

# Species occurrence and occupancy

Kevin J. Gaston and Fangliang He

## 11.1 Introduction

Arguably, one of the most fundamental units of biodiversity is the presence or absence of a species in a given site (a resource or habitat patch, an island, a mapping unit, etc.). This is evidenced by the fact that simple species  $\times$  sites ( $r \times c$ ) matrices, in which the presence/absence of different species (in rows  $r$ ) is given for a set of different sites (in columns  $c$ ), lie at the heart of probably the vast majority of biodiversity studies (Simberloff & Connor 1979; Bell 2003; Arita et al. 2008; Gaston et al. 2008a). Indeed, some of the most basic patterns which emerge from presence/absence  $r \times c$  matrices have long intrigued ecologists, such as species–area relationships, nestedness, and gradients in  $\beta$  diversity.

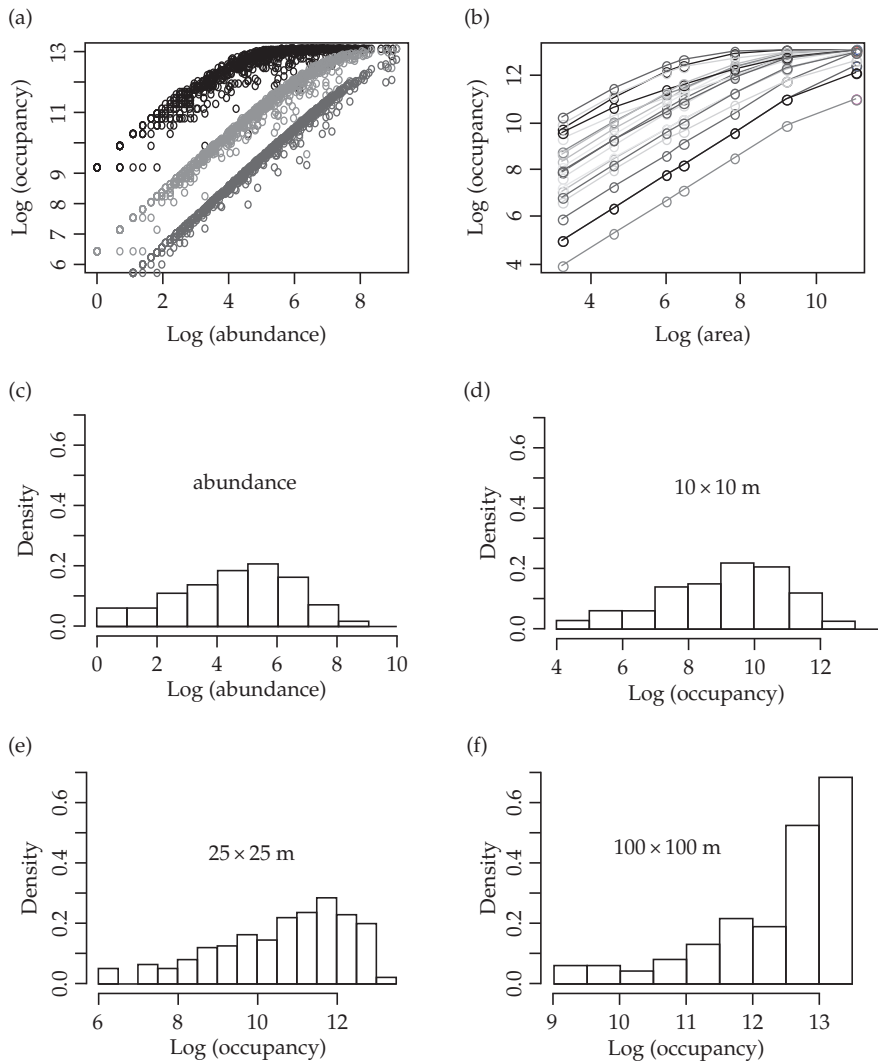
Summing the presences along a row of an  $r \times c$  matrix gives the level of occupancy of a given species, often most usefully expressed in terms of the proportion ( $p$ ) of the available sites that are actually occupied (or, equivalently, the probability that the species occurs in any one site). This may vary greatly from one species to another, and the question of why some species are widely distributed and others narrowly has long been a staple of ecological discourse (e.g. Darwin 1859; Harper 1981; Rabinowitz 1981; Gaston 1994; Kunin & Gaston 1997). The actual area over which a species is distributed (i.e. the summed areas of the sites at which it occurs) has been termed its ‘area of occupancy’ (Gaston 1991; Gaston & Fuller 2009), and is important for some, particularly applied, considerations, such as evaluating the risk of extinction that species face in the short term as a consequence of anthropogenic pressures (Mace et al. 2008).

The level of occupancy attained by a species is strongly influenced by the spatial resolution of the

occurrence data (i.e. by the size of a ‘site’). This is simply because the level of occupancy itself fails to capture other significant features of spatial distribution, particularly the way in which at a fine resolution occupied sites are spatially dispersed.

