The distribution of species: occupancy, scale, and rarity

FANGLIANG HE
University of Alberta
RICK CONDIT
Smithsonian Tropical Research Institute

Introduction

Species occupancy is typically measured as the number of cells occupied by the species in a study area. Because it is easy to document and interpret and it correlates with species abundance, occupancy is widely used for measuring species rarity and for assessing extinction risk on which conservation decisions are made (Gaston, 1994; Fagan *et al.*, 2002; Hartley & Kunin, 2003; Wilson *et al.*, 2004). Ecologists and conservation practitioners, however, have long realized that occupancy often fails to capture significant spatial features of distribution. It is possible that two species having the same occupancy can exhibit very different patterns (Fig. 3.1). Most species in nature are discretely distributed due to the patchiness of landscapes, or due to intrinsic reproductive or dispersal behavior of the species. An outstanding problem concerning species distribution in space is how to describe the patchiness of a species and to measure the effect of changing spatial scale (cell size) on the patchiness for the purpose of predicting distribution at fine scales from coarse scales.

There are two primary approaches to addressing this question. The first one is to use existing measures and methods to describe patchiness and scale effect. Many fragmentation indices in landscape ecology can be used for this purpose (Turner, Gardner & O'Neill, 2001; Wu *et al.*, 2003). These include edge length (perimeter), the number of patches, perimeter/area ratio and many other indices to capture the spatial features of species distribution. An interesting development on this front is the percolation models for edge length and the number of patches proposed by He and Hubbell (2003). These models unify edge length, the number of patches, spatial scale and abundance and show that the edge length and number of patches of a distribution can accurately be predicted by the abundance of the species and the degree of fragmentation. These models are useful for analyzing and comparing species distributions.
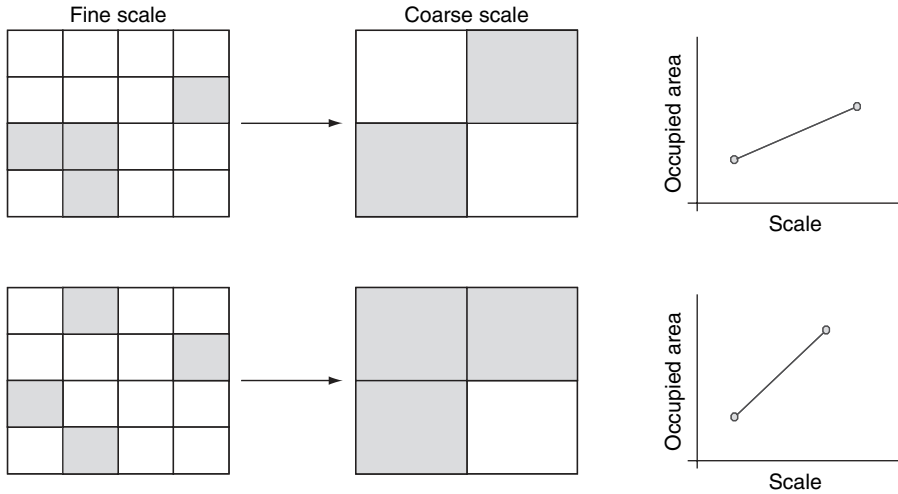
Figure 3.1 Two species of the same occupancy but different distributions, leading to the difference in the occupancy in the coarse-scale maps and the slope of the occupancy–area relationship. Note the coarse-scale maps are produced by combining four adjacent cells on the fine-scale maps. If all the four fine-scale cells are empty, the coarse-scale cell is designated empty, otherwise occupied.

The second approach is to search for scale invariant properties of species distributions. This includes investigating fractal distribution of species (Kunin, 1998; Kunin, Hartley & Lennon, 2000; Hartley *et al.*, 2004) and employing the negative binomial distribution to quantify and estimate species abundances from coarse-scale distribution maps (He & Gaston, 2000; He, Gaston & Wu, 2002). The fractal method has been shown to be effective in quantifying distribution fragmentation and useful for assessing extinction risk of species, particularly of rare species (Fagan *et al.*, 2002; Wilson *et al.*, 2004). However, these methods meet with limited success in measuring species distributions and estimating abundance, particularly for relatively abundant species (Warren, McGeoch & Chown, 2003; Witte & Torfs, 2003; Tosh, Reyers & Jaarsveld, 2004). More often, the negative binomial method underestimates abundances while the fractal method overestimates (He & Gaston, 2000; Kunin *et al.*, 2000; Warren *et al.*, 2003; Witte & Torfs, 2003; Tosh *et al.*, 2004). Thus, the question about how we may predict distributions across multiple spatial scales remains largely unsolved. The capability to infer distributions across scales is essential for predicting distribution at fine scales from coarse scales, for determining the conservation status of species, and for making management plans and reserve design (Hartley & Kunin, 2003).

The primary objective of this study is to examine the scaling of several common occupancy–area models and thus identify models that may be independent of scale. Scale in this study is referred to as the cell size at which a
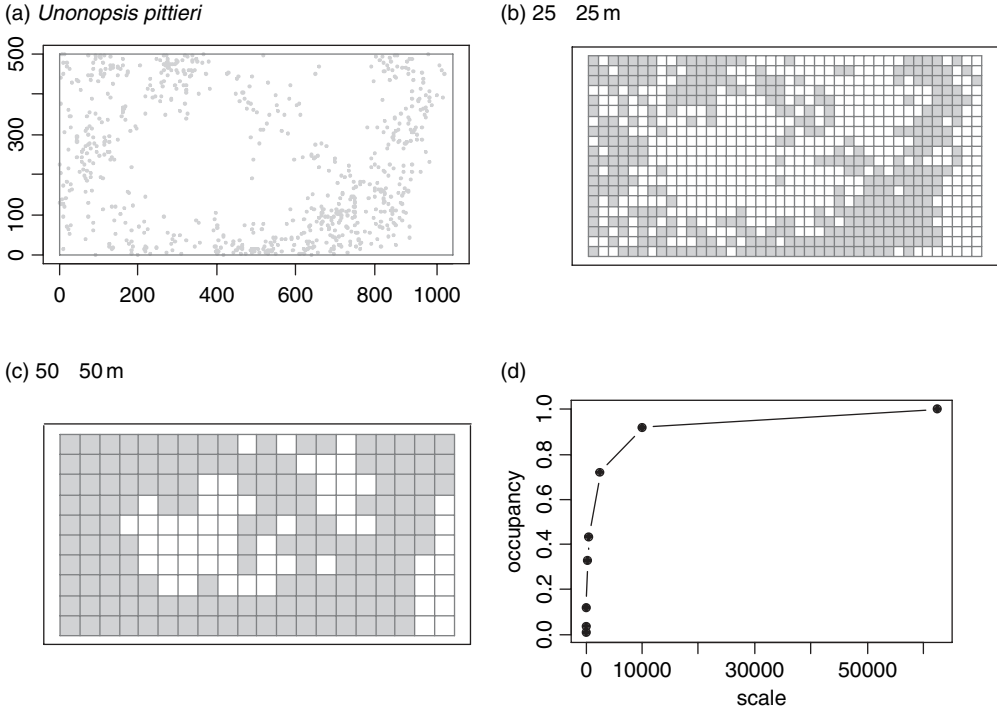
Figure 3.2 (a) Distribution of 718 stems of *Unonopsis pittieri* in a 50 ha (1000 × 500 m) plot on Barro Colorado Island, Panama. Panels (b) and (c) are the occurrence maps at two mapping scales, and (d) plots the occupancy–area curve, i.e. the relationship between the proportion of occupied area of each map and scale (cell size).

distribution is gridded, and occupancy the proportion of grid cells in which a species is found (Fig. 3.1). First, we introduce several occupancy–area models, and show that a general model unifies them all. We evaluate the scaling properties for three of the models, the power-law model, the Nachman model, and the logistic model, using simulated and empirical data. We then use the three models to estimate occupancy at a fine scale from occupancy at coarse scales, and we compare the slope of occupancy versus area between rare and common species. We conclude with a discussion of the use of the three occupancy–area models for describing species distribution and for conservation practice.

