial dependence makes it possible to use spatial patterns as the basis to generate hypotheses about processes and about feedback effects between those processes and patterns.
- The negative aspect is that spatial dependence can affect statistical tests through biased parameter estimations and its interference with the effective (versus nominal) number of degrees of freedom.

The overall result is some interesting and possibly inspiring tension, expressed as a question: "trouble or new paradigm?" (Legendre 1993).

Myths and misunderstandings

Given the above issues, it is not surprising that several myths and misunderstandings have emerged about what is important and what is not important in analysing spatially explicit data. These myths are not all independent of each other but run together in some cases, and we have tried to make some of them explicit.

1 Distance is the key factor

Myth: Near things are more closely related than far things and therefore nearby locations are more similar and more important to each other than distant ones (Tobler's Law; see Chapter 1).

Clarification: Sometimes this is true, but, as we have described elsewhere, ecological data are often patchy which can produce cycles in similarity as a function of distance. Think of an archipelago: the similar vegetation of the islands has marine communities interspersed between them. Second, connections can trump mere distance; flow of rivers, marine currents and prevailing winds, and the movements of animals, cargo containers and aircraft can enhance the similarity of farther sites, just as deep channels or high mountains can create isolation between sites that are otherwise apparently close.

A related misunderstanding is the optimistic view that we can rely on simple models like the 'textbook variograms' for our studies: a misunderstanding that the asymptotic form of the variogram is a good default for ecological data and so corrections can be based on variants of that form in which the corresponding correlogram quickly approaches zero. Because ecological data can be patchy, autocorrelation cycles between positive and negative as a function of lag distance, producing a variogram also with cyclic behaviour (Figure 12.1*a*, *b*) it is inadvisable to assume that the 'textbook' model is true. Corrections based on a model that is not true can produce results that are even farther from correct than the 'uncorrected' original! Mizon (1995) strongly advised against attempting to correct for autocorrelation because it could make the situation worse.

2 Distance to independence

Myth version A: If your samples are far enough apart, they are as good as independent.

Response to version A: This myth is based on the assumption that autocorrelation declines with distance to effectively zero, so that distant samples are effectively independent. This myth depends on the structure of spatial dependence (rapid decline to stable values near zero) and its stationarity throughout the study, but this

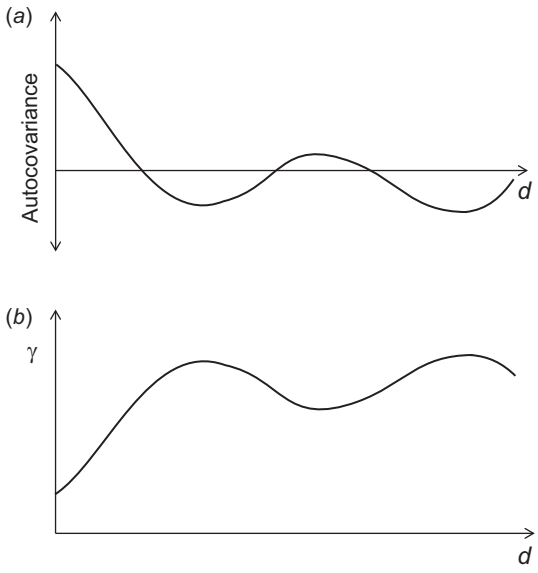

Figure 12.1 (*a*) Autocorrelation as a function of distance showing cycles in positive and negative autocorrelation resulting from patchiness (e.g. alternation of patches of high density and gaps of low density) that is typical of many kinds of ecological data. (*b*) A variogram showing the same phenomenon, as evaluated by a different statistic.

may be quite wrong for most ecological data, as just explained above in 'Distance is the key factor'. Cyclic behaviour of variables can make more distant sites more similar than close ones; and functional connectivity plays a part. In some cases, authors may be tempted to throw out data to increase the distance between samples used for analysis, but that is very wasteful of information, and not recommended (Legendre & Legendre 1998).

Myth version B: If autocorrelation at a given distance is nonsignificant, it can be ignored.

Myth version C: If the data are subsampled to remove locations that are too close to one another, the spatial autocorrelation will not be detected.

Response to versions B and C: Not so. This is another case of the asymmetry of significance. When we cannot show that our observations are sufficient grounds to reject the null hypothesis of randomness, we cannot necessarily conclude that the process is indeed truly random. First, in most circumstances, the total autocorrelation over all lag distances may be the important quantification of its overall strength, not the value at any one distance (see Cressie 1993; Chapter 1; Chapter 8 of this book). Second, the detection of 'significance' depends on the numbers of observations at any lag distance; larger distances tend to have fewer observations and therefore are less likely to show significance (Fortin 1999a; Chapter 1). Therefore, there may be bias against detecting significant autocorrelation at larger scales, depending on the characteristics of the sampling design and the study area, but that is not sufficient justification to ignore it.

3 The zone of ecological influence is the 'range' in spatial statistics

Myth: The important ecological zone of influence of spatial autocorrelation in a biological system can be determined from the observed spatial 'range', the distance at which observed spatial dependence reaches zero in the correlogram or the sill in the variogram calculated from the data.

Clarification: The spatial correlogram or variogram characteristics, such as range, are averages over locations and over directions, and so their values do NOT match the characteristics at actual geographical locations. Many of these spatial techniques provide averages of a particular index over all spatial locations where non-stationarity or simple variability may create differences (Figure 12.2), and over all directions, when simple variability or anisotropy may cause directional differences (Figure 12.3). Inherent variability is always a factor, and both non-stationarity and anisotropy may play roles in preventing easy inference from statistical summaries to individual locations.

4 Biological diversity is a variable just like any other

Myth: This myth is a bit more difficult to explain, but it suggests that biological diversity is a real characteristic of a site, like temperature, perhaps, that can be measured from location to location. The statistics of diversity measures may be complicated and confusing, but what these measures attempt to capture is a genuine

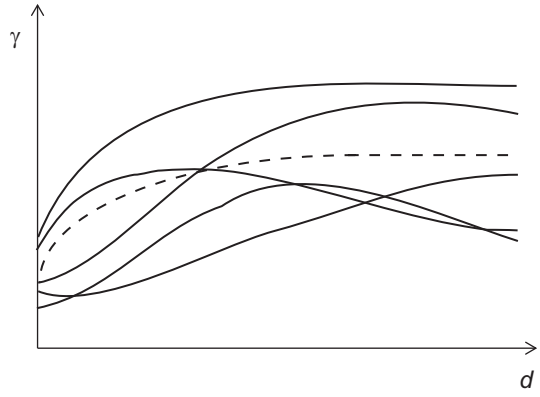

Figure 12.2 The observed variogram (dashed line) is an average over locations at which the spatial structure may be different because of inherent variability or because of non-stationarity.

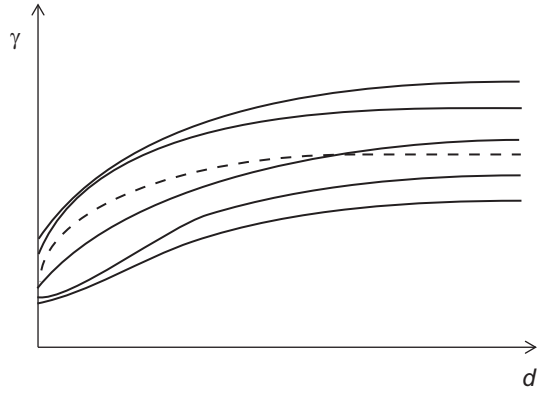

Figure 12.3 The observed variogram (dashed line) is an average over directions for which the spatial structure may be different because of inherent variability or because of anisotropy.

characteristic of the assemblages of organisms being studied. All that is required is to get the correct statistical formulation.

Response: The fundamental difficulty is that diversity is not a variable like temperature or elevation; it is a synthetic or 'constructed' variable, and one that is highly scale-dependent both spatially and temporally. It is also non-stationary both in time and in space, both of which need to be considered (White *et al.* 2011). The misunderstanding is that of reification: if we have created measures of diversity, there must be a real thing, diversity, for the measures to measure! There isn't really, is there?

5 Landscape metrics are immune to statistical problems

Misunderstanding: Landscape metrics, such as area to perimeter ratio, number of patches, and so on, can be used to compare the difference between regions because they are just measures of patch characteristics, not statistics. They provide a set of reliable techniques!

Clarification: The comparison of landscape metrics works if and only if the two regions being compared have the same categories in the same proportions. In addition, spatial autocorrelation will also make reliable comparisons difficult (see Fortin *et al.* 2003; Remmel & Csillag 2003; Remmel & Fortin 2013). If they are just indices, how can we determine whether any differences observed between two landscapes are significant, either statistically or ecologically? We generally use the former as a guide to the latter, and these require some assumptions or knowledge of their distributions under certain circumstances. Therefore, as the basis for reliable comparisons, these metrics are not a simple solution that avoid the problems of more familiar statistics in a spatial context.

6 Spatial statistics provide the solution (We're saved!)

Myth: Spatial statistics do not have assumptions such as those of the familiar inferential statistics, such as independence and asymptotic normality, and so there is none to be violated.

Response: Beware! The assumptions may not be as obvious or as restrictive, but assumptions there are, implicitly or explicitly: most importantly that n samples or n observations reliably give a true or effective sample size of n, but it is well known that, because of autocorrelation, n samples can give an 'effective sample size' quite different from the nominal value n (see Chapter 8).

7 The 'shotgun' works!

Myth: The more methods you use in analysis, the more you learn about the data.

Response: There are close relationships, both conceptual and mathematical, among many of methods for spatial analysis, which are described and discussed throughout this book (see summary at the end of Chapter 5, for example; also Dale *et al.* 2002). As one example, the variogram and Moran's I are not complementary methods, but very similar, giving results that are almost completely redundant. Many other pairs or sets of methods are only somewhat complementary, and so, while using only one method will not give all the answers that the data can provide, some caution and judgment is necessary in choosing a reasonable set of methods to provide a range of insights.

Challenges

These are not, of course, all the challenges to be faced in the spatial analysis of ecological data, but the following list provides some general categories of the kinds of difficulties encountered both frequently and in a broad range of study systems.

1 Lack of independence

We have spent much effort on explaining spatial autocorrelation throughout this book, but autocorrelation is not the only source of lack of independence encountered in spatial studies. Dependence also arises when spatial coefficients are computed at more than one distance lag. This lack of independence arises because the same data are used over and over in calculations; for example, the same data that are used to calculate any statistic at lag 2 are used again in different combinations to calculate the statistic at lag 6. More importantly, local statistics that are calculated at location (x, y) will not be independent of those calculated at nearby location $(x + d, y + d)$ for lags of distance r, where $d<r$. These two are similar sources of lack of independence in ecological studies, related to the 'independence of observations' category described by Legendre & Legendre (1998, 2012; Chapter 1). Sources of dependence in our ecological data include spatial dependence, spatial autocorrelation, temporal autocorrelation, phylogeny and familial relatedness, lack of independence among descriptors (for example, canopy height and total biomass in a study of forest spatial structure), and lack of linear independence among variables, producing non-zero covariance. Physical and other functional connectedness can also produce a lack of independence in a variety of different ways and with a variety of manifestations.

2 Distinguishing the key functional connections.

The previous entry leads us to this important challenge. In order to include spatial structure in the most useful and appropriate manner, ecologists need to be able to discern the most critical among the various forms of functional connectivity that are found in the systems being studied. Think of the many choices of what links to include in a spatial graph (Chapter 3). The correct choice will not always be obvious but will greatly affect the outcome of the spatial analysis and interpretation or of subsequent spatial modelling and parameter fitting.

3 Distinguishing the key spatial scales

Ecologists now address questions dealing with larger areas than were typical in the past, and we need to understand the array of spatial statistics that are available and how to disentangle the key scales at which processes occur (Fortin *et al.* 2012a). Different processes, and even different kinds of processes (e.g. biotic interactions vs abiotic conditions), may tend to act at different scales, spatial scales for some and temporal scales as well. It is important to be able to make these distinctions, but that is not an easy assignment. Confounding and technical trade-offs both present challenges for disentangling scales to determine which are the most important, and which factors contribute to them. Two families of methods have been proposed to determine which scales are the most important in a system: one for sampled locations, based on eigenvalues (such as MEM; Dray 2011; Dray

et al. 2012; Chapter 7), and one for fully censused grid data, based on wavelets (Keitt & Urban 2005; James & Fortin 2012; Chapter 7). Further work on this important area is probably required to clarify the best practices for application in ecological studies.

4 Local and global evaluations

We have already commented that most statistics for evaluating spatial structure were originally formulated in a 'global' form, to summarize a particular characteristic over the entire area being studied. Almost all can be formulated also in 'local' versions (Anselin 1995; Getis & Ord 1996; Boots 2002; Jetz *et al.* 2005) quantifying the characteristic of interest as measured for subsections of the study area. This approach provides flexibility and sophistication for spatial analysis and, in particular, the ability to evaluate non-stationarity: how the characteristics of the spatial structure vary across the study area. Non-stationarity can then be evaluated by comparing local statistics with the global analysis (cf. Kabos & Csillag 2002) and by comparing the local statistics at different locations in the study area. This approach is obviously related to the concept of scan statistics as described in Chapter 6, and may have particular relevance to the study of sensitive issues such as biological diversity, which may vary greatly with scale and location (Chapter 10).

There are both obvious and subtle advantages to using local spatial statistics for the analysis of ecological data, even when the questions of interest are actually global in scope. In particular, at least some of the myths and problems associated with spatial autocorrelation in ecological studies can be addressed by looking at a localized evaluation of spatial autocorrelation. The trade-off is that, the more localized the analysis becomes, the less precise any evaluation can be, because of the unavoidable reduction of sample size associated with the analysis. This question of precision becomes acute when we are trying to distinguish between local variability in the observed spatial structure and true non-stationarity in which the underlying characteristics of the structure actually change from place to place. At the same time, because multiple tests would require Bonferroni-type procedures for adjusting significance levels, which could become extreme for large numbers of local statistics, we have to recommend that they be used only for exploratory analysis. We have focused in this commentary on spatial variability, but it applies equally well to temporal variability and to spatio-temporal analysis as outlined in Chapter 11.

Comment on myths, misunderstandings and challenges

The concepts described in these myths, misunderstandings and challenges are not easy ones, but they are important for ecological researchers to understand. The subjects can be difficult: conceptually, technically and ecologically. Without seeming too circular, we suggest that spatial autocorrelation is itself a pattern for which it may be impossible to distinguish process. Spatial dependence in variable y looks more-or-less the same whether it is due to spatial autocorrelation in y itself ('inherent dependence'), the result of responding to some other underlying variable x, itself autocorrelated ('induced dependence'), or a combination of the two acting at the same time ('double' spatial dependence; Chapter 8). A good understanding of this phenomenon and a grasp of solutions to the problems it presents to researchers is required; the problem will not go away and cannot be ignored. No single solution solves all the problems because the solutions are context-, scale- and system-dependent. As with the choice of spatial methods, some care and judgment will be needed in finding a set of solutions to a multifaceted or complex problem with several possible confounded effects. In spite of the difficulties of this topic, and despite the myths and misunderstandings, this is an area in which real progress has been made in the past decade, both in a greater general awareness and understanding on the part of ecologists and scientists in related disciplines, and in the technical sophistication of finding solutions.