On average, of two species with the same occupancy at a fine scale, that with the more dispersed pattern of occurrences will have the greater occupancy at a coarser resolution (one simple widely used measure of this dispersion is the area contained within the geographically outermost occurrences of a species, its ‘extent of occurrence’; Gaston 1991; Gaston & Fuller 2009). Although it is only relatively recently that the analytical form of the relationship between spatial resolution and the observed level of occupancy of a species (Fig. 11.1; the occupancy–area relationship) has come to the fore, the potential significance of spatial resolution on observed patterns of occupancy has been recognized for much longer (e.g. Erickson 1945; Rapoport 1982).

The occupancy of a species is only a crude caricature of the distribution of its individuals. However, the level of occupancy attained across a set of sites is inevitably a function of the number of individuals summed across those sites. Only one site can be occupied when there is a single individual (assuming that sites are defined to be at least larger than an individual), and there is almost invariably some upper limit to the number of individuals that can physically co-occur in a single site, even ignoring all of the factors that can act to disperse individuals more widely (e.g. air and water currents, life history, behaviour). Much attention has been paid to the actual form taken by occupancy–abundance relationships (Fig. 11.1), being motivated initially mostly by the wish to be able to estimate the abundance of a species from its much more readily (and



**Figure 11.1** (a) Occupancy–abundance curves for 817 tree species from the 50 ha (1000  $\times$  500 m) Pasoh forest plot, Malaysia. The lowermost curve is constructed for the occupancy calculated by cell size  $a = 10 \times 10$  m, the middle curve is for  $a = 25 \times 25$  m, and the uppermost curve for  $a = 100 \times 100$  m. (b) Occupancy–area curves for 20 species randomly selected from the 817 species. (c) The distribution of the abundance of the 817 species, serving as a reference for the occupancy distributions of three cell sizes (d, e, f).

economically) estimated level of occupancy (e.g. Nachman 1981, 1984; Wilson & Room 1983; Kuno 1986; Ward et al. 1986). Although such considerations have in the main concerned agricultural pests (and the need for indicators of when to take control measures against them), more recently they have extended particularly to species of conservation concern (and the need to understand their

population sizes; e.g. Tosh et al. 2004; Sara 2008; Figueiredo & Grelle 2009).

By contrast, the frequency distribution of the occupancy levels of the species in an  $r \times c$  matrix (i.e. the distribution of the row totals), the species occupancy distribution (Fig. 11.1), has been subject to investigation by generations of ecologists (e.g. Willis 1922; Gleason 1929; McIntosh

1962; Williams 1964; Anderson 1977; Hengeveld 1990; Gaston & Blackburn 2000). The form of this distribution serves to move consideration beyond the question of why some species are more widely distributed than others to that of why typically a disproportionately high number of species are narrowly distributed, and the circumstances in which this is not the case (e.g. Hesse et al. 1937; Hanski 1982; Gaston & He 2002; McGeoch & Gaston 2002; Storch & Šizling 2002; Gaston 2003). Answers to this latter issue have become increasingly significant with recognition that at broad spatial scales the form of species occupancy distributions may be central to an understanding of many other macroecological patterns and processes (e.g. Gaston 2003; Šizling et al. 2009).

In this chapter we provide a broad review of the present understanding of these three key patterns in the study of biodiversity, namely occupancy–area relationships, occupancy–abundance relationships and species occupancy distributions. In so doing we touch on a wide variety of issues pertaining to the ecology of the occurrence and occupancy of species. Following the bulk of the literature on these patterns, and for reasons of clarity of presentation, throughout we largely ignore the influences of variation in the detectability of individuals of given species or between different species, although clearly these can have marked influences on estimated levels of occupancy and abundance, and in empirical work will often need to be given careful consideration (see Chapter 3).

## 11.2 State of the field

### 11.2.1 Occupancy–area relationships

It has long been known that the spatial distributions of the individuals of a species are seldom, if ever, strictly random or uniform. Rather, they are aggregated on multiple spatial scales, with individuals occurring in clusters that are themselves aggregated into larger clusters, which in turn are aggregated into larger clusters still, and so forth. Many methods are available for describing such intraspecific occupancy patterns (e.g. Taylor 1984; Krebs 1999).

One key approach is to characterize them in terms of simple models linking occupancy and the area,

*a*, of a site (here termed occupancy–area relationships, but elsewhere referred to as area–area, *p*–area or scale–area relationships; Kunin 1998; He & Gaston 2000a; Šizling & Storch 2004). Here sites are assumed to have some characteristic size or scale, usually because species occurrences are mapped onto an equal-area grid, from which cells at a fine resolution can easily be perfectly aggregated to form cells (and thus an occurrence map) at a coarser resolution. This is mathematically convenient, although clearly at odds with the marked heterogeneity of the natural environment and the complexities of the sizes of the set of habitat patches in which most species reside.

