8

# Spatial autocorrelation and inferential tests

## Introduction

Parametric statistics are based on the assumption of independence of the errors given the statistical model to be fitted, such as linear regression. When performing parametric tests with ecological data, the assumption of independence of the errors is often violated, increasing the Type I error and biasing the estimation of regression parameters (Diniz-Filho *et al.* 2003; Dormann *et al.* 2007; Bini *et al.* 2009; Diniz-Filho *et al.* 2009; Beale *et al.* 2010). The lack of independence in the errors (see also Chapter 12) can arise because objects (samples, observations, etc.) that are closer together sometimes have a tendency to be more similar than those that are further apart (see Chapters 1, 2 and 6). This lack of independence can be due to spatial, temporal and phylogenetic processes (Ives & Zhu 2006; Peres-Neto 2006). Here we focus only on spatial dependency that generates spatially autocorrelated data. One source of spatial autocorrelation in the data can be causal interactions within the measured variable itself; for example, in studying species distribution and abundance, the abundance of a single species may be spatially autocorrelated because of constraints on the organisms' mobility and dispersal. This kind of autocorrelation is sometimes called 'true autocorrelation', but we suggest it would be better referred to as 'inherent autocorrelation' (Chapters 1 and 2). This descriptor is to distinguish this phenomenon from 'induced spatial dependence' where the observed variable (e.g. species abundance) has a functional dependence on an underlying variable (e.g. soil moisture or nutrient content) that is itself autocorrelated (cf. Legendre *et al.* 2002). It may not be easy to distinguish between the two in ecological studies, and both may occur together in a single example (Wagner & Fortin 2005). In fact, for many types of ecological study, it is common for both the biological response variable and an underlying environmental factor to be spatially autocorrelated. The presence of spatial autocorrelation in the dependent variable (hereafter the response variable, $y$), in the independent variables (hereafter covariate or predictor variables, $x$), or in both types of variables in an inferential framework, will affect both the parameters' estimation accuracy and the power of the significance test. Yet, while the presence of spatial autocorrelation in the response variable can be accounted for by the relationship with the covariate variables, spatially autocorrelated covariates' variables are still going to affect the detection of the true effect of the predictors (Bini *et al.* 2009; Beale *et al.* 2010).

As with most ecological phenomena, autocorrelation is scale- or distance-dependent; the values of a particular variable may be more similar than expected at short distances (positive autocorrelation) but less similar than expected at greater distances (negative autocorrelation). Spatial autocorrelation can result in various spatial patterns (e.g. patchiness, gradients) which can occur over a wide range of spatial and temporal scales. By having spatial structure(s) we can estimate the corresponding spatial scale(s) of the process (es) that generated the pattern in the first place as presented in Chapter 7. Hence, determining the degree of spatial autocorrelation and spatial scale of ecological variables is primary to relating the ecological response

to the covariates, because autocorrelation can be viewed as a nuisance that causes trouble for statistical testing. Ways in which spatial autocorrelation is a statistical nuisance include: (1) with spatial autocorrelated variables (response as covariate) their variance is smaller than with uncorrelated data; and (2) then each observation (sample) does not provide a full degree of freedom but only some part that is inversely proportional to the strength of autocorrelation in the data. Note that the spatial autocorrelation values of less than 0.2 may be considered weak and often not significant. Hence it may be that when the data have very weak degree (e.g. less than 0.1) of spatial autocorrelation the inferential tests are robust enough to deal with the spatial structure of the data, but that is not guaranteed.

Spatial patterns can provide insights about the underlying ecological processes that generate them (Chapters 1 and 7). For instance, past spatial pattern of vegetation and topography can influence the type and intensity of current spatial patterns. This is known as the *spatial legacy* of past spatial structure on current spatial pattern (Chapters 1 and 7; James *et al*. 2007), which can be considered as a kind of ecological memory (see Chapter 11). Furthermore, the surrounding locations can affect the spatial pattern at a given location *i*. This is known as *spatial contingency* where both response and predictor variables can influence the spatial structure of the response variable. Because of all these various effects of space, the spatial patterns in the data are often used as a surrogate for unsampled predictor variables (Chapters 1 and 7; Peres-Neto & Legendre 2010). While this use of spatial patterns as predictor variables increases the percentage of variance explained, it does not provide predictive value because the fitted regression model will be valid only for the studied area. Therefore 'space' cannot explain ecological variability and should be used only as a last resort because it is not transferrable to other locations and conditions. To have transferability we should add environmental, landscape, ecological and phylogenetic data as best we can as predictors in a regression model.

The aim of this chapter is to present ways in which to deal with the significant and strong degree of spatial autocorrelation while using inferential tests. This chapter differs from much of the rest of the book, because of the importance of models and their relationship to underlying processes. As models are essential to understanding the underlying processes that operate in structuring spatial patterns, it is important to understand how the spatial structure affects the various models commonly used in ecology. First, we introduce how autocorrelation models derived from time series analysis can be modified for use with one-dimensional autocorrelated data. Second, we describe a series of solutions to account for the degree of spatial autocorrelation in parametric inferential models. Third, we indicate how restricted randomization procedure can help in testing the significance of spatially autocorrelated data. Fourth, we synthesize how the various spatial regression models account for the spatial structure in the data. Finally, we describe how spatial autocorrelation needs to be accounted for in sampling and experimental designs.

## 8.1 Models dealing with one-dimensional autocorrelated data

Before we can describe more general spatial regression models (Section 8.4), we need to present the simplest models for one-dimensional data (spatial or temporal). Historically, models dealing with serial correlation were developed first for time series and then modified for spatial series. The first models apply to observations of the dependent variable $y$ at a series of $n$ locations (e.g. $y_1, y_2, \ldots, y_i, \ldots, y_n$) as follows.

- Complete independence model (hereafter referred to as Model CI):

$$\begin{cases} y_i = \varepsilon_i; \\ \varepsilon_i \approx N\left(0, \sigma_\varepsilon^2\right). \end{cases} \tag{8.1a}$$

Here, $\varepsilon_i$ is an independent 'error' term following some statistical distribution, such as a normal distribution with a mean of zero and variance $\sigma_\varepsilon^2$. Figure 8.1*a* illustrates the fact that the expected spatial autocorrelation between observations is 0. The term $\varepsilon_i$ is usually assumed to follow a normal distribution, but that is not necessary for independence, and other distributions might be considered.

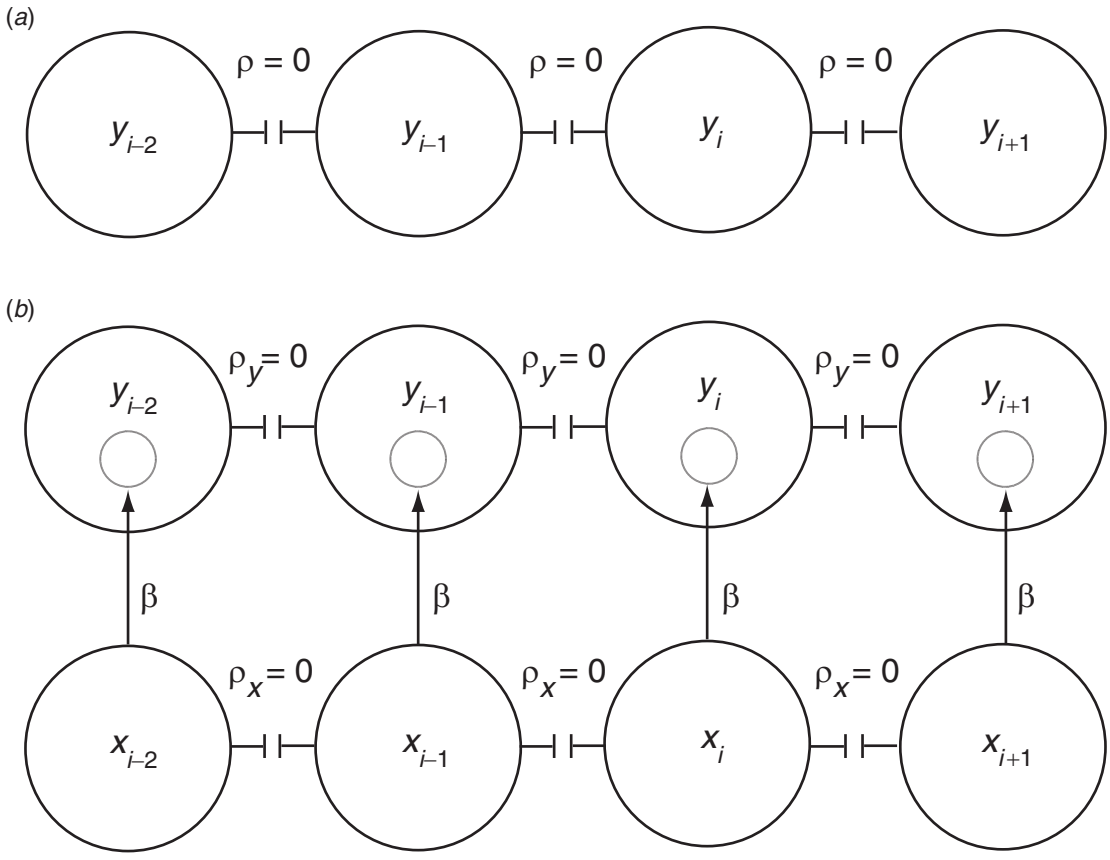



**Figure 8.1** (*a*) Model CI: the series of values is independent of each other and so the expected correlation between adjacent values is 0 (see Eq. (8.1*a*)). (*b*) Model SI: the series of *y* values is dependent on the *x*-series, which are independent of each other. The *y* values remain spatially independent and so the expected correlation between adjacent values is 0 (see Eq. (8.1*b*)).

An alternative model of independence can include functional dependence, but retain spatial independence.

- Spatial independence model (hereafter referred to as Model SI):

$$\begin{cases} y_i = \beta x_i + \varepsilon_i, \\ x_i = \xi_i. \end{cases} \tag{8.1b}$$

Both $\varepsilon_i$ and $\xi_i$ are independent 'error' terms following some statistical distribution, and $\beta$ is the linear regression parameter. The expected correlation between adjacent values of $y$ and between adjacent values of $x$ is 0, as illustrated in Figure 8.1*b*.

Then the spatial structure of the response variable can affect its own spatial pattern.

- Inherent autoregression model (hereafter referred to as Model IH):

$$y_i = \rho y_{i-1} + \varepsilon_i \text{ with } -1 \leq \rho \leq +1. \tag{8.2}$$

Here, $\varepsilon_i$ is an error term as above, and $\rho$ is the autocorrelation parameter that determines the strength of the autocorrelation. Figure 8.2 illustrates the model; the expected correlation between adjacent values is $\rho$.

The spatial structure of the response variable can also be affected by the spatial pattern of the predictor variables.

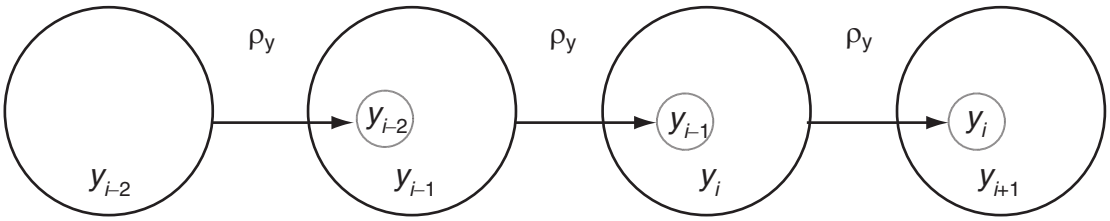



**Figure 8.2** Model IH: the values in the series are not independent, with each being directly dependent on its predecessor and thus indirectly dependent on all preceding values. The expected correlation between adjacent values cannot be 0, but $\rho$ (see Eq. (8.2)).

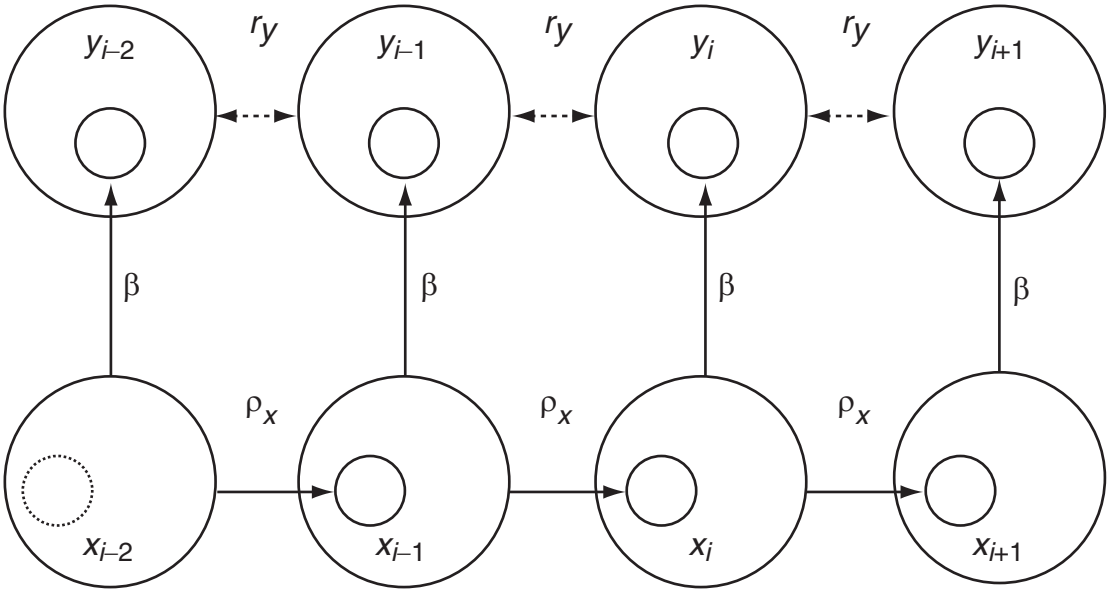



**Figure 8.3** Model ID: the values in the $y$-series are not independent, but their observed autocorrelation, $r_y$, is induced by their linear dependence on the $x$-series, which is governed by a Model IH process (see Eq. (8.3)).

- Induced autoregression model (hereafter referred to as Model ID):

$$\begin{cases} y_i = \beta x_i + \varepsilon_i \\ x_i = \rho x_{i-1} + \xi_i \end{cases} \quad \xi_i \approx N\left(0, \sigma_\xi^2\right). \tag{8.3}$$

In this model, both $\varepsilon_i$ and $\xi_i$ are normally distributed error terms, $\beta$ is the usual regression parameter and $\rho$ is again the autocorrelation parameter. The correlation between adjacent values of $y$ is not expected to be 0, but it is a function of the values of $\beta$ and $\rho$, as illustrated in Figure 8.3. Again, the second error term, $\xi_i$, need not follow a Normal distribution, but many discussions will assume that it does.

Finally, the spatial structure of the response variable can be due to both the spatial pattern of the response and of the predictor variables.

- Double autoregression model (hereafter referred to as Model IH-ID):

$$\begin{cases} y_i = \beta x_i + \rho_y y_{i-1} + \varepsilon_i \\ x_i = \rho_x x_{i-1} + \xi_i \end{cases}. \tag{8.4}$$

The symbols are as in the previous models, but there are now two autocorrelation parameters, one for $y$ and

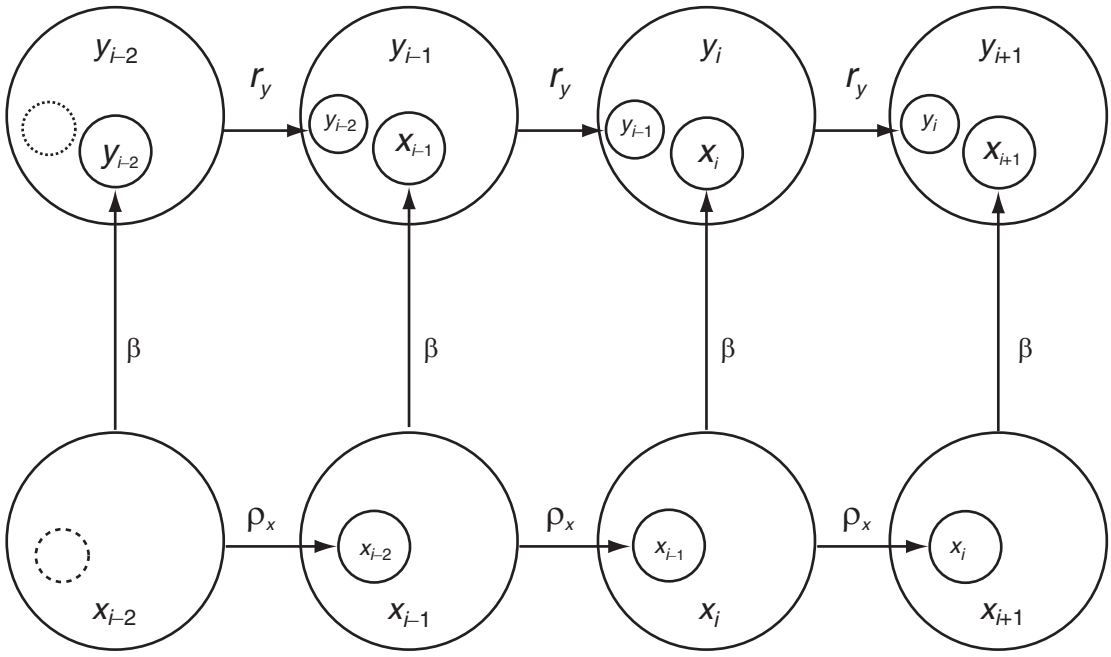



**Figure 8.4** Model IH-ID: correlation of $y$ values comes from two sources, inherent in $y$ itself and induced by dependence on $x$, giving a doubly autoregressive model.

one for $x$. This more complicated model is referred to as doubly autoregressive because it includes both inherent and induced forms of spatial dependence; it is illustrated in Figure 8.4.