All that being said, our next comment is a reminder that spatial autocorrelation is only one aspect of spatial structure, and only one form of lack of independence. Spatial dependence can be found not only in an apparently homogeneous plane or volume, but it also occurs on many kinds of networks and graph-theoretical structures. Think of all the structural complexities that

can be efficiently encoded in the sparse form of an aspatial graph. Connections, propinquity, flow and multiple interactions are all part of the picture, and when we can include such complexities in our conceptual development of research, the next challenge is to include them in the research itself, from the conceptual stage to design, and from observation and experimentation through to analysis and interpretation.

12.1 Back to basics

Both authors of this book favour a visual approach to problems, as is evident in the number of figures we have used (and this is us being restrained!). It is not surprising, therefore, that we advocate visual evaluation for every step of the analysis process. Draw a map of the locations and values; draw a spatial graph; draw the time series when you can. Plot the data, plot the results of analysis and, when fitting a model, plot the residuals. There are insights to be gained by this throughout the process.

The first step in any analysis is to plot the data. Many problems can be avoided by this simple step, combined with an awareness of potential problems and some thought. A common mistake in dealing with spatial data is that the x- and the y-coordinate axes (the columns or the rows in a data file) are not used in the right orientation and order during the analysis. The required input format can vary with the statistical software package; for example, the location of the origin (0, 0) can be either the upper left corner (as in a matrix array) or the lower left corner (as conventional in most field sampling). This difference in the location of the origin can result in analysing the mirror image of the data, which is not always a problem but in some applications may be critical. Plotting the data can provide obvious and useful information to guide the choice of which spatial statistics to use and which methods will detect the spatial structure of the data. This early informal evaluation of the spatial behaviour of the data is also useful later, in interpreting the results of analysis and in understanding the patterns and the processes that generated them.

Having analysed the data, plot the results. It is quite possible for a global analysis to detect very little

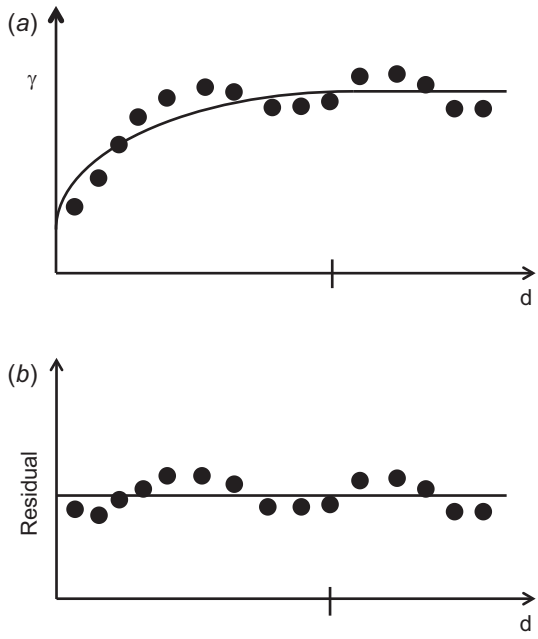

Figure 12.4 The importance of plotting residuals. (*a*) The asymptotic variogram (the line) fitted to the data (the dots) explains much of the observed variability, but it is a poor description in some ways, because it does not detect the cyclic behaviour in the data. (*b*) That behaviour becomes obvious when the residuals are plotted.

pattern in non-stationary data, but the non-stationarity could be revealed by spatially explicit local analysis. For example, in point pattern analysis, Ripley's K-function may show little consistent departure from CSR, where Getis' method of plotting the scores may reveal spatial trends (see Chapter 4). Plotting the results and thinking about what they show may persuade us to change the approach we use.

When fitting models such as spatial regressions, plot the residuals or map them where appropriate. The example in Figure 12.4 also illustrates the importance of plotting the residuals of a fitted model. The assumption is often that the residuals have a normal distribution with a constant variance and are independent of each other. The residuals plotted in Figure 12.4 clearly do not meet the condition of independence, here because of spatial autocorrelation that has not been removed by taking out the geographic trends.

In the same spirit of understanding how the data are analysed, some detailed knowledge of data storage and the analysis algorithm is always beneficial. Some statistical and GIS software packages are too user-friendly and the user is kept unaware of the details of analyses actually performed. Understanding the programs will make you more alert to potential miscoding that could have taken place in some, especially when using shared code from website sources. The spatial analyses for this book were performed using both software packages (PASSaGE, BoundarySeer, ClusterSeer, Splus+Spatial, GS+, R, SAM, IDRISI and ArcGIS), and programs specifically created in programming languages (VisualBasic, QuickBasic, Fortran, C++, R, Python, etc.).

The selection of the appropriate spatial extent and grain at which to study the process of interest can be tricky if no prior knowledge is available. To perform a meaningful study, information about the spatial and temporal domains of the process, as well as spatial and temporal response scales of the patterns, is needed. Such information can be obtained by carrying out a pilot study (see Legendre *et al.* 2002; among others). To facilitate the design of sampling when there is no time or resources for a pilot study, a variety of other sources may be useful. These include aerial photographs, remote sensing images, vegetation maps, digital elevation or hydrologic maps and bathymetric charts, as well as knowledge from previous studies on the variables, the species and the systems of interest from reports, papers and colleagues' expertise.

Another simple and obvious suggestion is that larger sample sizes often offer more options for analysis, as well as more power. This may be particularly important when the particularities of the data do not permit parametric analyses, and randomization methods seem to be the best approach to analysis. We know that in many areas of ecology it is rare to have sample sizes of 30 or more. From examining the problems of fitting models to data (of known structure) described in Chapter 8, it is clear that in some circumstances, a sample size of even 100 is really too small. To model the spatial structure of the data may require much larger samples than we are accustomed to having available. Similarly, for the detection of spatial pattern, a transect of 40 contiguous quadrats is definitely too short, because of the very limited number of lags or block sizes that can be examined and the very small sample sizes that contribute to the calculations at each lag. Under those circumstances, smaller and more numerous sample units will be more effective. For studies in two dimensions, the same considerations apply, but there is the important issue of the trade-off between extent and grain of sampling when resources are finite. As for many of these issues, a balance of considerations and limitations is required.

More does not always mean better when it comes to data. Indeed, more and more ecological studies are carried out at the landscape level at which novel and challenging questions can be investigated. Usually, such studies use either aerial photographs or remote sensing images, which can provide very large data sets (e.g. tens of thousands of pixels). Besides the obvious problems with remotely sensed data, such as spatial accuracy, image distortion, and misclassification (see Burrough & McDonnell 1998), the data usually cover a large area where several environmental factors and ecological processes can occur and therefore non-stationarity, rather than stationarity, should be assumed (see discussion in Section 1.6). In such a context, global spatial analyses (Chapters 4 to 6) of the data for the entire area should not be performed unless the data are first spatially stratified and partitioned into spatially homogeneous subareas (see Chapter 9). We recommend using local spatial statistics (Chapter 6) under these circumstances. The practical issues of dealing with very large data sets is not unique to ecological research, but is now very common in many areas of science including molecular biology, genetics and phylogeny. Those challenges of 'too much data' may be dealt with at a more general level than we can provide advice on here.

12.2 Numerical solutions: software programs and programming

Of all the advice we might give to graduate students in almost any branch of ecology, one of the best recommendations is to suggest acquiring at least some

programming skills. This may sound like an old-fashioned suggestion, given the wealth of software packages available, but that wealth of software is itself part of the reason for the advice. It is very dangerous and potentially very misleading to use software programs when the details of their calculations are not made explicit or cannot be understood without some knowledge of programming. For example, one popular analysis package calculates the variogram for a range of spatial displacements, h, but uses the total sample size n as the divisor for all spatial intervals, rather than n_h, the number of pairs at the spatial displacement h. The user needs to know of this change of divisor in order to interpret the results correctly.

The second reason for the advice 'learn to program' is that the skill opens the door for the researcher to explore methods or variants of methods of their own devising without relying on others for help, providing greater flexibility. In addition, there is usually a time lag between the creation and publication of a new method of analysis and its general availability in software packages. The ability to write or to modify analysis programs will allow the researcher to implement the most up-to-date methods. The popularity and widespread acceptance of the software environment R (http://www.r-project.org/), which offers both readily available and easy-to-use prepared programs and the flexibility of user modification, leads to the recommendation that R would currently be a good place for any student to start.

Another theme for this book, which may have been less obvious than some others, is the usefulness of numerical, as opposed to analytical, solutions to methodological problems. As an example, consider the details of edge correction for a technique such as Ripley's K-function analysis. For a simple square or rectangular plot, edge corrections are available, based on the position of the index point and the radius of the circle being used, t. If, however, the study region is not a simple rectangle, or has irregular or curving boundaries, a derived formula for an edge correction would be very complicated. A numerical solution such as described in Chapter 4 (Section 4.1.3) seems like a sensible alternative. Increasing computing power makes possible a range of computer-based solutions to problems, such as the 'model and Monte Carlo' approach to dealing with spatial autocorrelation that affects statistical tests (Chapter 8). It also changes the way in which models can be used by ecologists. We can use them to explore the effects of particular structures on our understanding of what is occurring (cf. Legendre *et al.* 2002), rather than trusting models as reasonable representations of the data themselves, and then basing the analysis on that trust. (Watch the assumptions!)

We have not spent much of this book discussing 'classical' experimental design and the analysis of variance (ANOVA) or similar techniques, except to provide some thoughts on the relationship of design and analysis to the spatial structure (mainly spatial autocorrelation) of the environment in which the experiment is carried out (Chapter 8). Without citing specific examples, the reader will probably not need much convincing that this can be a rather confusing area, particularly with complex designs, even before spatial considerations are included. Statistical textbooks may sometimes seem to disagree, or are not always clearly in agreement, which makes life difficult for ecologists, who are trying to analyse and interpret their data correctly. (One graduate student, after a seminar on some tricky statistical concepts, asked, 'Does this mean we have to be statisticians as well as ecologists?' The answer was, 'No,... but you do need to know where to get reliable advice'.) Faced with several possible alternatives for ANOVA (often all apparently equally justifiable), one approach is to use the analysis of artificial data to provide guidance. Generate several sets of 1000 (say) iterations of data equivalent to those you wish to analyse with small to large treatment effects (3, 5, 8, 12, 20, 30%, ...) and examine the behaviour of Type II error. That should provide a realistic guideline for the choice of analysis. As a more general suggestion, it may be useful in many circumstances, particularly when beginning a new type of project, to create and analyse artificial data of the form expected to be found in the study, to anticipate and then to solve analytical problems in advance.

In the same way that the researchers have to be precise about the null hypothesis to be tested, they

need to be certain that the spatial statistic they use is indeed doing what they want. Our commentary, above, on 'myths' explains some of the misunderstandings that have arisen and are fairly common sources of error for spatial analysis in ecological studies. Our hope is that the commentary provides some understanding that is helpful in avoiding such errors.

12.3 Statistical and ecological tests

Having just suggested that we do not have to be statisticians to be good ecologists, we will now acknowledge that what is required is a good understanding of the relationship between the results of our data analysis using statistics of various kinds and the subsequent ecological interpretation. The relationship between statistical testing and the ecological meaning of the outcome is not always direct, and it is not easy to determine the most appropriate statistical approach to a particular ecological question. For one thing, statistical significance does not always result from a significant ecological process. For example, there could be a significant difference between the degree of spatial patchiness between two populations due to the spatial structure of the habitats rather than to the species' ability to move in a fragmented landscape. The reverse is also true; a nonsignificant statistical result can still have an important ecological implication. This is especially true for the detection of spatial pattern in the presence of positive spatial autocorrelation: each sampling unit does not contribute a full degree of freedom to the statistical tests (Chapter 8), but we actually want to learn about the pattern itself. In the case of adjacent sampling units in the presence of spatial structure, each sampling unit does not bring a full degree of freedom but the similarity of adjacent samples tells us something else about the size of the spatial structure: that it is larger than the sample units. Consequently, when adjacent units have similar values, this 'redundant' information is informative about the scale of pattern.

Some parametric tests are more sensitive or less robust to the presence of spatial structure in the data than others. For example, the t distribution does not actually change much in shape between an effective sample size of 20 and one of 50. Therefore, tests based on that distribution may be more robust than others (see Chapter 8), and large changes in the effective sample size due to spatial autocorrelation may have little effect on the interpretation of the data, particularly if n is large to begin with. In contrast, the χ^2 distribution changes markedly with the number of degrees of freedom and some tests that use it are also sensitive to the total sample size. This difference may lead to a different choice of statistical test when there is a selection to be made.

The critical probability levels used by ecologists, such as $\alpha = 0.05$, are only there for guidance in making decisions, and are based on our willingness to reject the null hypothesis when it is true. In many cases, the fact that the nominal significance level of 5% is actually 9% or actually 2%, because of negative or positive spatial autocorrelation, may not have a big effect on our interpretation of the data. Again, larger sample sizes can help. In addition, where testing and interpretation are sensitive to the distribution of the variable of interest (e.g. the normal distribution), larger sample sizes may allow us to be more confident in the analytic distribution attributed to the variables we are using. Unfortunately, most of the time we do not know for sure that the variable actually follows the distribution we assume and we have to rely on the robustness to departures from the assumptions of the tests we use. Very large data sets have a different problem, which that almost any deviation from expectation will be found to be significant because of large sample sizes, even if the deviations are very small. Then, once again, the ecological interpretation may need careful assessment of factors other than just the statistical test.

We would be remiss if we did not end this section with the reminder that spatial analysis is more than just spatial statistics and statistical tests. Exploratory analysis, model development and parameter estimation are all important components, as well. In fact, we

have colleagues who argue that statistical tests are only a half-step to knowledge, which really requires modelling and parameter estimation.

12.4 Complementarity of current methods

Many of the methods used for spatial analysis are closely related to one another, either conceptually or mathematically (Dale *et al.* 2002, among many; see Box 5.1). This is true, in an even more general way, of the broader range of methods described in this book. Knowing these relationships enables us to choose and use sets of methods that provide complementary insights, as with descriptive methods and inferential statistics. The methods chosen can be complementary in the characteristics they detect or in the range of their treatment, such as global versus local. The methods may also complement each other by the interconversion of data type (e.g. point events versus sampling units) or by being cumulative (such as blocking of quadrat data) as in TTLQV or Ripley's *K*-function versus decumulative (e.g. using individual units) as in PQV or Condit's Ω-function. Finally, methods may complement each other by having one method that provides evaluations at individual points in time (such as boundary detection) and another that provides evaluations at changes through time (such as polygon change analysis). It would be impossible to give an exhaustive list of all the combinations of methods that researchers might use to answer sets of related questions of their data, but we will give some examples to illustrate the concept and to provide some guidance.