A large set of occupancy–area relationships take the general form

$$p = 1 - \left(1 + \frac{ca^z}{k}\right)^{-k} \quad (11.1)$$

where *a* is the area of the mapping unit, *c* and *z* are positive parameters, and *k* is a real number parameter in the domain  $(-\infty, -ca^z)$ , or  $(0, +\infty)$  (He & Condit 2007). For a given area and *k* greater than 0, as *k* increases occupancy also increases (Fig. 11.2). For a given area and *k* less than 0, occupancy declines as *k* assumes progressively greater negative values. This embraces many of the more familiar occupancy–area relationships (see section 11.2.2 for more details) as special cases (He et al. 2002), such that

- (i)  $k \rightarrow \pm\infty$  and  $c = z = 1$  gives the Poisson model

$$p = 1 - e^{-a} \quad (11.2)$$

- (ii)  $c = z = 1$  gives the negative binomial model

$$p = 1 - \left(1 + \frac{a}{k}\right)^{-k} \quad (11.3)$$

- (iii)  $k = -1$  gives the power-law model

$$p = ca^z \quad (11.4)$$

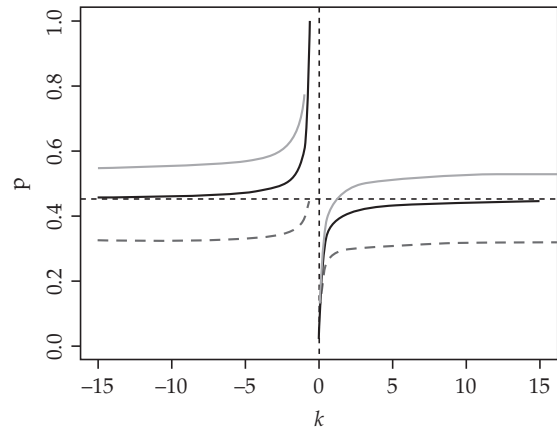
- (iv)  $k = 1$  gives the logistic model

$$p = \frac{ca^z}{1 + ca^z} \quad (11.5)$$

and

- (v)  $k \rightarrow \pm\infty$  gives the Nachman model

$$p = 1 - e^{-ca^z} \quad (11.6)$$



**Figure 11.2** Illustration of model (11.1) as a function of aggregation parameter  $k$ , where  $a = 0.6$ . Black curves:  $c = z = 1$ ; grey curves:  $c = 1$ ,  $z = 0.5$ ; dashed curves:  $c = 0.5$ ,  $z = 0.5$ . The dashed vertical and horizontal lines are the limits of the model. The occupancy  $p$  for the horizontal line is  $p = 1 - e^{-0.6} = 0.4512$ , which is model (11.2).

In all cases, occupancy tends to saturate as the resolution at which this is being documented becomes coarser (i.e. as  $a$  increases). The linearized forms of these models, which are therefore often of more value for analytical purposes, are respectively (i) Poisson model:  $\log(1 - p) = -a$ , (ii) negative binomial model:  $\log(1 - p) = -k \log(1 + a/k)$ , (iii) power-law model:  $\log(p) = \log(c) + z \log(a)$ , (iv) logistic model:  $\log(\frac{p}{1-p}) = \log(c) + z \log(a)$  and (v) Nachman model:  $\log[-\log(1 - p)] = \log(c) + z \log(a)$ . Note that the last three of these models share a common form in being functions of  $\log(a)$ . For a given set of species' occupancy data, the models can be fitted by using either a simple linear regression (with the exception of the negative binomial model, which requires a non-linear algorithm; He & Condit 2007) or a maximum likelihood method (He et al. 2002).

There has been limited comparative testing of these different models for intraspecific occupancy-area relationships, but it seems unlikely that any one of them will remain a consistently better descriptor of the patterns for different species (He & Condit 2007). However, whilst there is no evidence for its inherent superiority, versions of the power-law model ( $p = ca^z$ ) have been particularly widely used in examining occupancy-area relationships. This is because the nested clustering of individuals of a species has led to the view that occupancy patterns can best be understood in terms of a fractal paradigm (e.g. Lennon et al. 2007; Šizling & Storch 2007). Whether one strictly

agrees with this or not, at the very least this may be a useful way to think about the consequences of a habitat hierarchy and/or a hierarchy of the ecological requirements of species in which a region is divided into broad environments defined by climate, within which are areas defined by geomorphology and substrates, divided in turn into patches of macrohabitat each comprised of different subpatches of microhabitat, and so forth (Storch et al. 2008; see also Kolasa et al. 1989).

The fractal model is more typically described as

$$p = p_0 l^{D-2} \quad (11.7)$$

where  $p_0$  is a constant,  $l = \sqrt{a}$  (i.e.  $l$  is the length of one side of a square mapping unit), and  $D$  is the box-counting fractal dimension (Lennon et al. 2007). In a stochastic fractal, the probability that a particular sub-area of a larger occupied area is occupied is a constant (i.e. it is scale independent), and  $D$  is therefore a constant ( $p_0$  is also expected to be unity, although an empirical fit often allows it to vary; Lennon et al. 2007). Empirical studies have shown variable fits to a stochastic fractal model (e.g. Kunin 1998; Ulrich & Buszko 2003; Hartley et al. 2004; Šizling & Storch 2004; Lennon et al. 2007). Moreover, whilst a constant  $D$  may be a mathematically useful property, there is no necessary reason why it need in reality be constant (Šizling & Storch 2007) and  $D$  is often found to decrease with the increase of spatial scale ( $l$ ) (He & Condit 2007). A variety of approaches have thus been taken to relax this constraint (e.g. Hui & McGeoch 2007b; Lennon

et al. 2007; Storch et al. 2008). The approach of generalized fractals seems to be particularly promising (Šizling & Storch 2007; Storch et al. 2008). This retains the essential properties of fractals but does not necessitate strict self-similarity by randomly replacing an original set of areas by sets of smaller patches nested within them following a particular repeated process. Other related approaches to modelling occupancy–area relationships (and often related biodiversity patterns) have also been taken based on simple rules of the occupancy of nested grid cells (e.g. Harte et al. 2005; Conlisk et al. 2007).