Occupancy–area models

Plotting species occupancy against spatial scale yields a monotonically increasing curve whose slope is steep at first but gradually becomes flat as scale increases (Fig. 3.2d). Some of the common occupancy–area models are given below. In the following p is occupancy (the proportion of occupied area), a is scale (or cell size), and c, z and k are parameters. While c and z must be positive k can be any real number.

the power-law model (Kunin, 1998): $p = ca^z$,
the Poisson model (Wright, 1991): $p = 1 - e^{-a}$,
the negative binomial model (He & Gaston, 2000): $p = 1 - \left(1 + \dfrac{a}{k}\right)^{-k}$,
the Nachman model (Nachman, 1981): $p = 1 - e^{-ca^z}$,
the logistic model (Hanski & Gyllenberg, 1997; Leitner & Rosenzweig, 1997):

$$p = \frac{ca^z}{1 + ca^z}.$$

These models were originally developed for different purposes by different authors in the form of occupancy–abundance models. In those occupancy–abundance models, the occupancy p is of the same interpretation as in the above occupancy–area models. The only difference is that the scale a is replaced by mean density μ in occupancy–abundance models (see He *et al.*, 2002, for a detailed description of occupancy–abundance models). Although the derivation of the above occupancy–area models is largely empirical, the replacement of μ by a is justified because abundance and area are widely found to have a simple linear relationship (Preston, 1962). For instance, the Poisson occupancy–abundance model, written as $p = 1 - e^{-\mu}$, was first developed to describe the relationship between occupancy and abundance (Wright, 1991), the Nachman model was initially used for predicting agricultural pest density from incidence data (Nachman, 1981), and the logistic model was for modeling the species–area relationship (Hanski & Gyllenberg, 1997). These models are now commonly used to describe occupancy–abundance relationships (Gaston, 1994; He *et al.*, 2002; Holt, Gaston & He, 2002). Replacing abundance μ in occupancy–abundance models by scale a, we obtain the occupancy–area curves listed in the above.

Interestingly, the five models can all be written with one general form (He *et al.*, 2002):

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

(a)   when $c = z = 1$, the general model becomes the negative binomial model;
(b)   when $k = -1$, it is the power-law model;
(c)   when $k = 1$, it is the logistic model;
(d)   when $k \to \pm\infty$, it is the Nachman model;
(e)   when $k \to \pm\infty$ and $c = z = 1$, it is the Poisson model.

The application of the power-law model to describing species distribution is first proposed by Gaston (1994) and Kunin (1998) and has since become widely used. Beyond this power-law model, however, little is known about the scale-invariant property of other models. In the following, we examine this property for the Nachman and the logistic models together with the power-law model because they have similar linear forms, as given below.

the power-law model: $\log(p) = \log(c) + z\log(a),$ (3.1)

the Nachman model: $\log[-\log(1-p)] = \log(c) + z\log(a),$ (3.2)

the logistic model: $\log\left(\dfrac{p}{1-p}\right) = \log(c) + z\log(a).$ (3.3)

These three models have the same right-hand terms but differ in their left-hand term. Although the slopes of the three models are not equal, they do change in the same direction, i.e. for a given occupancy if z value is high for the power model, the z values for the Nachman and the logistic models are also high. For a given occupancy at a fine scale small slopes indicate more aggregated distribution, while large slopes suggest more scattered distribution, as illustrated by Fig. 3.1.

Testing the models

Three sets of data were used to test the performance of models 3.1–3.3: simulated fractal distributions, the local distributions of 301 tree populations in a 50 ha stem-mapped plot from Panama, and the regional distributions of 407 rare plant species of the United Kingdom.

Random and fractal simulations

Neutral landscape models (Gardner *et al.*, 1987) were used to generate species distributions in an area of size $= 256 \times 256$ cells. The probability of being occupied or empty of each cell was determined by p. The neutral distribution models used here include simple random maps and fractal maps. In the simple random maps occupied/empty state of a cell is independent of any other cells, while in the fractal maps occupancy of a cell is spatially correlated with its neighboring cells. The fractal distributions were generated using the random midpoint displacement algorithm (Saupe, 1988). Spatial correlation of a map (i.e. the variance between locations separated by distance x) approximately equals x^{2H}. Varying H from 0 to 1 corresponds to varying the distribution from extremely scattered to highly aggregated (With, Gardner & Turner, 1997; Turner *et al.*, 2001).