In Models CI and SI, the values of $x$ are independent of each other and Model IH gives rise to what is called a first-order autoregressive structure, which is probably the most frequently studied form of modelling autocorrelation. In that case, the correlation of any two variates, $y_i$ and $y_j$, depends only on the separation between them (here the separation distance can be measured by the number of intervening steps):

$$\mathrm{Cor}(y_i, y_j) = \rho^{|i-j|}. \tag{8.5}$$

For example, it is easy to show, by substituting for $y_{i-1}$ in Eq. (8.2), with the second version of that equation, that

$$y_i = \rho(\rho y_{i-2} + \varepsilon_{i-1}) + \varepsilon_i = \rho^2 y_{i-2} + \rho\varepsilon_{i-1} + \varepsilon_i, \tag{8.6}$$

so that $\mathrm{Cor}(y_i, y_{i-2}) = \rho^2$. The first-order autoregressive structure discussed here is the simplest of a more general $k$th-order autoregressive model (Cressie 1991):

$$y_i = \sum_{j=1}^{k} \rho_k \, y_{i-k}. \tag{8.7}$$

In Model ID, the dependence observed in $y$ is induced by its linear dependence on $x$ and $x$'s inherent autoregressive structure. With more effort, we can show that the correlation produced by this model is of a similar form:

$$\mathrm{Cor}(y_i, y_j) \propto \rho^{|i-j|} \quad \text{for } i \neq j. \tag{8.8}$$

The $n \times n$ variance–covariance matrix for the $n$ $y$s is expected to be as shown in Table 8.1 for Model IH and as in Table 8.2 for Model ID.

The value of the variance is different for the two models, being proportional to $\sigma_\varepsilon$ for Model IH and to $\sigma_\xi$ for Model ID, but the basic structure is the same. If both these variances are 1, for example, with $\rho = 0.4$ and $\beta = 0.6$, the expected correlation of adjacent observations for Model ID is 0.12, considerably less than the 0.48 in Model IH. The covariance structure for Model IH-ID (doubly autoregressive) is, not unexpectedly,

**Table 8.1** Matrix of covariance among positions for a first-order autoregressive inherent correlation structure

$$\begin{vmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & & \rho^{n-3} \\ \rho^3 & \rho^2 & \rho & 1 & & \rho^{n-4} \\ \vdots & & & & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{vmatrix} \sigma_\varepsilon^2/(1-\rho^2)$$

**Table 8.2** Matrix of covariance among positions for a first-order autoregressive induced correlation structure. The constant is $\kappa = \beta^2\rho/\left(\beta^2\sigma_\xi^2 + (1-\rho^2)\sigma_\varepsilon^2\right)$

$$\begin{vmatrix} 1 & \kappa\rho & \kappa\rho^2 & \kappa\rho^3 & \cdots & \kappa\rho^{n-1} \\ \kappa\rho & 1 & \kappa\rho & \kappa\rho^2 & \cdots & \kappa\rho^{n-2} \\ \kappa\rho^2 & \kappa\rho & 1 & \kappa\rho & & \kappa\rho^{n-3} \\ \kappa\rho^3 & \kappa\rho^2 & \kappa\rho & 1 & & \kappa\rho^{n-4} \\ \vdots & & & & \ddots & \vdots \\ \kappa\rho^{n-1} & \kappa\rho^{n-2} & \kappa\rho^{n-3} & \kappa\rho^{n-4} & \cdots & 1 \end{vmatrix} \sigma_\xi^2$$

considerably more complicated, but under the same parameter values ($\beta = 0.6$; $\rho_y = \rho_x = 0.4$), the expected correlation of adjacent values is approximately 0.5.

The models discussed so far seem to have been formulated for temporal rather than spatial series. This impression results from the fact that the models make explicit the dependence of $y_i$ on $y_{i-1}$ (Model IH) and thus on all preceding values of $y$. This may seem inappropriate for spatial data, even in only one dimension, where we would expect the dependence to be equal in both directions. Remember, however, that in this apparently directional model, the correlation between any two variates depends only on the distance between them, not on the direction (Eq. (8.5)). Consider the following thought experiment: we provide a data series generated by Model IH, and a second series which is the first series reversed; what criteria could be used to distinguish the original? In fact, Cressie (1991) commented that while the autocorrelation structure given by Eq. (8.5) can be generated by Model IH, it can also be associated with spatial data not generated in that way (Cressie 1991, p. 14). For example, if the amount of pollen dispersed by a single isolated tree declines with distance, $d$, such that the amount is proportional to $(1/2)^d$, measurements every metre might resemble the output of Model IH, but Model IH might not be a good description of how the observed pattern arises. Therefore, we can continue to use these models to provide insight into the characteristics of spatial dependence, for which we may not know the underlying process that generates the correlation structure we detect.

If we only have a single spatial data set, such as one series of $n$ observations of a variable recorded for a single time, it is like having only a single realization of an underlying model or process and we cannot calculate the covariance of two individual values, such as Cov($y_3$, $y_5$), for example. We can, however, calculate an observed covariance for all values separated by distance 1, by distance 2, and so on, as estimates of the underlying covariances, and that is what a covariogram or correlogram does (see Chapter 6). The details of the calculation are as follows. The estimated covariance for lag $d$ is:

$$C_y(d) = \frac{\sum_{i=1}^{n-d} y_i\, y_{i+d} - \sum_{i=1}^{n-d} y_i \sum_{j=d+1}^{n} y_j/n}{n-d-1} = \frac{\sum_{i=1}^{n-d} (y_i - \overline{y}_{(1,n-d)})(y_{i+d} - \overline{y}_{(d+1,n)})}{n-d-1}, \tag{8.9}$$

where the sample means used in the second part of the equation are for subsets of the data based on the first and the last $n - d$ values in the series, which we will refer to as the 'regional' means. For example, in Figure 8.5, for $n = 12$ and a lag of 3, the sample means of $y_4$ to $y_{12}$ and of $y_1$ to $y_9$ would be used. It is tempting to calculate the correlation for lag $d$ using the regional sample variances (for example, of $y_4$ to $y_{12}$ and of $y_1$ to $y_9$ again in Figure 8.5), as we would for any usual calculation of a correlation coefficient:

$$r_y^*(d) = \frac{C_y(d)}{\sqrt{s^2_{y(1,n-d)}\, s^2_{y(1+d,n)}}}, \tag{8.10}$$

**Table 8.3** Results of 10 000 simulations for Models CI to IH-ID ($\rho = 0.4$; $\beta = 0.6$)

| Critical value (%)[a] | 0.05 | 0.5 | 2.5 | 5 | 95 | 97.5 | 99.5 | 99.95 |
|---|---|---|---|---|---|---|---|---|
| Expected count | 5 | 50 | 250 | 500 | 500 | 250 | 50 | 5 |
| Model CI | 4 | 43 | 227 | 472 | 531 | 262 | 66 | 3 |
| Model SI | 6 | 54 | 272 | 628 | 442 | 198 | 42 | 6 |
| Model IH | 116 | 419 | 913 | 1317 | 1575 | 1086 | 490 | 126 |
| Model ID ($y$) | 32 | 181 | 497 | 857 | 617 | 369 | 124 | 24 |
| Model IH-ID ($y$) | 309 | 648 | 1252 | 1600 | 1635 | 1210 | 624 | 237 |
| Model ID or IH-ID ($y'$)(residuals) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

[a] The values tabulated are the number of trials in which the test statistic was more extreme than the critical value associated with the probability given by the column heading. The null hypothesis (of mean 0) is true and, therefore, these rates represent Type I error.

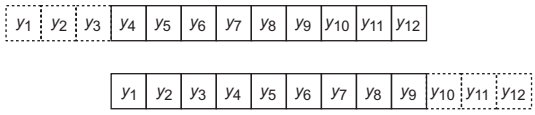



**Figure 8.5** The calculation of autocovariance or autocorrelation for $n = 12$ and lag $d = 3$: the 'regional' sample means are calculated from $y_4$ to $y_{12}$ and from $y_1$ to $y_9$. Using the regional values of sample variance results in poor estimates of the autocovariance or autocorrelation.

but this leads to poor estimates (Legendre & Legendre 2012; Jenkins & Watts 1968). A better estimate is:

$$r_y(d) = \frac{C_y(d)}{s^2_{y(1,n)}} = \frac{C_y(d)}{C_y(0)}. \tag{8.11}$$

This is the formulation for calculating autocorrelation that we will use in investigating some of the characteristics of a range of autocorrelation models, including those already described.

The discussion so far has focused on data in one spatial dimension, but many of the comments apply equally well to data in two or more dimensions. For example, in two dimensions, we usually estimate the covariance using distance classes, as described above, but in some cases, where anisotropy is a concern (see Chapters 1 and 6), direction classes may be used as well. In two or more dimensions, however, although the concepts related to spatial autocorrelation are the same as those in one dimension, the technical aspects of setting up models of spatial autocorrelation in additional dimensions tend to be more complicated. For example, Whittle (1954) showed that in one dimension, bilateral dependence (where $y_i$ depends on both $y_{i-1}$ and $y_{i+1}$) can be reduced to unilateral dependence, which is more tractable, but in two dimensions, a similar reduction is not easily achieved.

To investigate further the characteristics of models of autocorrelation in one dimension, we need to generate not one, but a large number of realizations of these models on a computer, using $n = 500$ and $\rho = 0.4$. For each realization, we calculate the sample mean, $\overline{y}$, and sample variance, $s^2$, and then the $t$ statistic to test the null hypothesis that the true mean of $y$ is zero: $H_0 \equiv (\mu_y = 0)$:

$$t = \frac{\overline{y} - 0}{s/n}. \tag{8.12}$$

For 10 000 realizations of each model, we count the number of times the test statistic is less than the 0.05%, 0.5%, 2.5% and 5% critical values of $t_{n-1}$ and the number of times it is greater than the 95%, 97.5%, 99.5% and 99.95% critical values. For Models ID and IH-ID, which include induced autocorrelation, we can examine the results both for the original variable $y$ and for its residual after the linear dependence on $x$ is removed: $y' = y - (a + bx)$, where $a$ and $b$ are estimated from the data using the standard linear regression techniques. Table 8.3 shows some typical results. For the independence models (Model CI or SI), the rates are close to the expected counts (as they should be), but for Model IH and Model ID ($y$), the rates are much higher. This comparison illustrates the effect of positive spatial autocorrelation on standard

statistical tests: they become too liberal, producing more apparently significant results than the data actually justify (Cliff & Ord 1981, among many). This effect is one of the main topics of this chapter and the subject of much of the discussion that follows, but we should interject a comment about the last line of Table 8.3. Given that in Model ID ($y$), the autocorrelation appears in $y$ because of its linear dependence on $x$, which has inherent autocorrelation, a reasonable prediction would be that removing that dependence would just remove the induced autocorrelation and its effect from $y$, so that the last line should resemble the first. The fact that it does not may be a bit surprising, and we will return to the topic of removing linear (and other) dependence later in this chapter. The concept is closely related to the process used in time series analysis called 'pre-whitening' in which trends in the data are removed, supposedly leaving only pure error or 'white noise'.

The main message of Table 8.3, putting aside the last line of interesting zeros, is that positive spatial autocorrelation, whether inherent or induced, produces 'too many' significant results, and a lot too many. For a two-sided test with $\alpha = 0.05$, as is often used in ecological studies (i.e. using the 2.5% and 97.5% critical values), Model IH gives almost 2000 apparently significant results, four times as many as the nominal rate of 500 in 10 000 trials. For Model ID ($y$), with the same value of $\rho$, it is at least twice the nominal rate, and for Model IH-ID ($y$), it is about eight times the nominal rate. Clearly, the magnitude of the effect could lead to serious errors in decision-making based on the test results. An intuitive understanding of this effect can be based on the fact that because the $n$ observations are not completely independent of each other, we do not get a full $n$ units of information, but something less. The 'effective sample size' is the equivalent number of independent observations that would provide the same amount of information as $n$ non-independent observations. Here the effective sample size, $n'$, is less than $n$ and so when we use $s/n$ in the divisor of the $t$ statistic, rather than $s/n'$, we are dividing $s$ by a number that is larger than it deserves to be, thus underestimating the variance of the mean. This then produces a test statistic too great in magnitude. In going through the literature on this topic, we find many articles that illustrate the problem, but few that identify 'appropriate remedial action' (cf. Haining 1991). There are, however, several approaches that might be taken to deal with this problem, ranging from some simple 'quick fixes', which may be misleading, to sophisticated solutions that depend on a particular underlying model, to a few robust techniques of more general applicability.

We will not keep the reader who is intrigued by the line of zeros at the bottom of the Table 8.3 in suspense (although they are pertinent to a later discussion). The zeros arise because by using $y' = y - (a + bx)$, with $a$ and $b$ estimated from the data, the observed mean of $y'$ is forced to be zero, causing the test statistic also to be zero in every case, and thus never outside the critical values. The topic of the effect of removing the dependence of one variable on another will return in another context.

In the next section, we proceed with our discussion of the main problem for inferential models, the fact that spatial autocorrelation changes the rates at which statistical tests detect significant results.

## 8.2 Dealing with spatial autocorrelation in inferential models

### 8.2.1 Simple adjustments

The simplest approach might be to acknowledge the existence of spatial autocorrelation and to adjust the Type I error rate, $\alpha$, to a more conservative value: such as $\alpha = \alpha/5$. For example, in Dale & Zbigniewicz (1997), $t$-tests were carried out on plant density data from transects of 1001 contiguous $10 \times 10$ cm quadrats. They used a 1% significance level, rather than the usual 5% because the critical value for 1% for large $v$ is 2.57 which is close to the 5% critical value for $v = 5$. Therefore, even if the autocorrelation in the data reduced the effective sample size by an order of magnitude or more, from 1001 to as low as 5, an $\alpha$ value of less than 0.05 was assured. This may not be the best approach, because we do not know by how much the nominal error rate needs to be adjusted to give a true error rate of the desired

value. Depending on the true autocorrelation structure underlying the data, there is a real danger of using a test that is now much too conservative.

One feature of the commonly invoked first-order autoregressive structure (Model IH) is that autocorrelation declines exponentially with distance. This suggests that for distances greater than some particular value, the autocorrelation is effectively zero and observations further apart can be treated as independent. If this is true, and there is a great abundance of data, it should be possible to use a widely spaced subset of the data to ensure independence. For example, Ostendorf & Reynolds (1998), in an analysis of landscape patterns in two dimensions, determined that autocorrelation did not extend beyond 20 pixels in their data, and therefore used what they considered to be a non-autocorrelated subset of pixels 20 units apart, 1/400 of the pixels available. This approach has two major draw-backs: first, it seems very wasteful of data (Legendre & Legendre 2012) and second, the concept of a 'distance to independence' may be mistaken for real spatial data, where non-zero autocorrelation may have an effect even if it is not significantly different from zero. We will elaborate in the next section.

### 8.2.2 Adjusting the effective sample size

As we described above, autocorrelation in data modifies the effective sample size to be something other than the number of samples, $n' \neq n$. Positive autocorrelation reduces the effective sample size, and the estimated degree of spatial autocorrelation can therefore be used to determine how much smaller the effective sample size is than the number of observations (Clifford *et al.* 1989; Cressie 1991; Dutilleul 1993b). Let us begin by considering tests concerning the mean, as described in Section 8.1. In the absence of spatial autocorrelation, the variance of the mean is estimated as the sample variance divided by the sample size:

$$var(\overline{y}) = s^2/n. \tag{8.13}$$

In the presence of spatial autocorrelation, $n$ in Eq. (8.13) is replaced by $n'$, the effective sample size, which is what we wish to determine. In general, as Cressie (1991) explained, the variance of the mean of the observations, $y_1, y_2, \ldots, y_n$, can be adjusted to correct for autocorrelation using the covariances of the $y$s, $\mathrm{Cov}(y_i, y_j)$, (Cressie 1991, equation (1.3.4)):

$$\mathrm{var}(\overline{y}) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}(y_i, y_j). \tag{8.14}$$

Therefore, we can get an estimate of the effective sample size by equating the right-hand sides of Eqs. (8.13) and (8.14) and transposing to isolate $n'$:

$$n' = n^2 s^2 \div \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(y_i, y_j) = n^2 \div \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cor}(y_i, y_j). \tag{8.15}$$

For example, for a first-order autoregression correlation structure (Table 8.1) with parameter $\rho$, we have $\mathrm{Cor}(y_i, y_j) = \rho^{|i-j|}$, and so the effective sample size is:

$$n' = \left[1 + 2\frac{\rho}{1-\rho}(1 - 1/n) - 2\frac{\rho^2}{(1-\rho)^2}\frac{1-\rho^{n-1}}{n}\right]^{-1} \times n. \tag{8.16}$$

For large $n$, this becomes

$$n' \cong n\frac{1-\rho}{1+\rho} = n\,\Theta, \tag{8.17}$$

where $\Theta$ is the approximate correction factor: $\Theta = (1-\rho)/(1+\rho)$. For example, if $n = 1000$ and $\rho = 0.4$, then $n' = 429$. Numerical simulations by computer, using artificial data with this autoregressive structure confirm the correctness of using $n' = n\Theta$ for one- and two-sample $t$-tests and for ANOVA $F$-tests for comparisons among means. Computer simulations also show that the same correction will work if applied to paired sample $t$-tests (Dale & Fortin 2002). Of course, for this to be useful, the autoregressive model has to be a good description of the autocorrelation structure, or the approach has to be robust to departures from that model. We will discuss the question of robustness further, because it is extremely important if the method of adjusting the effective sample size is to be used in real data analysis.