### 11.2.2 Occupancy–abundance relationships

The general model for occupancy–area relationships described above (equation 11.1) is a simple modification of, and was originally derived from, a general model for occupancy–abundance relationships, making the assumption that mean density ( $\mu$ ) in the latter can be replaced by area ( $a$ ) in the former (He & Condit 2007). Thus the general model for relationships between occupancy and abundance takes the form

$$p = 1 - \left(1 + \frac{c\mu^z}{k}\right)^{-k} \quad (11.8)$$

or

$$\mu = \frac{k}{c} \left[ (1 - p)^{-1/k} - 1 \right]^{1/z} \quad (11.9)$$

(He et al. 2002). The associated variance–mean relationship, of considerable ecological interest itself (Gaston 2003), takes the form

$$\sigma^2 = c\mu^z \left( \frac{c\mu^z}{k} \right) \quad (11.10)$$

where  $\sigma^2$  is the (spatial) variance in abundance across areas (Holt et al. 2002a).

The special cases of this general model, derived under the same conditions as before (section 11.2.1), are generally more familiar in the context of occupancy–abundance relationships, many having been standard ecological fare for a very long time. The Poisson model ( $p = 1 - e^{-\mu}$ ) reflects the case in which individuals are distributed in space at random, and is widely adopted as a null model for occupancy–abundance relationships (Wright 1991).

However, as we have seen, the circumstances under which this actually occurs are sufficiently scarce that the utility of such a null is very doubtful, and it is usually rapidly dismissed as a fit to real ecological data except when species are very scarce (e.g. Pielou 1977; Taylor et al. 1978; Greig-Smith 1983; Gaston 1994; Brown et al. 1995; Hinsley et al. 1996; Venier & Fahrig 1998). The usual explanation for this outcome is that there are numerous abiotic and biotic reasons why species are unlikely to be randomly distributed. However, it should not be forgotten that because such a distribution is only one among a large continuum of possibilities it may itself be exceedingly unlikely to occur on a simple probabilistic basis (Taylor 1961).

The negative binomial model ( $p = 1 - \left(1 + \frac{\mu}{k}\right)^{-k}$ ) is the most frequently used to describe aggregated spatial patterns of the occurrence of a species (He & Gaston 2000a). However, although a large number of possible causal derivations have been identified, mostly based on the compounding of random processes (e.g. Boswell & Patil 1970; Taylor 1984), its suitability as a general descriptor of the spatial distributions of species has been much debated. Even where the negative binomial provides a reasonable fit to observed distributions of abundances, appropriate values of  $k$  are dependent on the mean density ( $\mu$ ; Finch et al. 1975; Taylor et al. 1978, 1979; Nachman 1981; Taylor 1984; Perry & Taylor 1985, 1986; Shorrocks & Rosewell 1986; Hassell et al. 1987; Rosewell et al. 1990; Feng et al. 1993). The negative binomial model has a variance–mean relationship of the form  $\sigma^2 = \mu + \left(\frac{\mu^2}{k}\right)$  (Routledge & Swartz 1991; Perry & Woiwod 1992; Gaston & McArdle 1994).

The power-law model ( $p = c\mu^z$ ) (Leitner & Rosenzweig 1997) follows a (positive) binomial distribution describing a regular distribution of species. It has a variance–mean relationship of the form  $\sigma^2 = c\mu^z(1 - c\mu^z)$ , with  $c\mu^z < 1$ . The power model is suitable for species of less aggregated or regular distribution, with the level of occupancy for a given abundance being greater than that for the Nachman model (below).

The logistic model ( $p = \frac{c\mu^z}{1+c\mu^z}$ ) (Hanski & Gyllenberg 1997) describes the situation in which the occurrence of individuals of a species follows a geometric distribution (He & Gaston 2000a). Departure

from this distribution is captured by  $c$  and  $z \neq 1$ . It has a variance–mean relationship of the form  $\sigma^2 = c\mu^z(1 + c\mu^z)$ , which is larger than the variance in Taylor’s power model (see below), suggesting that the logistic model is appropriate to describe patterns for species having stronger aggregation than that under the Nachman model.

The Nachman model ( $p = 1 - e^{-c\mu^z}$ ) (Nachman 1981, 1984) is suggested as an empirical generalization of the Poisson model with  $c$  and  $z \neq 1$  representing a departure from the Poisson distribution (He & Gaston 2000b). The Nachman model predicts a variance–mean relationship of the form  $\sigma^2 = c\mu^z$ . This is Taylor’s power function (Taylor 1961). It is the model that has been used most widely to describe intraspecific mean–variance relationships in empirical abundance data, although the generality of its suitability has repeatedly been challenged (Taylor 1984; Sawyer 1989; Routledge & Swartz 1991).