To illustrate scale effect on distribution, we used a random (lattice) map and a fractal (lattice) map with $H = 0.1$ and calculated occupancy using a pixel size of 1, 2, 4, 8, ..., up to 256 (Fig. 3.3). The relationship between occupancy and scale is shown in Fig. 3.3 for each of the models. Although the power-law occupancy–area curve is linear at small scales, there is apparent curvature when a map approaches saturation (Fig. 3.3a). This curvature has traditionally been explained by the process of space filling near saturation due to finite study area. This concave-down curvature suggests that coarse-scale maps are not useful for estimating occupancy at fine scales. The departure from linearity will overestimate occupancy when extrapolating to smaller scales. The approximate linearity of the
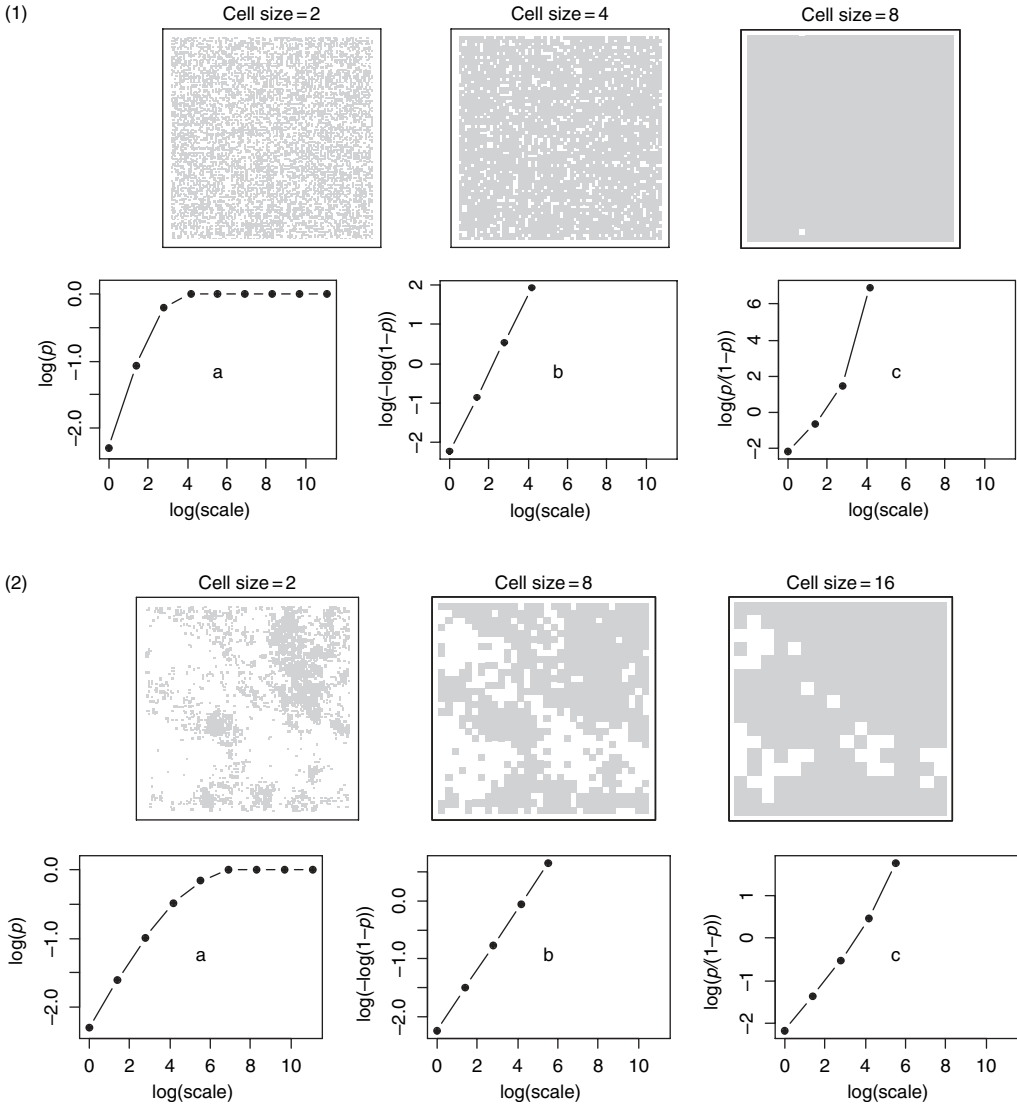
(1)

(2)

Figure 3.3 Two simulated distributions at extent 256×256 cells and their occupancy–area curves. (1) First row shows the simple random maps at three scales: cell size $= 2 \times 2$, 4×4 and 8×8. Coarse-scale maps are aggregated from fine-scale maps. The second row shows occupancy–area curves corresponding to the power-law model (a), the Nachman model (b), and the logistic model (c). (2) The fractal distribution simulated with $H = 0.1$ and its three corresponding occupancy–area curves.

Nachman model (Fig. 3.3b) suggests it is a better model for describing the scale dependence, and the linearity seems to hold reasonably well even at near saturation of distribution. The logistic model (Fig. 3.3c) shows concave-up, meaning it will underestimate occupancy if used to extrapolate to smaller scales. The relative

performances of the three models remain the same for fractal maps simulated using other H values (results are not shown).

Species distributions at the local scale

We now test the performance of models (3.1–3.3) using real distributions of tree species in a tropical rain forest on Barro Colorado Island (BCI) of Panama (Hubbell & Foster, 1983; Condit, Hubbell & Foster, 1996; also see http://ctfs.si.edu/datasets). In 1981 a 50 ha (1000×500 m) forest plot on BCI was established. In the plot, all free-standing trees and shrubs ≥ 1 cm diameter at breast height were enumerated, individually located on a reference map, and identified to species. Five field censuses have so far been surveyed. The data from the 1990 census are used in this study where there are 229 048 stems belonging to 301 species with the most abundant species having 36 060 stems.

We converted the distribution of each of the 301 species into an occurrence map for a given scale as for *Unonopsis pittieri* shown in Fig. 3.2. Occurrence maps at eight spatial scales, $a = 2 \times 2$, 5×5, 10×10, 20×20, 25×25, 50×50, 100×100 and 250×250 m, were produced. Models (3.1–3.3) were then fitted to the occupancy data for each of the 301 species. Note that although there are eight occurrence maps (at the eight scales) for each species, coarse-scale maps, e.g. at 100×100 and 250×250 m, may be saturated. Saturated maps contain no effective information on species distribution and are thus excluded in the analysis.

Two criteria were used to measure the goodness-of-fit of a model. The first one is the R^2 of the log-transformed linear models. The second criterion is the mean squared errors between the log-transformed observed and predicted occupied areas, defined as $MSE = \dfrac{1}{n} \sqrt{\sum_{i=1}^{n} [\log(x_i) - \log(\hat{x}_i)]^2}$, where n is the number of scales, x_i is observed occupied area (in m^2) at the ith scale, and $\hat{x}_i$ is the predicted occupied area. This prediction was obtained from each of the three models (3.1–3.3) by multiplying the predicted occupancy $\hat{p}$ by 500 000 m^2 (the total study area).

Distributions of three species and their occupancy–area curves are shown in Fig. 3.4. The Nachman model describes well the scale dependence of species distribution and is clearly superior to the power and logistic models. The shapes of the curves in Fig. 3.4 precisely reflect those simulated in Fig. 3.3: concave-down for the power-law model, more or less linear for the Nachman model, but concave-up for the logistic model. It is worth noting that this general finding also applies to all the BCI species except those rare species with abundance <50. The three models work approximately equally well to those rare species.

To provide an overall judgment on how well the three models fit the BCI data, the histograms of the R^2 and MSE for the log-log linear models (3.1–3.3) are
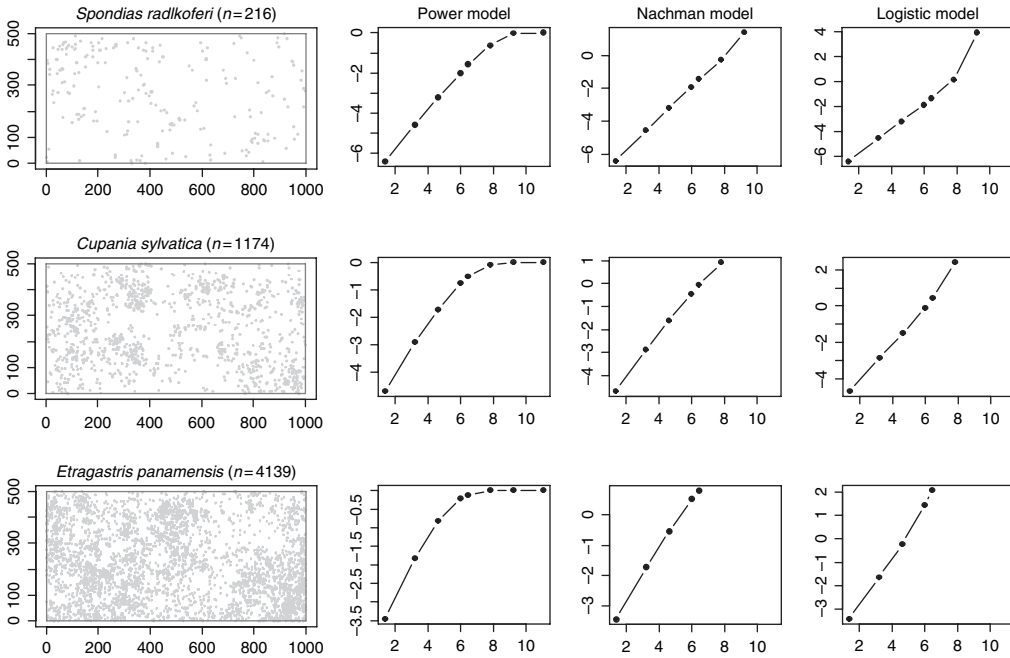
Figure 3.4 Distributions of three BCI species and their occupancy–area curves. The shape of the power model is typically concave-up, while the logistic model concave-down. The Nachman model seems to be approximately linear. The x-axis of the occupancy–area curves is log(scale), the y-axis is log(p) for the power model (second column), log[−log(1 − p)] for the Nachman model (third column), and log[p/(1 −p)] for the logistic model (fourth column). Here p is occupancy.

