

## Scaling Down: On the Challenge of Estimating Abundance from Occurrence Patterns

William E. Kunin,\* Stephen Hartley, and Jack J. Lennon

Centre for Biodiversity and Conservation, University of Leeds,  
Leeds LS2 9JT, United Kingdom

Submitted November 15, 1999; Accepted June 30, 2000

---

*Keywords:* species distributions, fractals, negative binomial, spatial autocorrelation, conservation.

---

One of the key goals of ecology is to explain the distribution and abundance of species. It has become apparent, however, that even to describe these matters is fraught with difficulty. The abundance or rarity of a species can be described in many ways, ranging from global geographical range through regional ubiquity and population structure down to local population size or cover. These different aspects of abundance can be thought of as different scales of analysis and indeed can be reflected by mapping a population at varying scales of resolution. There is, however, a fundamental asymmetry in such maps; a fine-scale map of a given area can be used to generate a coarse-scale map, but a coarse-scale map does not provide enough information to produce a fine-scale map. This is unfortunate, as a considerable amount of coarse-scale information about species distributions is available in many countries (e.g., Perring and Walters 1962; Jalas and Suominen 1987), but management decisions in conservation biology and other applied disciplines often require fine-scale (e.g., population) data.

Fortunately, the situation is not as bleak as it may seem at first glance. Even if fine-scale maps cannot be generated from coarse ones, it may be possible to infer some of the important properties of the fine-scale map from coarse-scale information. One of us (W. E. Kunin) recently demonstrated that some properties of a species distribution behave in similar ways at different scales, making it possible (at least in principle) to extrapolate coarse-scale information down to finer scales (Kunin 1998). “Scale-area

curves” were generated in which the area deemed to be occupied by a species was plotted (on logarithmic axes) as a function of the scale of analysis, and these curves were then extrapolated down to predict occupancy at finer scales of resolution. That paper considered two methods of extrapolation: a linear extrapolation (implying a fractal distribution) tended to overestimate fine-scale abundance, whereas an empirically fitted quadratic curve produced estimates without apparent bias. While that paper was restricted to the analysis of fairly coarse-scale distributional patterns (at 2–50 km linear resolution), it was suggested that similar methods might be applied to predict much finer distributional information, possibly to the point of generating crude population estimates.

He and Gaston (2000) make considerable progress toward realizing this goal. They demonstrate that a multiple-scale version of a negative binomial distribution (NBD) can be used to provide fairly accurate population estimates for tree species within an intensively surveyed 50-ha Malaysian forest plot (from the Pasoh Forest Reserve), using occupancy data from two grids of different resolutions. While their paper is presented in part as a challenge to the methods proposed in Kunin (1998), it serves primarily to confirm the key argument of that paper: that coarse-scale information can be used to predict fine-scale abundance.

Before proceeding further, a brief explanation of He and Gaston’s method is required. In general usage, the NBD is applied in a scale-specific manner, with the aggregation parameter ( $k$ ) estimated from the way in which  $N$  individuals are distributed across a set of quadrats or naturally occurring sample units (e.g., Bliss and Fisher 1953). For a given  $k$  and  $N$ , the number of empty quadrats expected from a NBD can be calculated (see Wright 1991; Hartley 1998). He and Gaston’s method turns this relationship around, and by assuming a common  $k$  for different scales, they are able to use the grid cell occupancy observed in two maps produced at different scales to solve simultaneously for  $N$  and  $k$ . This allows a scale-area curve (which they rename an “area-area curve”) to be drawn, predicting grid cell occupancy at any scale (e.g., fig. 1). In some ways, He and Gaston’s new method is similar to the two explored

\* E-mail: w.e.kunin@leeds.ac.uk.

in Kunin (1998); but rather than extrapolating a straight line (Kunin's method 1) or fitting a post hoc quadratic curve (Kunin's method 2), they fit and extrapolate a non-linear curve derived from a scale-invariant NBD. In doing so, they are able to perform a feat that was suggested in Kunin (1998) but has not hitherto been attempted: estimating population size ( $N$ ) from local occupancy patterns. The success of this venture is very good news; it provides the strongest evidence to date that the scaling asymmetry can be crossed and that useful information can be brought back from the other side.

In this note, we will explore some of the strengths and limitations of the box-counting fractal method used by Kunin (1998), of the NBD model proposed by He and Gaston (2000), and of the tests used to compare them. We conclude with suggestions for how best to use each model and possible future directions for combining the virtues of both.

#### Difficulties with the Fractal Model

We see two particular difficulties in the practical application and testing of the fractal model. The first of these problems has to do with an ambiguity as to how occupancy should best be measured. For scales much coarser than the scale of an individual, there is little practical difference between considering individuals as points or as space-filling objects. At very fine-scales of analysis, however, the difference becomes crucial: if individuals are treated as space filling, the total area occupied by a species drops off slowly at fine scales; if they are treated as points, the area drops off very steeply. Kunin's (1998) fractal extrapolation model treats individuals as space-filling objects that can occupy more than one cell (see, e.g., Kunin 1998, fig. 1B, 1-mm scale). The "observed" values plotted in He and Gaston's figure 2B, however, are generated from point data in which each tree can occupy only a single cell, however fine the grid resolution. Even if the fractal model made perfectly correct estimates of the area occupied by space-filling trees, they would appear well above the  $X = Y$  diagonal line when plotted against the observed occupancy of point-sized trees.