Similar to the occupancy–area models introduced in section 11.2.1, the fitting of occupancy–abundance models to empirical data can be done using linear (or non-linear) regression methods or maximum likelihood methods (He et al. 2002; He & Condit 2007). Occupancy–abundance relationships can be documented on the basis of intraspecific or interspecific data (Gaston et al. 2000). Note that commonly in the latter case, although less so in the former, these are documented with abundance averaged only over occupied sites ( $\mu'$ ) rather than over all sites ( $\mu$ ). From an ecological perspective this has been argued to remove the likelihood of an ‘artefactual’ positive occupancy–abundance relationship arising (occurrence in a smaller number of sites will tend to lead to lower  $\mu$  simply because the number of individuals is divided by the total number of possible sites). However, in terms of exploring the relationships between different patterns in occupancy and linking these to models in the broader statistical literature, it makes more sense to express occupancy–abundance relationships in terms of  $\mu$ , as we have done here.

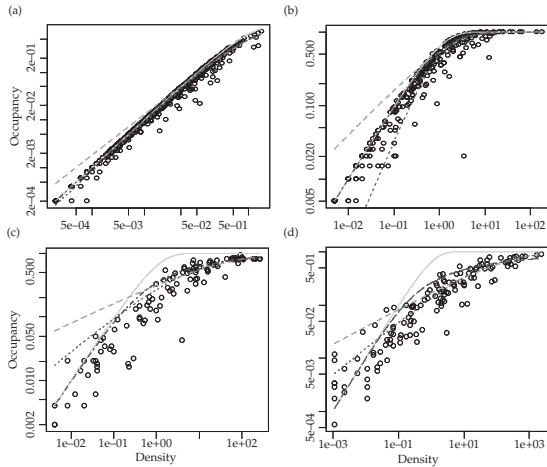
In the case of intraspecific data, an occupancy–abundance relationship is the equivalent of plotting occupancy at a coarser spatial resolution against occupancy at the resolution at which spatial units can only be occupied by a single individual of

the species. This can be done with the data points derived either across the same set of sites at different points in time (e.g. Gaston et al. 1998a, 1999b; He & Gaston 2003; Freckleton et al. 2005; Borregaard & Rahbek 2006), or across different sets of sites in different parts of the species’ geographic distribution (e.g. Venier & Fahrig 1998; He & Gaston 2003). In both cases, positive relationships are most frequently documented, although not invariably so, with a likely expected weakening for more widespread species (for which marked changes in abundance may be required, both statistically and ecologically, to result in detectable changes in occupancy). The fit of different forms of occupancy–abundance relationships has, however, seldom been tested for intraspecific data (but see He & Gaston 2003; Sileshi et al. 2006).

If intraspecific occupancy–area relationships were sufficiently similar between species, then positive interspecific occupancy–abundance relationships would tend inevitably to follow, as these would equate to plotting the occupancy achieved by a single species at a given coarser spatial resolution according to a particular scaling rule as its abundance varied. In fact, there is doubtless much latitude in how similar intraspecific occupancy–area relationships need to be in order for positive interspecific occupancy–abundance relationships to result. It is thus unsurprising that such positive interspecific relationships have been documented across a wide range of spatial scales, environments (terrestrial, freshwater, and marine), and higher taxa, almost invariably using data points derived for different species across the same set of sites over the same time period (Fig. 11.3; e.g. Gaston et al. 1998b; He & Gaston 2003; Selmi & Boulinier 2004; Heino 2008; Leger & Forister 2009). Indeed, a meta-analysis has shown that typically one-third of the variance in abundance between species is explained by differences in occupancy (Blackburn et al. 2006). The fit of different forms of occupancy–abundance relationships has, again, seldom been tested, although the general model (equation 11.8) tends often to fit very well (Holt et al. 2002a; Gaston et al. 2006).

The formation of occupancy–abundance relationships has most often been explained from the spatial perspective—any factors that would affect the





**Figure 11.3** Interspecific occupancy–abundance relationships and model fits for four data sets. (a) 817 tree species from the 50 ha (1000 × 500 m) Pasoh forest plot, Malaysia, at cell size = 10 × 10 m (the same data as Fig. 1d); (b) 301 tree species from the 50 ha plot, Barro Colorado Island, Panama, at cell size = 50 × 50 m; (c) 109 bird species from Hertfordshire, England, at cell size = 2 × 2 km (Smith et al. 1993); (d) 131 bird species from south-east Scotland, at cell size = 2 × 2 km (Murray et al. 1998). The four models are Poisson (solid line), Nachman (dashed), logistic (dotted), and negative binomial (longdash). As judged by AIC, the logistic model is the best for (a) and (d), while the negative binomial is the best for (b) and (c).