**Table 8.4** First-order serial autocorrelation structure. The main diagonal gives the correlation at distance zero, the first off-diagonals those at distance 1, and so on

$$\begin{vmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & & 0 \\ 0 & 0 & \rho & 1 & & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

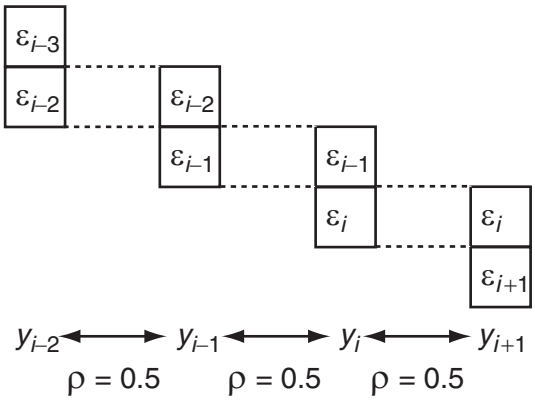



**Figure 8.6** Model $MA_1$: $y_i = \varepsilon i + \varepsilon_{i-1}$, a first-order moving average model. Half the information in $y_i$ is contained in $y_{i-1}$ and the other half in $y_{i+1}$ (see Eq. (8.19)). Unlike Model IH, there is no dependence beyond lag 1. The correlations are: $\text{Cor}(y_i, y_{i-1}) = 0.5$ and $\text{Cor}(y_i, y_j \,|j \neq i; j \neq i \pm 1) = 0$.

Our next step is to investigate further the equation for the effective sample size based on the correlation matrix, **R**, and its elements $r_{ij}$ estimated from the correlation calculated for each lag distance, $r(d)$. Because we cannot estimate the correlation of individual pairs of variates, our best estimate is to calculate the correlation for each lag, $d$, for which there are $(n - d)$ pairs. In the correlation matrix of individual pairs of variates (see Table 8.4 as an example) there are $n$ 1s on the main diagonal, and then two of each of the other entries, on either side of the diagonal. Therefore the estimate based on these correlations is:

$$n'(\mathbf{R}) = \frac{n^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}} = \frac{n^2}{n + 2\sum_{d=1}^{n-1}(n-d)r(d)}. \qquad (8.18)$$

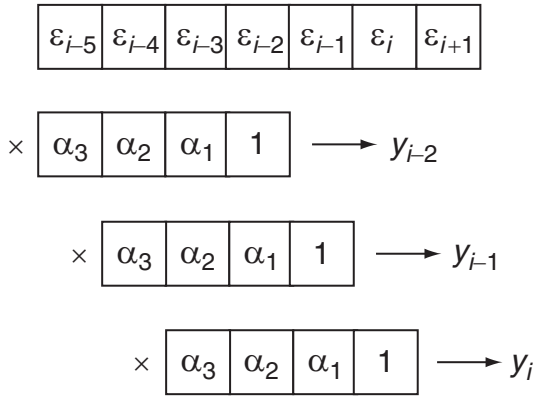



**Figure 8.7** Illustration of a general moving average (MA) model, here of order 4. A template with weights moves along the data series and calculates a weighted average at each position to create the value of $y$.

To evaluate the effective sample size, $n'(\mathbf{R})$, we can examine other autocorrelation structures. For example, a first-order serial autocorrelation structure has a correlation of $\rho$ between adjacent observations, but it is 0 for all other pairs (Table 8.4). The effective sample size is $n' = n^2/(n + 2(n - 1)\rho)$. A first-order moving average model (hereafter referred to as Model $MA_1$) is computed as follows:

$$y_i = \varepsilon_i + \varepsilon_{i-1}. \qquad (8.19)$$

It produces $\rho = 0.5$, and for large $n$, $n' = n/2$. Figure 8.6 gives an intuitive illustration of why this is so. If half the information in $y_i$ is contained in $y_{i-1}$ and the other half is contained in $y_{i+1}$, then only every second one of the $y$s are needed to recover all the information in the series, and $n' = n/2$.

Model $MA_1$ is the simplest member of a class of autocorrelation models called 'moving average' models, which, with $k$ being the model's order, have the general form (Chatfield 1975):

$$y_i = \varepsilon_i + \sum_{j=1}^{k} \alpha_j \varepsilon_{i-j} \qquad (8.20)$$

Figure 8.7 illustrates the appropriateness of the name, with a moving window creating a possibly weighted

**Table 8.5** Effective sample sizes based on autocorrelation values, n′(ρ), for artificial data from 'moving average' (MA) models of autocorrelation

| $\rho(d)$, $d = 0, 1, 2, \ldots$ | $\Sigma\Sigma\rho_{ij}/n^a$ | $n'(\rho)$ |
|---|---|---|
| 1, 0.5, **0**[b] | 2 | 250 |
| 1, 0.67, 0.33, **0** | 3 | 167 |
| 1, 0.8, 0.6, 0.4, 0.2, **0** | 5 | 100 |
| 1, 0.25, 0,−0.25, **0** | 1 | 500 |
| 1, 0.5, 0,−0.17,−0.33,−0.17, **0** | ⅔ | 750 |
| 1, 0.4,−0.2,−0.4,−0.2, **0** | ⅕ | 2500 |
| 1,−0.83, 0.67,−0.5, 0.33,−0.17, **0** | ≈0 | ? |
| 1,−0.75, 0.5,−0.25, **0** | ≈0 | ? |

[a] The actual sample size is $n = 500$.

[b] **0** means that correlations at this lag and all larger lags are zero.

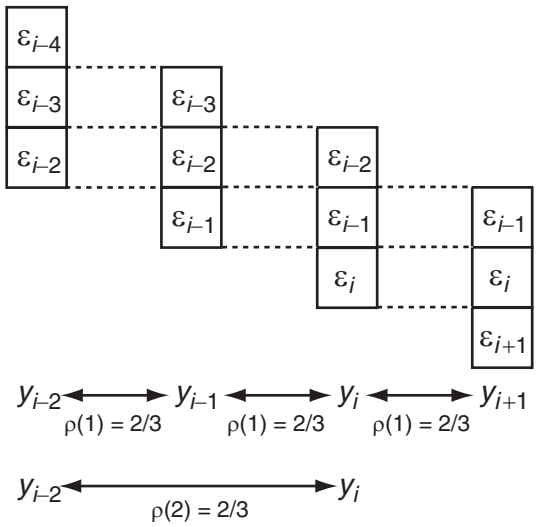



**Figure 8.8** A particular second-order moving average model: $y_i = \varepsilon_i + \varepsilon_{i-1} + \varepsilon_{i-2}$, $\mathrm{Cor}(y_i, y_{i-1}) = ⅔$ and $\mathrm{Cor}(y_i, y_{i-1}) = ⅓$.

average of the $\varepsilon$s. The question of directionality is not as puzzling for these models as for the autoregressive. For example, for Model $MA_1$, we could re-label all the $y_i$s as $y_{i-1}$, producing forward-looking rather than backward-looking dependence, without changing anything else in the characteristics of the data produced.

Extending autocorrelation beyond the first neighbours using the second model of this moving average series, the correlation between adjacent observations is $\rho_1$, $\rho_2$ for pairs at one remove, and 0 for all other pairs. The effective sample size is:

$$n' = n^2/[n + 2(n-1)\rho_1 + 2(n-2)\rho_2]. \tag{8.21}$$

The values $\rho_1 = 0.67$, $\rho_2 = 0.33$ can be generated by the model (Model $MA_2$; see Figure 8.8):

$$y_i = \varepsilon_i + \varepsilon_{i-1} + \varepsilon_{i-2}. \tag{8.22}$$

For these values and large $n$, $n' = n/3$.

Table 8.5, similar to one presented in Dale & Fortin (2002), shows the results for a range of autocorrelation structures, generated in the same fashion as Models $MA_1$ and $MA_2$. With the exception of the last two lines, computer experiments show that the effective sample sizes in the last column are 'correct' in that, with 10 000 realizations as in Table 8.5, the rates at which the true null hypothesis is rejected are close to the nominal values, when the derived effective sample size is used. The most interesting autocorrelation structures are those which have some negative autocorrelation added. (For example, the model $y_i = \varepsilon_i + \varepsilon_{i-1} - \varepsilon_{i-2} - \varepsilon_{i-3} - \varepsilon_{i-4}$, cited in the third to last line of Table 8.5, produces apparently cyclic behaviour in the data, see Figure 8.9). Sufficient negative autocorrelation actually increases

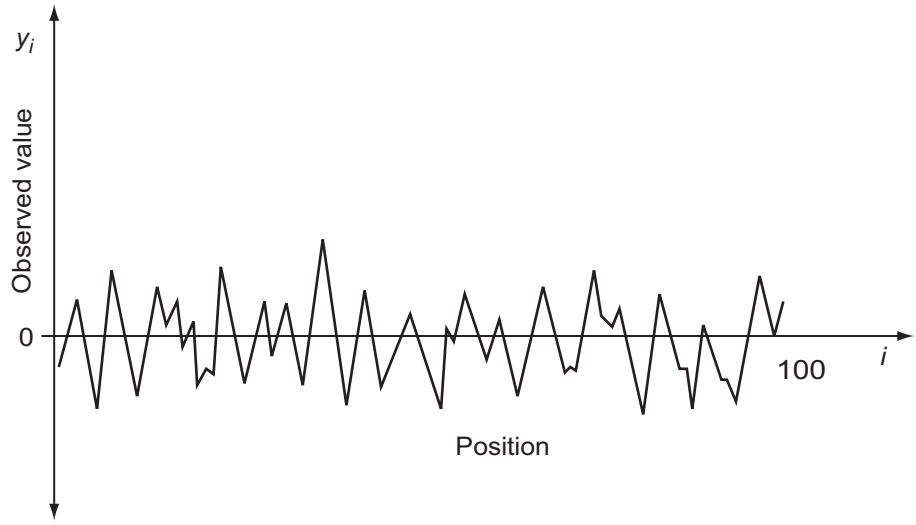



**Figure 8.9** Cyclic behaviour of the variable $y$ induced by the moving average model $y_i = \varepsilon_i + \varepsilon_{i-1} - \varepsilon_{i-2} - \varepsilon_{i-3} - \varepsilon_{i-4}$. The behaviour of $y$ is, however, aperiodic, with autocorrelation expected to be 0 at lags of 5 and beyond.

the effective sample size to be greater than $n$. This means that strong positive autocorrelation at small scales does not necessarily compromise statistical tests if there is cyclic behaviour that produces negative autocorrelation at larger scales. In fact, the test statistic may require inflation rather than deflation to achieve significant results at the correct nominal rates. This fact has implications for testing ecological data in which repeating structure (spatial pattern) is very common. One implication is that Dale & Zbigniewicz (1997), who attempted to correct for autocorrelation effects by using $\alpha = 0.01$ rather than the usual $\alpha = 0.05$, may have greatly overcorrected based on the short-range positive autocorrelation but leaving out of consideration the longer range negative autocorrelation.

The second related point is that autocorrelation at all lags must be included in the calculation of effective sample size, even if the individual value does not seem to be itself statistically significant. It is certainly tempting, and may seem logical to suggest using only significant values, but that can lead to errors. In fact, a large number of small negative values for autocorrelation at greater distances may be able to counteract the effects of larger positive values at short distances. This situation may arise frequently in ecological data, if the system being studied exhibits patchiness, which can give rise to cycles of positive and negative autocorrelation.

Based on the 'correctness' of the effective sample sizes for a range of models given in Table 8.5, it is tempting to suggest that the solution is to calculate the autocorrelation matrix, **R**, from the data and then to use its values to find the correct effective sample size. Alas, the computer runs that gave rise to Table 8.5 demonstrated the real problem with trying to adjust tests of data with spatial autocorrelation using estimates from the data themselves. The real problem is that the realizations of a very simple structure, such as that generated by $y_i = \varepsilon_i + \varepsilon_{i-1}$, can have very different estimates of $n$. For example, in a set of 1000 runs with $n = 500$ of that simple model (the first line of Table 8.5), while the average effective sample size calculated from the data was 250 (which looks fine), the range for individual realizations was from 48 to 492. Clearly, this approach cannot be used for even artificial data with a simple underlying structure and how much more dangerous might it

be for real data with an unknown and possibly complex structure.

Another feature became apparent from the computer runs, which is that while $\Sigma\Sigma\rho_{ij}$ cannot be less than 0, from a simple algebraic argument, in rare cases, estimates from data can be. The argument that the overall sum cannot be less than zero is:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma^2 = \sum_{i=1}^{n}\mathrm{Var}(y_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(y_i, y_j)$$
$$= \mathrm{Var}\left(\sum_{i=1}^{n} y_i\right) \qquad (8.23)$$

which, being a variance, cannot be negative. For the models in the last two rows of Table 8.5, the expected sum of the autocorrelations is very close to zero (for example, see Figure 8.7, which gives rise to the last row of the table).

In the absence of justification for proceeding from the data to an estimate of effective sample size, it may seem that we should abandon this approach, but that is not true, as we will show. Before we return to a discussion of general solutions to the problem, however, we will describe other kinds of models that produce spatial autocorrelation, as important background information, and then we will present some particular examples of corrections available in the literature.

### 8.2.3 More on induced autocorrelation and the relationships between variables

So far, we have used the terms autocorrelation and autoregression without drawing a clear distinction between correlation and regression. In general, correlation refers to the positive or negative relationship between two quantitative variables, both possibly measured with error, where it is not known that one has a direct causal effect on the other (Sokal & Rohlf 1995). For example, in the model of induced autocorrelation (Model ID), adjacent *x*s have non-zero correlation although they have no direct causal effect on each other. Correlation is a measure of the covariance of the two variables, relative to their variances:

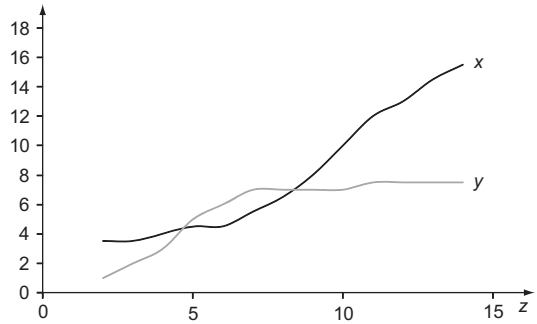



**Figure 8.10** Both $x$ and $y$ are correlated with $z$ ($r_{xz} = 0.96$; $r_{yz} = 0.90$) as illustrated in part and positively correlated with each other ($r_{xy} = 0.76$); but they are negatively correlated with each other when their dependence on $z$ is controlled for ($r_{xy\cdot z} = -0.91$).

$$r_{xy} = \frac{\mathrm{Cov}(x,\ y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}. \qquad (8.24)$$

Given three quantitative variables, $x$, $y$ and $z$, the partial correlation of $x$ and $y$ with $z$ held constant is:

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{\left(1 - r_{xz}^2\right)\left(1 - r_{yz}^2\right)}}. \qquad (8.25)$$

Figure 8.10 illustrates an artificial example, in which both $x$ and $y$ are positively correlated with $z$ (and apparently with each other: $r_{xy} = 0.76$), but they are negatively correlated with each other when the relationships with $z$ are removed: using the formula given above, $r_{xy\cdot z} = -0.91$.

In linear regression, by contrast, we are evaluating the strength of the linear dependence of a dependent variable on an independent variable, with the underlying hypothesis being a model such as:

$$y_i = \alpha + \beta x_i + \varepsilon_i. \qquad (8.26)$$

When the linear dependence of $y$ on $x$ is controlled or removed, what is left is the residual:

$$y_i' = y_i - (\hat{\alpha} + \hat{\beta}x_i). \qquad (8.27)$$

In the artificial example of Figure 8.10, when the linear dependence of $x$ and $y$ on $z$ is removed, negative

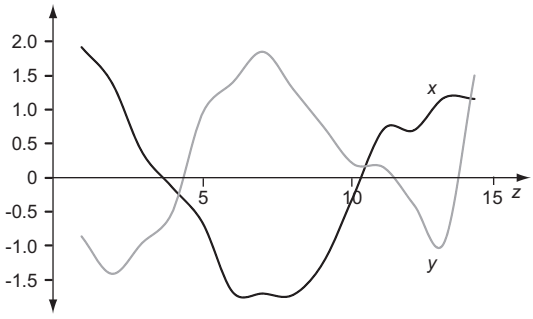



**Figure 8.11** The residuals of $x$ and $y$ when the linear dependence on $z$ is taken into account.

correlation of the residuals is obvious (Figure 8.11) and the correlation coefficient of $x'$ and $y'$ is $-0.91$. It is tempting to speculate that if autocorrelation in $x$ and $y$ could be attributed to their dependence on $z$, its effects could be similarly controlled by accounting for their dependence on $z$.

With real data, we observe spatial dependence in a variable of interest, $y$, but we may not always know its origins. It may be inherent, or it may be induced, or it may be both. We may not be able to distinguish among the possibilities. For Model IH, where autocorrelation is induced by an underlying autoregressive structure, the correlation at lag $j$ is:

$$\mathrm{Cor}(y_i, y_{i-j}) = \rho^j, \tag{8.28}$$

and for Model ID it is:

$$\mathrm{Cor}(y_i, y_{i-j}) = \frac{\beta^2 \rho^j \sigma_\xi^2}{(1-\rho^2)\sigma_\varepsilon^2 + \beta^2 \sigma_\xi^2}. \tag{8.29}$$

In both cases the autocorrelation declines exponentially with increasing distance, but that may not be true of all models of spatial autocorrelation. For the doubly autoregressive model (Model IH-ID, refer to Figure 8.4), the correlation is much more complex, but it also declines exponentially with distance. Therefore, we may not even distinguish the possibilities of Models IH and ID from the more complicated Model IH-ID, because they all have an exponential decay of autocorrelation with distance. In addition, given a single data set, it may not be possible to positively identify it as being derived from an autoregressive or from a moving average model and, for field data, that distinction may not hold.

In describing what we will call 'the trace method' for correctly testing the correlation coefficient of two variables in the presence of autocorrelation, we begin to look at the evaluation of the relationships between variables under those conditions. We will investigate this topic further by looking at a few models in which the relationship between $x$ and $y$ arises from their dependence on a third variable, $z$. There is obviously a number of ways in which this can happen.