The simplest spatial data may be a series of events in a single spatial dimension, like waterfalls along a river or termite nests along a line transect. To analyse such data, we can use the statistic W_m to detect nonrandomness and then h_m to detect clumping of events (see Chapter 4), or we can use the one-dimensional Ripley's *K*-function analysis to detect scales of overdispersion or underdispersion. If clumping is detected, we can plot the scores of the Ripley's analyses to find the locations of greatest clumping or those locations can be found using the probability calculations described in Dale (1999). That approach finds the sections of the sample transect where the probability of finding as many events as are observed is the lowest based on the null hypothesis of randomness. Whichever approach is used, the process of analysis involves the use of several complementary methods that detect different characteristics of the pattern.

As another example, consider a forest plot that has been mapped, recording the positions, species and diameters of all the tree stems. A first analysis could be the modified Ripley's *K*-function analysis to determine the scales at which the stems (of any species and any size) are aggregated or overdispersed. This could be followed by a univariate version of Condit's Ω analysis, based on rings rather than circles, to determine whether there are any distance classes of particular interest. If the overall pattern of the stems is patchy, Getis' score mapping or circumcircle score mapping could be used to examine the data for nonstationarity and to identify the positions of patches and gaps. The next analysis might be a mark correlation analysis to determine the aggregation or segregation of tree sizes as a function of distance (see Figure 12.5).

Any of a number of the multispecies analysis approaches described in Chapter 4 could be used to examine the interspecific associations. Having already used distance-based analysis on the data, however, a complementary approach would be to use the neighbour networks to look at the join counts for species pairs as in the Dixon method, which compares observed and expected counts. Finally, the data could be converted to raster format of counts or other quantitative data, using square units of a size chosen by the results of the previous analyses, and subjected to multiscale ordination) (MSO) to investigate the existence of multispecies pattern (Wagner 2004; Dray *et al.* 2012). This analysis scheme is illustrated in Figure 12.5.

Given abundance data for a single species from a transect of contiguous quadrats, 3TLQV or the Mexican hat wavelet could be used to determine the scales of pattern in the data, followed by NQV analysis to determine the sizes of the smaller phase in those scales of pattern. If the abundances are patchy, wavelets or a moving split-window (MSW) could be used to find the

edges between the regions of high density and those of low density. If an environmental factor such as altitude was recorded for the same sampling units, any of the covariance methods described in Chapter 5 (3TLQC or wavelet covariance) could then be used to determine the scales at which the species abundance covaries (positively or negatively) with the environmental variable (see Figure 12.6).

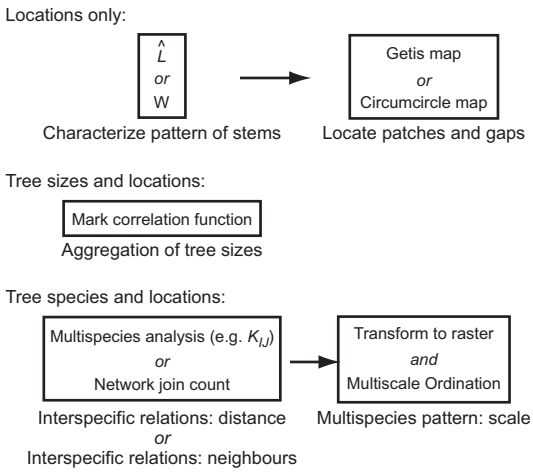

Figure 12.5 Complementary approaches to the analysis of marked event data (stems) in two dimensions: Ripley's K-function or Condit's Ω to characterize the nonrandomness of the events' positions, with the Getis method or circumcircle scores to plot the locations of the centres of patches or of gaps for a given scale.

For the quantitative single-variable data collected at spaced locations, there is also a range of methods that can be used to evaluate different spatial characteristics in the data (Chapter 6). For example, an omnidirectional correlogram or variogram can be used to evaluate the isotropic autocorrelation structure of the data as a function of distance. LISAs may then be used to plot localized areas of high and low spatial association. A complementary approach, not using actual distance, would be to look at the correlation of first-order neighbours in a spatial or functional graph of neighbours, then second-order neighbours and so on. Then, spatially constrained clustering can be used to identify aggregations of similar values, or triangulation-wombling can be used to detect boundaries (Chapter 9). Depending on the purpose of the study, having evaluated the nature of the spatial autocorrelation as a function of distance, interpolation techniques like Kriging can be used to provide an estimate of the variable over the whole study area. Lastly, if that analysis indicates areas of high variance, and thus poor-quality prediction, subsidiary sampling may be indicated to improve the quality of the interpolation. This sequence of analysis steps is shown in Figure 12.7.

As a last example, consider the hourly position records for a radio-collared animal and a habitat map on which those positions can be located. A first analysis would be to quantify the radial (distance) and angular autocorrelation as a function of lag, as described in Chapter 11. Local autocorrelation scores or local tortuosity measures would allow us to detect

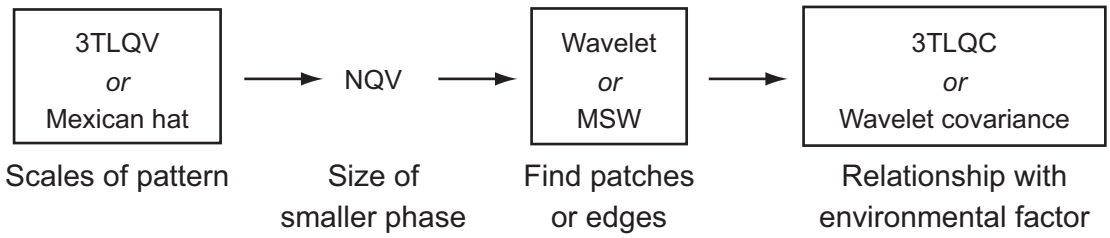

Figure 12.6 Complementary methods for analysing density data in a string of contiguous quadrats: 3TLQV or the Mexican hat wavelet analysis to detect the scales of pattern in the data, with Galiano's NQV to detect the size of the smaller phase, and local wavelet analysis or a moving split-window to detect patches or edges. If an environmental factor is also recorded, 3TLQC or wavelet covariance analysis can be used to detect the scales of positive or negative association of density with that factor.

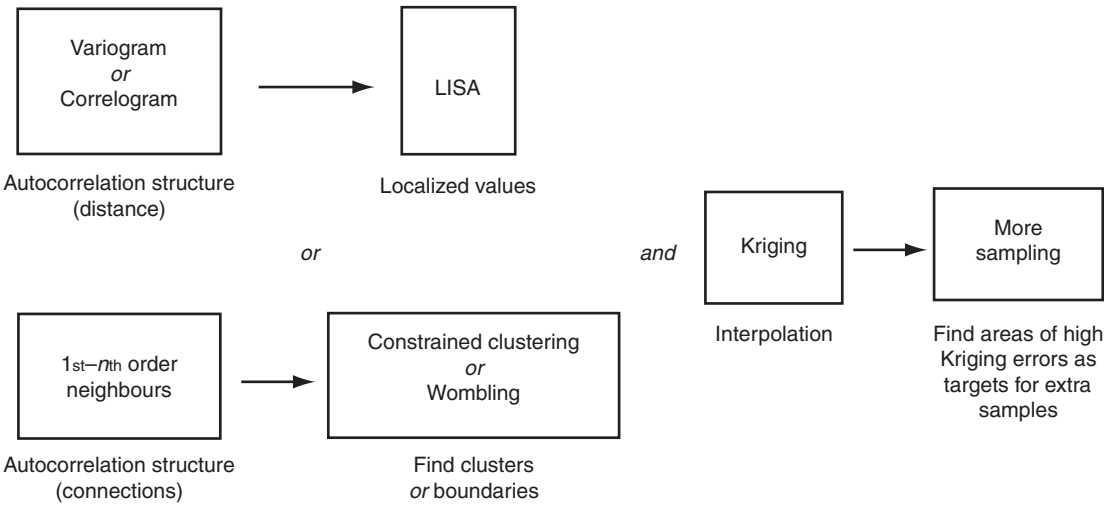

Figure 12.7 Complementary analysis of irregularly spaced records of a quantitative variable: variogram or correlogram analysis to characterize the overall autocorrelation structure as a function of distance, with LISA methods to detect local characteristics. Neighbour networks could be used to characterize the autocorrelation structure, based on connections rather than distance, with constrained clustering to find local clusters of similar values or wombling to detect boundaries between regions of different values. Kriging can be used to locate areas with high estimated variances, indicating a need for greater sampling intensity.

non-stationarity in the data. If stationarity is a reasonable assumption (or at least piecewise stationarity), we might then model the data to achieve a reasonable basis for Monte Carlo generation of artificial data for comparison (cf. Chapter 8). We could then compare the actual habitat use with either these Monte Carlo 'data' or with randomized positions of the original path of movement on the habitat map to evaluate nonrandomness in habitat use. This scheme for analysis is illustrated in Figure 12.8.

We cannot go through all possible combinations of analysis techniques that might be used, but these examples provide an idea of what we mean when we talk about complementary techniques. Figure 12.9, which is based on the 'relationship' diagrams in Dale *et al.* (2002), shows the relationships among the groupings of methods we described in our examples. It is clear from that figure that there are many other combinations of complementary analyses that could be pursued. In some cases, the choice of subsequent methods will depend on the results of the preceding step.

12.5 Analyses in both space and time

In this section, we will look at three examples of data frequently encountered by ecologists in their work, and treat them in more detail than in other parts of this book. They are permanent sample plot data, such as those used to monitor and predict forest growth; spatially linked time series such as metapopulation densities or tree rings; and animal–vegetation relationships, such as are of interest in wildlife management or conservation biology. One feature that the three have in common is that they involve temporal aspects as well as spatial.

12.5.1 Analysis of permanent sample plot data

Permanent sample plots (PSPs) are commonly used in forestry (1) to determine ecologically and economically important characteristics of forest stands, (2) to study the processes of recruitment, growth, competition, and mortality that occur within them, and (3) to make predictions about their future development.

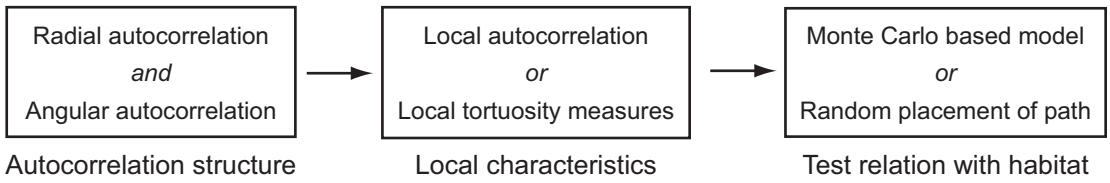

Figure 12.8 Complementary methods for analysis of radio-collar position data referenced to a map of habitat types: radial and angular autocorrelation analysis will characterize the autocorrelation structure, which might then be modelled. Local scores of measures such as tortuosity can then detect areas of behavioural intensity. The relationship between the path characteristics and the habitat can be tested by comparison with paths generated by a Monte Carlo method based on the model generated in the earlier steps, or by randomization of the path's position relative to the habitat map.

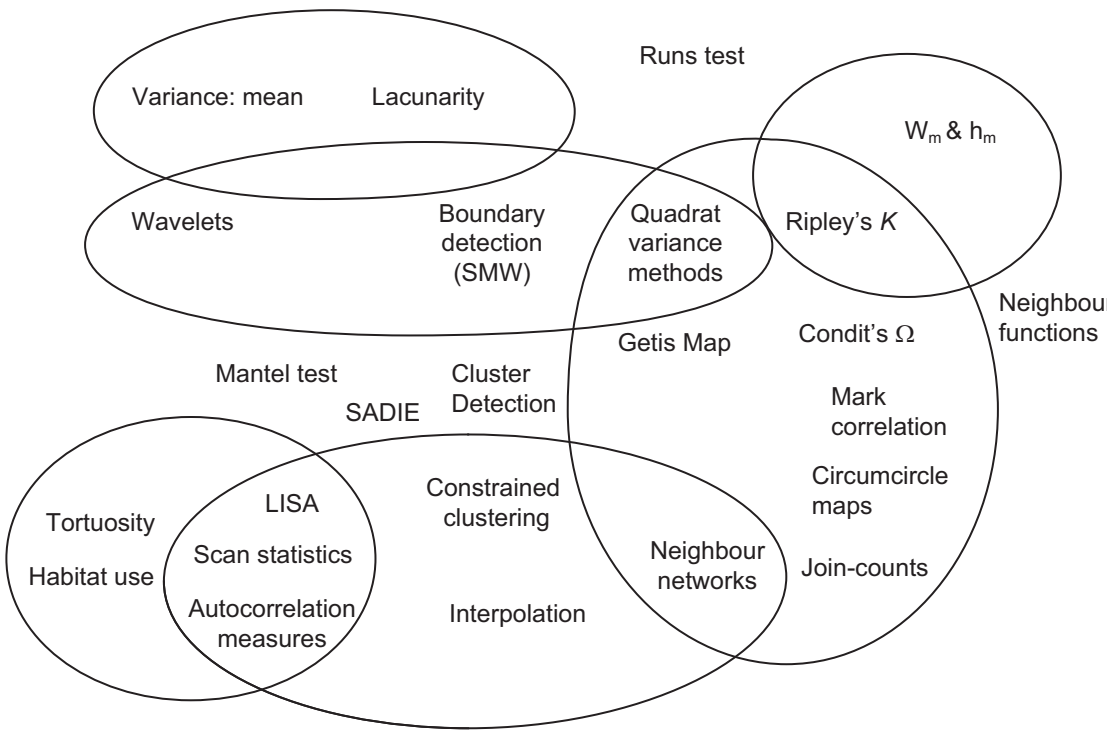

Figure 12.9 Relationships among the spatial analysis methods showing subsets of previous figures. Relative positions reflect the degree of similarity among the methods.

Most approaches to PSP analysis are non-spatial, or aspatial, because they do not include locations in the analysis in any way, whereas others are truly and explicitly spatial, where the stem location is important as are the locations of other stems, usually relative to the location of the stem of interest. Similarly, in some cases the actual year of observation is a factor (explicitly temporal), but only the relative timing of observations for other kinds of analysis (implicitly temporal).

Furthermore, given a single continuous study area, the results from an analysis can provide an evaluation of the entire area, giving what we call a 'global' analysis. A 'local' form of analysis, on the other hand, provides analytic results for subsections or point-focused portions of the study area. Many of the basic methods for spatial analysis exist in both the global form (usually) the original, and in a local version for more spatially localized interpretation.

12.5.1.1 Data

The basic data recorded in permanent sample plots are the locations of tree stems (usually a certain size, determined by height or diameter at breast height (dbh), or by age), together with their species, height, diameter, age and status (live, dead, healthy, diseased, etc.). There are several factors that affect the usefulness or appropriateness of the data for spatio-temporal analysis, including plot size (usually a hectare or less) and the number of stems included in it (standards in British Columbia, Canada, require at least 50 stems), edge effects, limits on stem size for inclusion and the closeness of the match between the spatial and temporal scales of the processes of interest and those of the data collected. In sampled data, often, there can be serious problems with inconsistencies and inaccuracies of the records, missing observations, and various kinds of recording errors.