spatial distribution, aggregation, or randomness in the distribution of the individuals of a species would contribute to occupancy–abundance relationships (Holt et al. 2002a). These might include niche breadth, niche position, and habitat selection (for reviews see Gaston et al. 1997, 2000; Gaston 2003). A number of empirical studies have sought to test the importance of such mechanistic drivers (e.g. Hughes 2000; Cowley et al. 2001; Simková et al. 2002; Heino 2005; Kolb et al. 2006), although the great variation in how this has been done limits the ease with which broad conclusions can be drawn. A plausible but different argument is that occupancy–abundance relationships can arise from the colonization–extinction equilibrium of metapopulation dynamics. According to metapopulation dynamics, occupancy (or the incidence function) of a species can be expressed as a function of colonization and extinction such that

$$p = \frac{C}{C + E} \quad (11.11)$$

where  $C$  and  $E$  are colonization and extinction rates, respectively (Hanski 1994, 1997). Whilst  $C$  and  $E$  may be various functions of mean abundance ( $\mu$ ), the most general forms are  $C = a\mu^b$  and  $E = e\mu^{-d}$ , i.e. colonization increases and extinction decreases with the number of individuals (Hanski 1994). Substituting these two rates into (11.11) leads to

$$p = 1 - \left(1 + \frac{c\mu^z}{1}\right)^{-1} \quad (11.12)$$

where  $c = a/e$  and  $z = b + d$ . Equation 11.8 is a generalization of equation 11.12, pointing to the connection between metapopulation dynamics and occupancy patterns.

A more interesting approach is to equate equations 11.8 and 11.11, leading to

$$\frac{C}{E} = \left(1 + \frac{c\mu^z}{k}\right)^k - 1 \quad (11.13)$$

This model postulates a relationship between spatial aggregation as represented by  $k$  and colonization–extinction processes. Consider a metapopulation with fixed  $\mu$ . Under this condition, when  $k$  approaches 0 (i.e. the population is extremely aggregated), equation 11.13 equals 0, i.e.  $C/E = 0$ . This means that spatial aggregation results if colonization is extremely low or extinction is extremely high. This result is consistent with observations that species aggregation can be closely associated with poor colonization ability (Tilman et al. 1997a).

### 11.2.3 Species occupancy distributions

In isolation, understanding of the form of occupancy–area relationships provides limited insight into the relative frequency with which, at a given spatial resolution, the species in an assemblage will attain different levels of occupancy. A diversity of assumptions have thus been made about the variation amongst species in the scaling of occupancy with area to enable prediction of the shape of species occupancy distributions (e.g. Hui & McGeoch 2007a,b; Šizling & Storch 2007; Storch et al. 2008). In the main, although not exclusively, attention here has focused on very fine spatial resolutions (i.e.  $a$  is small), at which

point the species occupancy distribution equates to the species abundance distribution (assuming that only a single individual can fit within any given spatial unit). This is because of the great interest in being able to extrapolate from relatively coarse species occupancy data, which are for many species relatively easily obtained (although not necessarily for rare ones), to abundance data, which are much more difficult and expensive to acquire (e.g. Kunin 1998; He & Gaston 2000a; Tosh et al. 2004; Sara 2008; Conlisk et al. 2009; Figueiredo & Grelle 2009). Although species abundance distributions are discussed at greater length elsewhere in this volume (Chapter 10), it is notable that the predictions from occupancy–area relationships can provide particularly good fits to empirical abundance data (e.g. Storch et al. 2008).

Species occupancy distributions will change with spatial resolution, such that those at coarser resolutions may often be of rather different shapes from species abundance distributions. This can be simply illustrated by assuming that (i) the species abundance distribution follows a log-normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \quad (11.14)$$

where  $x$  is abundance and  $\mu$  and  $\sigma$  are the mean and variance of  $\log(x)$ , and (ii) the occupancy–area relationship takes a Poisson form

$$\frac{y}{A} = 1 - e^{-ax/A} \quad (11.15)$$

where  $a$  and  $A$  are the spatial resolution and extent, respectively, and  $y$  is the occupied area. Our interest is to find out the distribution of  $y$  and how the resolution would affect the distribution. By the technique of random variable transformation from  $x$  to  $y$ , it is easy to show that  $y$  has distribution

$$h(y) = \frac{A}{a} f\left(-\frac{A}{a} \log\left(1 - \frac{y}{A}\right)\right) \frac{1}{A - y} \quad (11.16)$$