We then add autocorrelation to $x$ and $y$, or to $z$, in turn, either in the moving average (MA) or autoregressive (AR) form. We can examine tests of correlation between the residuals of $x$ and $y$, after their linear dependence on $z$ is controlled, for non-zero correlation or for inflation of the rates of apparent significance in all nine combinations. The results are given in Table 8.6.

Let us begin with the simple model:

$$\begin{cases} y_i = \beta_x z_i + \varepsilon_i \\ x_i = \beta_y z_i + \eta_i. \\ z_i = \xi_i \end{cases} \tag{8.30}$$

Table 8.6 provides several clear messages.

(1) Controlling for the dependence on $z$ does not remove the effects of spatial autocorrelation in $x$ and $y$, in that the rates are inflated (MA) or the correlation is greater than zero (AR).
(2) The autocorrelation in $z$ is not an important factor in this context.
(3) The moving average and autoregressive models produce qualitatively different behaviour. (Models that combine AR and MA terms will also have a non-zero correlation.)
(4) Interpretation of a significant correlation will be especially difficult if it is not obvious whether the AR or MA model is the better description of the data.

This discussion is especially important for the ecological context, in which we can reasonably expect some form of inherent autocorrelation in most of the biological variables we measure *and* some form of induced autocorrelation in those variables due to autocorrelation inherent in the underlying abiotic

**Table 8.6** Results of correlation tests of the residuals of x and y, x′(z) and y′(z), with different combinations of autocorrelation. The values of β were all positive. When r > 0, the rate of rejection of the null hypothesis is no longer relevant because it is no longer true, hence designated as 'n.a.' in the table.

| | Autocorrelation in $z$ | | |
|---|---|---|---|
| | 0 | MA | AR |
| Autocorrelation in $x$ and $y$ | | | |
| 0 | $r_{x'y'} = 0$ rates nominal | $r_{x'y'} = 0$ rates nominal | $r_{x'y'} = 0$ rates nominal |
| MA | $r_{x'y'} = 0$ rates inflated | $r_{x'y'} = 0$ rates inflated | $r_{x'y'} = 0$ rates inflated |
| AR | $r_{x'y'} > 0$ rate n.a. | $r_{x'y'} > 0$ rate n.a. | $r_{x'y'} > 0$ rate n.a. |

factors in the environment. Usually, we will not be able to determine the relative strength of these two sources. When we then examine the relationship between two ecological variables, it is therefore probable that both will exhibit double autocorrelation.

### 8.2.4 Correlation and related methods

The correlation coefficient measures the strength of the linear relationship between two variables. For the correlation coefficient between two independent variables each with an autoregressive correlation structure with parameters $\rho_1$ and $\rho_2$, Bartlett (1935) showed that its variance is approximately:

$$s_r^2 = \frac{1 + \rho_1\rho_2}{n(1 - \rho_1\rho_2)}. \tag{8.31}$$

This result suggests an effective sample size adjustment of:

$$n' = n\,\Psi = n\frac{1 - \rho_1\rho_2}{1 + \rho_1\rho_2}. \tag{8.32}$$

If there is autocorrelation only in one of the series, for example if $\rho_2 = 0$, then no correction for autocorrelation is required (cf. Bivand 1980).

Clifford *et al.* (1989) suggested a method for using the *t*-test to assess the significance of the correlation coefficient in the presence of spatial autocorrelation, with the sample size, $n$, replaced with an effective sample size. In the Clifford *et al.* (1989) version, it is $n'_r = 1 + s_r^{-2}$, using an estimate of the variance of the correlation coefficient, which can be based on the autocorrelations of $x$ and $y$. Using the autocorrelations at lag $k$, $r_x(k)$ and $r_y(k)$, with $n(k)$ being the number of pairs in distance class $k$, here $n - k$, the effective sample size $n'$ is

$$\begin{aligned} n'_r &= 1 + \frac{n^2}{n + 2\sum_{k=1}^{n-1}(n-k)r_x(k)r_y(k)} \\ &= 1 + n\frac{1}{1 + \frac{2}{n}\sum_{k=1}^{n-1} n(k)r_x(k)r_y(k)}. \end{aligned} \tag{8.33}$$

This looks similar to the estimates of the effective sample sizes for $x$ and $y$ considered separately; it can also be written as:

$$n'_r = 1 + \frac{n^2}{\mathrm{tr}(\mathbf{R}_x\mathbf{R}_y)}, \tag{8.34}$$

where $\mathbf{R}_x$ and $\mathbf{R}_y$ are the autocorrelation matrices for the two variables, estimated from the data, and 'tr' is the trace of the matrix, the sum of the major diagonal.

Dutilleul (1993b) refined this method and provided a generalized and exact form of it, thus correcting problems that can occur with small sample sizes. This approach calculates the variance of the sample covariance from the (auto)covariance matrices and also provides an adjustment for the number of degrees of freedom: $n' = 1 + s_r^{-2}$.

The procedure is to calculate the covariance (not correlation!) matrices for $x$ and $y$ as estimates based on the distance classes $d = 0, 1, 2, \ldots$; call the matrices $\mathbf{S}_x$ and $\mathbf{S}_y$. Let $\mathbf{B}$ be the matrix with $b_{ii} = 1/n - 1/n^2$ on the main diagonal and $b_{ij} = -1/n^2$ elsewhere. Then:

$$n' = 1 + \frac{\mathrm{tr}(\mathbf{BS}_x)\mathrm{tr}(\mathbf{BS}_y)}{\mathrm{tr}(\mathbf{BS}_x\mathbf{BS}_y)}, \tag{8.35}$$

where 'tr' refers to the trace, which is the sum of the elements on the major diagonal. The same correction can also be formulated using matrices of Moran's autocorrelation coefficient rather than the variance–covariance matrices (P. Legendre, personal communication). With both $x$ and $y$ modelled as first-order simultaneous autoregressive processes on a lattice, Dutilleul (1993b) provided some examples of the effect of this correction. For a $10 \times 10$ lattice ($n = 100$) and $\rho x = \rho y = 0.1$, the effective sample size is 80; if the autocorrelation parameters are of opposite signs, $\rho x = -\rho y = 0.1$, the effective sample size is 119, illustrating again the 'positive' effects of negative spatial autocorrelation. Investigations by Legendre *et al.* (2002) confirm the effectiveness and robustness of Dutilleul's correction.

#### 8.2.4.1 Correlation

We investigated the robustness of spatial autocorrelation corrections based on the 'trace' approach for several bivariate tests, including those for the correlation coefficient, based on a range of artificial data generated by AR and MA models (Dale & Fortin 2009). We included the potentially difficult cases of artificial data which exhibited waves, from fairly strict and consistent to very irregular. The effective sample size for the bivariate test, $n'_r$ derived from the individual realizations of a model, is much less variable than those of the individual variables $n'_x$ or $n'_y$. In fact, it is possible for $n'_r$ to be much less than $n$, even while $n'_x$ and $n'_y$ are considerably greater because of negative autocorrelation. The effective sample size $n'_r$ cannot be derived from the values of $n'_x$ and $n'_y$ alone. It is the combination of the autocorrelation structures that determine the effective sample size; and so we cannot characterize spatial autocorrelation simply by its 'strength' or 'intensity' in the individual data series.

The 'trace' correction is robust for most MA or simple AR models. More complex AR models can have incorrect rates of rejection of the null hypothesis, but only when the models produce reinforced wave structures and when the two data sets have the same period or periods that are integer multiples of each other. These results confirm that this correction for spatial autocorrelation in this particular bivariate test is broadly successful but not infallible.

#### 8.2.4.2 Partial correlation

Partial correlation is a technique for evaluating the linear relationship between two variables, when a third variable may be involved. Given three quantitative variables, $x$, $y$, and $z$, the partial correlation of $x$ and $y$, controlling for $z$ is

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\left(1 - r_{xz}^2\right)\left(1 - r_{yz}^2\right)}. \tag{8.36}$$

This coefficient measures the relationship of $x$ and $y$ as if $z$ were held constant, and so only the interaction of $x$ and $y$ remains, and the effective sample size for the partial coefficient should be the same as for the correlation coefficient. Therefore, the 'trace' correction should work for this statistic over the same range of autocorrelation structures as for simple bivariate correlation. Our own computer simulations confirmed this and this result is presented in Alpargu & Dutilleul (2006).

#### 8.2.4.3 Linear regression

The calculations involved for simple linear regression are similar to those for the correlation coefficient, and so it seems likely that the same effective sample size correction derived for one can also be used for the other. The fact that this can be done was shown by Alpargu & Dutilleul (2003) and our computer runs confirm that the robustness conditions for the 'trace' correction for this use of the $F$-test is the same as for the $t$-test for the correlation coefficient. We can

speculate also that the same correction will also work for logistic regression (cf. Cerioli 2002).

#### 8.2.4.4 Tests for proportions

Tests for proportions are carried out using contingency tables of counts and goodness-of-fit statistics such as Pearson's $X^2$ or the log-likelihood ratio, $G$. For example, the positive or negative association of two species can be tested based on the counts of their presence : absence in sample units. These counts are summarized in a $2 \times 2$ contingency table, and then the goodness-of-fit statistic is calculated and compared to the $\chi^2$ distribution with one degree of freedom. The question here, however, is how to account for the lack of independence among sampling units due to their spatial locations. Returning to the one-dimensional situation, if the sample units are contiguous (such as quadrats in a string or transect), the data will have spatial autocorrelation and their dependence might be well described by a Markov model. For example, suppose that each of two species is recorded in a string of contiguous sampling units as being in one of two possible states: 0 for absence and 1 for presence. A reversible Markov model of the sequences of presence and absence for species $s$ would be based on the underlying transition probabilities, $\tau_s(i, j)$, the probability that it makes the transition from state $i$ to state $j$ between one quadrat and its neighbour. The overall probability of state $i$ for species $s$ is $\pi_s(i)$. These two probabilities can be estimated from the data as $T_s(i, j)$, the frequency of transition from state $i$ to state $j$, and $p_s(i)$ as the overall probability of state $i$ for species $s$. The approach to accounting for the spatial dependence in the data is to use these estimated probabilities to determine how much the test statistic calculated from the $2 \times 2$ contingency table, derived from such data, should be deflated to give the correct rejection rate; that is to determine a value, $\Phi$, by which to decrease the test statistic.

For $2 \times 2$ contingency tables derived from data in which the serial correlation is due to a reversible Markov process, Tavaré (1983) provided a deflation factor for the test statistic, based on the non-unit eigenvalues, $\lambda$, of the transition probability matrices which can be estimated from the frequencies as

$$\lambda_s = \sum_{i=1}^{n} \frac{T_s(i, j) - p_s(i)}{2 - 2p_s(i)}, \quad \text{with} \quad s = 1 \text{ or } 2. \tag{8.37}$$

The deflation value is then

$$\Phi = \frac{1 + \lambda_1\lambda_2}{1 - \lambda_1\lambda_2} \tag{8.38}$$

(Tavaré 1983; Tavaré & Altham 1983; see Upton & Fingleton 1989, p. 92). The structural similarity to the Bartlett correction factor is striking, and Tavaré & Altham (1983) suggested that under the null hypothesis, $r^2/s_r^2$ is asymptotically $\chi_1^2$. Computer experiments using artificial data with autoregressive structure confirm their suggestion.

The test statistic is calculated in the usual way but it is divided by the deflation factor before being compared to the reference distribution, $\chi_1^2$ in the case of a $2 \times 2$ contingency table. In the particular case of the goodness-of-fit statistic, deflating the statistic by $\Phi$ and reducing the effective sample size to $n = n\Phi^{-1}$ are mathematically equivalent.

Multiway $2^k$ tables arise in situations such as the testing of multispecies association, based on the presences and absences of $k$ species in sampling units (Dale *et al.* 1991). Again, if the data are from contiguous quadrats in transects, the spatial dependence may be well described by Markov models. Porteus (1987), following the Tavaré approach, provided the corresponding formulation for multiway tables, again based on an underlying Markov process, but the formula becomes quite complicated and concerns about the complexity and appropriateness of the Markov model as a description of the multispecies spatial structure may come into play.

A more general correction for spatial autocorrelation using $2 \times 2$ contingency tables was published by Cerioli in 1997, based on the correlation structure calculated from the data, rather than on known or inferred properties of an underlying model. It looks very similar to the general approach of Clifford *et al.* (1989) and Dutilleul (1993b) and seems to be robust. For the $2 \times 2$ case, the correction is essentially the

same as the correction for the correlation test, although the approach used to derive it was different. Doing our best to harmonize the notation, Cerioli's effective sample size for a 2 × 2 table is:

$$n' = n\left(\frac{1}{1+\varphi}\right) \text{ with} \tag{8.39}$$

$$\varphi = \frac{2}{n}\sum_{k=1}^{D} r_x(k) r_y(k) n(k). \tag{8.40}$$

Here, there are $D$ distance classes and $r_x(k)$ and $r_y(k)$ are the calculated autocorrelations of $x$ and $y$ in the $k$th distance class, with $n(k)$ pairs in the class. This gives:

$$n' = \frac{n^2}{n + 2\sum_{k=1}^{D} r_x(k) r_y(k) n(k)}. \tag{8.41}$$

Apart from the addition of 1, missing here, in the context of one-dimensional data series this is identical to the correction of Eq. (8.39).

This same approach has since been extended to $R \times C$ tables (Cerioli 2002). For $R$ rows and $C$ columns, designate the proportions of observed values in the $i$th row and $j$th column as $p_{i\bullet}$ and $p_{\bullet j}$. The correction also requires $r_{I(x,i)}(k)$, the autocorrelation coefficient for $x$ belonging to category $i$ and distance class $k$, as indicated by the index function $I(x,i)$, and so on:

$$\varphi = \frac{2}{n}\sum_{i=1}^{R}\sum_{j=1}^{C}\left\{(1-p_{i\bullet})(1-p_{\bullet j}) \times \sum_{k=1}^{D} r_{I(x,i)}(k) r_{I(y,j)}(k) n(k)\right\} \tag{8.42}$$

and then

$$n' = n\left(\frac{(R-1)(C-1)}{(R-1)(C-1)+\varphi}\right) = n\left(\frac{1}{1+\varphi/(R-1)(C-1)}\right). \tag{8.43}$$

We evaluated this correction for goodness-of-fit tests using a range of spatial models of binary data. The data were produced from the continuous values produced by AR or MA models by converting them to 1s and 0s. This mimics presence : absence data in strings of contiguous quadrats, and gives a 'zero–one process' based on a primary process that underlies it (see Lomnicki & Zaremba 1955). Similarly, dividing the values of a continuous AR or MA model into several more-or-less equal parts generates data to evaluate corrections for the $R \times C$ goodness-of-fit test.

The evaluation showed that the correction for goodness-of-fit tests is mainly robust, failing only for models in which reinforced waves in the data created strongly cyclic autocorrelation. Our conclusion is that, except in unusual circumstances, this correction based on the structure of the data can be used to correct for autocorrelation for $R \times C$ goodness-of-fit tests.

Following the evaluation of an entire contingency table using a goodness-of-fit test, standardized residuals are often calculated to determine which cells of the table contributed most strongly to a significant result. Where $o$ is the observed frequency and $e$ is the expected, the Freeman–Tukey standardized residual is calculated as $z = \sqrt{o} + \sqrt{o+1} - \sqrt{4e+1}$. For a 2 × 2 table, the standardized residual can be compared to $\sqrt{\chi_1^2/4}$ to determine which values make important contributions to the overall significance (Sokal & Rohlf 1981). Dale *et al.* (1991) suggested that whatever the inflation factor, $\Phi$, is for the overall statistical test, the standardized residuals can be corrected by dividing by $\sqrt{\Phi}$. The argument for this procedure is that the sum of the squares of the residuals is approximately the test statistic and if it is to be deflated by $\Phi$, the residuals should be deflated by its square root.

The main problem for this approach in an ecological context is that its applicability depends on how well a Markov model describes the characteristics of the data. Often, that will not be a good description of the data, and the test is not particularly robust to departures from the underlying assumption. For the purposes of this discussion on correcting for spatial autocorrelation in goodness-of-fit tests, the conclusion would be that the Tavaré approach should only be used when there is some confidence that the particular kind of Markov model is a good description of the characteristics of the data.

The concerns about Markov models aside, the situation for bivariate and multivariate testing is cause for optimism. For bivariate data, if the autocorrelation is 'well-behaved' with any cycles irregular or decaying, the trace corrections are reliable. Ecologists should be concerned that the autocorrelation may be persistently cyclic because of patchiness or waves in the data. Our advice is to check the autocorrelation structure, particularly the product of the autocorrelations: if the product is regularly cyclic and decreases only slowly, the corrections may not work. On the other hand, if any autocorrelation cycles are irregular or their amplitude decreases with distance, the correction procedures probably can be applied. The exceptions include some instances of the 'seasonal model' with strict sine wave fluctuations in the data, for which the autocorrelation may be itself non-diminishingly cyclic, depending on the frequency and relative phases of the two data series. In general, if the product of the autocorrelations is 'well-behaved', declining with lag distance or with the amplitude of cycles declining rapidly with distance, the corrections described here may safely be used. In fact, Cerioli (1997) warns explicitly that long-range independence is crucial to his results, so that the correction is not reliable if the autocorrelation does not diminish with lag distance.

The correction for bivariate tests is much more robust than attempts at corrections for univariate tests. The reason is that they use the product of the autocorrelation values for the correction, and these have magnitudes less than one, which results in their product being small. This makes the procedure more robust. In fact, this procedure guarantees that if there is spatial autocorrelation in only one of the variables, no correction is needed because the product is always zero (cf. Tavaré 1983; Cerioli 1997). Corrections for univariate tests remain elusive.