Although PSPs are often small (e.g. 0.25 ha), providing the handicap of too few data for some of the approaches we might prefer to use, often they are established in regional groupings that are relatively close to each other. These are not usually intended as true replicates, it being known that there are site differences, and sometimes the intention is to cover a broad range of growth conditions. While some of the familiar methods for spatial analysis, such as those based on Ripley's K-function, cannot be used for data from spatially discontinuous areas, requiring a complete census of a study area, others, such as many of the modelling approaches can be used equally well for these dispersed sets of PSPs, provided that the fact of site location is included in the analysis, either implicitly or explicitly.

12.5.1.2 Summaries (aspatial)

One obvious use of PSP data is to create summaries of demographic processes such as growth rate as a function of age or size and mortality risk as a function of age or size. These summaries are usually averages over both space and time, implicitly assuming stationarity both spatial and temporal. Processes such as reproduction are also summarized, for example by yearly ratings of cone crops, admitting the expected year-to-year variation, whatever the trends with age and size. Competition is also included as a demographic process, by a 'mean field' measure of growth as a function of overall initial density and by plotting live plant size as a function of live plant density, giving the well-known 'self-thinning' curves. Although these measures are averages over space and time, they can be made spatially local by calculating them at geographically determined sub-sites, or temporally local, for example by decade, when the data are sufficiently numerous to allow this kind of subdivision.

12.5.1.3 Spatio-temporal graphs

A natural way to approach the presentation and analysis of PSP data is the use of spatio-temporal graphs (STGs; see Chapters 3 and 11). These graphs are mathematical objects consisting of sets of nodes (points), here the tree stems, joined in pairs by sets of edges (lines). For spatio-temporal graphs, the nodes have locations in space and in time; there are a number of different ways in which edges can be 'located', but their end-points must have locations in space and time, too, because they are nodes. Within a single time period, edges can indicate pairs of stems that are neighbours according to one of several possible neighbour network rules (topological, such as the Gabriel graph, or by distance threshold; see Chapter 3) or they can be directional, indicating competitive suppression or phenological precedence. Between different observation times, edges can indicate identity (joining stem i at time t to stem i at time $t + 1$), or parent–offspring relations. Changes in the positions or the directions of edges from one time period to another can be used to track changes in the structure or related processes in

the plot being studied. In general, any hierarchy of links in a spatio-temporal graph allows the testing of hierarchies of hypotheses: in space, in time, and in space and time. For example, the relative importance of neighbours defined by different network rules or of 1°, 2° and 3° neighbour.

In situations of several (or many) PSPs being established in a region, there is also the potential for STGs at another level, in which the individual plots are considered to be nodes, with edges indicating neighbours, pollen flow, plant community similarity, temporal precedence, or any of a large number of possibilities. This gives the option of creating what we have called a 'graph-of-graphs' (see Chapter 3), with the nodes of the graph at one level, each having a graph at another level associated with it. Here, both levels are spatio-temporal graphs.

12.5.1.4 Competition

One focus of the analysis of PSP data is on measuring the effects of neighbouring stems on stem survival and growth, through competition. This seems simple enough, and can be approached through a linear regression of annual growth increment as a function of the size, height and distance of neighbours, however defined (perhaps as the neighbours in a Delaunay triangulation, or as the stems within some distance threshold, e.g. the tree's own height). This approach is not really spatial, although distances are included, because the actual locations are not used. A number of factors that might need consideration are absent from this approach.

- The potential of secondary or tertiary effects of neighbours, such as indirect facilitation by the suppression of primary neighbours.
- Non-stationarity of processes or response variables.
- Spatial and temporal autocorrelation and their effects on significance assessment. For example, the distances to neighbours, to be used as 'independent' variables in a regression, are not independent of each other when a number of stems are used in the analysis. Stems A and B may be each other's neighbours, as well as sharing stems C, D, and E as neighbours, too.
- Unmeasured spatially explicit factors such as soil type, moisture, and so on.
- Effects of different threshold distances whether deliberate or as a result of other decisions on neighbour network definition.

12.5.1.5 Polygons

The dual of the network of Delaunay triangles is a set of Voronoi polygons (Chapter 3). Each polygon is associated with a single node (tree stem) of the network, and is that part of the plane closer to that node than to any other. The polygon, therefore represents the space and other resources that the stem (node) can pre-empt in competition with its neighbours. A number of studies (e.g. Mack & Harper 1977) have related the size of such polygons to the growth and survival of the plant it belongs to. A second factor that might be worth investigating is the plant location's eccentricity in the polygon (how far it is from the actual middle of the polygon), which is affected by the proximity of its closest neighbour sharing a polygon boundary. Being very close to another stem may reduce the positive influence on survival of having a large polygon. Voronoi polygons can also be used to determine the area associated with each of a set of PSPs.

12.5.1.6 Scales of pattern

A popular exploratory analysis of point location data is to determine the scales of pattern, almost always measured as physical distance, exhibited by the data using measures such as Ripley's *K*-function and variants upon it. These methods require the full census of the study area being analysed, sets of sample plots that are not contiguous cannot be approached in this way. The basic method performs a global analysis on unlabelled point pattern data to determine the scales of greatest underdispersion (clumping) and overdispersion (apparent repulsion) of the point events. Where the point events are marked by category variables, such as diseased or not diseased, marked point process analysis determines the most common scales of segregation and of aggregation of the different categories of the events. When the events have

quantitative labels, such as the tree height or stem dbh, mark correlation analysis determines the most common spatial scales of positive and of negative correlation of these values. For all this family of methods, the researcher must decide whether the hypotheses being tested are really of interest. The simple conclusion that the stems or their labels are not completely random in space, may not be helpful. Some refinement of the questions being asked may be required.

These exploratory forms of analysis are almost always carried out as global summaries over the entire area of study, but they can all provide local measures of the characteristic of interest, or spatially explicit results in the form of maps of 'hot spots' of high values, such as areas densely populated by trees, and 'cold spots' of low values, such as open areas of canopy gaps. For PSP data, these characteristics can then be followed as a function of time to monitor important changes and to make predictions about future behaviour. These analyses apply only to sample areas that are continuous and complete, so that for a set of PSPs distributed in a region, they are appropriate for individual plots, with the attendant problems of small numbers of stems, and potentially large edge effects; they are not appropriate for the set of plots or for any kind of regional analysis. However, questions of scale remain important, because the best understanding of the ecological characteristics of the plots may require a good match between the spatial structure of the data and the spatial structure, including scale, of the processes that give rise to the observations.

12.5.1.7 Spatial regression

In the section about competition above, it was suggested that regression equations that were used to describe and predict competitive outcomes were not truly spatial even when they included the distances to neighbours as 'independent' variables (Chapter 8). Consider the height increment of an individual stem i in year t, $G_i(t)$, first as a function, B, of its own characteristics, such as height, canopy volume, and so on, at time $t-1$, $C_i(t-1)$:

$$G_i(t) = \alpha + B(C_i(t-1)) + \varepsilon_i(t). \tag{12.1}$$

We can add in the neighbour effects with a function E, as

$$G_i(t) = \alpha + B(C_i(t-1)) + \sum_j N_{ij}\left[E(C_j(t-1), d_{ij})\right] + \varepsilon_i(t) \tag{12.2}$$

where summation is over all j, with N_{ij} as an indicator function taking the value 1 when stems i and j are neighbours by the current chosen rule, and 0 otherwise.

Autocorrelation of neighbours may also be included either in the response variable, G,

$$G_i(t) = \alpha + B(C_i(t-1)) + \sum_j N_{ij}\left[E(C_j(t-1), d_{ij}) + A_R(G_j(t))\right] + \varepsilon_i(t), \tag{12.3}$$

in the predictor variable, C,

$$G_i(t) = \alpha + B(C_i(t-1)) + \sum_j N_{ij}\left[E(C_j(t-1), d_{ij}) + A_P(C_j(t-1))\right] + \varepsilon_i(t) \tag{12.4}$$

or in both

$$G_i(t) = \alpha + B(C_i(t-1)) + \sum_j N_{ij}\left[E(C_j(t-1), d_{ij}) + A_R(G_j(t)) + A_P(C_j(t-1))\right] + \varepsilon_i(t). \tag{12.5}$$

The As are autocorrelation functions based on the definition of neighbours, and may take a variety of forms. True spatial regression includes the effect of location based on the coordinates of the focal stem:

$$G_i(t) = \alpha + B(C_i(t-1)) + \sum_j N_{ij}\left[E(C_j(t-1), d_{ij})\right] + L(x_i, y_i, t) + \varepsilon_i(t). \tag{12.6}$$

Here, $L(x_i, y_i, t)$ is a possibly nonlinear function of the coordinates of the focal stem, and of the observation

time. This permits the possibility of non-stationarity in either time or space, and in some circumstances may mimic the removal of linear trends in the data due to regional effects from moisture, temperature, or soil type.

Both spatial effects, those of neighbours, in two forms of autocorrelation, and those of 'absolute' location can be included together:

$$\begin{aligned} G_i(t) &= \alpha + B(C_i(t-1)) \\ &\quad + \sum_j N_{ij}\Big[E(C_j(t-1), d_{ij}) + A_R(G_j(t)) \\ &\quad + A_P(C_j(t-1))\Big] \\ &\quad + L(x_i, y_i, t) + \varepsilon_i(t). \end{aligned} \tag{12.7}$$

Of course, the sources of spatial effects are not fully independent and in variability partitioning among the sources of 'environment', 'space', 'space + environment' and 'error', they would be included together and not clearly separated from each other. That is 'a fact of life' for many kinds of spatial analysis, and should not deter us from making this kind of partition as appropriate as the circumstances and the data allow.

12.5.1.8 Hierarchical Bayesian analysis

Section 12.6.2 of this chapter describes the hierarchical Bayesian approach (HBA) in greater detail and in greater generality; here we will just discuss the application of this approach to PSP data. In general, for ecological applications, a hierarchical Bayesian analysis employs a hierarchy of three levels: data, process and parameters. Following the notation used in the section above, we can explicate this approach as follows.

The data are such observations as

$H_i(t)$ = height of stem i at time t

$G_i(t)$ = growth as height increment of stem i from time $t-1$ to time t

$S_i(t)$ = survival of stem i from time $t-1$ to time t

d_{ij} = distance from stem i to stem j, where j is a neighbour under the working definition, sometimes coded by N_{ij} which is an indicator function taking the value 1 when stems i and j are neighbours, and 0 otherwise.

The process is growth constrained by competition, and might be modelled by an equation similar to Eq. (12.7) above:

$$\begin{aligned} G_i(t) &= \alpha + \beta_H H_i(t-1) \\ &\quad + \beta_N \sum_j H_j(t-1) d_{ij}^{-2} N_{ij} \\ &\quad + L(x_i, y_i) + \varepsilon(t), \end{aligned} \tag{12.8}$$

where $L(x, y) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2 + \cdots$, giving a multinomial response 'surface', and N_{ij} is the 'network neighbour' indicator function determined by the chosen definition of the neighbour network. The 'error' term, $\varepsilon_i(t)$, is usually taken as normal with a mean of 0 and variance σ^2.

The parameter layer is then the distributions of the parameters in the models: σ, α, β_H, β_N, and $\beta_0, \beta_1, \beta_2, \ldots$.

Again, the analysis proceeds under the assumption of temporal stationarity and of the ability of the locational function, L, to account for any 'non-stationarity' in space. To complete the discussion of the analysis of PSP data, some further comments on detecting non-stationarity are in order.

12.5.1.9 Detection of change and non-stationarity

Most approaches to data analysis assume stationarity in both space and time, but the purpose for setting up PSPs in the first place is to determine and explain differences among sites, and to detect changes through time that may be attributable to factors external to the forest itself, such as climate change.

1. Spatial

Most PSPs are individually too small for the determination of spatial change, although that is not always the case. For large PSPs, various versions of point pattern analysis or spatial regression, followed by the explicit mapping of 'hot spots' and 'cold spots' for the variable

of interest, will be useful for detecting spatial changes. Similarly, with the HBA approach, explicit mapping of parameter values in space will provide a good indication of differences.

Where there is a regional collection of smaller PSPs, explicit mapping of the values of variables or parameters of interest, one value per PSP, can accomplish the same task for discontinuous data. Spatially constrained clustering of such values can provide a more formal evaluation of the spatial structure of these data.

2. Temporal

To detect changes in time, the approaches available will also depend on the amount and quality of the data available. Much can be done with long series of growth increments, such as can be obtained by the analysis of annual tree ring data. Under those circumstances, change point analysis for time series can be used, in addition to any 'hot spot – cold spot' methods for depicting change in variables or HBA parameters.

3. Spatio-temporal

Obvious combinations of the methods listed above can be applied to detect changes in time and space, given sufficient data of sufficient quality. In addition, cluster analysis constrained in both space and time may be very useful for detecting spatio-temporal clusters of similar values, and thus the boundaries between them. Similarly, the framework of a spatio-temporal neighbour network, as a STG, can be the starting point for the detection of boundary elements and then linked boundaries.

Comments

This outline is not intended to provide detailed instructions on these various analyses, but only to indicate the range of approaches that can be used in the analysis of PSP data. As always, the choice of methods will depend on the questions being asked or the hypotheses being tested and, in part, on the quality of the data. Also, as is usually the case, the available methods are not fully independent of each other, nor can their results be treated as independent findings, but only mutually confirming when they agree. In fact, there are some clear relationships among the methods described, such as that between the multiple spatial regression approach and the hierarchical Bayesian analysis, which use similar equations although different techniques for their evaluation and interpretation.

12.5.2 Spatially linked time series

One special topic that deserves some extra clarity and explicit instructions is that of spatially linked time series. Data of this kind arise when a single variable, such as rainfall, has been recorded periodically at several sites of known locations over a period of time. Other examples include the density of an organism of interest estimated at a number of sites over a period of several years, or the growth rings of trees of the same species in several forest stands within a region. Often, more than one variable is recorded, such as growth and spring temperature, or the abundances of both the predator population and the prey, or a whole set of meteorological measurements, including temperature, precipitation, humidity, wind speed, and so on. The goal is to determine the characteristics of the spatial relationships among the time series, or the temporal relationships among the observations in the spatially indexed data.

The basic hypothesis is usually that samples closer together in space or time probably are more similar than samples farther apart. In both space and time, however, it is also possible for more widely spaced samples to be similar (Jacquez 1996). In space, a simple alternation of patches and gaps produces spatially cyclic behaviour and temporal cycles will produce the same structure as a function of time, as we discussed in (Chapter 11). Data sets exhibiting cycles are not the only kind of spatially linked time series that are of interest in ecological applications and, even for cyclic data, synchrony versus asynchrony is not the only characteristic of interest in their analysis.

Very generally, we can identify at least two approaches to the analysis of such data: methods based on correlation, and methods based on model development and evaluation. Of course, there are actually lots more, but these are the most common and the most obvious choices.