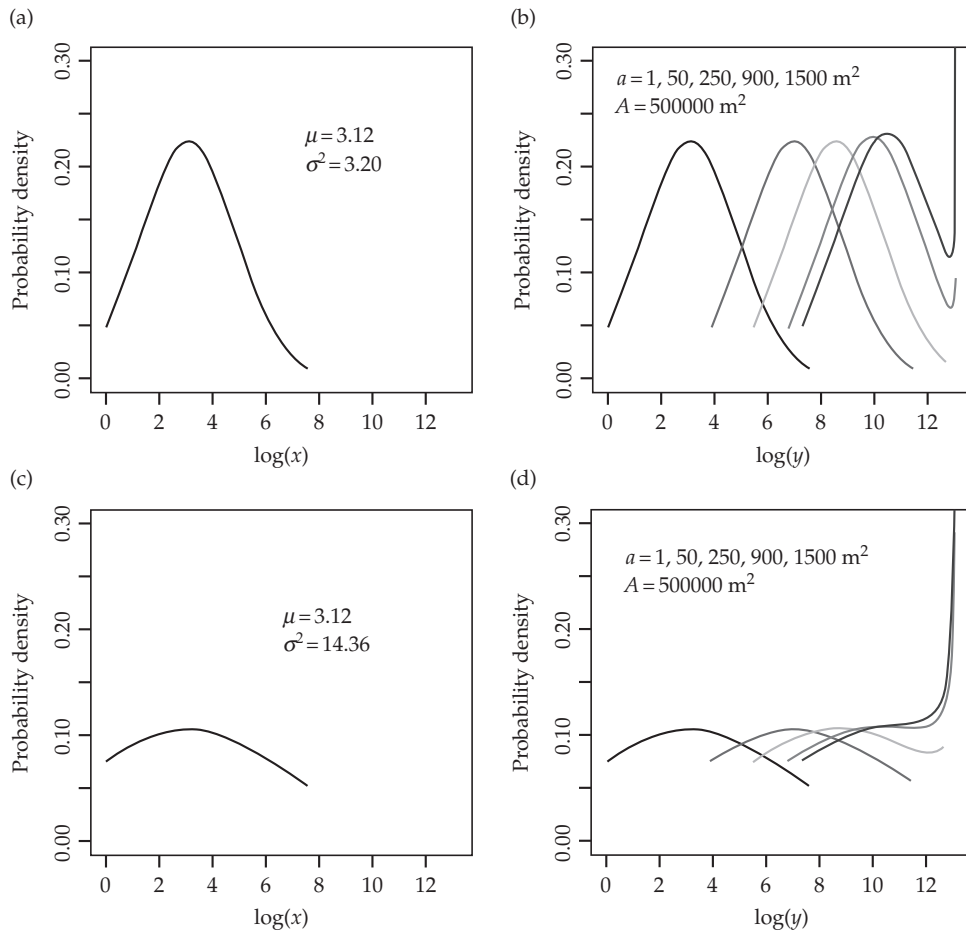
Note here that  $f(\cdot)$  is the log-normal distribution of equation 11.14. From Fig. 11.4, it is apparent that when  $a$  is small, the species occupancy distribution (equation 11.16) is virtually the same as the species abundance distribution. However, as  $a$  increases the species occupancy distribution shifts toward the

right. When  $a$  increases further, the distribution of more and more species saturate the entire study area, making the occupancy distribution upturn at the far right (right panels of Fig. 11.4). This upturn merely reflects the effect of spatial resolution, with little biology. Note that if other occupancy–area models are used in place of the Poisson model the species occupancy distribution (equation 11.16) can have multiple modes.

Typically, the number of species with low occupancy will increase as spatial resolution increases, and also will increase with the spatial extent of a study and sample number; the opposite will be true for species with high occupancy (McGeoch & Gaston 2002). However, because the details of occupancy–area relationships differ between species, the picture may become rather more complex than this. Indeed, although the potential for marked variation in species occupancy distributions with spatial resolution has been shown to arise both from relatively simple occupancy–area models (Hui & McGeoch 2007a, 2007b) and empirically (Collins & Glenn 1997; Guo et al. 2000; van Rensburg et al. 2000; McGeoch & Gaston 2002), it remains an open question to what extent the wide range of observed forms of species occupancy relationships can be explained on this basis alone. McGeoch & Gaston (2002) observe that across a diversity of spatial resolutions and extents documented forms include unimodal distributions with peaks at low, intermediate, or high levels of occupancy; bimodal distributions with dominant peaks at low, high, or both low and high levels of occupancy; and uniform or apparently random distributions. At the scale of the entire geographic ranges of species, species occupancy distributions tend invariably to be strongly right skewed (Gaston 1996a, 2003).

The importance of spatial resolution for the shape of species occupancy distributions may be reflected in a number of other mechanisms that have been postulated as explanations, including those based on habitat and environmental heterogeneity (see McGeoch & Gaston 2002). The principal alternative set of mechanisms are rooted in colonization/extinction dynamics, which may yield a wide variety of species occupancy distributions, dependent on the assumptions that are made as to





**Figure 11.4** The effect of spatial scale on the relationship between species abundance distributions and species occupancy distributions. All the distributions are plotted on a log scale. The left panels are the log-normal species abundance distributions (model 11.14). The first set of log-normal parameters ( $\mu = 3.12$ ,  $\sigma^2 = 3.20$ ) were calculated from the tree species abundances of the 50-ha plot on Barro Colorado Island, Panama (for those trees of size  $\geq 10$  cm). In the second row, the variance is arbitrarily set to 14.36 to illustrate how variance affects occupancy distribution. The right panels are the respective species occupancy distributions (model 11.16) corresponding to the log-normal distribution. In both cases, the total study area is set to be 50 ha. The density curves from left to right correspond to the cell sizes  $a = 1, 50, 250, 900$ , and  $1500 \text{ m}^2$ .

how rates of site colonization and extinction change with overall levels of occupancy and local abundance (Hanski 1982; Gotelli 1991; Warren & Gaston 1997).