## 8.3 Randomization procedures

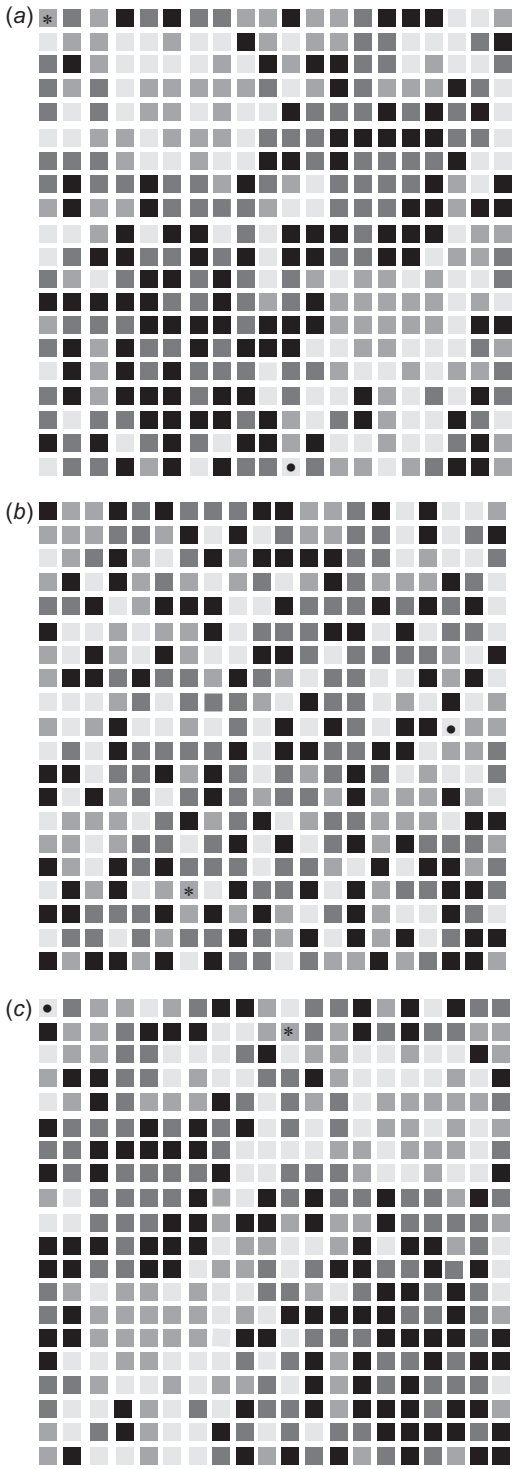
### 8.3.1 Restricted randomization and bootstrap

The randomization approach to general significance tests of hypotheses uses the original data themselves to generate a reference distribution (Manly 1997, 2006; Legendre & Legendre 2012). A test statistic calculated from the original data is compared with the distribution of the same statistic calculated after the data have been permutated or randomized in some way (see Chapter 1). In our applications, restricted randomizations are those in which the actual structure of the data (spatial, temporal, genetic, and so on) is retained as much as possible, rather than having it erased as it would be by complete randomization (Figure 8.12; cf. Fortin & Jacquez 2000; Fortin *et al.* 2012b). These may therefore be useful in testing autocorrelated data. Figure 8.12 illustrates the difference between complete and restricted randomization. There are 400 cells in a $20 \times 20$ grid and each cell is classified into four density classes, which we will reduce to two, high and low, for analysis. There are 180 black (high density) cells in the original data (Figure 8.12*a*) and using a join count approach based on 'rook's move' neighbours (Chapter 3; cf. Pielou 1977b), the observed number of black–black joins ($J_{BB} = 220$) is considerably greater than the number expected:

$$E(J_{BB}) = 760 \times \frac{180 \times 179}{400 \times 399} = 153.4. \qquad (8.44)$$

This reflects the high degree of autocorrelation in the data, with obvious patches of low and high density. When the data are completely randomized by exchanging randomly chosen pairs of grid cells (Figure 8.12*b*), the spatial structure is destroyed, and the number of black–black joins falls close to the expected value ($J_{BB} = 154$). In Figure 8.12*c*, the data have been randomized with a toroidal shift of 10 on the $x$ axis and 1 on the $y$ axis, which has preserved much of the spatial structure ($J_{BB} = 212$). This is one of the most commonly used restricted randomization techniques.

The applicability of any randomization procedure will depend on the nature of the data and the purpose of testing. For example, if we have data from a transect of contiguous quadrats, in which tree-canopy density and understorey cover are recorded, we could test the significance of their correlation, given their individual spatial structures, by shifting the relative positions of the two data sets and recalculating the correlation for




all possible relative positions (see Figure 8.13). This is sometimes referred to as 'caterpillar' randomization (after the tractor tread, not the insect larva) and it is the equivalent of 'toroidal shift' randomization but in only one dimension (cf. Upton & Fingleton 1985). If the original observed value is greater than 95% of the values from the shifted data sets, we would conclude that the observed value is significant. Longer transects with more quadrats will have more possible 'shifted' relative positions of the two data sets and greater sensitivity. Very short transects may not have sufficient numbers of relative positions to allow this kind of test. We can use restricted randomization for testing the correlation between variables in the same set of samples, but not for comparing the mean densities for two different transects. That is because there is no original natural pairing of the data in the two sets which can then be shifted in a restricted randomization test.

Legendre *et al.* (1990) described a contiguity-constrained permutation technique, which is a variant of spatially restricted randomization, for testing the significance of differences among regions in an ANOVA framework. Their numerical simulations show that, with their permutation technique, ANOVA is not very sensitive to spatial autocorrelation and provides a test with a correct Type I error.

Another application of randomization procedures is to derive variance estimators using jack-knifing estimation, as described by Lele (1991) and by Heagerty & Lumley (2000). In particular, it derives the reference distribution by re-sampling the data, leaving out one observation in each iteration. This jackknife approach enables confidence intervals to be derived for parameters of interest, for example, those of a regression model, which can then be tested for significant difference from zero. Cohn (1999) recommended a bootstrap procedure for comparisons of multivariate structures in the presence of

**Figure 8.12** A comparison of complete and restricted randomization. In the original data (*a*) there is significant autocorrelation. With complete randomization, shown in (*b*), that structure is destroyed, but randomization by toroidal shift (shown in (*c*)) retains most of the structure. The dots identify individual squares where shifts in locations from (*a*) to (*c*) occurred.

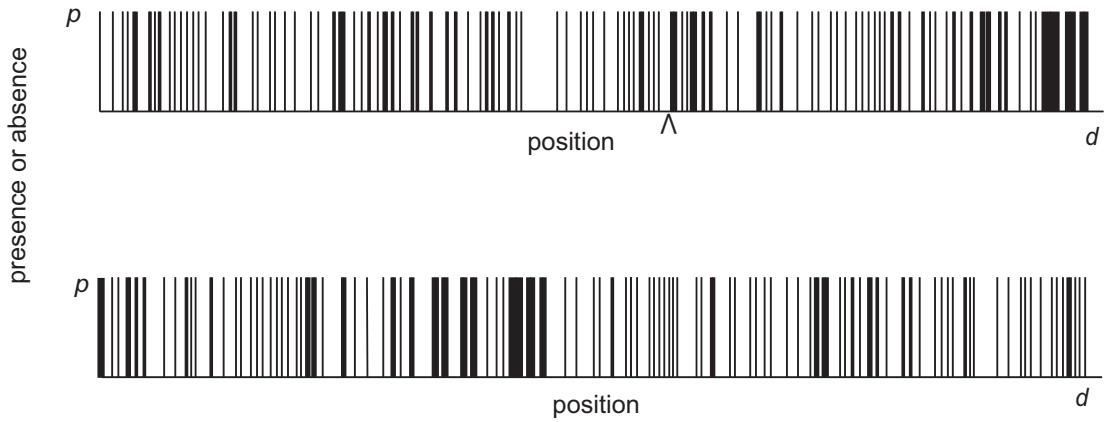



**Figure 8.13** The 'caterpillar' randomization of one-dimensional data: the original data series (upper) is broken at a randomly chosen point (indicated by the arrowhead) and then recombined (lower). If there are two data sets to be related, either set can be shifted in this way to provide new relative positions.

serial correlation and we suggest (without rigorous proof) that it should apply well for other forms of autocorrelation in two dimensions. Bootstrapping is another randomization technique, which uses re-samples of the data, allowing each datum to be used more than once. Bjørnstad & Falck (2001) proposed a bootstrap algorithm to create a confidence envelope for a non-parametric estimate of a spatial covariance function. There has been a number of advances in the use of bootstrapping in spatial situations (Loh 2008; Lin *et al*. 2011), including the use of block bootstrapping which, like restricted randomizations, attempts to preserve the dependence structure within the data (see Lahiri 1999), but we know of no applications of these in ecological studies.

### 8.3.2 Markov Chain Monte Carlo

Another approach solving the autocorrelation problem is to develop a relatively simple model of the autocorrelation structure and then to use a Monte Carlo simulation (Markov Chain Monte Carlo or MCMC methods) to generate artificial data sets to compare with the observed data. This approach begins by finding a parametric model of the spatial dependence in the data by standard model selection procedures. The model can then be used in a Monte Carlo approach to find good confidence intervals for the test statistics (Manly 2006). This last approach is the one we advocate in the absence of a robust 'analytic' solution for a particular set of circumstances. It may seem somewhat indirect, but it does allow for tests of significance. It is also the procedure recommended by Mizon (1995) in the context of economic analysis. He suggested that we start with a very general model, which well describes the data no matter how many terms are necessary, and then test for valid reductions of it; that is we can determine which explanatory variables can be omitted. Overspecified models, those with more variables than they really need, do not lead to invalid inferences; they are merely inefficient (Mizon 1995). We could start, therefore, with a very general autoregressive model of the data, considering observations as far as 20 steps apart, such as:

$$y_i = \sum_{j=1}^{20} \beta_j \, y_{i-j} + \varepsilon_i, \tag{8.45}$$

where $\varepsilon_i$ is $N(0, \sigma^2)$.

By eliminating many of the variables, we might end up with a model in which only two or three of the $\beta$s (Fisher 1932) were significantly non-zero. (In this

**Table 8.7** Coefficients of models fit to realizations of $y_i = 0.4y_{i-1} - 0.2y_{i-4} + \varepsilon_i$

| $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | n′(approx.)[a] |
|---|---|---|---|---|---|---|---|
| 0.47 | −0.04 | −0.04 | −0.18 | 0.33 | •[b] | • | 22 |
| 0.56 | −0.22 | • | • | • | • | • | 44 |
| 0.30 | 0.19 | −0.04 | −0.21 | • | • | • | 59 |
| 0.30 | • | • | • | • | −0.26 | 0.08 | 77 |
| 0.21 | 0.05 | −0.08 | −0.32 | • | • | • | 128 |

[a] The effective sample size, based on the underlying model, is $n' = 65$ (approximately).
[b] Symbol '•' indicates a nonsignificant term.

instance, it is appropriate to omit the nonsignificant terms.) As an illustration of this idea, we generated ten realizations of the model $y_i = 0.4y_i - 1 - 0.2y_i - 4 + \varepsilon_i$, with $n = 100$. We then fit the model given in Eq. (8.45) and determined the best-fitting submodel using a maximum likelihood backward selection procedure. Table 8.7 gives examples of the results.

Table 8.7 shows a range of possible outcomes for the sample size of $n = 100$; many are 'close' to the original underlying model, particularly at lag 1, but the effective sample size ranges from one-third to almost double the 'true' value. Clearly, however, these results show that the best-fit model is not always the model that generated the data. Using a larger sample size, $n = 500$, the best fit models are more similar to the originals, showing the advantage of larger sample sizes, but they still exhibit considerable variability. The 'model and Monte Carlo' (MCMC) procedure does seem to be a good approach in the absence of more direct solutions.

## 8.4 Spatial regressions

In the absence of spatial structure in the data, we can use a linear regression to model variability of a response variable, ***y***, as a linear function of the **X** predictors:

$$y \sim \beta\mathbf{X} + \varepsilon, \tag{8.46}$$

where $\beta$ is a vector of regression coefficients of fixed predictors **X**, and $\varepsilon$ are the random errors. The random errors, $\varepsilon$, are assumed to be independent such that $\text{var}(\varepsilon) = \sigma^2\text{I}$, where I is the $n \times n$ identity matrix. This implies that off-diagonal elements in the covariance matrix are zeros.

Yet when either the response or the predictor variables are spatially autocorrelated (Figure 8.14) it is likely that the errors are spatially autocorrelated which precludes the use of linear regression. To determine whether the errors are autocorrelated we can plot the residuals and compute Moran's $I$ correlogram. Spatially structured residuals may be due to (1) the lack of important predictors in the model, (2) incorrect model specification, or (3) a mismatch between the spatial patterns of the response and predictors.

Hence the first solution to account for spatial structure is to add environmental and ecological predictors in the model. If this is not sufficient and there is still spatial pattern in the residuals then we can use a generalized linear mixed model (GLMM; Bolker *et al.* 2009) where a random effect is added to account for the effects of location at **s**. In a GLMM the added random effect, z*i*, accounts for the variation in $y$ at location *i*, given the effects of the other sampling locations that are not explained by predictor variables:

$$yi \sim \beta\text{X}_i + \text{z}_i + \varepsilon. \tag{8.47}$$

GLMMs (as well as mixture models) have been used successfully to deal with spatial effects (among others, Hothorn *et al.* 2011; Hughes *et al.* 2014).

When a GLMM approach is not enough to account for the spatial structure of the data, then spatial regression can be used. A simple definition is that *spatial regressions* are models that include 'space' explicitly as

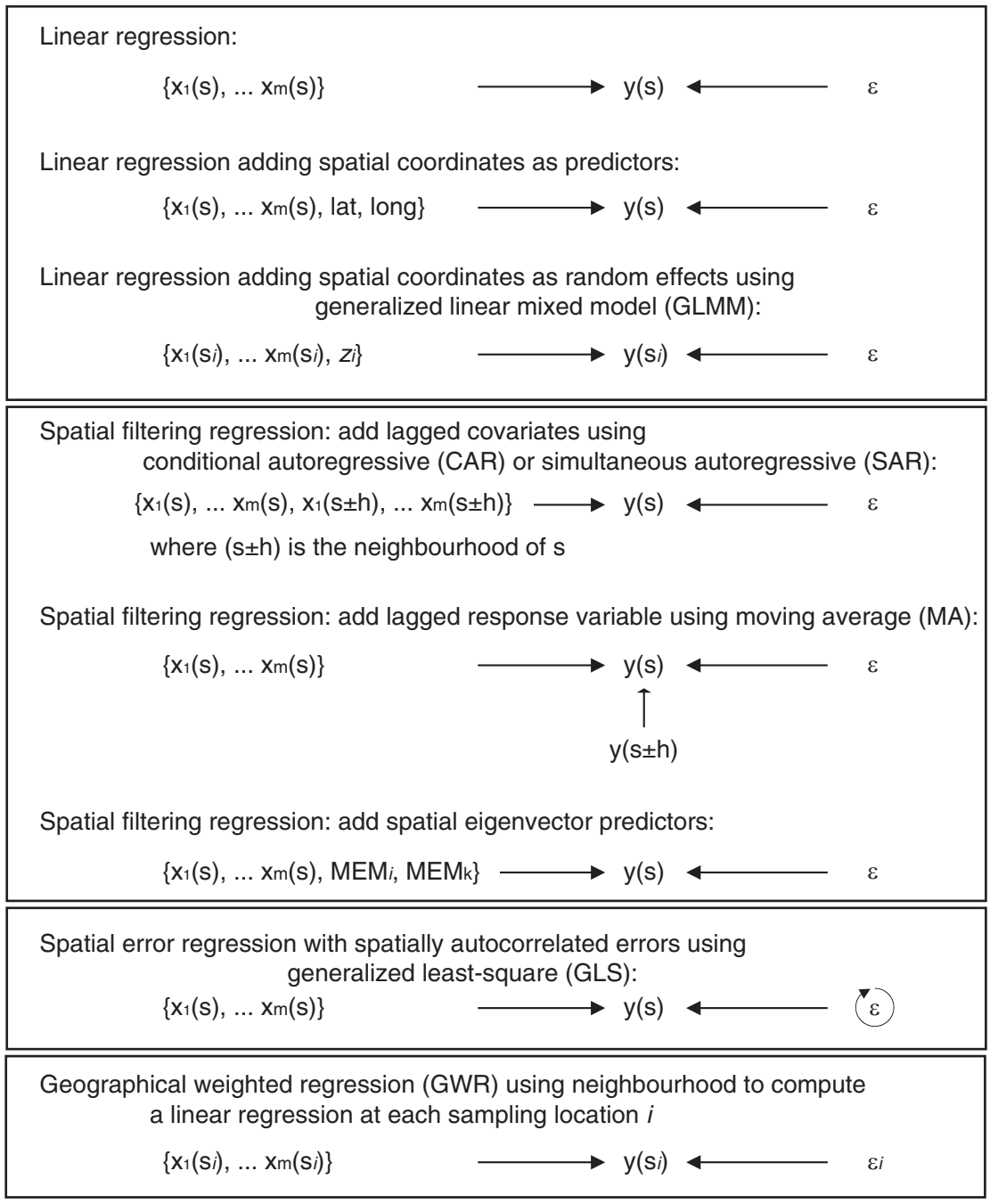



**Figure 8.14** Key regression approaches to account for spatial autocorrelated data within a regression framework: (1) linear regression adding spatial predictor variables or random effect, (2) spatial filtering models, (3) spatial error models, (4) geographical weighted regression.