plotted in Fig. 3.5. It is clear that the Nachman model has the largest overall R^2 and lowest *MSE*. The second best model is the logistic model and the power-law model is least satisfactory.

Species distributions at the regional scale

The performance of models (3.1–3.3) was also tested and compared using 407 rare plant species from the UK (Kunin, 1998; Hartley *et al.*, 2004). Each species was mapped at seven scales with cell size = 1×1, 2×2, 5×5, 10×10, 20×20, 50×50 and 100×100 km. The occupancy–area curves for three of the species are shown in Fig. 3.6; most of the curves show a slight concave-down shape, which differs from the simulations (Fig. 3.3) and the local-scale BCI data (Fig. 3.4). Previous studies have shown that these UK plant species could be adequately modeled by the power model (Kunin, 1998; Hartley *et al.*, 2004). Our results in Fig. 3.7 support this conclusion. As judged by R^2 and the *MSE*, all the three models describe the occupancy–area curves sufficiently well. There is no definite superior model although the Nachman seems to perform slightly better than the other two models. The indistinguishable results are probably due to the fact that the UK plants are rare species.
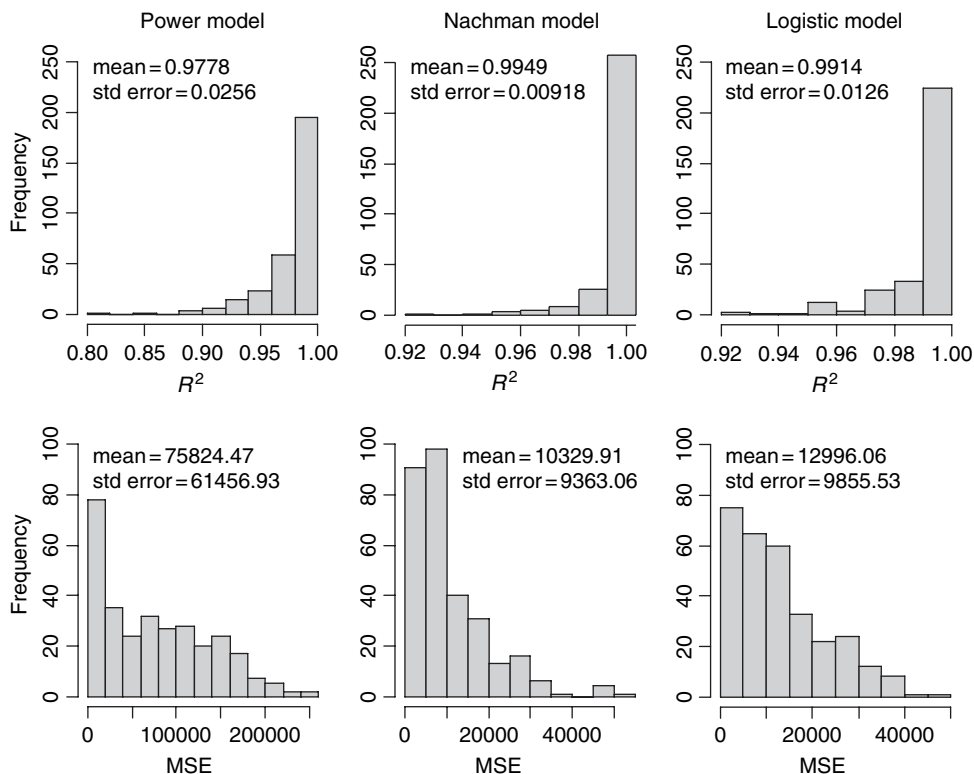
Figure 3.5 Histograms of R^2 and *MSE* (mean squared errors) indicating the goodness-of-fit of the three models (3.1–3.3) fitting to each of the 301 species distributions. High R^2 or low *MSE* suggests a good fit of a model.

Estimating occupancy at fine scales from the occupancy at coarse scales

We now turn to use models (3.1–3.3) to estimate occupancy at fine scales from the occupancy at coarse scales for both the BCI and UK data. Two predictions are made here. The first prediction uses the two maps at the coarsest unsaturated scales to predict occupancy at a very fine scale. For most BCI species, this means to use maps at 50×50 and 100×100 m to predict occupancy at 2×2 m. For highly abundant species (>3000 trees) the coarsest unsaturated maps are at 25×25 and 50×50 m. For the UK rare species, the coarsest unsaturated maps are 50×50 and 100×100 km. They were used to estimate occupancy at 1×1 km. We also repeated the estimation but used the three (instead of two) coarsest maps to estimate occupancy at 2×2 m in the case of the BCI species and occupancy at 1×1 km in the case of the UK species.