12.5.2.1 Methods based on correlation

One-variable series: let us start with a time series of a single variable, observed at a number of sites; at location, i, there is a series of observed values, $z_i = z_{i1}, \ldots, z_{it}, \ldots, z_{iT}$. Data may be transformed before analysis, such as a log transformation of population density data. For any two sites (and suppose there are n in total, so there are $n(n-1)/2$ pairs), an obvious first step is to calculate the familiar Pearson correlation coefficient of the observations paired by observation date: z_{it} and z_{jt} for all t. The statistical significance of such coefficients will be affected by the lack of independence within the set of all possible pairs, and by the temporal autocorrelation in the data series, although Chapter 8 provides instructions on how this latter difficulty can be addressed. Some authors recommend removing the first-order temporal autocorrelation effects by fitting the equation

$$z_{t+1} = \alpha + \beta z_t, \tag{12.9}$$

and then using the residuals and the parameters' estimates for further analysis (Hanski & Woiwod 1993):

$$z'_t = z_t - (\hat{\alpha} + \hat{\beta} z_{t-1}). \tag{12.10}$$

We might then proceed to calculate the correlation coefficient of these residuals: z'_{it} and z'_{jt} for all t, but this first-order filtering may not be necessary, and may even prove confusing.

A variant on the simple correlation approach, for equal time spacing, is to combine the individual observations into matching temporal blocks of 1, 2, 3, ... observations, using a one-part gliding box template, to determine the temporal scale of correlation.

For any two spatio-temporal series, a general approach is the calculation of the cross-correlation coefficient between the two series, initially with a time lag of zero, but the time lag of the analysis can be manipulated in iterative recalculation in order to determine the best temporal offset. Where the two time series are z'_{it} and z'_{jt}, both of length T and with the 'bar' notation indicating the site mean, their cross-correlation is:

$$r_{ij}(z') = \frac{\sum_{t=1}^{T}(z'_{it} - \overline{z}'_i)(z'_{jt} - \overline{z}'_j)}{\sqrt{\sum_{t=1}^{T}(z'_{it} - \overline{z}'_i)^2 (z'_{jt} - \overline{z}'_j)^2}}. \tag{12.11}$$

The cross-correlation between series for the same variable from different sites (or the first-order residuals), $r_{ij}(z)$ or $r_{ij}(z')$, can be studied for their dependence on the distance between sites, d_{ij}. One measure of the two series' synchrony is the y intercept of the linear regression of this correlation as a function of distance, but its usefulness depends on the validity of a linear decay of correlation with distance. It certainly makes sense to plot the correlation coefficients as a function of distance as a first step in determining the spatio-temporal structure of the phenomenon.

A related approach is to compare 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 such a comparison is the Mantel test, which we have explained elsewhere has problems of its own (Chapter 7). As an alternative to plotting all site–pair correlations as a function of distance, a 'modified correlogram' plots the mean correlation coefficient of randomly chosen pairs of sites within specified distance classes, with each site used only once. This can have the drawback of giving results that may depend on the allotment of the pairs chosen, and the use of the mean obscures the variability in the data. Using each site only once will reduce the effectiveness of the information available.

Where each site has two time series of data, z_{it} and w_{it}, for example winter snowfall and annual growth increment in spruce trees, the cross-correlation coefficient between the two series can be calculated for each site, i:

$$r_i(z, w) = \frac{\sum_{t=1}^{T}(z_{it} - \overline{z}_i)(w_{it} - \overline{w}_i)}{\sqrt{\sum_{t=1}^{T}(z_{it} - \overline{z}_i)^2 (w_{it} - \overline{w}_i)^2}}. \tag{12.12}$$

To examine how this relationship changes with distance, we need the two-variable cross-correlation of z and w to be calculated for each pair of sites, i and j:

$$r_{ij}(z,w) = \frac{\sum_{t=1}^{T}(z_{it}-\overline{z}_i)(w_{jt}-\overline{w}_j)}{\sqrt{\sum_{t=1}^{T}(z_{it}-\overline{z}_i)^2(w_{ij}-\overline{w}_j)^2}}. \quad (12.13)$$

A symmetric estimate of the cross-correlation would require two calculations, $r_{ij}(z,w)$ and $r_{ij}(w,z)$ or $r_{ij}(z,w)$ and $r_{ji}(z,w)$, but the asymmetric version is usually sufficient. The changes in this coefficient as a function of distance can then be plotted using a weighted average of these values (see Tobin & Bjørnstad 2003), to determine the spatial structure of the time series' relationship. As always, a spatially explicit result in the form of a map showing the locations of unexpectedly high and unexpectedly low values of the cross-correlation coefficient could prove useful in pinpointing anomalies.

There are several variations on these basic methods, such as using non-parametric Spearman's correlation coefficient, rather than Pearson's, to evaluate synchrony. This measure 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. An alternative for more than two sites is to use Friedman's measure of concordance for several series (Conover 1980).

12.5.2.2 Spatial time series modelling

Developing and testing models of spatial time series provides greater choice and flexibility for analysis than the correlation approach just described (cf. Kamarianakis & Prastacos 2003). The most general model for a single variable is that called STARIMA, which is the space-time version of ARIMA, which, in turn, is the autoregressive integrated moving average model that generalizes the ARMA models we described in Chapter 8 (Box & Jenkins 1970). As the name suggests, this combines three elements: autoregression, an 'integrated' component (differencing), and moving average. These models are often denoted as ARIMA(p,d,q) models with the classification based on the orders of the three components: p for the order of autoregression, d for the differencing order, and q for the order of the moving average component.

For example,

$$z_t - z_{t-1} = \mu + \rho(z_{t-1} - z_{t-2}) + \varepsilon_t \quad (12.14)$$

is ARIMA(1,1,0) because it has a first-order autocorrelation term and a first-order differencing term, but no moving average component. On the other hand,

$$z_t = z_{t-1} + \varphi\varepsilon_{t-1} + \varepsilon_t \quad (12.15)$$

is ARIMA(0,1,1) because it has no autocorrelation, first-order difference, and a first-order moving average term. Combining all three components and rearranging gives ARIMA(1,1,1):

$$z_t = \mu + z_{t-1} + \rho(z_{t-1} - z_{t-2}) + \varphi\varepsilon_{t-1} + \varepsilon_t. \quad (12.16)$$

Clearly, these are models for single time series at individual locations, and we have a set of such series from different locations. The way this is done is to use a neighbour network of some kind to link up the locations in an orderly way and then to consider spatial lags, l, defined on this network, as well as temporal lags, k, in time. Then a single model, no matter how complicated it may be, is fit to the entire data set (all locations, all times); or several versions that can be evaluated for adequacy using AIC or some equivalent measure. Then, individual sites can be evaluated by studying the pattern of their deviations from the model that best fits all sites. Similarly the spatio-temporal pattern of the residuals from the model fitting exercise can be studied to provide an understanding of the overall spatio-temporal structure, and its locational details.

To provide a simpler example, following Kamarianakis & Prastacos (2003), consider a model with no difference terms (ARMA) with W_l being a $n \times n$ weight matrix of non-zero elements indicating where locations i and j are lth order neighbours, and Z_t is a

$n \times 1$ vector of observations. Then we end up with something like this:

$$Z_t = \sum_{k=1}^{p} \sum_{l=1}^{L(k)} \rho_{lk} W_i Z_{t-k} - \sum_{k=1}^{q} \sum_{l=1}^{L(k)} \varphi_{lk} W_i \varepsilon_{t-k} + \varepsilon_t. \quad (12.17)$$

The first double sum is the autoregression term and the second is the moving average term. Including a differencing term as a third component allows even greater flexibility and may be used to account for either seasonal or non-seasonal effects. This ARIMA approach is especially appealing for spatio-temporal systems in ecology, because it is based on spatial networks that define neighbouring sites, thus having a close and familiar relationship with spatial networks and spatio-temporal graphs we have described earlier (Chapters 3 and 11). In addition, depending on the system being studied, the approach also overlaps with the preceding discussion on the analysis of tree growth data from permanent sample plots. Given n sites, each of the n time series from those sites is modelled simultaneously as a linear combination of past observations at the site and the conditions at the sites that are defined as neighbours in the spatial network (Pfeifer & Deutsch 1980).

We will recommend this approach, while admitting that there are many others available in the range of spatio-temporal models. Without detailing or even listing them all, other approaches include spatio-temporal variograms, spatio-temporal maps by indicator Kriging, spatio-temporal covariance decomposition (Egbert & Lettenmaier 1986) and analysis based on spectral approaches (such as Cressie & Huang 1999) leading to wavelet analysis for this kind of data (Cressie & Wikle 2011). In addition, the modelling itself (not just the model) may follow different techniques, whether the more traditional or Bayesian (cf. Smith & Robinson 1995).

12.5.3 Spatial analysis of animal–vegetation according to data types

The third special topic that we have chosen as deserving some extra clarity and explicit instruction is that of how to approach the analysis of animals in relation to vegetation. Again, this topic takes us, at least sometimes, beyond simple aspatial and spatial approaches into the area of spatio-temporal analysis.

As an initial and general classification, we can consider three categories of vegetation data based on their spatial dimension: (1) points or point events, such as the locations of tree stems in a forest, (2) lines or linear structures, such as riparian strips, hedgerows or vegetation boundaries, and (3) patches or polygons, such as fields and woodlots in a landscape, or areas of evergreen or deciduous canopy in a mixed forest (Figure 12.10). Not surprisingly, the same dimension-based classification works for the animal data: (1) point events, such as nest sites or individual telemetry locations, (2) linear structures, such as foraging paths or telemetry locations linked through time, and (3) patches or polygons, such as territories or home ranges. All combinations of these are possible, giving nine categories for discussion, but we will not treat them equally, because some data combinations are more frequently encountered (e.g. animal locations or movement paths among vegetation polygons), whereas others are rare (e.g. animal polygons and linear vegetation). Usually, although not always, the plants are considered a fixed physical structure in which the animal behaviour, including preferences and directional movement, takes place. There are examples, however, where the defining vegetation characteristics change through time, such as individual plants coming into flower and then seed, or patch dynamics in a landscape. For all nine dimensional category combinations of data, the extra dimension of time can always be an additional component for more complex analysis and interpretation.

12.5.3.1 Plants and animals as point events (categories 1×1)

With data in this form, we are looking at the relationship among two or more point patterns and examining the segregation and aggregation of those different kinds of point events. This can take the form of an analysis based on Ripley's K-function (Chapter 4), when the data are from a complete census of the

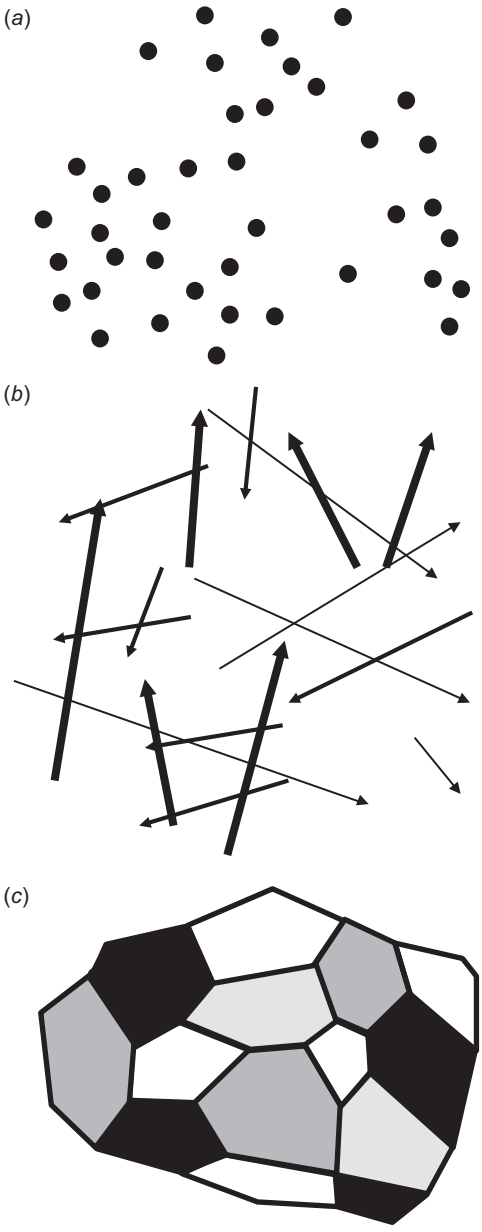

Figure 12.10 Three forms of vegetation data: (*a*) point data, such as plant stem positions, (*b*) linear features or structures, such as fallen logs with width, direction and length (*c*) patches or polygons, such as vegetation patches in a landscape mosaic. Animal data also fall into these three categories, giving nine possible combinations of the two.

events of interest, in which case we are determining the scales of greatest segregation and aggregation, or it can follow the protocol of a join count analysis based on some neighbour network as a spatial graph (Chapter 3). In the latter case, the actual distances between the events, the lengths of the edges in the spatial graph, can be used to evaluate the compactness or dispersion of any nonrandomness detected. Similarly, for a full census of events, spatial scan statistics can be used to detect clusters of any single type of point event or clusters of mixed labels (Chapter 6).

12.5.3.2 Plants as points with animal lines (categories 1 × 2)

In many instances, where the plants are (fixed) point events and the animal data is in the form of linear structures such as paths, the lines will go from point to point, joining them in pairs sequentially, and a spatial graph with the plants as the nodes is an obvious choice as the basis for analysis. A familiar example would be records of foraging activity in which an insect pollinator moves from plant to plant. Chapter 3 explains a number of interesting and helpful ways that spatial graphs can be used to determine the characteristics of such data. More challenging for analysis will be the examples where the animal lines do not run between the plants' points, but are independent or perhaps even avoid them. This is a situation in which a 'point and fibre' approach can be used, as described in Chapter 3 (Section 3.3), where the spatial relationships between the two different kind of structures are evaluated (ore bodies and lineaments in Australia). The test we describe in that Section (3.3) is the simple hypothesis of independence, with the two sides of its rejection, positive spatial association between the two sets, point events and lines, and negative association being the alternatives.

12.5.3.3 Plants as points with animal polygons (categories 1 × 3)

The most obvious picture of this situation is a spatial point pattern of the plants' locations with animal

feeding habits or behavioural tendencies allocating those locations into spatial polygons, like territories or home ranges, one for each animal or pair or family unit, whether the polygons are exclusive or not.

For exclusive non-overlapping polygons, the sizes and shapes of the polygons, the length of shared boundary for each pair of polygons, and some characterization of their 'contents' (such as number of stems, number of species, species diversity, biomass, etc.) are suggestions for the measures that should be useful in an ecological context. Most of the same list may apply when the polygons can overlap, except for the length of shared boundary for each pair. The measure of the relationship between overlapping polygons becomes much more complicated, because it can be as simple as physical area shared or number of stems in the common area or as complicated as measures of species diversity or species list distinctiveness for the region of overlap.

In either case, the focus of the analysis is to compare the characteristics of the individual animal-determined polygons: the range, average and variability of physical size, topographic structure and vegetation diversity. Also of interest will be the comparisons of these characteristics for pair of polygons that are first-order neighbours in some spatial network, and a similar list of characteristics of the boundaries where they abut or of the shared areas where they overlap.