**Table 8.8** Effect of three sources of spatial autocorrelation on the estimation of parameters for non-spatial and spatial regression

| Dependent/ response variable ($y$) | Independent/ covariate variable ($x$) | Residuals | Non-spatial regression: parameter estimations | Spatial regression: parameter Estimations |
|---|---|---|---|---|
| Random | Random | Random | Accurate | Accurate |
| Random | Spatially autocorrelated | Random | Accurate | Accurate |
| Random | Spatially autocorrelated | Spatially autocorrelated | Inaccurate | Accurate |
| Spatially autocorrelated | Random | Spatially autocorrelated | Accurate | Accurate |
| Spatially autocorrelated | Spatially autocorrelated | Spatially autocorrelated | Inaccurate | Accurate |

a factor. Spatial regressions are therefore extensions of the familiar statistical technique for characterizing the dependence of one variable on another or on a set of other variables, with the modifications required to take account of spatial dependence in the data. Note that a model that merely divides a population into spatially inexplicit subsets is not really spatial, although the term 'spatial model' is sometimes applied to that approach. Anselin (2009) provided an especially clear introduction to spatial regressions. Furthermore, there are numerous papers comparing the advantages and disadvantages of the various existing spatial regression models (e.g. Dormann *et al.* 2007; Bini *et al.* 2009; Beale *et al.* 2010).

There are various types of non-spatial and spatial regressions (Figure 8.14) that can be used to account for the spatial structure in the data by estimating the regression parameters (Table 8.8). The basic idea is that physical position affects the value of a variable of interest. For example the value of the variable, $y$, can be affected by its position, $\mathbf{s}$, determined with respect to 1, 2, or 3 spatial axes. That is, $\mathbf{s} \rightarrow y$ (read as '$y$ depends on $\mathbf{s}$' or '$\mathbf{s}$ affects $y$') or perhaps, slightly more formally, $y = f(\mathbf{s}) + \varepsilon$ (read as '$y$ is some function of $\mathbf{s}$ plus an independent error term'). The position $\mathbf{s}$ can be determined by a frame of reference created by spatial axes, the 'absolute position' (if there is such a thing), or relative to a local neighbourhood, $\mathbf{N}$, call the neighbourhood position $\mathbf{s}'(\mathbf{N})$. Another possibility is that position combines with other variables to affect the value of a dependent variable $\{x_1(\mathbf{s}), \ldots, x_m(\mathbf{s})\}$ and $\mathbf{s} \rightarrow y$ or $y = f(\mathbf{s}) + g(\{x_i(\mathbf{s})\}) + \varepsilon$. Because this dependence occurs in a spatial context, $y$ can be affected by spatial autocorrelation in $y$ itself, in the predictor variables (the $x$s), or in both. As we have already described, apparent spatial dependence observed in $y$ can be caused by autocorrelation inherent in $y$ itself, dependence induced in $y$ by its dependence on $x$ in which there is inherent autocorrelation, or by double autocorrelation: in $y$ itself and in $x$ on which it depends (Chapters 1 and 2).

In both non-spatial and spatial regression, the value of $y$ is determined by position, combined with dependence on the $x$s. In the spatial regression, we want to use the position and the $x$s to explain the observed value of $y$. We can begin with $y(\mathbf{s})$ as a function of $\{x_1(\mathbf{s}), \ldots, x_m(\mathbf{s})\}$ and an error term, $\varepsilon$. The non-spatial regression model can be modified into a spatial regression one by adding:

(1) lagged covariates, such that $y$ also depends on values of the $x$s at locations around $\mathbf{s}$ as well as on values at $\mathbf{s}$ itself: $\{x_1(\mathbf{s} \pm \varsigma), \ldots, x_m(\mathbf{s} \pm \varsigma)\}$, where $(\mathbf{s} \pm \varsigma)$ is the neighbourhood of $\mathbf{s}$ with maximum displacement $\varsigma$;
(2) lagged response variable, so that $y$ also depends on $y(\mathbf{s} \pm \varsigma)$;
(3) autocorrelation in the error term.

Another way to think about the spatial regressions is to classify them according to the data used (Haining 2003).

**(1) Location continuous; data continuous: (a surface)**

Large-scale variation can be modelled by smooth functions including linear trends and polynomial surfaces (i.e. spatial interpolation; see Chapter 6). The value of $y$ at a location, $\mathbf{s} = (s_1, s_2)$, is more-or-less determined by the value of the smooth surface at **s**. A simple example would be a linear response of $y$ to the axes of position, producing a plane surface:

$$y = \alpha + \beta_1 s_1 + \beta_2 s_2 + \varepsilon. \quad (8.48)$$

Small-scale variation can also be accounted for by spatial autocovariances. Valid autocovariance models include versions of semi-variogram models: spherical, exponential, Gaussian, 'hole', and so on (Chapter 6). Generally, in these models, autocovariance declines with distance and approaches zero at varying rates, for example:

$$\mathrm{Var}(y(s_i) - y(s_j)) = \sigma_z[1 - \exp(3|s_i - s_j|/a]. \quad (8.49)$$

**(2) Locations discrete (such as areas or points); data continuous**

For this combination of locations and data, we typically consider neighbouring observations in a regular grid or lattice, or in an irregular system of points or areas (shown in a spatial graph, as described in Chapter 3). Here we can use models such as Markov models, conditional autoregressive (CAR), simultaneous autoregressive (SAR), or moving average (MA) models to account for neighbourhood autocorrelation (Haining 2003; Dormann *et al.* 2007).

**(3) Locations discrete; data discrete**

This combination requires spatial versions of the logistic model for binary data such as presence : absence data (the autologistic model), or of the binomial or Poisson models for counts or other integer-valued observations (Dormann *et al.* 2007).

Most classifications of spatial regressions are based on the way by which space is incorporated and we will use the following structure to present the various spatial regression models in the next sections (Figure 8.14).

- Spatial filtering models, where the spatial signal in the response or covariate variables is accounted for using kernel approaches (conditional autoregressive (CAR), simultaneous autoregressive (SAR), moving average (MA) models; also known as spatial lag regression models), Moran's eigenvector approaches (MEMs, Chapter 7; Dray *et al.* 2006; Dray 2011) as orthogonal spatial predictors, or spatial cross-regressive model (SARMA; Anselin 2001). The implicit assumption in spatial filtering is that the spatial structure in the data is mostly due to spatial autocorrelation (its strength decays with increasing spatial distance). When the spatial structure is a composite of spatial autocorrelation and spatial heterogeneity (i.e. subregions) then other spatial regression models need to be used (e.g. geographically weighted regression, mixture models, hierarchical Bayesian models).
- Spatial error models (SEMs) which account for the spatial structure in the errors (spatial errors) such as generalized least-squares regression (GLS; Pinhiero & Bates 2000) and regression Kriging (Hengl *et al.* 2004; Hengl *et al.* 2007).
- Geographically weighted regression (GWR; Fotheringham *et al.* 2002) that models as many regressions as there are sampled locations.

### 8.4.1 Spatial filtering using autoregressive models

So far, our presentation of autoregressive models has dealt implicitly with continuous variables, often with a normally distributed error term, for which the calculation of correlation was a logical approach. We need to consider also the concept of autocorrelation as it applies to discrete variables, the most simple being data that consist of sequences of 0s and 1s. The introduction of autocorrelation in such data can be achieved by having the value at a particular location being dependent on the values at preceding locations; for example, the probability of a 1 could decrease with the length of the preceding run of 1s. These structures are called Markov models and were introduced in Section 8.2.4.

Also, our presentation so far implicitly and explicitly used directionality in the description of models: for example with $y_i$ as a function of $y_{i-1}$. This direction of apparent dependence is logical in time series, but does not seem to have the same intuitive appeal for spatial data, particularly when we consider two dimensions rather than just one. The unidirectionality of the one-dimensional models is apparent in the first-order autoregressive model (Model IH):

$$y_i = \rho y_{i-1} + \varepsilon_i, \tag{8.50}$$

where $\rho$ is the first-order degree of autocorrelation. The question is whether, in spatial models, both the 'forward' ($\varphi$) and 'backward' ($\rho$) neighbours should be considered; for example:

$$y_i = \rho y_{i-1} + \varepsilon_i + \varphi y_{i+1}. \tag{8.51}$$

Clearly, to implement this structure in a practical way requires the simultaneous solution of a set of $n$ equations for the $n$ values of the $y$s. This is not as direct as starting with a random value for $y_0$, and then generating $y_1$ to $y_n$ each from the preceding value using Eq. (8.50).

In this one-dimensional example, each location has two neighbours, but in two dimensions, more neighbours are considered, whether in a regular lattice or not, and their effects on the location's value may have different weights depending on the neighbours' positions and distances. The weights are often given in a 'proximity matrix', $\mathbf{W}$, with $w_{ij}$ being greater than zero if the value at location $i$ is not independent of the value at location $j$ (see Chapter 3). For example, in a regular square lattice, all 'queen's move' (or 'king's move' if it is only one step) neighbours might receive equal weighting (say ⅛) in a proximity matrix, with all others being 0. There are many different ways in which autocorrelation can be introduced in two-dimensional data, and we will describe only two, and the simplest versions of two approaches: simultaneous autoregressive models (SAR) and conditional autoregressive models (CAR) (Haining 2003). These approaches are most easily explained using matrices (indicated by bold font), and our explanation of SAR borrows heavily from Bailey & Gatrell (1995) who described this model with great clarity.

Simultaneous autoregressive models (SAR) are based on the concept illustrated in Eq. (8.51), in which the equation defining $y_i$ contains $y_{i-1}$ and $y_{i+1}$, each of which has their own defining equations containing other $y$s. Therefore there is a system of *simultaneous* equations to be solved. We begin with the model in which the measured variable $y$, given as a vector, $\mathbf{y}$, is linearly dependent on some independent underlying variables, $x_1$, $x_2$, $x_3$, . . ., given in matrix $\mathbf{X}$:

$$\mathbf{y} = \beta\mathbf{X} + \mathbf{u}, \tag{8.52}$$

where $\mathbf{u}$ is a vector of possibly non-independent errors with a mean of zero and variance–covariance matrix $\mathbf{C}$. Spatial autocorrelation is introduced into the model by having the errors in $\mathbf{u}$ autocorrelated:

$$\mathbf{u} = \rho\mathbf{W}u + \varepsilon, \tag{8.53}$$

where $\varepsilon$ is a vector of independent standard normal error terms. The matrix $\mathbf{W}$ is the neighbour weights (standardized to row totals of 1). In this case, $\mathbf{W}$ is not necessarily symmetric, making it possible to include the effects of water currents, prevailing winds, or other factors that might impose directionality on the autocorrelation effects. The model is now:

$$\mathbf{y} = \beta\mathbf{X} + \rho\mathbf{W}(\mathbf{x} - \mathbf{Z}\beta) + \varepsilon \tag{8.54}$$

and the variance–covariance matrix (associated with $\mathbf{u}$) is:

$$\mathbf{C} = \sigma^2[(\mathbf{I} - \rho\mathbf{W})^{\mathrm{T}}(\mathbf{I} - \rho\mathbf{W})]^{-1} \tag{8.55}$$

(see Bailey & Gatrell 1995 for the derivation). Bailey & Gatrell (1995) gave a simple example for $n = 3$ and $\rho = 0.5$ (see Figure 8.15*a*): if

$$\mathbf{W} = \begin{bmatrix} 0.0 & 0.3 & 0.7 \\ 0.2 & 0.0 & 0.8 \\ 0.7 & 0.3 & 0.0 \end{bmatrix}, \tag{8.56}$$

then

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1.81 & 0.97 & 1.31 \\ 0.97 & 1.69 & 1.15 \\ 1.31 & 1.15 & 1.90 \end{bmatrix}. \tag{8.57}$$

In this example, the variances (the elements on the main diagonal) are not all the same, but they are all greater than 1.0 because the variances are 're-inforced' by the correlation with the other values. The variance

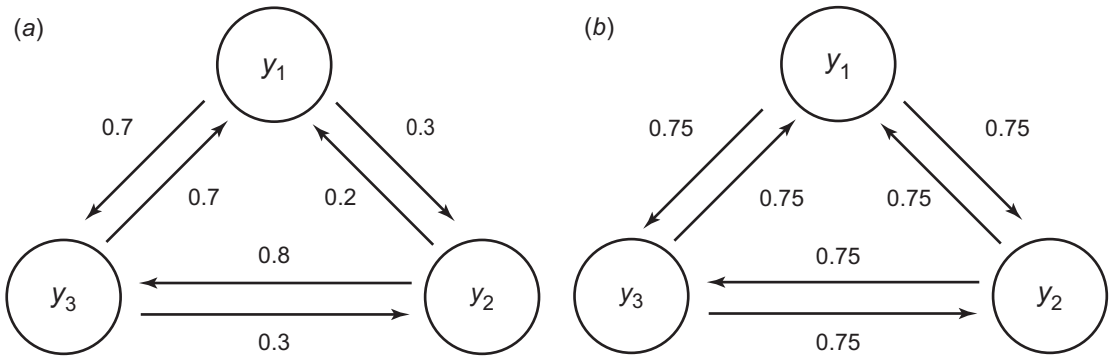



**Figure 8.15** (*a*) Illustration of the asymmetric weight matrix given in Eq. (8.56). (*b*) Illustration of the symmetric weight matrix given in Eq. (8.61).

of the third unit is the greatest because it has the largest values of the neighbour weights.

Although the SAR model is used extensively, for some technical reasons, many statisticians emphasize the use of the CAR model instead. The conditional autoregressive model is not based on the linear dependence of the value at a particular location on the values of its neighbours, but the probability that it takes a particular value is *conditional* upon the neighbour values:

$$P(y_i = y) = P(y_i = y|\{y_j; w_{ij} > 0\}). \tag{8.58}$$

It is not that much different from the SAR model, but it requires that the weight matrix, $\mathbf{V}$, must be symmetric. The model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \tag{8.59}$$

where $\mathbf{u}$ is a vector of errors with a mean of zero and a variance–covariance matrix that includes the autocorrelation parameter $\varphi$:

$$\mathbf{C} = \sigma^2(\mathbf{I} - \varphi\mathbf{V})^{-1}. \tag{8.60}$$

A simple CAR example for $n = 3$ is illustrated in Figure 8.15*b* (the values are not required to be equal):

$$\mathbf{V} = \begin{bmatrix} 0.0 & 0.75 & 0.75 \\ 0.75 & 0.0 & 0.75 \\ 0.75 & 0.75 & 0.0 \end{bmatrix}, \tag{8.61}$$

with $\varphi = ⅓$, this gives

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1.2 & 0.4 & 0.4 \\ 0.4 & 1.2 & 0.4 \\ 0.4 & 0.4 & 1.2 \end{bmatrix}. \tag{8.62}$$

Again, the values on the main diagonal are greater than 1.0, but here they all have the same magnitude.

These two kinds of models are closely related and any SAR process is a CAR process with:

$$\mathbf{V} = \mathbf{W} + \mathbf{W}^{\mathrm{T}} - \mathbf{W}^{\mathrm{T}}\mathbf{W}, \tag{8.63}$$

but the converse is not true (see Ripley 1981).

### 8.4.2 Spatial filtering using moving average models

Haining (1978) advocated the advantages of another type of spatial model: the moving average (MA) models. For lattice or grid data, it seems natural to consider a moving average based on rook's move neighbours:

$$y_{ij} = \varepsilon_{ij} + (\rho_b\varepsilon_{i-1,j} + \rho_f\varepsilon_{i+1,j} + \rho_u\varepsilon_{i,j-1} + \rho_d\varepsilon_{i,j+1})/4 \tag{8.64}$$

or, more simply:

$$y_{ij} = \varepsilon_{ij} + \rho(\varepsilon_{i-1,j} + \varepsilon_{i+1,j} + \varepsilon_{i,j-1} + \varepsilon_{i,j+1})/4. \tag{8.65}$$

The MA models have the advantage that autocorrelation can be made to decline sharply with distance and

become more-or-less zero at close range, whereas in autoregressive models it tends to persist over greater distances. In general form, the MA model is

$$\mathbf{y} = \mathbf{X}\beta + \rho\mathbf{W}\varepsilon + \varepsilon, \tag{8.66}$$

and the variance–covariance matrix is

$$\mathbf{C} = \sigma^2[(\mathbf{I} - \rho\mathbf{W})^{\mathrm{T}} + (\mathbf{I} - \rho\mathbf{W})]. \tag{8.67}$$

This looks similar to the equation for the SAR model, but there is no inverse in the formula. For the same symmetric proximity matrix, **W**, the variance–covariance matrices that arise for the three different models will, in general, be different.

These models can be viewed as an aid to understanding; the fact that we can estimate the parameters of a model and get a good agreement with the data does not mean that we know the underlying process. For real data, we probably do not really know even the proximity matrix, **W**. The other useful characteristic of such models, however, is that we can use them to generate artificial data of known structure, with which to compare what we have observed in the data we are trying to analyse. The use and comparison of these types of models seems to be a good approach to studying this phenomenon.

### 8.4.3 Spatial filtering using Moran's eigenvector maps

The previous type of spatial filtering models (SAR, CAR, MA) mostly deal with spatial structure based on local neighbourhood scale, but intermediate and large scales can also affect the spatial pattern of the data. To be able to account for multiple spatial scales in a regression framework, Moran's eigenvector maps (MEMs) can be used as spatial predictors as introduced in Chapter 7. Because there are as many orthogonal MEM as there are sampling locations ($n$), we first need to select a subset ($k$ out of $n$) of the MEM that best explains the variation in the response variable $y$. Then this subset of $\mathrm{MEM}_k$ can be added as spatial predictors in a linear model as follows:

$$\mathbf{y} = \beta_x\mathbf{X} + \beta_k\mathbf{MEM}_k + \varepsilon, \tag{8.68}$$

where $\beta_x$ and $\beta_k$ are the coefficients of regression for the **X** predictors and the $\mathrm{MEM}_k$, respectively. The MEM can also be used in an ordination analysis (e.g. RDA, CCA) as a table of spatial predictors or to partial out their effects in a partial ordination technique (e.g. partial RDA, partial CCA). The selected $\mathrm{MEM}_k$ can be used for insights about which spatial scales the spatial patterns of the data respond to and help to identify potential processes that can act at these key spatial scales. Yet as mentioned in Chapter 7, while the percentage of variance explained by models that include MEM as spatial predictors, the MEM are not process-based factors that can be used for prediction at different regions or time periods. Furthermore, Beale *et al.* (2010) found biased parameter estimates and inflated Type-I error rates when using MEM in spatial filtering models.