The power-law model predicts occupancy well for rare BCI species at 2×2 m (Fig. 3.8), but overestimates for abundant species because of the curvature in the
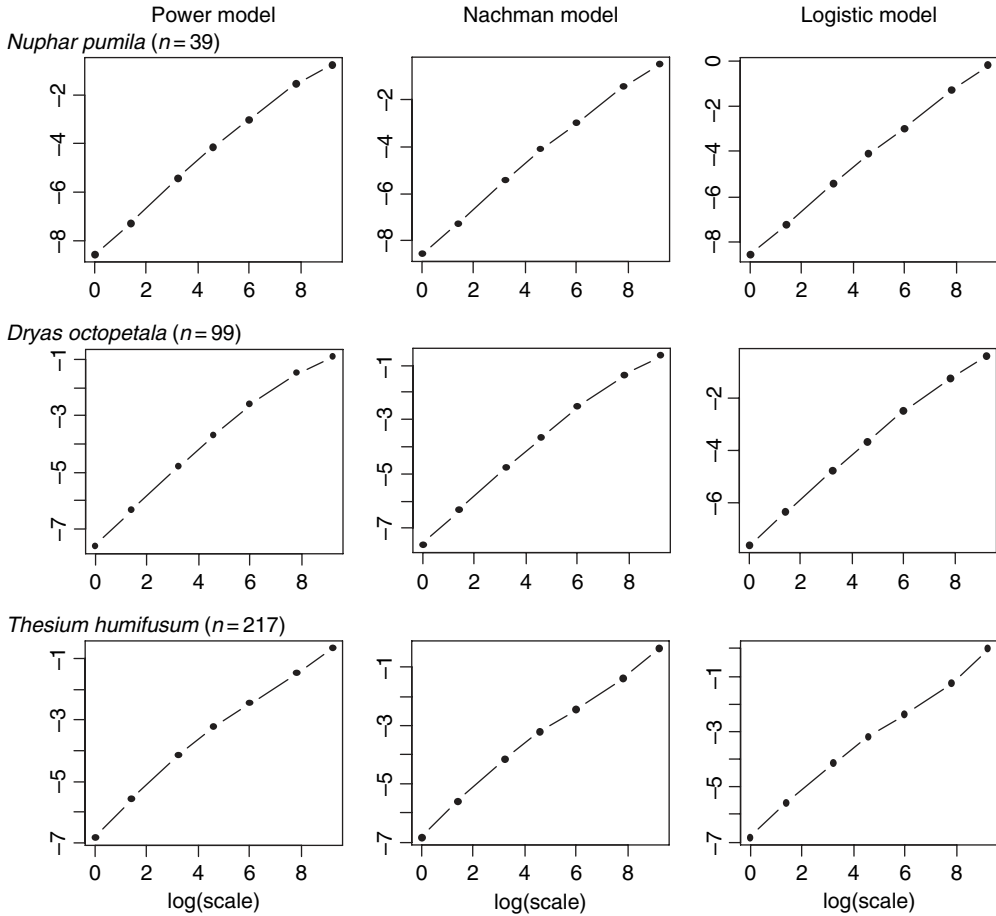
Figure 3.6. Occupancy–area curves for three plant species from the UK. The number of occupied cells of the base map (i.e. the map at the finest scale of 1×1 km) is 39, 99 and 217 for *Nuphar pumila*, *Dryas octopetala* and *Thesium humifusum*, respectively. The first column is the power model, the *y*-axis is $\log(p)$. The second column is the Nachman model with *y*-axis $\log[-\log(1-p)]$. The third column is the logistic model with *y*-axis $\log[p/(1-p)]$.

occupancy–area curves (Figs. 3.3 and 3.4). In contrast, the logistic model underestimates due to the opposite curvature (Figs. 3.3 and 3.4). The Nachman model lies in the middle between the power and logistic models and is superior to both, although there is some degree of overestimation. Estimation is much improved by using three maps (Fig. 3.8, bottom row) although this does not change the relative performance of the three models. This result suggests that we should aggregate as many maps as possible (up to saturation) from an observed distribution in order to accurately scale down occupancy at fine scales. Every map up to saturation helps increase accuracy. The Nachman model has the smallest deviation from the observed occupancy (Table 3.1).
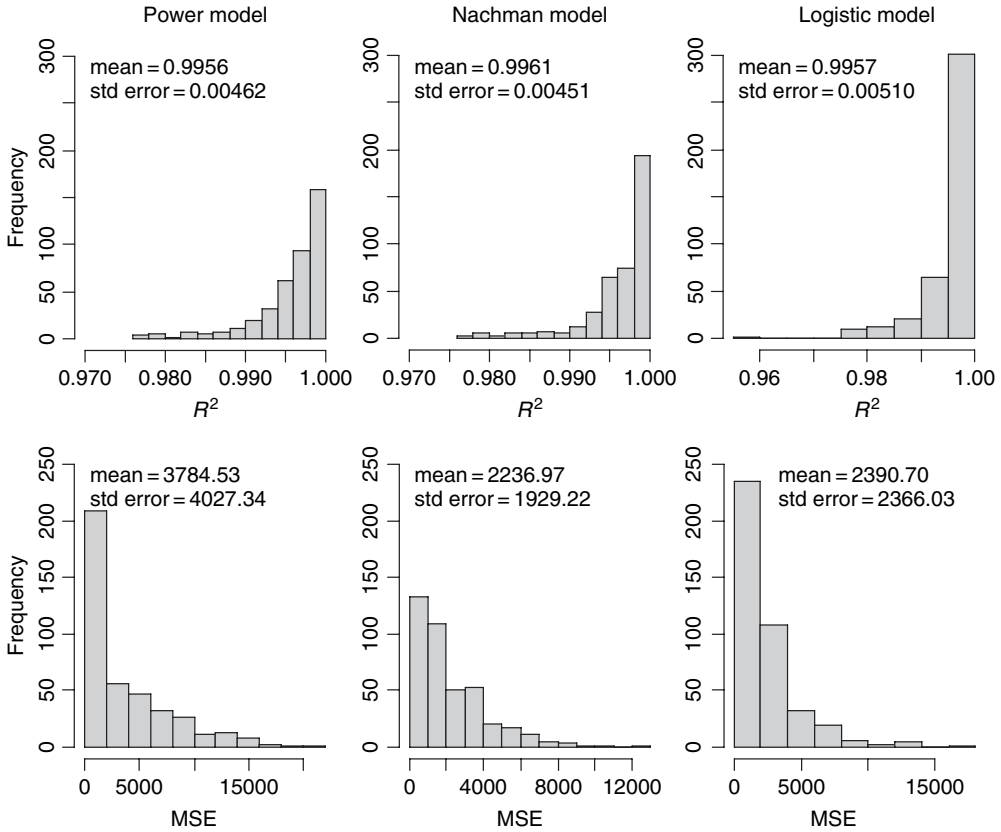
Figure 3.7 Histograms of R^2 and *MSE* indicating the goodness-of-fit of the three models (3.1–3.3) fitting to each of the 407 UK plant distributions. High R^2 or low *MSE* means a good fit of a model.

Although a similar result is also observed from the UK data, the difference in the estimation accuracy among the three models is much smaller (Table 3.1, Fig. 3.9). None of the models satisfactorily estimates the occupancy at $1 \times 1\,\text{km}$ from the two coarsest maps, although their performance is substantially improved by using three coarsest maps (Fig. 3.9). Overall, the power model still relatively overestimates the occupancy compared with the Nachman model, while the logistic model underestimates it.