12.5.3.4 Plants as lines, animals as points or lines (categories 2×1 and 2×2)

These combinations are likely to be rare in ecological applications, but we can imagine studies in which the plant structure is a set of linear features such as hedges or riparian strips, or even individual stems or fallen logs, along which individual animals or animal point events, such as crossings or periods of browsing, are recorded. In such cases, the Ripley's K-function for one dimension is an obvious choice for evaluating scale (Chapter 4) of the point events as they occur along the length of one-dimensional structures. If the actual spacing of the points along the lines is of interest, the broken line statistics, W_m and h_m, can be used to look for patterns of nonrandomness of these lengths. If more than one type of point event has been recorded, simple run statistics can also be used to examine the data for segregation or aggregation of two or more classes of events.

Where both the plants and the animals give rise to lines, there are a number of options for analysis, depending on the questions of interest, starting with the use of fibre analysis with two categories of fibre, P and A. These can be compared for their characteristics determined separately, as described in Chapter 4 (scales of overdispersion or underdispersion, unexpectedly few or many self-crossings, statistics describing fibre lengths, angles of orientation, tortuosity, autocorrelation, etc.) or for jointly considered characteristics of the whole pattern: scales of segregation or aggregation, unexpectedly few or many crossings of the two categories, pairwise measures of overlap of lengths, correlations of neighbour lengths, orientation, tortuosity, and so on. Two additional approaches suggest themselves, depending on the questions of interest. The first is a measure of the degree of 'linear coincidence', the proportion of the A fibres that follow closely the positions and alignment of the P fibres. This evaluates the extent to which the animal paths are correlated with the positions of the plant features. The second is the statistic of the intersections of A fibres with P fibres, A–P intersections, in comparison with the A–A and P–P intersections. In addition to the count statistics thus created, the result is also a spatial point pattern with three categories of point events (A–P, A–A, and P–P) which can be subjected to the usual kinds of spatial analysis in its own right. If we accept that the vegetation features have some sort of priority, we can also analyse the A–P intersections as point event on the fibres of the plant linear features, following the methods already described.

For both kinds of statistics, we can employ a hierarchical approach to their evaluation by a series of restricted randomizations:

- repositioning both entire sets of A fibres and P fibres without rotation;

- repositioning both entire sets of A fibres and P fibres allowing rotation;
- repositioning all individual A fibres and P fibres without rotation;
- repositioning all individual A fibres and P fibres allowing rotation.

Comparing the counts from this series of restricted randomizations with those from the original data will permit an evaluation of the 'significance' of the observed counts, and some ideas on the factors that affect their frequency.

Although this data combination, of lines and lines, may be rare, the richness of avenues for analysis suggest that it would be worthwhile pursuing the combination in future studies for the sophisticated hierarchy it allows.

12.5.3.5 Plants as linear structures with animal polygons (categories 2 × 3)

This is an interesting possibility, with the approach to analysis depending on whether the linear features are sufficiently large to contain (small) polygons associated with individual animals (like territories or home ranges), or they are much smaller than polygons, which can therefore contain several of the plant structures. The small polygons in larger linear features probably can be treated in more or less the same way as point events in linear features, because of the relative sizes, and the small lines in large polygons are going to be similar in analysis to the case of points in polygons, already described, with a few modifications perhaps to allow for the fact that line events, even small ones, can interact spatially in more ways than point events. Again, measures such as counts, diversity, and spacing will help characterize these polygons and will form the basis for a comparison of pairs of polygons, and the effect of the distance between them (physical or network-based) on their similarity or difference.

12.5.3.6 Vegetation polygons and animals as points or lines (categories 3 × 1 and 3 × 2)

For data that combines polygons for plants (usually polygons of vegetation types) with animal point events, the choice of analysis will depend on the animal data: simple locations or locations with categories attached based on behaviour or activity. At the most basic, the analysis can be non-spatial, asking questions about the frequency of (different categories of) animal point events in the different vegetation classes; such as, 'Do red deer occur (or browse) more frequently in open mixed woods or in shrub heaths?'. When the spatial relationships among the polygons are included, questions about the degree of similarity or autocorrelation between neighbouring polygons can then be asked and answered. If the animal data are lines, not points, and short enough to be contained in one polygon, the same kinds of analysis can be used.

If, however, the lines are long enough to include more than one vegetation polygon, or if the point events have labels indicating temporal order, thus creating the edges of a spatial graph of that sequence, a number of other options become available. Here, the lines can go between different polygons, and a sequential path or other spatial graph can be viewed as linking the polygons by means of the graph edges. This data structure then allows counts of various kinds of transitions: edges within vegetation type A, B, C, and so on; edges between types A–B, A–C, B–C, and so on. This allows several approaches to analysis, including join count statistics (Chapter 6), spatio-temporal analysis including spatio-temporal graphs (Chapter 11), habitat selection analysis or similar approaches through spatial regression (Chapter 8), and hierarchical Bayesian analysis (Chapter 12, Section 12.6.2).

A special case of this data combination is commonly encountered when the vegetation data (possibly accompanied by environmental data such as elevation) are available as a grid of pixels, derived from a number of possible sources including satellite images, and the animal data are point locations based on sightings or telemetry. Although each point location is found in only one pixel, series of points or paths include many pixels. Pixels differ from other kinds of polygons because adjacent pixels can be somewhat different but still included in the same general category, depending on the classification system being used, and may then be amalgamated into larger

polygons of vegetation classes, based on one of many different algorithms available. While this is a special case of the 'polygon-point' data combination, the comments above on how these may be analysed apply in this instance.

As long as the linear events are sufficiently small in comparison to the size of the plant-defined polygons, then (like point events), they can lend themselves to analysis by spatial regression to determine the relationship of these animal events to the characteristics of the polygons. Following the discussion of spatial regression in Section 12.5.1, we can modify the material there to consider the counts of individual point events in polygon i in observation period t, $F_i(t)$. Hence, we need first as a function, B, of the polygon's quantitative characteristics, such as area, age, canopy height, canopy volume, and so on, at the same time, $C_i(t)$:

$$F_i(t) = a + B(\mathrm{C_i}(t)) + \varepsilon_i(t). \tag{12.18}$$

Because these are polygons, not tree stems, we do not add true neighbour effects, as we did for PSP data, but we can add in the spatial autocorrelation due to neighbouring polygons, however defined, with some dependence on the inter-centroid distance, $d_{ij,}$ with summation over all j, and N_{ij} as an indicator function taking the value 1 when polygons i and j are neighbours by the chosen neighbour rule, and 0 otherwise.

Spatial autocorrelation of neighbours may be included either in the response variable, F,

$$F_i(t) = \alpha + B(C_i(t)) + \sum_j N_{ij}[A_{SR}(F_j(t), d_{ij})] + \varepsilon_i(t), \tag{12.19}$$

in the predictor variable set, C,

$$F_i(t) = \alpha + B(C_i(t)) + \sum_j N_{ij}[A_{SP}(C_j(t), d_{ij})] + \varepsilon_i(t), \tag{12.20}$$

or in both

$$F_i(t) = \alpha + B(C_i(t)) + \sum_j N_{ij}[A_{SR}(F_j(t), d_{ij}) + A_{SP}(C_j(t), d_{ij})] + \varepsilon_i(t). \tag{12.21}$$

The As are autocorrelation functions based on the definition of neighbours for the current context, and these functions may take a variety of forms, including a possible dependence on the distance between the polygons' centroids (or edges), distance d_{ij}, as indicated in the equations. This flexibility allows us to consider combinations of different characteristics for the form of spatial dependence or spatial autocorrelation in the predictor and response variables. Observed spatial dependence in the response variable can be the result of its own spatial autocorrelation, only from its response to the predictor variable which has its own autocorrelation, or from both sources. In addition, any of the autocorrelation functions can be of different forms, for example, being linearly decreasing with distance (comparable to the semi-variance linearly increasing), or falling off with distance to zero (comparable to the semi-variance reaching a sill), as shown in Figure 12.11. In either of those cases, the distance between the centroids of polygon neighbours will affect the autocorrelation function A_{SR} and A_{SP}. As a simple example, the response variable could be counts of boreal toads in forest stands, with the predictor set be a single variable of soil moisture. Spatial dependence in the toad counts could be made up of its own autocorrelation which has a spatial range defined by dispersal distance and of its dependence on soil moisture, which has a linear response to distance of a much larger scale.

True direct spatial regression includes the effect of location based on the coordinates of the polygon:

$$F_i(t) = a + B(C_i(t)) + L(x_i, y_i, t) + \varepsilon_i(t). \tag{12.22}$$

Here, $L(x_i, y_i, t)$ is a possibly nonlinear function of the polygon's centroid coordinates, and of the observation period. This permits the possibility of non-stationarity in either time or space, and in some circumstances may mimic the removal of linear trends in the data due to regional effects from moisture, temperature, or soil type.

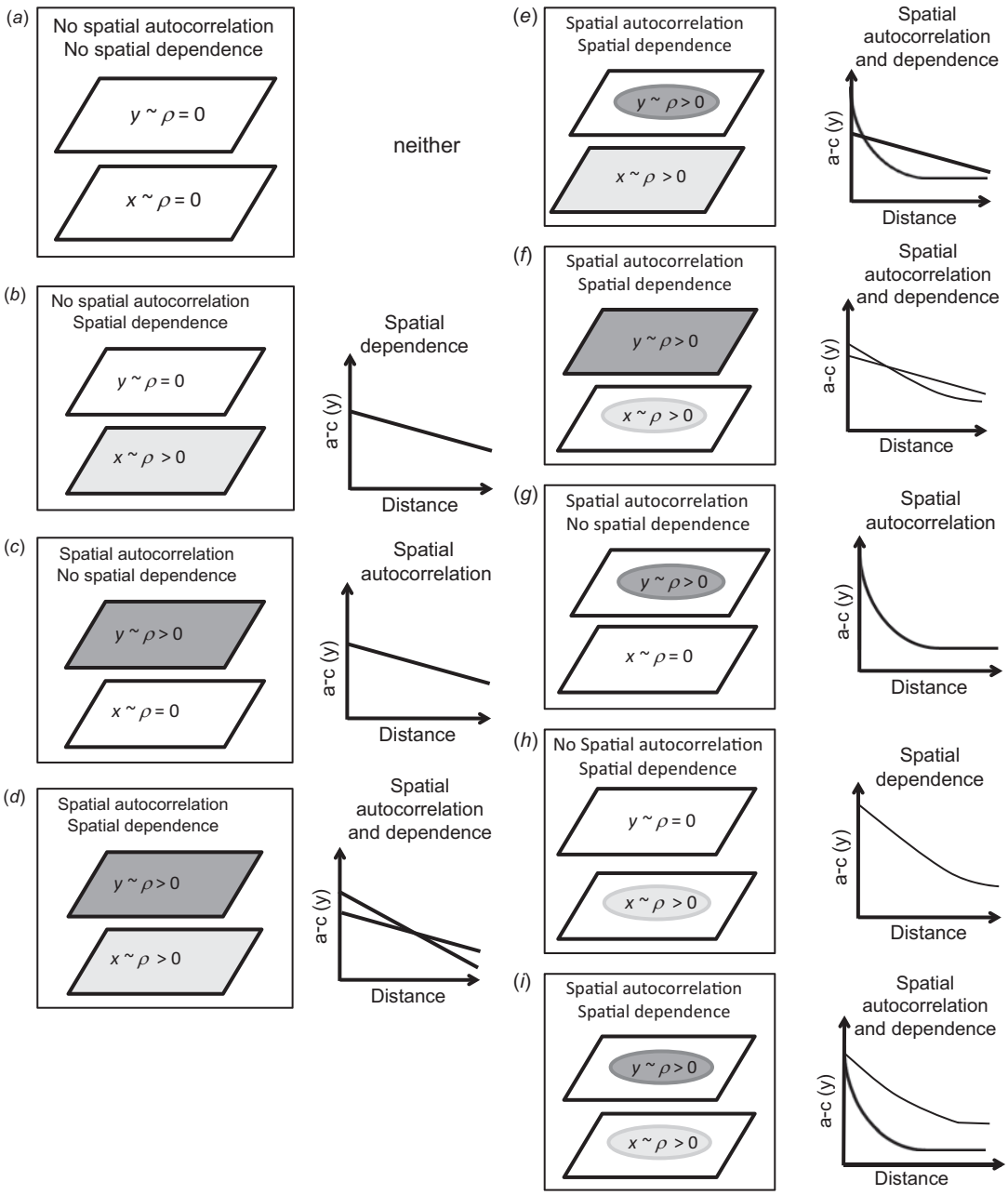

Figure 12.11 Combinations of different forms of spatial autocorrelation (ρ of the response variable *y*) and spatial dependence (ρ of the predictor variable *x*), whether continuing to decrease linearly with distance, or reaching a plateau. The left panels show the spatial structure of the ecological and environmental data in a landscape. The right panels show the degree of spatial autocorrelation (as indicated by 'a-c') of the response variable (as indicated by a-c (*y*)). (*a*) no spatial autocorrelation and no spatial dependence; (*b*) only linear spatial dependence; (*c*) only linear spatial autocorrelation; (*d*) both linear spatial dependence and autocorrelation; (*e*) patchy spatial autocorrelation and linear spatial dependence; (*f*) linear spatial autocorrelation and patchy spatial dependence; (*g*) spatial autocorrelation; (*h*) only patchy spatial dependence; and (*i*) patchy spatial autocorrelation and dependence.

Both kinds of spatial effects, those of neighbours, in two forms of autocorrelation, and those of 'absolute' location can be included together:

$$F_i(t) = \alpha + B(C_i(t)) + \sum_j N_{ij}[A_{SR}(F_j(t), d_{ij}) + A_{SP}(C_j(t), d_{ij})] + L(x_i, y_i, t) + \varepsilon_i(t). \quad (12.23)$$

The last 'wrinkle' that can be added is to include temporal autocorrelation for polygon i from either the polygon's characteristics (the predictor variables, C_i), or in the animal event counts (the response variable, F_i), as defined by the functions $A_{TC}(\mathrm{C}i(t-1))$ and $A_{TF}(F_i(t-1))$. Of course, both can be included if both inherent and induced temporal autocorrelation are expected to act in the system under study. This addition provides the flexibility of various combinations of spatial and temporal autocorrelation.

12.5.3.7 Polygons and polygons (categories 3 × 3)

This is the last of the data combinations, and probably the least studied up to this point. If there is a large disparity between the spatial sizes, the data may be reduced to points and polygons, but the more interesting, and challenging structure will be the case where the polygons from the two data sources are similar in size. Here there will be interesting questions about how well, or how badly, the two kinds of polygons 'match'. Do the two sets pair up? How well do areas of polygons of the two kinds correspond? Do pairs of polygons have similar spatial 'footprints'? If the complete footprints do not match, how well do boundaries or portions of boundaries align?

One part of the analysis would be to use the centres of gravity of the two sets of polygons as the nodes for two spatial graphs, with edges determined by a simple rule such as the Delaunay triangulation or a distance threshold where appropriate. The joint characteristics of these two spatial graphs can then be used to evaluate some aspects of the relationship between the set of animal-based polygons and the plant polygons, however defined, using polygon change analysis. This combination of data, and the most fruitful analysis for it, requires and deserves further investigation.