### 11.3 Prospectus

Understanding of the form, determinants, and implications of occupancy–area relationships, occupancy–abundance relationships, and species occupancy distributions has increased markedly in

recent years and doubtless will continue to do so. In this regard, we make four main sets of suggestions:

- (i) Predicting abundance from occupancy: Substantial success has been achieved in predicting the abundance of a species from its spatial occupancy for study areas of small extent (e.g. of a few hectares or a small number of sites; He & Gaston 2000a; Sileshi et al. 2006; Conlisk et al. 2009). However, this is of only limited practical benefit, given that it is relatively easy to estimate abundance more

directly at such scales. Of far more importance would be an ability to predict abundance from occupancy at scales of hundreds or thousands of square kilometres, where direct estimation is extremely resource demanding and sometimes virtually impossible. Unfortunately, the methods presently available are of limited value for this purpose, commonly leading to under-estimation of numbers of individuals at such scales even by several orders of magnitude. A major challenge is how best to develop methods that are more appropriate. We doubt that occupancy data alone will be sufficient for this purpose, but suggest that some combination of broad-scale occupancy data and fine-scale abundance data from one or a few small areas are probably needed. Together, these will provide a hierarchical understanding of species distribution, which may provide a much more secure basis for the upscaling of abundance to large geographic areas.

- (ii) **Detectability:** As stated at the outset, for ease of presentation throughout this chapter we have ignored issues of differences in the detectability of species in sites, and particularly the frequent under-estimation of levels of occupancy and abundance that result from individuals being overlooked during field sampling. Increasingly sophisticated methods of improving the estimation of both variables are available (see Chapter 3). From the perspective of the estimation of abundance from occupancy, of particular interest is the modelling of detectability probabilities based on time series data on occupancy (MacKenzie et al. 2002, 2003; Royle et al. 2005). Consideration of such approaches highlights that the effective evaluation both of existing and novel methods of predicting abundance may require better estimations of actual occupancy than have in some cases thus far been employed.
- (iii) **Temporal dynamics:** More generally, the contents of this chapter have provided a reasonable reflection of the ecological literature at large in focusing primarily on occupancy–area relationships, occupancy–abundance relationships and species occupancy distributions based on estimations of the occupancy of a

species at a given time or averaged over a given period, and in the main ignoring the temporal dynamics of occupancy (but see Gaston & Lawton 1989; Webb et al. 2007; Heino 2008; Zuckerberg et al. 2009). However, particularly in light of the rapid changes in the distributions of species being wrought as a consequence of anthropogenic pressures (including climate change), a fuller understanding of these patterns is undoubtedly required. Of particular value would be much improved knowledge both of how the different patterns change through time and of how changes in the occupancies of species modify the observed patterns.

- (iv) **Mechanisms:** Among the many factors that can potentially affect species occupancy, species abundance and the spatial distribution of that abundance are the most immediately significant. Other factors do so through their effects on these variables. Given the numbers of possible such variables, and the indirect and scale dependent nature of their effects, their influence may often be difficult to detect other perhaps than under highly controlled circumstances (e.g. microcosms; Gaston & Warren 1997; Warren & Gaston 1997; Holt et al. 2002b, 2004a). This may well in large part explain the great controversies that have raged around the possible influences of niche breadth and position and dispersal abilities on levels of occupancy. In consequence, we do not expect that a single coherent mechanism will be determined that gives rise to occupancy patterns.

## 11.4 Key points

1. Occupancy is a fundamental measure of the distributions of species in landscapes. It is a function of their abundance and spatial distribution. A locally abundant species typically has higher global occupancy, and vice versa. On the other hand, spatial aggregation reduces occupancy. These two general qualitative patterns are described by the occupancy–abundance and occupancy–area relationships. Although many potential mechanisms give rise

to these relationships, colonization/extinction dynamics play a key role in determining occupancy.

2. Occupancy is scale dependent. The scale not only refers to the spatial resolution at which a species is mapped, but also to the spatial extent over which the species is studied. Because of the effect of scale, the measurement and quantification of occupancy can become complex. Comparison of occupancies for two or more species is irrelevant unless the spatial scales are compatible. Ideally, this should be done using occupancy–area curves, which describe the change of occupancy across the full range of spatial resolutions.
3. The most important application of the occupancy–abundance relationship is to estimate species abundance from occupancy. Present methods have met with a certain degree of success for relatively small spatial extents. Their usefulness in estimating abundance at the regional scale remains to be tested, although it is

likely that they will substantially under-estimate regional abundance. It will be a major breakthrough if methods can be developed to estimate abundance from regional occurrence data.

4. Occupancy is the basic unit underlying several macroecological patterns, including species–area curves and gradients in  $\beta$  diversity. Assuming  $p_i(a)$  is the occupancy probability of species  $i$  in a site of size  $a$ , then the species–area curve is simply the sum of probabilities across all species.  $\beta$  diversity can also be formulated by a conditional probability through occupancy: given a species is present in an area, what is the probability that the same species will occur in another area at distance  $d$  away?

## Acknowledgements

KJG was supported by The Royal Society and The Leverhulme Trust, and FH by the NSERC (Canada).