Comparing the slope between rare and common species

If there is a difference in distribution between rare and common species, the difference should be manifest in the slope of an occupancy–area curve. In other words, the z values of the three models (3.1–3.3) should be correlated with abundance. This correlation is inevitable for abundant species that cover more than 25% of the study area (25% because four fine cells are combined into one coarse cell when scaling up distribution). However, in BCI the most abundant
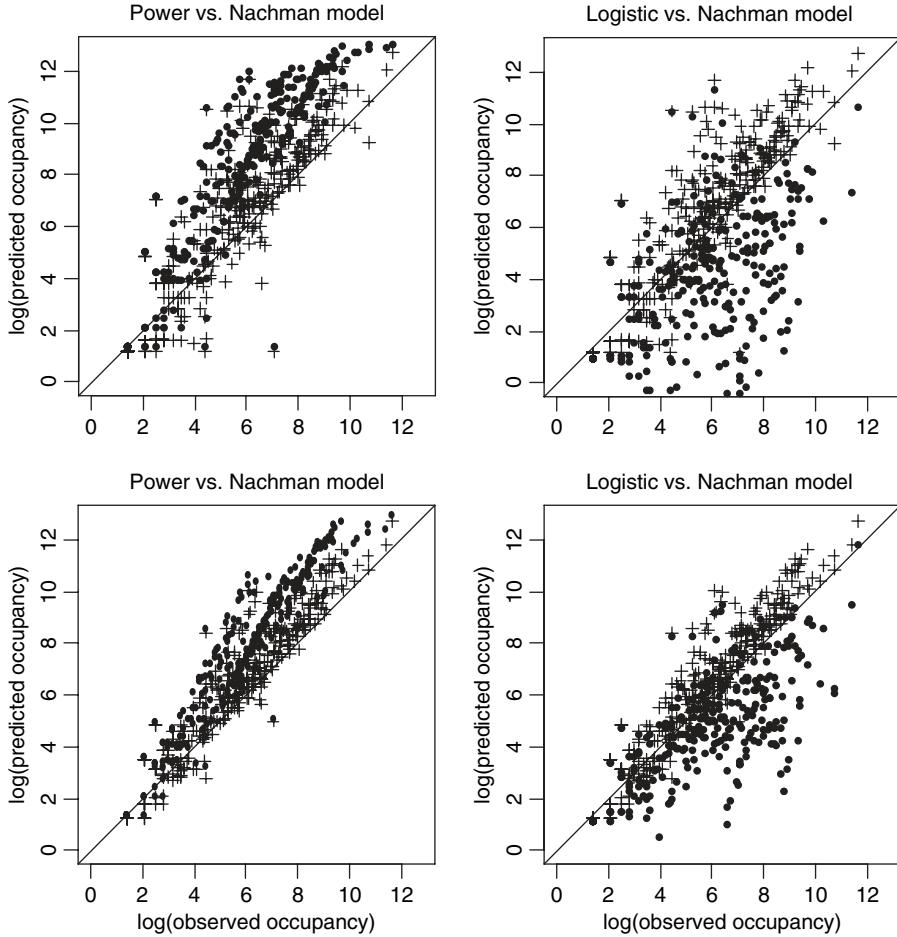
Figure 3.8 Log-log plot for the estimated occupied area versus the observed area at 2×2 m scale for the 301 BCI species. The first row is the estimation made from the two coarsest unsaturated maps. The second row is the estimation made from the three coarsest unsaturated maps. Each point represents a species. The crosses are the occupied area estimated from the Nachman model, the filled dots are either from the power model or from the logistic model. If the estimation is accurate, all the points should lie on the diagonal line.

species *Ampea appendiculata* only has 22.89% occupancy at scale 2×2 m. All the UK plants are rare species, their occupancies are much smaller. Therefore, the slope of an occupancy–area curve for the BCI and UK species should, to a degree, reflect the difference in distribution between rare and common species.

Because abundances for the UK species are unknown, to be consistent we use the occupancy at the finest scale as a proxy for abundance in both data. This means the occupancy at 2×2 m for the BCI species and occupancy at 1×1 km for the UK species. The relationships between z and abundance of the species

Table 3.1 *The "goodness-of-estimation" of the power, Nachman, and logistic models for estimating the occupancy at 2 × 2 m for the BCI and at 1 × 1 km for the UK species, estimated from two or three coarse-scaled maps*

The goodness is assessed by the mean squared errors $\text{MSE} = \frac{1}{n}\sqrt{\sum (\log(x_i) - \log(\hat{x}_i))^2}$, where x_i is the observed occupied area for the ith species, $\hat{x}_i$ is the estimated area and n is the number of species ($n = 301$ and 407 for the BCI and UK data, respectively).

	MSE (BCI species)		*MSE* (UK species)	
	2 maps	3 maps	2 maps	3 maps
Power model	0.1588	0.115	0.0835	0.0484
Nachman model	0.0942	0.0629	0.0814	0.0438
Logistic model	0.201	0.113	0.109	0.0501

show that slopes indeed differ between rare and common species (Fig. 3.10). The figure also shows the slopes of the BCI species expected by random distribution of the occupied cells over the 50 ha plot. The power and Nachman models consistently predict higher slopes for rare species than for common species. The result for the logistic model is less consistent. For the BCI species the logistic model shows that slopes are relatively constant for most species, but for the UK species they decrease with abundance.

Discussion

Knowledge about spatial distribution is essential for studying macroecological biodiversity patterns (Gaston, 1994; Gaston & Blackburn, 2000), for predicting the distribution of species in areas where observations are not available (Heikkinen & Högmander, 1994; MacKenzie *et al.*, 2002; Raxworthy *et al.*, 2003), for understanding environmental determination of species distribution (Currie, 1991; Lennon, Greenwoord & Turner, 2000; He, Zhou & Zhu, 2003; Hurlbert & Haskell, 2003), for assessing the effect of landscape fragmentation and climate change on extinction (Thomas *et al.*, 2004; Fagan *et al.*, 2005), and for planning biological conservation priority (Myers *et al.*, 2000). However, the scale dependence of spatial distributions is sometimes considered as a statistical and ecological nuisance that prevents the prediction of species distribution across scales, thus hampering our ability to determine species' conservation status across scales.

Our study has enforced this adverse impression and found that the description of species distributions remains elusive and few species are scale independent in distribution. There exists no universal, genuinely scale-independent model applicable to all species across scales, although the Nachman model is
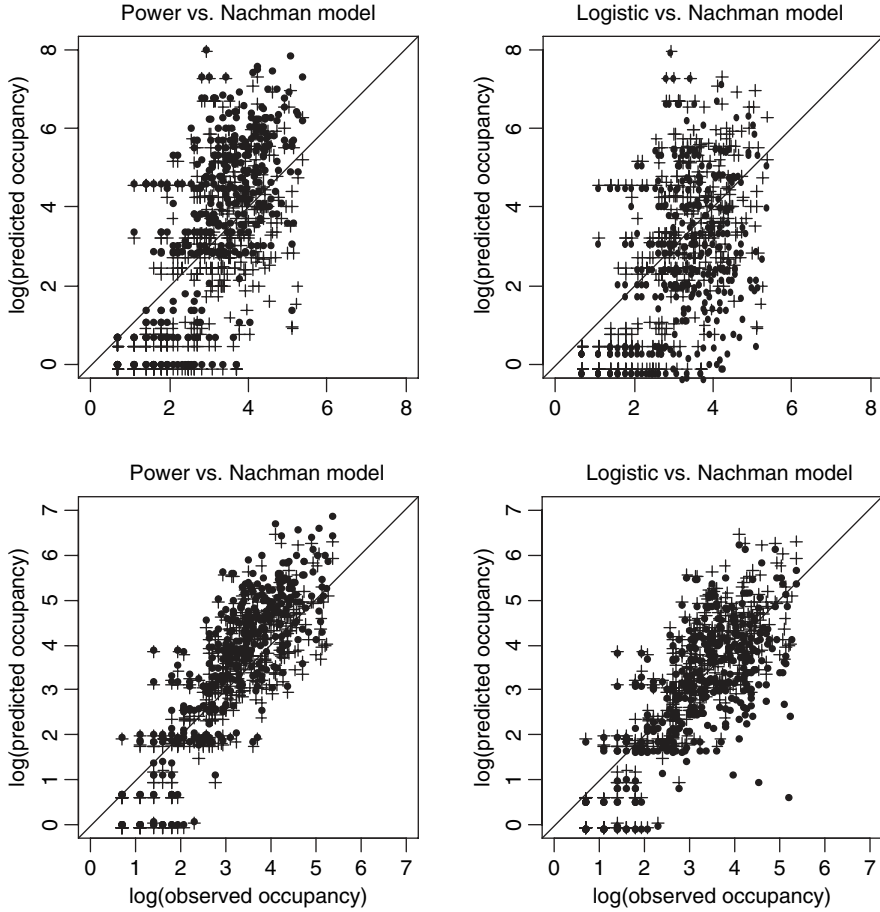
Figure 3.9 Log-log plot for the estimated occupied area versus the observed area at 1×1 km scale for the 407 UK species. The first row is the estimation made from the two coarsest unsaturated maps. The second row is the estimation made from the three coarsest unsaturated maps. Each point represents a species. The crosses are the occupied area estimated from the Nachman model, the filled dots are either from the power model or from the logistic model. If the estimation is perfect, all the points should lie on the diagonal line.