12.6 Future work

12.6.1 Ongoing development

Throughout this book, we have pointed out many areas of methodological research we have described that require and deserve further efforts. A few of these are polygon change analysis, spatio-temporal graphs theoretical analysis, entropy approaches to categorical data, and the analysis of multispecies point patterns. This list is not exhaustive, but gives some idea of the range of topics that are worth further investigation. For ecologists, the 'bottom line' is still to clarify the relationship between process and pattern. This may take the form of the challenge to find and make explicit the right ecological hypotheses to test and to find the right spatial measure to test them. For example, at least some of the confusion or controversy about β-diversity analysis may be resolved by determining the ecological hypotheses that proposed measures or methodological approaches are expected to clarify. In many instances, our concerns about methods, and their possible weaknesses, could be solved by knowing more about the biology of the system being studied. In particular, very detailed knowledge about spatial processes would enable us to refine the methods we use. This parallels the suggestion that pilot studies and other prior information can help with better decisions when we design surveys or experiments to investigate ecological phenomena.

Another issue that is yet to be fully resolved is the testing of spatial indices and measures for statistical significance. As discussed in Chapter 1, the search for significance tests is greatly complicated by the various forms of lack of independence, both in the data themselves and the calculations on which our statistics are based (TTLQV, Moran's I, Geary's c, and semi-variance being extreme examples of the reuse of data). The sensible use of restricted randomizations seems to be the solution, again going beyond (we hope) the obvious and simple null hypothesis of 'there is no pattern'. Similarly, it is becoming more and more of interest to be able to compare the spatial pattern of two different areas or of the same area at different moments. When it is the case of the same area at different times, it is

more likely that the same underlying process is at play; but in the case of two different areas, it is not so. Moreover, a spatial pattern is only one realization of a process and so the pattern may be 'significantly' different at two locations, despite having resulted from the same process. These issues related to the comparison of spatial pattern can be only partially addressed by means of stochastic spatial modelling (Fortin *et al.* 2003; Remmel & Fortin 2013). As we have noted, statistical testing may often be viewed as the first step in a process that leads to greater understanding through subsequent modelling and parameter estimation for those models.

In Chapter 8, we spent a considerable amount of effort discussing models of spatial autocorrelation, particularly as a learning tool. While the patchiness, which we determined is a common characteristic of ecological data, can be modelled in the ARMA structure, it is not clear how biologically realistic such models would be. We have to ask what the ecological processes are that would give rise to the kinds of spatial autocorrelation we described. Dieckmann *et al.* (2000) provided some helpful discussion of this issue, but we will provide a couple of examples. In most statistics textbooks, the analysis of variance is presented using a model such as

$$X_{ij} = B_i + T_j + \varepsilon_{ij}. \tag{12.24}$$

The observed value (suppose it is crop yield) is interpreted as the sum of a block effect, B_i, a treatment effect, T_j, and an error term, ε_{ij}. When the error term is attributed to variation in soil nutrients, soil moisture, and light availability, the autocorrelation in yield is very similar to the induced structure described in Chapter 8. In a well-known plant competition experiment, Franco & Harper (1988) found that the sizes of first neighbours were negatively correlated while those of second neighbours were positively autocorrelated. That is, large plants had smaller neighbours and small plants had larger neighbours, probably as a result of competition. In this case, a first-order autoregressive model (Chapter 8) with ρ negative, would be a biologically realistic model of the spatial structure. As we suggested above, in many ecological examples we expect the variable of interest to have both inherent and induced autocorrelation, but the biological and physical processes that give rise to it may not always be clear, and so realistic models may remain a challenge for us all.

The final piece of commentary for this section follows on from the previous discussion of inherent and induced autocorrelation and the fact that given only the data for the dependent variable, they may not be distinguishable (cf. Chapter 8). For a number of reasons, it is either necessary or merely desirable to be able to distinguish between the two, and therefore, finding a study design that permits this distinction is well worth the effort. Figure 12.11 illustrates the interactions of three forms of autocorrelation in the predictor (x) and response (y) variables (none, global, localized) and the effects of their combinations on the autocorrelation. Clearly, a sampling design that allows separate evaluations of the spatial structure of predictor and response variables will provide a better understanding of the system being investigated.

12.6.2 The hierarchical Bayesian approach

This approach is increasingly popular for a range of applications in spatial analysis, and therefore seems to require a section of its own, to provide a somewhat detailed explanation of the concepts. Obviously, we cannot provide all the details here of the applications, but enough to provide a helpful indication of its application so far, and of its potential for the future. We begin with a short commentary on the Bayesian approach to probability in general, and then move to a discussion of hierarchical models and their application for spatial analysis.

12.6.2.1 Bayesian analysis

In comparison with the classical statistics that are usually taught to undergraduates, the Bayesian approach has some notable differences. The term 'probability' becomes more closely associated with degrees of belief or credibility than with long-term averages from repeated trials or with the physical properties of simple systems such as tossed coins.

Instead of calculating the probability of events, such as the probability of a standard pair of fair dice turning up a total of 7, from hypotheses, models or model parameters the emphasis is on determining how the observed events or the data affect the probabilities of the hypotheses or model parameters. As an example, given a pair of apparently standard dice that produce six 7s in a row, what is the perceived probability that the dice are indeed fair? For that reason, the approach typically begins with some assumed prior probability distribution for the parameter of interest and then determines how the observed data changes our understanding and provides a measure of that value in the form of a posterior probability.

Given a hypothesis with parameter θ (e.g. for the random roll of a fair die the probability of turning up 1), and data D (e.g. a sequence of rolls of that die in which exactly three of ten are 1), we are used to calculating the conditional probability of D given θ, $P(D|\theta)$, for example $P(3\ 1\text{s in } 10 \text{ trials} \mid \theta = 1/6)$:

$$P(D|\theta) = \binom{10}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7. \tag{12.25}$$

This conditional term also appears in the equation relating it to the joint probability of D and θ:

$$P(D, \theta) = P(\theta)P(D|\theta). \tag{12.26}$$

The Bayesian approach requires us to find an expression for $P(\theta|D)$:

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{P(D)}. \tag{12.27}$$

This looks easy to solve, but has a few wrinkles that make it more difficult than it looks. The same equation is sometimes expressed in terms of likelihood, denoted L, which is a sort of inverse of probability so that the likelihood of a parameter or hypothesis, given the data, is proportional to the probability of the data given the hypothesis:

$$\theta \xrightarrow{P(D|\theta)} D$$
$$D \xrightarrow{L(\theta|D)} \theta.$$

The relationship is simple:

$$P(D|\theta) \propto L(\theta|D). \tag{12.28}$$

The posterior probability can then be expressed based on these likelihoods or probabilities (some texts use the probability form but refer to $P(D|\theta)$ as the 'likelihood term') but $P(D)$, the marginal probability, has to be calculated over the entire range of possible values for parameter θ, designated by Θ:

$$P(\theta|D) = \frac{P(\theta)L(\theta|D)}{\int_\Theta P(\theta)L(\theta|D)d\theta}. \tag{12.29}$$

The two wrinkles to be addressed are determining an appropriate prior probability distribution for the parameter, $P(\theta)$, because it can affect that posterior probability that is calculated in this way, and calculating the integral, which may not have an analytic solution. The first wrinkle may be solved by using the most 'neutral' prior assumption, for example equal probabilities for all values (a uniform distribution) or by using additional information about the system being studied. The second can be solved by numerical methods, such as the commonly used Markov Chain Monte Carlo (MCMC) iterative procedure, or the more recent Integrated Nested Laplace Approximation (INLA) to determine the appropriate value (see Beguin *et al.* 2012).

12.6.2.2 *Introductory example* (paraphrased, with some liberties, from Eddy 2004)

A cue ball is rolled from one end onto a billiard table and it stops at a position which follows a uniform distribution along the table's long axis (Figure 12.12). Its position is noted but is treated as an unknown.

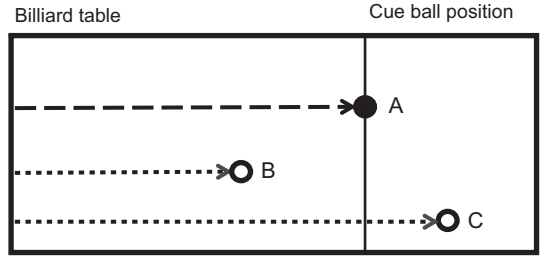

Figure 12.12 The physical arrangement of a billiard table and billiard balls as an illustration for Bayesian statistics.

A second ball is then rolled in the same way a number of times, and the result is scored as B if it comes to rest left of the cue ball's marked position and as C if it is to the right. The probability of A depends on the cue ball's position, and occurs with probability q; B occurs with probability $(1 - q)$. Eddy (2004) provided the following problem: after eight trials, A = 5 and B = 3, and although q is unknown, we want to calculate the probability that three more Bs occur before a single additional A occurs; call that outcome W. Based on a uniform prior distribution for q, which can be justified by the way the trials were set up, we have

$$P(q|A = 5, B = 3) = \frac{P(q)P(A = 5, B = 3|q)}{\int_\theta P(q)P(A = 5, B = 3|q)dq}. \tag{12.30}$$

With some substitution and rearrangement this gives

$$P(W) = \frac{\int_0^1 q^5(1-q)^6 dq}{\int_0^1 q^5(1-q)^3 dq}. \tag{12.31}$$

The convenient fact about this example is that the two integrals have analytical solutions: (5!6!/12!) and (5!3!/9!). This allows us to determine the expected $P(W) = 1/11$. Based on this approach, and the fact that the expected $P(W)$ is $(1 - q)^3$, the estimate of q is $1 - (1/11)^{1/3} = 0.550$; clearly this is not the same as the maximum likelihood estimate of q, which is $5/8 = 0.625$.

This example provides a useful practical and conceptual introduction to how a simple Bayesian analysis proceeds, with the use of inverse probability and assessment of a posterior distribution based on a prior distribution. There is much debate about the advantages and disadvantages of Bayesian methods in the literature, but a brief and clear assessment is provided by Wasserman (2004) in the book with the alluring title *All of Statistics*.

12.6.2.3 Hierarchical analysis

For the most basic situation, in general, the understanding is that

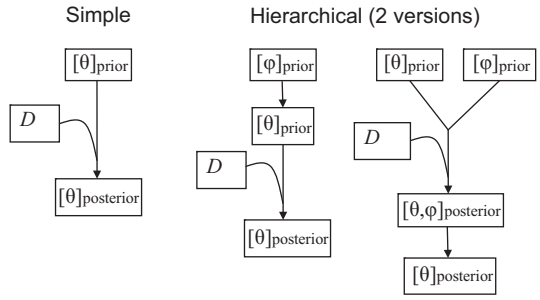

Figure 12.13 In simple Bayesian analysis with parameter θ (left), a prior distribution of a parameter is combined with the data to produce a refined posterior distribution for the same parameter. Hierarchical Bayesian analysis includes another parameter, φ, (two versions on the right); the parameter of immediate interest, θ, has a distribution that depends on another parameter, here φ.

$$P(\theta|D) \propto L(\theta|D)P(\theta).$$

(The posterior probability is sometimes given a slightly different notation to distinguish it, for example π instead of P.) This simple model becomes hierarchical when the parameter θ is itself conditional upon another parameter, call it φ. The posterior distribution then depends on the prior distribution of θ given φ and φ has its own prior, too (Figure 12.13). That is, there is a 'hyper-prior' with 'hyper-parameters' (Gill 2002). The posterior probability becomes:

$$P(\theta, \varphi|D) \propto L(\theta|D)P(\theta|\varphi)P(\varphi).$$

For example, the data themselves could consist of counts of individuals in sampling units like quadrats, with the counts following a Poisson distribution with parameter λ. That parameter could vary across time or space, and could itself follow another distribution, such as a Gamma distribution with parameter β (see example below by Beckage *et al.* 2007).

The hierarchical form can also be expressed by reference to an intervening process, rather than the parameters of parameters. Many of the ecological papers cited here use the following structure, with square brackets indicating the probability distribution (see Cressie & Wikle 2011).

Data model: $[D|Y, \theta_D]$ or just $[D|Y, \theta]$.

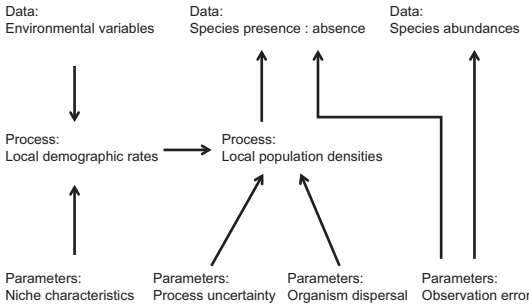

Figure 12.14 The schematic structure of a three-level Bayesian approach to spatial population dynamics (based on Pagel & Schurr 2012).

Process model: $[Y|\theta_P]$ or just $[Y|\theta]$.

Parameter model: $[\theta_D, \theta_P]$ or just $[\theta]$.

$$\text{Then, } [Y, \theta|D] = \frac{[D|Y, \theta][Y|\theta][\theta]}{[D]}. \qquad (12.32)$$

In summary, the three-level hierarchical scheme for ecological applications can be shown as follows.

$$\begin{array}{ll} \textit{data} & [D|Y, \theta] \\ & \updownarrow \\ \textit{process} & [Y|\theta] \\ & \updownarrow \\ \textit{parameter} & [\theta] \end{array}$$

This three-level model can be made more complicated by having several categories or compartments in each level. Pagel & Schurr (2012) provided an example in their description of a process-based dynamic range model that combines the niche concept with spatial population dynamics. It looks something like Figure 12.14 (Pagel & Schurr 2012): a hierarchical model can also have more layers or more levels of dependence, but the use of at least two levels in ecological models for analysis has become quite common, and we can provide a summary of a few of these applications.

12.6.2.4 Empirical hierarchical models

One alternative to specifying the parameter model, as we have described for the Bayesian hierarchical model approach, is to estimate the parameters from the data and then use those estimates in the models of the layers above. This approach can be called the empirical hierarchical model approach, for obvious reasons, but it is sometimes referred to as the empirical Bayesian model, which could be quite confusing (see Cressie & Wikle 2011; among many). The estimates can be derived by any of several methods, but they are substituted into the process and data models as if they were *known*, which can give misleading results, especially if there are transformations of the estimates, which have variability associated with them (because they are estimates!).

12.6.2.5 Hierarchical Bayesian analysis (a summary of hierarchies in ecological applications)

The hierarchies in these kinds of analysis are not usually 'structural' in the sense of an inclusive sequence like leaf, branch, tree, stand, site, region, etc. They tend to represent series of increasing statistical resolution or of causal refinement. This summary is intended to explain the kinds of hierarchies that have appeared in ecological examples, with comments, where appropriate, on the inclusion of spatial structure in the analysis. It is not intended as a detailed summary of the models or of the results, nor is it a complete catalogue of all studies to date.

Example 1.
Change point analysis of seedling counts in transects of contiguous quadrats in gaps and under canopy (Beckage *et al.* 2007). This example provides a good link between this material on Bayesian approaches and the more 'traditional' methods described elsewhere on patch: gap analysis for quadrat-transect data.