### 8.4.4 Spatial error regression

When the residuals of a linear regression are spatially autocorrelated this implies that the $\mathrm{var}(\varepsilon) = \sigma^2\Sigma$, where $\Sigma$ is a spatially structured $n \times n$ matrix. These values are proportional to the distance between the sampled data based on autocovariance functions. Hence the off-diagonal elements of the covariance matrix are not zeros. In such conditions, we should reanalyse the data using a spatial error regression method such as generalized least-squares regression (GLS; Pinhiero & Bates 2000; Perez *et al.* 2010) and regression Kriging (Hengl *et al.* 2004; Hengl *et al.* 2007).

Generalized least-squares regression can parameterize this spatially structured covariance matrix of the errors. One way to parameterize this matrix is to use an autocorrelation function: an inverse distance function, an autoregressive function, or a variogram model. Hengl *et al.* (2004, 2007) proposed using variogram models for the covariance matrix and named their model a regression Kriging model. Yet in essence, GLS and regression Kriging are the same method. The advantage of using variogram models is that anisotropic spatial structures can be modelled as well as isotropic ones.

Comparing several non-spatial and spatial regression models, Beale *et al.* (2010) concluded that overall GLS performs well in absolute bias and root mean square error (RMSE). Hence using GLS to

account for spatially autocorrelated data in a regression framework is a good way to obtain reliable parameters' estimate.

### 8.4.5 Geographically weighted regression

When the spatial structure of the data is due to both spatial autocorrelation and spatial heterogeneity (several different regions) then other kinds of spatial regression need to be used. The geographically weighted regression (GWR; Fotheringham *et al.* 2000, 2002; Fotheringham 2009) is one of these models. It computes a regression model for each sampling location using neighbouring sampling locations only to parameterize the model. Hence, the usual regression equation expressing the dependence of $y$ on a set of $x$s:

$$y = \alpha + \sum_k \beta_k x_k + \varepsilon, \tag{8.69}$$

is modified so that the parameters are not stationary, but are able to vary with location, **s**:

$$y(\boldsymbol{s}) = \alpha(\boldsymbol{s}) + \sum_k \beta_k(\boldsymbol{s}) x_k + \varepsilon. \tag{8.70}$$

Thus, each observation of $y$ has a separate model with its own parameter estimates, based on a weighting function (also referred to as a kernel) that allows closer samples to have a greater effect on the estimates. The shape and the size of the kernel can vary. Usually a bell-shaped Gaussian kernel is used, so that nearby observations have more weight than ones farther away. Then the size of the kernel can be the same (fixed kernel) throughout the study area or it can vary (adaptive kernel) according to location to produce the same number of observations used in the estimation of the parameters. Because the estimates are local, they can be mapped to produce a spatially explicit result.

This technique illustrates the interplay of independent predictor variables (the $x$s) and location in explanatory modelling. For example, we can combine the effects of the two kinds of variables, the spatial coordinates, $\mathbf{s} = (s_1, s_2)$, with potentially nonlinear effects, and the environmental variables, the $x$s, with possibly non-stationary effects, in a model to explain the $i$th observed value of $y$:

$$y_i = \sum_{p=0} \sum_{q=0} \alpha_{pq} s_{1\mathrm{i}}^p s_{2\mathrm{i}}^q + \sum_{k=1} \beta_k(s_i) x_{ki} + \varepsilon_i. \tag{8.71}$$

If the spatial response is simple and linear and the response to the $x$s is stationary, with the subscript $i$ omitted, this model simplifies to:

$$y = \alpha_{00} + \alpha_{10} s_1 + \alpha_{01} s_2 + \alpha_{11} s_1 s_2 + \sum_{k=1} \beta_k x_k + \varepsilon. \tag{8.72}$$

Of course, because the $x$s may also depend on location, spatial effects and environmental effects may be (probably are!) confounded, even if the parameters are stationary so that the $\beta$'s do not depend on **s**.

Explanation and the effort to distinguish space effects from 'environment' effects originated in ecological studies, but the 'environment' can be any set of factors other than spatial location. The total variation in the data can be partitioned into four by identified sources of the variation:

(a) explained by environment variables alone;
(b) explained by confounding of environment and spatial variables together;
(c) explained by spatial variables alone;
(d) explained by neither.

GWR has been used in several studies (e.g. Foody 2004; Osborne *et al.* 2007; Windle *et al.* 2010). Yet because there are as many models as there are sampling locations, GWR should be used as an Exploratory Spatial Data Analysis (ESDA) to identify subregions within the study area where the relationship between the response and the predictors have comparable parameter values. Hence GWR cannot be used as a predictive method. Furthermore, Griffith (2008) stressed that GWR is an overfitted approach and hence it should be used with caution.

### 8.4.6 Remove spatial autocorrelation from the residuals

In recent years several alternatives have been developed, the first using Bayesian spatial models, and the second using wavelets, in order to remove spatial autocorrelation from the residuals of models. Aing *et al.* (2011) described the use of a Bayesian

hierarchical occupancy model for assessing the occurrence of river otter (*Lontra canadensis*) using snow-track data from helicopter surveys. These data clearly include some autocorrelation from more than one source, and the model used included 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 using an intrinsic conditional autoregressive (CAR) model, with evaluation of parameters by the MCMC procedure. They found that a spatial model gave more accurate estimates for the detection parameters and better credibility intervals for spatially autocorrelated data.

Ver Hoef *et al.* (2006) proposed adding anisotropic autocovariance to model dendritic networks such as stream networks (Peterson *et al.* 2013) using directional functions to capture stream anisotropy:

$$\mathbf{y} = \mathbf{X}\beta + \mathrm{z}_{\mathrm{TU}} + \mathrm{z}_{\mathrm{TD}} + \mathrm{z}_{\mathrm{E}} + \varepsilon, \tag{8.73}$$

where the $\mathrm{z}_{\mathrm{TU}}$ is the tail-up vector based on a moving-average function (tail in the upstream direction), $\mathrm{z}_{\mathrm{TD}}$ is the vector based on a moving-average function (tail in the downstream direction), and $\mathrm{z}_{\mathrm{E}}$ is a correlation structure based on Euclidean distance between the sampling locations. This model has much potential for the study of stream 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. The particular class of models considered is that of latent Gaussian models, with a latent Gaussian field, *x*, at least partially observed through the data variable, *y*. The field *x* can include non-spatial or spatial random effects, and *y* can follow any of several distributions in the exponential family such as Poisson, normal, and so on. This allows the approach much latitude for application, and Beguin *et al.* (2012) provided more details on this, as well as an example investigating woodland caribou in Eastern Canada's boreal forest. They used both the conditional autoregressive model (CAR) and the Matérn model (see Minasny & McBratney 2005 for details of that model), and 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. The approach also allowed a good evaluation of the uncertainty of the parameter estimates.

Haas *et al.* (2011) described another application of the INLA approach to ecological data, using the intrinsic conditional autoregressive error model in a study of the effects of forest diversity on disease risk during the invasion of a generalist plant pathogen. Their results suggest that disease risk is less where species diversity is higher, once the effects of host density and landscape heterogeneity are accounted for. Clearly the correct treatment of spatial autocorrelation in such data is essential for the correct interpretation of the disease data and the insight into process they provide.

Another recently developing approach to solving autocorrelation in the residuals of linear regression and of more general linear models is based on wavelets. In a series of papers, Carl and colleagues (Dormann *et al.* 2007; Carl *et al.* 2008; Carl & Kühn 2008, 2010) have proposed and developed a simple technique called the 'wavelet revised model'. The wavelet of choice (they used the Haar wavelet) is used to remove a localized mean component from the data (the 'smooth' component), which on subtraction leaves the 'detail' component. For a range of scales of wavelet, say $2 \times 2$ blocks of the original data, then $4 \times 4$, $8 \times 8$, and so on, the 'detail' components are added together. This then forms the basis for subsequent analysis with the autocorrelation substantially removed. In concept, this approach is very similar to the method proposed by Bartlett in 1948 for the localized analysis of periodograms, which removes differences due to the local means, the results of which are then summed. There are several advantages to this wavelet approach, including the fact that it requires no a priori knowledge or assumptions about the spatial structure, and it deals well with non-stationarity in the autocorrelation structure because the wavelets make the 'corrections' that are applied to the data local in scope.

At the time of writing, there are obviously several exciting developments taking place, and thus an ongoing challenge of making choices in the future as those choices are further worked out and clarified. The main question to consider is to what extent the good performance of any particular approach is robust or is sensitive to departures from the models and assumptions that underlie their development. That seems to be a strength of methods where corrections, of whatever kind, are derived from the characteristics of the data, rather than from a priori structural details. That provides an ability to respond to the characteristics of the data themselves, and may provide greater confidence in the outcomes of what may be complex or not-so-transparent methods.

### 8.4.7 Example of the use of non-spatial and spatial regressions

To illustrate how non-spatial and spatial regressions behave when applied to spatially autocorrelated data, we investigated how the relationship we can detect between bird distributions and landcover types varies according to the regression method used. Specifically, we studied how the local variability in data on bird occurrence is related to the local variability in landcover types over a study area of 920 000 km$^2$ in southern Ontario (Canada), based on 920 sampling units of 10 × 10 km each. The local variability was assessed by computing the rate of change among adjacent sampling units using the lattice-wombling algorithm (see Chapter 9 for mathematical details). Rates of change are the first partial derivatives of a quantitative variable in both the $x$ and $y$ directions. The rates of change for the bird data were based on the occurrence of 60 species (mostly passerines; Figure 8.16*a*) as given in the Ontario breeding bird atlas (OBBA 2001–2005; Cadman *et al.* 2007). Of the 60 species, 27 species are associated with woods and forest habitat; 11 with grassland, agricultural and open habitat; 10 with shrub and early successional habitat; seven with urban and suburban habitat; and five with wetlands (see Polakowska *et al.* 2012). The Ontario Land Cover data (OMNR 1998) were used to quantify the rates of change in the amounts of landcover types per sampling unit (agriculture, mixed forest, deciduous forest, coniferous forest, water bodies, urban; see Figure 8.16*b*).

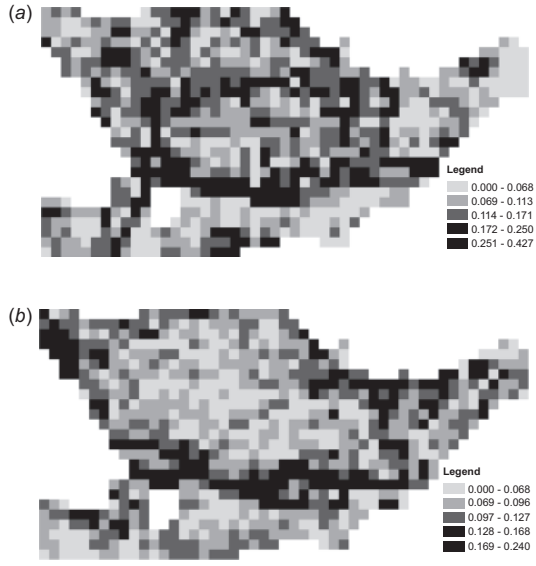



**Figure 8.16** Maps of the rates of change using 10 × 10 km sampling unit in southern Ontario based on the average rates of change of: (*a*) 60 species, (*b*) six landcover types.

Before performing the regressions, we tested for significant spatial autocorrelation in the data using Moran's $I$, with 19 equidistant classes of 14.5 km each. The rates of change of the birds and of the landscover types both have significant spatial autocorrelation at the first distance lag (0.476 and 0.546 respectively) and the shape of their correlograms indicates a large patch and a trend (similar to the one illustrated in Figure 6.3). Hence both the response (bird data) and the predictor (landcover types) variables are spatially autocorrelated.

Also as a preliminary step, we calculated a Pearson correlation between the two variables and applied Dutilleul's algorithm to determine the effective sample size, $n'$, accounting for the degree of spatial autocorrelation of each variable. The Pearson correlation between the two variables is 0.668 and significant, even based on the effective sample size $n'$ of 58.18, which is much less than the nominal $n$ of 920 sampling units. It is clear that there is a significant

relationship between the variables but that the presence of spatial autocorrelation needs to be accounted for in these regression analyses.

The non-spatial regressions we used are ordinary least-square linear regression (OLS), OLS including $x$ and $y$ coordinates as predictors, and the generalized linear mixed model (GLMM). The linear regression between the two variables is significant and the $R^2$ is 44.53% (Table 8.9). The Moran's $I$ spatial correlogram of the residuals of the linear regression shows spatially autocorrelated structure where the shape of the correlogram indicates a trend and the value of spatial autocorrelation at the first distance lag is 0.159 and significant. Including $x$ and $y$ coordinates as predictors in the linear regression increases the $R^2$ to 45.52% and the degree of spatial autocorrelation in the residuals decreases to 0.146. Then adding in random effects to account for $x$-$y$ coordinates using a GLMM, the degree of spatial autocorrelation in the residuals reverted to 0.159. Using the 'watershed unit' as a random effect improved the AIC value but the degree of spatial autocorrelation in the residuals increased to 0.239. Including both the $x$–$y$ coordinates and the watershed units as random effects, improved both the AIC and the degree of spatial autocorrelation. The coefficient of regression through the various versions we tested ranged from 0.66 (GLMM) to 1.011 (OLS).

The spatial regressions we tested are GLMM with spatial error models (SEM), generalised least-square (GLS), GLS with SEM, simultaneous autoregressive (SAR), OLS with MEM as predictors, and geographical weighted regression (GWR). Adding a spatial error component to the model of GLMM improved the AIC value but not the degree of spatial autocorrelation in the residuals. The GLS gave values comparable to those of the OLS. The GLS with a linear SEM produced better results both in AIC and in the degree of spatial autocorrelation in the residuals. This improvement was less when the SEM used was spherical because the degree of spatial autocorrelation in the residuals increased. With the SAR, there was no significant spatial autocorrelation in the residuals and the coefficient of regression was 0.7842. Using MEMs as spatial predictors with OLS was an improvement over the OLS alone, but not by much. Finally, the GWR analysis gave the best AIC (and adjusted pseudo-$R^2$) with the no significant spatial autocorrelation in the residuals, but at the cost of having 920 regressions (Figure 8.17). The coefficients of the spatial regressions ranged from 0.7 to 1.106.


**Figure 8.17** Map of the geographical weighted regression (GWR) pseudo-$R^2$ for each 10 × 10 km sampling unit of the relationship between the average rates of change of bird and landcover types in southern Ontario.

It is clear from these results that, depending of the type of regression used (non-spatial and spatial), the estimation of the coefficient of regression of the fixed effect (landcover types) will vary, just as Bini *et al.* (2009) showed. While the best model overall is the SAR, a spatial filtering model, its weakness is that the applicability of the neighbourhood values as spatial predictors is limited to the study area. The next best model seems to be the GLS + SEM(linear), which provides a compromise between the AIC and the degree of spatial autocorrelation in the residuals. More work is needed to bring a better understanding of the effect of spatial autocorrelation on the estimation of regression coefficients. Certainly, when spatial regressions are used, the estimated coefficients of regression have lower values than those estimated by the non-spatial models. This makes sense because both variables are spatially autocorrelated and probably affected by the same latent variables not included explicitly in the analyses. Therefore, when it is possible, we should first include other predictors in the analyses before using spatial regression, and then use spatial regression if spatial autocorrelation is present in the residuals.

**Table 8.9** The relationships detected between the average rates of change based on 60 bird species and eight landcover types in southern Ontario using non-spatial (linear regression, 'OLS'; generalized linear mixed model, 'GLMM'; and generalized least-squares, 'GLS') and spatial regressions (simultaneous autoregressive model, 'SAR'; LM with Moran's eigenvector maps as spatial predictors, 'LM + MEMs'; and geographical weighted regression, 'GWR'). All the parameters given are significant unless indicated by 'n.s.'

| | Intercept | $\beta$ | $\lambda$ | SEM range | SEM nugget | Adj $R^2$ | AIC | Residuals Moran's $I$ at first lag |
|---|---|---|---|---|---|---|---|---|
| OLS | 0.0172 | 1.0106 | | | | 0.4453 | | 0.159 |
| OLS + $x$–$y$• | 0.0388 | lc = 0.9662 | | | | 0.4552 | | 0.146 |
| | | $x = -0.00025$ | | | | | | |
| | | $y = -0.00002$ | | | | | | |
| | | n.s. | | | | | | |
| GLMM \| $x$–$y$† | 0.0169 | 1.0127 | | | | | −3357 | 0.159 |
| GLMM \| w†† | 0.0491 | 0.6808 | | | | | −3537 | 0.239 |
| GLMM \| $x$–$y$ + w††† | 0.0234 | 0.9551 | | | | | −3373 | 0.170 |
| GLMM \| w + SEM (linear) ♮ | 0.0509 | 0.6609 | | 6.5089 | 0.0895 | | −3637 | 0.245 |
| GLMM \| w + SEM (spher.) ♯ | 0.0511 | 0.6584 | | 9.8985 | 0.0996 | | −3651 | 0.246 |
| GLS | 0.0172 | 1.0106 | | | | | −3358 | 0.159 |
| GLS + SEM(linear) | 0.0224 | 0.9237 | | 6.3345 | 0.0907 | | −3550 | 0.110 |
| GLS + SEM(spher.) | −0.0102 | 0.7008 | | 347 768 | 0.0 | | −3596 | 0.234 |
| SAR | 0.0316 | 0.7842 | 0.6 | | | | −3622 | 0.011 n.s. |
| OSL+MEMs‡ | 0.0209 | 0.9514 | | | | 0.4897 | | 0.150 |
| GWR Global‡‡ | 0.0172 | 1.0106 | | | | 0.7781 | −4234 | −0.05 n.s. |
| GWR min. | −0.1158 | −1.7460 | | | | 0.0491 | | |
| GWR max. | 0.3110 | 3.0760 | | | | 0.8848 | | |

lc indicates landcover types.