most robust to scaling effect among the three models we tested. From the practical point of view, the Nachman model is superior in two aspects. First, it consistently shows a linear relationship between occupancy and scale up to a certain level for almost all of the species from both the local BCI trees and the regional UK rare plants. Simulation results shown in Fig. 3.3 also reveal the linearity of the Nachman model even at near map saturation (compare the two nearly perfect lines in Fig. 3.3b against the evident curvature in Fig. 3.3a and 3.3c). Second, the Nachman model is apparently the best model for estimating occupancy at fine scales from data of coarse scales. The power-law model
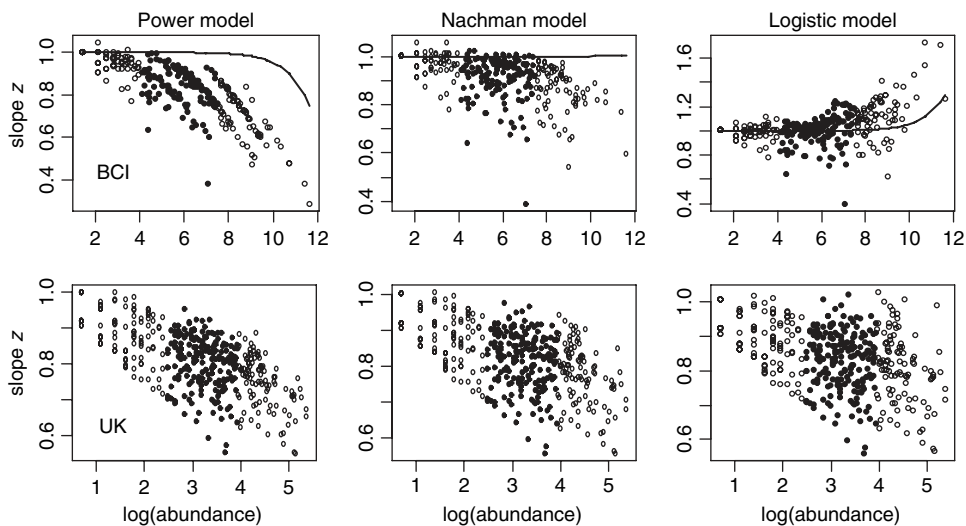
Figure 3.10 Slope z of the power-law, Nachman and logistic models (3.1–3.3) versus occupancy at 2×2 m (BCI, the first row) and 1×1 km (UK, the second row). The curves for BCI are the slope expected from random distribution. They were calculated by: (i) randomly allocating the same number of occupied cells at scale 2×2 m for each species over the study area, (ii) aggregating up by combining four fine cells to form one coarse scale cell, and (iii) fitting each occupancy–area model to the two scale occupancy–area data. Because only the overall occupancy is known to us the same expected slope cannot be calculated for the UK species. The open circles on the two sides are the species of the 1st and 3rd quantiles of abundance.

substantially overestimates the occupancy at fine scales, while the logistic model underestimates it. The Nachman model lies in the middle despite noticeable overestimation (Figs. 3.8 and 3.9). It is important to note that estimation is significantly improved if occupancy is estimated from three or more maps (see the bottom panels in Figs. 3.8 and 3.9) or if the maps on which the estimation is based are relatively fine-scaled.

An interesting result of this study is that slopes of the three models (3.1–3.3) consistently decrease with abundance except the logistic model for abundant BCI species (Fig. 3.10). Three implications of this result are in order. First, it suggests that there are differences in spatial distribution between rare and common species. Rare species show a tendency of scattered distribution, while common species tend to be more aggregated. Second, the fact that rare species have significantly larger slopes than the common species as predicted by the power-law and Nachman models suggests that rare species fill up space more quickly than common species when scaling up. As a consequence, determination of species rarity based on occupancy at coarse scales can be substantially biased because species of the same occupancy at the coarse scale may have very different occupancies at the fine scale (i.e. the scaling down in Fig. 3.1), conforming to the

result of Hartley and Kunin (2003). Third, the dependence of the slope on abundance implies that the slope of a scaling model alone is not sufficient to describe spatial distribution. A good index of spatial pattern ought to be independent of the abundance of species; it has been a methodological challenge to develop such a measure in ecology, and scaling models are no exception.

Whether a species is considered fractal is debatable, depending on how fractal is defined and measured for an empirical distribution which is always of finite size (Cutler, 1993). An important point is that an adequate fitting of the power-law model does not necessarily imply fractal in distribution. Our study is not designed to test the self-similarity properties of species distribution but to investigate the utility of the three occupancy–area models in describing distribution. However, our results show that many species do have similar occupancy–area shapes as those produced from fractal simulations as shown in Fig. 3.3. The breakdown in linearity near map saturation is inevitably caused by space filling due to finite study area. This saturated region is routinely suggested to be excluded from the calculation of fractal dimensions (Halley *et al.*, 2004) although such practice is not the best interest to many applications. From the practical point of view, it is also important to model the curvature as it is an inherent component of the data (Šizling & Storch, 2004).

In summary, we have studied scale-invariant property of species distribution for three occupancy–area models. The findings of the study should shed a light on our understanding of species distribution and provide potentially useful tools for determining the rarity of species for the purpose of conservation. The main results of the study include:

1. No occupancy model is universally superior in describing occupancy data. Among the three models intensively tested in this study, the Nachman model appears to be most robust to scaling effect. It is superior to the power-law and the logistic models in describing species distribution and for predicting occupancy across scales.
2. All empirically observed distributions share a similar occupancy–area shape that increases linearly at small scales but saturates at coarse scales due to space filling. Empirical occupancy–area curves have similar shapes as those simulated from fractal distributions. However, this similarity should not be interpreted as an implication of fractality because the occupancy–area models alone are not appropriate for measuring self-similarity.
3. An important application of occupancy–area models is to estimate species abundance from occupancy. The accuracy is substantially improved if three (or more) rather than two scale maps are used to estimate occupancy at any fine scale. In other words, every coarse-scale map up to saturation is useful for scaling down occupancy at fine scales. The Nachman model which is most robust to scaling effect is a best model for estimating abundance.