Level 1. Independent seedling counts, X_i, with Poisson parameters λ_c (under canopy) and λ_g (in gaps). Two versions (canopy versus gap) of $P(X|\lambda)$.

Level 2. Link λ_c and λ_g in several different transects with similar, but not identical, values; the λ_c and λ_g follow Gamma distributions, Γ_c (under canopy) and Γ_g (in gaps), 'borrowing strength' across different sample units. Two versions of $P(\lambda|a,b^{-1})$.

Level 3. The parameter b^{-1} of the Gamma distribution itself follows an Inverse Gamma distribution with parameters γ and ζ. Two versions of $P(b^{-1} \mid \gamma,\zeta)$.

The parameters at successively higher levels of this kind of hierarchy are sometimes called 'hyperparameters'.

Correlations between seedling counts in adjacent quadrats (spatial autocorrelation) were permitted by including a Gaussian random field prior for random effects on the λs for quadrats in the same transect.

Example 2.
Harbour seal haul-out counts in Prince William Sound (Cressie *et al.* 2009).

The data are the counts of seals at specific sites around the Sound, observed at a number of dates over a number of years, and so the predictor variables include site, but also year (for trend), date (for seasonal effects), and tide status (for phase).

The hierarchy in this example, and in the two that follow, is based on the following construction: The data, D, are dependent on ecological processes, E, and the parameters, P_D, for the relationship between the data and the processes, a data model. In turn, P_E are the parameters that characterize E, giving a process model. The third level is the parameter models for the P_E, with parameters (or 'hyperparameters') of their own. Each of the levels can be further decomposed into components. Spatial (or temporal) structure can be included by relaxing the assumptions of independence.

Level 1. Data: independent site counts, Y_{ij} for site i and time j, with negative binomial parameters λ_{ij} and κ. This gives, with square brackets indicating the distribution:

$$[D|E,P_D] = [\{Y_{ij}\}|\{\lambda_{ij}\},\kappa]. \tag{12.33}$$

Level 2. Process: giving the log counts a normal distribution $\log(\lambda_{ij}) = \mu_{ij} + \varepsilon_{ij}$, with the mean related to temporally dependent site factors by the vector of weights, $\boldsymbol{\theta}$, and the error terms being normal i.i.d. (independent and identically distributed random variables) with variance σ^2, we get:

$$[E|P_E] = [\{\lambda_{ij}\}|\{\theta\},\sigma^2]. \tag{12.34}$$

Level 3. The parameter model for $\mu_{ij} = \theta_{0i} + \theta_{1i}\,year + \theta_{2i}\,date + \ldots$

The times and sites are initially treated as independent but dependence can be introduced by having the outcome conditional upon the same true underlying process (Cressie *et al.* 2009). Thus, if the data are decomposed into three different data sets (e.g. three different observation days), we can compare having three potentially different parameter sets, $P_{D,1}$, $P_{D,2}$ and $P_{D,3}$ with the condition that they are actually all equal: $P_{D,1} = P_{D,2} = P_{D,3} = P_D$. Similarly, the ecological processes can be decomposed into different numbers of components, which might be site-specific in their composition.

A Level 4 of the hierarchy is certainly possible, if the elements of θ, which are the parameters of the linear dependence model, have their own distributions.

Example 3.
The spread of the House Finch in Eastern USA (Wikle 2003), with counts of birds in geographically defined squares over a number of census years.

This example follows the structure of (1) data model; (2) process model; and (3) parameter model.

Level 1. $X_t(s_i)$ is the count of birds at time t at location s_i. X has a Poisson distribution with parameter $\lambda_t(s_i)$.

Level 2. $\text{Log}(\lambda_t(s_i)) = \mu_t + k'_{it}\,u_i\,\varepsilon_{ij}$, where u is a vector of a gridded latent spatio-temporal dynamic process, and k' maps u onto the spatial locations. Year-to-year autocorrelation is provided by $\mu_t = \mu_{t-1} + \eta_t$.

The vector u provides a kind of reaction–diffusion model, creating a Spatio-Temporal Autoregressive (STAR) model with diffusion parameter δ and error term γ.

Level 3. A model for parameter δ that includes dependence on the correlation matrix of sites, with a spatial dependence parameter like the familiar autocorrelation parameter ρ.

Space is included explicitly and by necessity in this analysis.

Example 4
Using non-homogeneous species abundance data to reconstruct interaction networks (Aderhold *et al.* 2012). These networks represent the full set of interactions among the species at a particular site, including competition, predation, parasitism, mutualism, and facilitation. The nodes of the network (a directed

graph) are the component species and the edges between pairs of nodes represent specific interactions (e.g. species A is a parasite of species B). These interactions are detected in a spatially explicit context, and can be analysed using the hierarchical approach.

Level 1. Abundances are observed for N species on a grid of locations, s_i, with coordinates (x_1, x_2), giving $\{Y_k(s_i)\}$ for $k =$ to N. The network of interactions is a directed graph with the species as nodes and a directed edge from each member of the 'parent set' of k, called π_k. There are change points in each of the two directions dimensions, called $\{\xi_1\}$ and $\{\xi_2\}$.

Level 2. Latent effects in close proximity are usually similar, but there may be several change points along each of the two orthogonal directions. The abundance of any species depends on the abundances of its parent set, with a set of parameters determining a linear response. For each segment h, there is a segment-specific linear regression:

$$Y_k(s_i) = b_{k0}(h) + \sum_{m \in \pi_k} b_{km}(h) Y_m(s_i) + \varepsilon(s_i). \tag{12.35}$$

The latent variable, h, depends on the location and on the locations of the change points.

Level 3. Spatial autocorrelation in the response variable, the density of the focal species, is explicitly included, based on the values at the four nearest neighbours weighted by the inverse of the Euclidean neighbour distance:

$$A_k(s_i) = \sum_j d_{ij}^{-1} Y_k(s_i) \div \sum_j d_{ij}^{-1}, \tag{12.36}$$

where s_j is in the neighbour set of s_i, currently the four nearest sample locations. Then,

$$\begin{aligned} Y_k(s_i) = & b_{k0}(h) + \sum_{m \in \pi_k} b_{km}(h) Y_m(s_i) \\ & + b_{kA}(h) A_k(s_i) + \varepsilon(s_i). \end{aligned} \tag{12.37}$$

Notice that the effect of the autocorrelation can vary in strength depending on location, indexed by variable h. The details of this approach, and its working out, are provided in the paper cited. This example has a somewhat complex set-up, but it shows the flexibility of the hierarchical approach in including both change points and spatial autocorrelation.

12.6.2.6 Other examples

We have already mentioned two further examples of this approach in Chapter 8. Aing *et al.* (2011) used a Bayesian hierarchical occupancy model for assessing the occurrence of river otter from snow-track data. The model had three levels: occupancy by otters; availability of tracks for detection, conditional upon occupancy; and records of track's presence or absence, conditional upon their availability. Spatial dependence was included with an intrinsic Conditional Autoregressive (CAR) model and the MCMC procedure. They found that a spatial model gave better estimates for the detection parameters and better credibility intervals for the spatially autocorrelated data.

Beguin *et al.* (2012) introduced an alternative to MCMC procedure, Integrated Nested Laplace Approximations (INLA), for fitting Bayesian hierarchical spatial models with fairly general covariance structures. They used both the CAR model and the Matérn model (see Minasny & McBratney 2005) in a study investigating woodland caribou in Eastern Canada's boreal forest, and they found that the INLA method and the Matérn model provided a number of advantages over the MCMC approach, being both accurate and rapid, and effectively removing spatial autocorrelation from the model residuals, and giving a good evaluation of the uncertainty of the parameter estimates.

12.6.2.7 Comment on hierarchical Bayesian spatial analysis

We have included a fairly lengthy treatment of the hierarchical Bayesian approach to analysis because of its applicability to spatial analysis in many contexts, and because of its flexibility in the techniques used to include spatial structure in ecological analysis. Harrison *et al.* (2011) presented a sophisticated example of butterfly metapopulation dynamics in which spatial structure is included implicitly in the

dispersal between habitat patches, as affected by inter-patch distances. Gelfand (2012) provided some further insights and discussion that focus on the use of hierarchical models for examining explicitly spatial data. The points made by Gelfand (2012) are that the approach is both useful and very flexible, and that the approach will prove its worth even more convincingly as more research includes spatially referenced data.

12.6.3 Hypothesis testing with spatio-temporal graphs

For hypothesis testing, we can describe the use of spatio-temporal graphs as a framework to deal with hierarchical hypotheses and for variations on that approach. Earlier in Chapter 3, we described the use of a series of spatial graphs on the same set of nodes (mutually nearest neighbours, nearest neighbours, Minimum Spanning Tree, etc.; see Table 3.1), to test a hierarchy of hypotheses concerning which neighbours are the most influential for a given focal node. The choice of rules for determining the edges that link pairs of a given set of nodes in a spatial graph allows a very broad range of connectivity in the graph. These vary from sparse (e.g. only mutually nearest neighbour pairs, or a very short distance threshold) to dense and highly connected (e.g. a large distance threshold or a defined complete graph with edges between all possible pairs). This range, and the choice of rules, allows a very precise form of hierarchical hypothesis testing.

The basic approach is to compare the results of a test or measurement based on a less restrictive 'relative null hypothesis' to those based on a more restrictive 'relative alternate hypothesis' for each step in the series of graph definitions. For example, in the strictly spatial context, the Delaunay graph contains all the nearest neighbour graph edges, and many more; if the sparse graph gives an improved prediction of the characteristics associated with individual nodes as a function of the neighbour properties, then we will prefer the more restrictive hypothesis that the closest neighbours are the most important (compared to all first-order neighbours in the Delaunay graph). More fully, to test the effects of various kinds of neighbours, we can compare the null hypothesis as determined by the complete graph, which is equivalent to the statement that all the other nodes have equal influence on the characteristics of the focal node, with the alternative hypothesis depicted by the graph of first-order neighbours in a Delaunay triangulation, equivalent to the statement that those neighbours provide the most important explanatory factors. A second alternative hypothesis is that only the single nearest neighbour of any node is a good predictor of the outcome.

This concept can be reduced to a schematic form as:

$\{H_0 = \text{complete graph}\} \equiv$ all nodes influence the focal node equally,

$\{H_1 = \text{Delaunay graph}\} \equiv$ only those first-order neighbours influence the focal node,

$\{H_2 = \text{nearest neighbour graph}\} \equiv$ only the single closest node affects the focal node.

Spatio-temporal graphs provide the same opportunity for the development and testing of hierarchical hypotheses and variants that are sophisticated and precise although not necessarily strictly hierarchical. The structured sets of hypotheses can be within time or within space alone, or may include the interaction of time and space (time $\times$ space). In the last case, there is a secondary structure in the interaction hypotheses that is based on the amount and nature of the dependence between the temporal and spatial structures. Consider a widespread population of plants for which the phenological sequences of the appearance of leaves and flowers are observed and recorded over a number of years at a range of sites. The similarity of the timing of events for any one plant can then be compared with others based on a hierarchy of neighbours in both space and time.

$\{H_0 = \text{complete ST graph}\} \equiv$ no structure.

All possible edges in time and space, the complete spatio-temporal graph: this is equivalent to the hypothesis that similarity is independent of time lags and geographic distances.

$\{H_1 = \text{graph of close neighbours in time and space}\}$ $\equiv$ 'regional' structure.

Many edges between first-order neighbours in space and time: this is equivalent to the hypothesis that similarity is broadly dependent on nearness in time and geographic distance. The spatial structure can be chosen from several graphs with different edge densities (MST, Gabriel graph, and so on).

$\{H_2 =$ graph of only the nearest neighbours in time and space$\} \equiv$ 'local' structure.

Only edges between closest neighbours in space and time: this is equivalent to the hypothesis that similarity is found only at the shortest time lags and shortest spatial distances.

Being spatio-temporal, however, there are other alternatives that can be tested.

$\{H_S =$ graph with only spatial edges$\} \equiv$ Only space matters; there is no history.

Edges occur between spatial neighbours, but not between instances of the same individual at different times. The control of variation is based on location, regional, or neighbourhood effects; there are no individual effects, and no history.

$\{H_T =$ graph with only temporal edges$\} \equiv$ only time matters; there are no spatial effects.

Only temporal edges between instances of the same entity at different times are used, possibly with several time lags (t versus $t-1$, $t-2$,...). The variation has no spatial component, no neighbour or locality effects, but persists from one time period to the next: pure history.

These general schemes for examples should convey some of the possible precision and flexibility that this approach affords the researcher through testing in well-defined hierarchies of hypotheses. The added advantage is that there will often be a simple visual presentation of the relationships among the various hypotheses in the hierarchy.

12.7 Other future directions

Throughout this book, we have ended many discussions with the comment that the area is one for future research and further investigation for ecological applications. This was true of fibre analysis, autocorrelation correction and spatial graphs to name just three. Clearly there is much to be done in almost every area of our chapters. Let us emphasize a few.

- Spatial graphs: this is a very promising area for research and application for ecological studies, with lots to do and lots to learn. They can be used to test hypotheses and to study spatial dynamic behaviour of processes.
- Spatio-temporal graphs: here too, there are a lot of good questions and methodologies to be understood and absorbed (Paley *et al.* 2007; Del Mondo *et al.* 2010), and then used in ecological research; and beyond these graphs, there is also lots more in spatio-temporal analysis generally (see Cressie & Wikle 2011).
- Fibre analysis and related areas of study can go beyond the treatment of zero-dimensional point events to include the analysis of linear objects, whether straight or curving, simple or branched, to analyse a wide range of spatial structures of interest to ecologists. There is a broad and substantial literature on the analysis and modelling of point patterns and point processes, on which developments for linear objects can be based. There is a wide range of ecological hypotheses that can be tested with the approach, and the linear events can also be combined with point events in a number of ways for even greater scope and flexibility.
- Spatial analysis as applied to biological diversity is another area that deserves much more attention and development. A multiscale spatially explicit (and temporally indexed) approach should answer many questions about the structure of diversity and lead to a much better understanding of its causes and response to a range of factors.
- Spatial structure is more than spatial autocorrelation and more than either exploratory analysis or the application of spatial statistics on their own. While it may be easy to incorporate spatial thinking in our conceptual models of how ecological systems work, incorporating spatial structure into our ecological studies can be very difficult and prone to complexities, not to mention the potential to be very expensive in many different resources,

including time, money, and space itself! This is the real challenge for future work: applying the knowledge, techniques and understanding about spatial analysis that we have built up over the past decades.

- For all of the above, there is an opportunity to exploit a hierarchical approach to testing a nested series of null and alternative hypotheses, but the details have yet to worked out based on what is of greatest ecological interest and importance.

Despite the fact that we have been selective rather than all-inclusive in the topics covered in this book, it is far from short! Maybe this is the last work in 'book' format of this kind we will do, as such a rapidly growing field becomes more and more difficult to summarize in a sufficiently detailed but succinct manner, and as it becomes easier and easier to fall behind as the number of relevant studies grows at an ever-increasing rate.