$\lambda$ indicates the degree of spatial autocorrelation among first-order neighbors.

SEM range indicates the spatial range of the spatial error model.

SEM nugget indicates the nugget of the spatial error model.

AIC indicates the Akaike Information Criterion.

• $x$–$y$ correspond to the $x$–$y$ coordinates of the sampling units of 10 × 10 km.

† | $x$–$y$ indicates that the $x$–$y$ coordinates of the sampling units of 10 × 10 km are used as random effects.

†† w indicates that watershed units are used as random effects.

††† both $x$–$y$ and watershed units are used as random effects.

♮ SEM(linear) indicates that the spatial error model used is linear variogram.

♯ SEM(linear) indicates that the spatial error model used is a spherical variogram.

‡ MEMs used: 2, 7, 19, 30, 60, 72, 90, 97, 118, 131, 141, 152,159, 170, 188, 205, 289, 301, 350, 505, 552.

‡‡ GWR is computed for each sampling units and the average of the parameters are show in the row 'GWR Global'. Then the minimum (min.) and maximum (max.) values are showed in the two subsequent rows.

## 8.5 Considerations for sampling and experimental design

The material of this chapter, more than in any other, relies heavily on the use of mathematical models. The reason for this feature is that it is the only way to gain insight into the effects of different forms of spatial autocorrelation on the testing procedures we use. As stated above, and well demonstrated, single realizations of even simple models can appear very different, especially for small values of $n$, and real data must have similar problems, even if the underlying processes are actually stationary (and worse if they are not!). Much of the treatment of spatial autocorrelation in the statistical literature is predicated on the simplest autoregressive model, which produces an exponential decline in autocorrelation as a function of distance (Figure 8.18). In the geostatistical approach, it produces the 'typical' variogram depicted in Figure 6.8. One concern for our discussion of the application of models to the consideration of autocorrelation is that the usual models are not good descriptions of the structure of ecological data and that there is a tendency for ecological variograms not to converge to an asymptote, but to rise and fall nonrandomly as a function of distance, as would result from patchiness in the variable of interest. ("The natural world is a patchy place"! Dale 1999). That suggests, however, that simple corrections based on the first-order autoregressive model may often be incorrect or misleading, although the matrix trace approach to correction avoids the weakness of model-based corrections. It is important to consider then how our understanding of spatial autocorrelation and its effects can be used to inform our decisions in designing sampling regimes or the spatial layout of experiments.

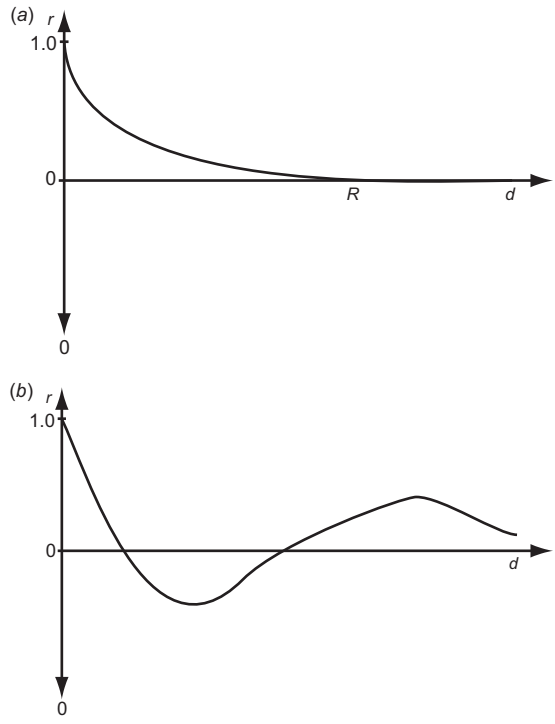



**Figure 8.18** Two models of how autocorrelation may vary with distance. (*a*) Autocorrelation falls to zero at some distance R, known in geostatistics as the range. This is equivalent to the 'exponential' model in geostatistical analysis. (*b*) Autocorrelation falls below zero and then fluctuates above and below zero.

### 8.5.1 Sampling

The detection and testing of spatial autocorrelation is an important step in understanding ecological data and one that must be considered in analysis and interpretation (Chapter 1). It is probably also apparent to the reader that, if the underlying autocorrelation structure were known in advance, the design used for sampling or the design of an experiment could be adjusted to minimize the effect of spatial autocorrelation on the outcome of the study, before any analysis was conducted. For example, if the autocorrelation declines rapidly with distance and becomes effectively zero beyond some distance, $R$, (as in Figure 8.18*a*) samples or experimental units with spacing $R$ or greater might be treated as independent. This is the concept of 'distance to independence', which (alas) probably almost never applies in ecological studies (see Chapter 12). Ecological variables are typically

patchy, often at more than one scale, so that their autocorrelation cycles between positive and negative with increasing distance (as in Figure 8.18*b*). Under those circumstances and with knowledge of the locations of regions of low values and regions of high values, sampling could then be stratified or the experimental units could be placed in positions of known characteristics. On the other hand, if autocorrelation declines only very slowly with distance, the distance between samples or experimental units may not be that important. In the intermediate case, where autocorrelation declines appreciably at the scale of the extent of the study, experimenters may wish to use a design that provides a balanced set of distances between units assigned the same treatment (van Es & van Es 1993; Dutilleul 1993a; Legendre *et al.* 2004).

Whatever the behaviour of the autocorrelation, the important first step is to determine its characteristics before designing the sampling scheme or the experiment, and to do so, a pilot study is required (Legendre *et al.* 2002). In general, a pilot study will involve taking a number of samples according to some scheme that will allow an evaluation of autocorrelation, including its behaviour as a function of distance and whether it is anisotropic. We will consider these to be point samples of some kind, whether measurements of altitude, moisture or pH, estimates of population density, or the presence of a particular substrate. There are several different ways in which these point samples can be arranged, and they all have advantages and disadvantages. Random placement has the advantage of simplicity, but the disadvantage of a lack of control over the spacing of the points, the coverage of the study plot, and the range of lags available in each direction. A regular grid of equally spaced points has an inherent direction and scale of its own, which may interact with the actual pattern in the variable under study. The 'wagonwheel' design of radiating lines of sample points will produce a trend in sampling intensity and in circumferential distances from the centre of the wheel to the edge. A design based on the Fibonacci spiral avoids trend and directionality problems, but the full spiral could require very high intensity of sampling and has a trade-off between the number of samples

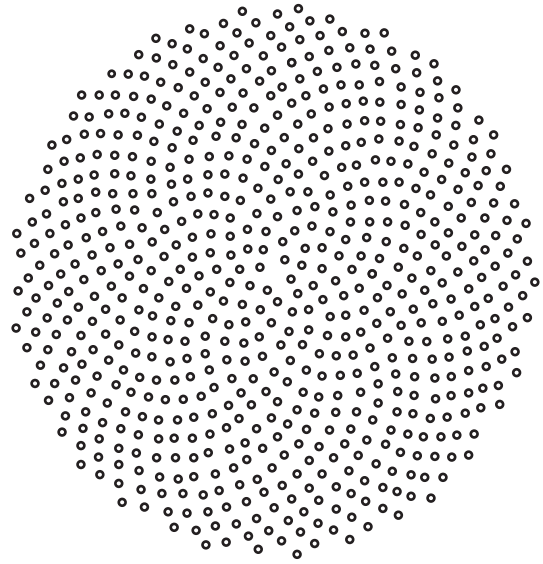


**Figure 8.19** Fibonacci spiral defined by angle $\theta$ and radius $r$, with scaling parameter $\kappa$ and the 'golden mean' parameter $\tau$. $\theta_i = 2\pi/\tau; \quad r = \kappa\sqrt{i}; \quad \tau = (1+\sqrt{5})/2.$

in the study area and the distances it can evaluate. The full spiral is illustrated in Figure 8.19, but the sampling designs that seem to combine most of the advantages with few disadvantages are partial Fibonacci designs in which only a systematic or randomly chosen subset of the full spiral is sampled. Whichever subset is chosen, such a design offers the advantages of no directional bias, a good range of inter-sample distances, and a lack of trend in sampling intensity.

We shall end Chapter 8 with a summary of the findings of two studies on the effects of spatial structure on the design and analysis of field surveys and on the design of field experiments, by Legendre *et al.* (2002, 2004). Both studies used simulations of an environment with one of several different kinds of spatial structure (gradient, waves, etc.) in the underlying environmental variable as well as autocorrelation in it and in the variable of interest to address questions about the effect of these structures on design and analysis.

In the study on surveys (Legendre *et al.* 2002), the simulation structure can be described using $E_{ij}$ as the

environmental variable, $\rho_{E,ij}(R_E)$ as its autocorrelation component with range $R_E$, and $\varepsilon_{ij}$ as a standard normal error term:

$$E_{ij} = S_{ij} + \rho_{E,ij}(R_E) + \varepsilon_{ij}, \tag{8.74}$$

and

$$V_{ij} = \beta\, E_{ij} + \rho_{V,ij}(R_V) + \eta_{ij}. \tag{8.75}$$

Here, $V_{ij}$ is the variable of interest, $\rho_{V,ij}(R_V)$ is its autocorrelation component with range $R_V$, $\beta$ is a measure of its linear dependence on the environmental variable and $\eta_{ij}$ is a standard normal error term. Among the questions asked were the following.

(1) Can the effect of spatial autocorrelation on tests be reduced by the survey design?
(2) Which designs provide the greatest power given a particular combination of spatial structure and autocorrelation range?

Legendre *et al.* (2002) found that spatial autocorrelation in both the environmental and response variable affects the standard tests, spatial autocorrelation in only one does not. A broad-scale spatial structure in the underlying environmental variable, however, combined with spatial autocorrelation in the response variable, inflated Type I error, just as spatial autocorrelation in both variables would. The major piece of advice to be derived from this study was to use a pilot study to identify the underlying structure. If there is a gradient, its effect can be accounted for by using it as a covariate. If there are different zones (really representing non-stationarity), a covariable that distinguishes the zones should be used. Another finding of this paper was the reassurance that Dutilleul's method correcting the $t$-test of the correlation coefficient was found to be robust to various spatial structures tested.

### 8.5.2 Experimental design

The study of experiments (Legendre *et al.* 2004) had a list of questions parallel to those for the survey study, and the approach used was similar. The only modification is that the variable of interest now includes a treatment effect for experimental units, $\tau$, which can be 0 not clear whether: 0, low, or 0 (low), medium or high:

$$V_{ij} = \beta\, E_{ij} + \rho_{V,ij}(R_V) + \eta_{ij} + \tau_{ij}. \tag{8.76}$$

There are several general lessons to be learned from this study.

(1) If either spatial autocorrelation or repetitive structures like waves are present in the underlying environmental variable, randomly positioned experimental units should not be used. The use of blocks is recommended.
(2) For a set number of experimental units in the presence of spatial autocorrelation, using more, smaller blocks spread throughout the study provides greater statistical power.
(3) Short-range spatial autocorrelation (related to the size of the experimental units and the blocks of them) affects ANOVA tests more strongly than long-range spatial autocorrelation. For example, where the blocks were three, six, or more units in linear extent, autocorrelation with a range of four units (as opposed to 16 or 40 units) caused the greatest decrease in statistical power.

The approach used in these two studies is clearly a useful one and there is probably a lot more that we can find out about mitigating the effects of spatial autocorrelation on statistical tests by its further application. The 'bottom line', however, is that it is essential to have a good assessment of the nature and range of spatial autocorrelation before designing or carrying out a survey or experiment (see Figure 8.20).

## 8.6 Concluding remarks

The presence of spatial autocorrelation is an important characteristic of ecological data and it often has several sources that may not be distinguishable. Spatial autocorrelation affects the outcomes of statistical tests and it cannot be ignored. Even apparently nonsignificant amounts of autocorrelation can have a serious cumulative effect.

Furthermore, because the natural world is often patchy, autocorrelation may fluctuate between

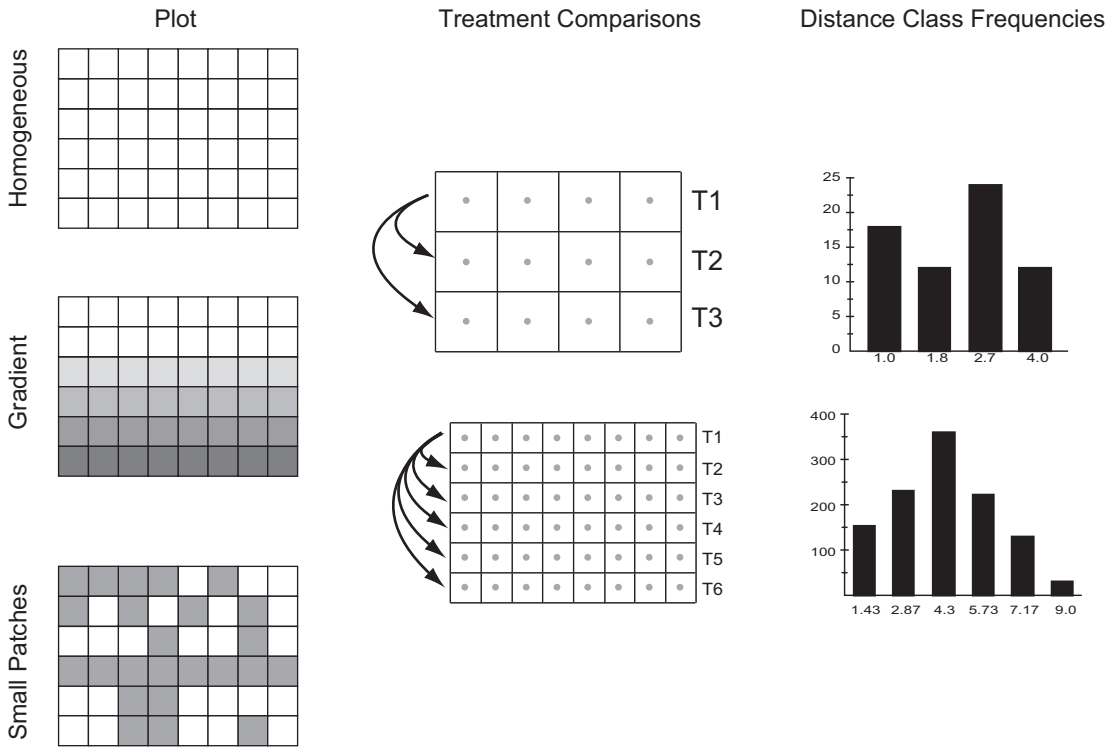



**Figure 8.20** Illustration of how the inherent spatial pattern of the data (e.g. homogeneous, gradient, small patches) prior to the experiment can influence its outcome depending on the experimental design: the number of treatments and the geographical distance between the experimental units. (Figure modified from van Es & van Es 1993.)

positive and negative, with increasing distance, which poses its own kind of challenge for correction and interpretation. In addition, we might question whether the similarity of adjacent samples (contiguous quadrats in a transect, for example) really represents redundancy in the data, the way it would in a simple model of autocorrelation. Having two samples in the same patch tells us something more than sampling the same individual twice.

In considering solutions to the effects of this phenomenon on statistical tests, we can offer some observations. The first and most important is that the problem seems to be mainly solved or solvable for bivariate and multivariate tests using the properties of the data themselves, thanks to the work of researchers like Dutilleul and Cerioli. It remains a major difficulty for univariate tests, although a number of solutions have been proposed. Thinning the data is not a good idea because it is wasteful of information and because it is based on the concept of distance to independence, which may well be wrong (see Chapter 12). Adjusting the effective sample size is a possible solution if the chosen model is a good description of the data's structure; we cannot usually calculate the effective sample size from the data. Randomization methods may work, but they must be applied carefully and with an awareness of the possible problems. Complete randomization of the data cannot and does not control the effects of spatial autocorrelation on statistical tests of significance. Restricted randomizations that maintain much of the spatial structure of the data are the most appropriate, as are related approaches such as block bootstrapping. The 'model and Monte Carlo' method has several features that recommend

it, but it is not perfect. As we showed with artificial data, the best-fitting model may not be the original model that generated the data; for field data, we will not know the relationship between the model fit and the underlying structure. Larger sample sizes should improve the accuracy of the modelling exercise, but they increase the risk of encountering non-stationarity if the extent of the sampling is also increased.

Studying the relationship between autocorrelated variables requires careful consideration of the structure of the variation, particularly in ecology, where autocorrelation may be both inherent and induced in a single variable, and where it may not be well described by one of the standard models. This is why more and more ecologists are using spatial regression models. Yet we need to understand well the source of the spatial structure to account for it appropriately. Of particular importance is the fact that removing the variables' dependence on an underlying factor does not avoid the general problems associated with analysis in the presence of spatial autocorrelation.

In considering the design of sampling or experimentation, the important first step is to determine the characteristics of the spatial autocorrelation before doing anything else. Knowing that structure will enable the researcher to develop a design that avoids, or at least reduces, the effects of autocorrelation on subsequent analysis. Pilot studies are a necessity and may be combined with other prior knowledge to produce effective designs.

In 1993, Legendre posed the question "Spatial autocorrelation: trouble or a new paradigm?" as the title of a paper that provided a wide-ranging discussion of this topic. The answer, 20 years later, is that it is both. It is troublesome, not in the sense of being just a nuisance, but because it is not an easy phenomenon to deal with, and it is certainly part of the current approaches to ecology, which include spatial structure (they must) or provide spatially explicit results. We must remember, however, that it is this sort of lack of independence through space and time that makes any prediction possible. We would be in trouble, indeed, if ecological phenomena were spatially and temporally independent (if that were possible). Clearly, the characteristics, effects and corrections for spatial autocorrelation, and other sources of spatial dependence, require and are worth a good deal more effort and thought, before we can suggest that we understand them truly and thoroughly.