4. Distributions are different between rare and common species. The slopes of the log-log occupancy curves are significantly higher for rare than for common species, suggesting that rare species fill up space more rapidly than common species when scaling up. Therefore, determination of species rarity based on occupancy at coarse scales can be substantially biased. If rarity of a species has to be determined and coarse-scale maps are the only data available, the Nachman model must be used to calculate the occupancy at finer scales.

5. The difference among the power, the Nachman and the logistic models for the UK plants seems not to be as evident as for the BCI data. This perhaps reflects the fact that all the UK plants are rare species. Therefore, further testing and comparison of the three models using regional distribution data for both rare and common species are still needed in order to ensure the superiority of the Nachman model at the regional scale.

Acknowledgments

References

Condit, R., Hubbell, S. P. & Foster, R. B. (1996). Changes in tree species abundance in a Neotropical forest: impact of climate change. *Journal of Tropical Ecology*, **12**, 231–256.

Currie, D. J. (1991). Energy and large scale patterns of animal and plant species richness. *American Naturalist*, **137**, 27–49.

Cutler, C. D. (1993). A review of the theory and estimation of fractal dimension. In *Nonlinear Time Series and Chaos*. Vol. I: *Dimension Estimation and Models*, ed. H. Tong, pp. 1–107. Singapore: World Scientific.

Fagan, W. F., Unmack, P. J., Burgess, C. & Minckley, W. L. (2002). Rarity, fragmentation, and extinction risk in desert fishes. *Ecology*, **83**, 3250–3256.

Fagan, W. F., Aumann, C., Kennedy, C. M. & Unmack, P. J. (2005). Rarity, fragmentation, and the scale dependence of extinction risk in desert fishes. *Ecology*, **86**, 34–41.

Gardner, R. H., Milne, B. T., Turner, M. G. & O'Neill, R. V. (1987). Neutral models for the analysis of broad-scale landscape patterns. *Landscape Ecology*, **1**, 19–28.

Gaston, K. J. (1994). *Rarity*. London: Chapman and Hall.

Gaston, K. J. & Blackburn, T. M. (2000). *Patterns and Process in Macroecology*. Oxford: Blackwell Science.

Halley, J. M., Hartley, S., Kallimanis, A. S., Kunin, W. E., Lennon, J. J. & Sgardelis, S. P. (2004). Uses and abuses of fractal methodology in ecology. *Ecology Letters*, **7**, 254–271.

Hanski, I. & Gyllenberg, M. (1997). Uniting two general patterns in the distribution of species. *Science*, **275**, 397–400.

Hartley, S. & Kunin, W. E. (2003). Scale dependency of rarity, extinction risk, and conservation priority. *Conservation Biology*, **17**, 1559–1570.

Hartley, S., Kunin, W. E., Lennon, J. L. & Pocock, M. J. (2004). Coherence and discontinuity in the scaling of species' distribution patterns. *Proceedings of the Royal Society of London, Series B*, **271**, 81–88.

He, F. & Gaston, K. J. (2000). Estimating species abundance from occurrence. *American Naturalist*, **156**, 553–559.

He, F. L. & Hubbell, S. P. (2003). Percolation theory for the distribution and abundance of species. *Physical Review Letters*, **91**, Art. No. 198103.

He, F., Gaston, K. J. & Wu, J. (2002). On species occupancy-abundance models. *Écoscience*, **9**, 119–126.

He, F., Zhou, J. & Zhu H. T. (2003). Autologistic regression model for the distribution of vegetation. *Journal of Agricultural, Biological, and Environmental Statistics*, **8**, 205–222.

Heikkinen, J. & Högmander, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, **43**, 569–582.

Holt, A. R., Gaston, K. J. & He, F. (2002). Occupancy-abundance relationships and spatial distribution: a review. *Basic and Applied Ecology*, **3**, 1–13.

Hubbell, S. P. & Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In *Tropical Rain Forest: Ecology and Management*, ed. S. L. Sutton, T. C. Whitmore & A. C. Chadwick, pp. 25–41. Oxford: Blackwell Scientific Publications.

Hurlbert, A. H. & Haskell, J. P. (2003). The effect of energy and seasonality on avian species richness and community composition. *American Naturalist*, **161**, 83–97.

Kunin, W. E. (1998). Extrapolating species abundance across spatial scales. *Science*, **281**, 1513–1515.

Kunin, W. E., Hartley, S. & Lennon, J. J. (2000). Scaling down: on the challenge of estimating abundance from occurrence patterns. *American Naturalist*, **156**, 560–566.

Leitner, W. A. & Rosenzweig, M. L. (1997). Nested species-area curves and stochastic sampling: a new theory. *Oikos*, **79**, 503–512.

Lennon, J. J., Greenwoord, J. J. D. & Turner, J. R. G. (2000). Bird diversity and environmental gradients in Britain: a test of the species–energy hypothesis. *Journal of Animal Ecology*, **69**, 581–598.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nachman, G. (1981). A mathematical model of the functional relationship between density and spatial distribution of a population. *Journal of Animal Ecology*, **50**, 453–460.

Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part I. *Ecology*, **43**, 185–215.

Raxworthy, C. J., Martinez-Meyer, E., Horning, N. *et al.* (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837–841.

Saupe, D. (1988). Algorithms for random fractals. In *The Science of Fractal Images*, ed. H. O. Peitgen & D. Saupe, pp. 71–113. New York: Springer-Verlag.

Šizling, A. L. & Storch, D. (2004). Power-law species-area relationships and self-similar species distributions within finite areas. *Ecology Letters*, **7**, 60–68.

Thomas, C. D., Cameron, A., Green, R. E. *et al.* (2004). Extinction risk from climate change. *Nature*, **427**, 145–148.

Tosh, C. A., Reyers, B. & Jaarsveld, A. S. (2004). Estimating the abundances of large herbivores in the Kruger National Park using presence-absence data. *Animal Conservation*, **7**, 55–61.

Turner, M. G., Gardner, R. H. & O'Neill, R. V. (2001). *Landscape Ecology in Theory and Practice*. New York: Springer-Verlag.

Warren, M., McGeoch, M. A. & Chown, S. L. (2003). Predicting abundance from occupancy: a test for an aggregated assemblage. *Journal of Animal Ecology*, **72**, 468–477.

Wilson, R. J., Thomas, C. D., Fox, R., Roy, D. B. & Kunin, W. E. (2004). Spatial patterns in species distributions reveal biodiversity change. *Nature*, **432**, 393–396.

With, K. A., Gardner, R. H. & Turner, M. G. (1997). Landscape connectivity and population distributions in heterogeneous environments. *Oikos*, **78**, 151–169.

Witte, J. P. M. & Torfs, P. J. J. F. (2003). Scale dependency and fractal dimension of rarity. *Ecography*, **26**, 60–68.

Wright, D. H. (1991). Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, **18**, 463–466.

Wu, J., Shen, W.-J., Sun, W.-Z. & Tueller, P. T. (2003). Empirical patterns of the effects of changing scale on landscape metrics. *Landscape Ecology*, **17**, 761–782.