The fact that the fractal model is based on area rather than points makes the estimation of population sizes difficult (which creates a clear advantage for a method such as the NBD that estimates population size,  $N$ , directly). Indeed, the best course of action may be to avoid using an area-based model of this sort at any scale fine enough for the distinction between space-filling and point-based models to be important (see also Pfeifer and Obert 1989). If one nonetheless wishes to use the fractal model to predict  $N$ , then it is essential to choose an appropriate base scale; one where each occupied cell contains, on average,

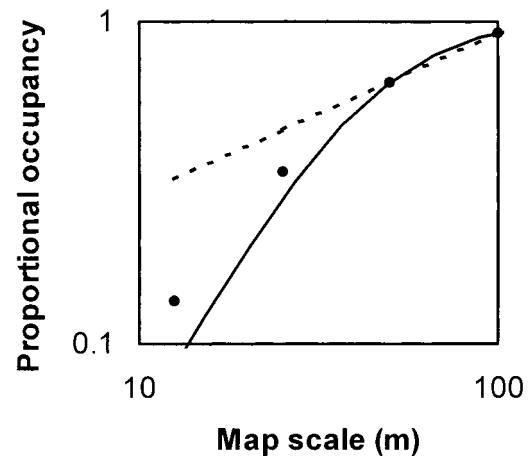
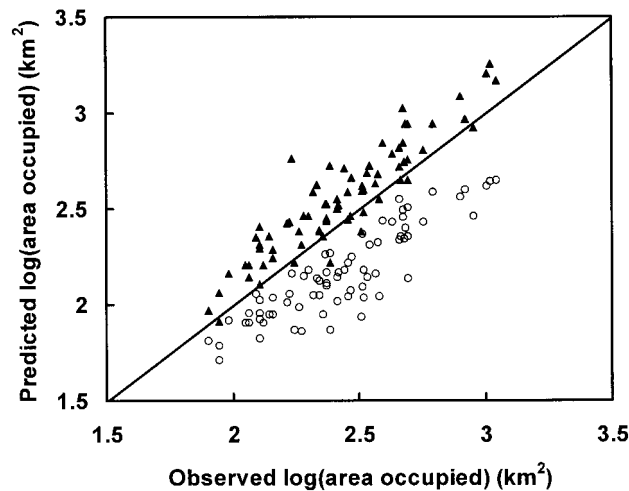


Figure 1: Scale-area plots for *Dacryodes rubiginosa*. Points represent the observed abundance of the species at 100-, 50-, 25-, and 12.5-m linear resolutions (data from F. He, personal communication). Lines represent fine-scale projections by means of the negative binomial distribution (solid line) and the fractal (dotted line) models, based on 100- and 50-m data.

precisely one individual. At this scale, cell occupancy can be equated with population number. If too coarse a scale is used, the average occupied cell will contain more than one individual (especially if individuals are clustered together); if too fine a scale is used, most individuals will occupy more than one cell (especially if individuals are large). Unfortunately, there is no easy recipe for finding this best scale a priori; it should depend on the average size of an individual and the degree of aggregation in the population. As species differ greatly in both of these respects, each species requires its own scale of analysis. Despite Kunin's (1998) speculation to the contrary, the ideal grid resolution for estimating population size from occupancy will generally be considerably coarser than the mean size of an individual, as a randomly placed space-filling plant is unlikely to fall precisely inside the boundaries of a cell its own size.

If one wishes to apply a common scale of analysis to estimate populations of multiple species, as He and Gaston (2000) do, the problem of choosing a scale is made even more difficult. Whatever scale is chosen will probably be wrong for most species. It seems likely that the 1.32-m<sup>2</sup> scale used by He and Gaston is rather too fine for the average tree. The figure was derived (at W. E. Kunin's suggestion) by assuming that the mean individual size could be calculated by dividing the total survey area by the total number of individuals (of all species), but it does not take into account the multistoried nature of a tropical forest or the lack of registration between individuals and grid cells (referred to above). Based on figures kindly made



**Figure 2:** Predicted and observed area occupied at 2-km scale for 73 scarce British plant species. Predictions from the fractal model (*solid triangles*) and the negative binomial distribution (NBD) model (*open circles*) are based on occupancy at 10- and 50-km scales (data from Stewart et al. 1994). Mean  $\pm$  SE of  $\log_{10}(\text{predicted/observed}) = 0.123 \pm 0.014$  (fractal),  $-0.265 \pm 0.015$  (NBD).

available to us by Fangliang He, it seems that the appropriate scale for converting grid cell occupancy into tree population estimates at Pasoh is between 5 and 12.5 m in linear resolution. While these scales seem reasonable for trees, the fact that they can only be discovered a posteriori underscores the weakness of the fractal model for making population estimates.

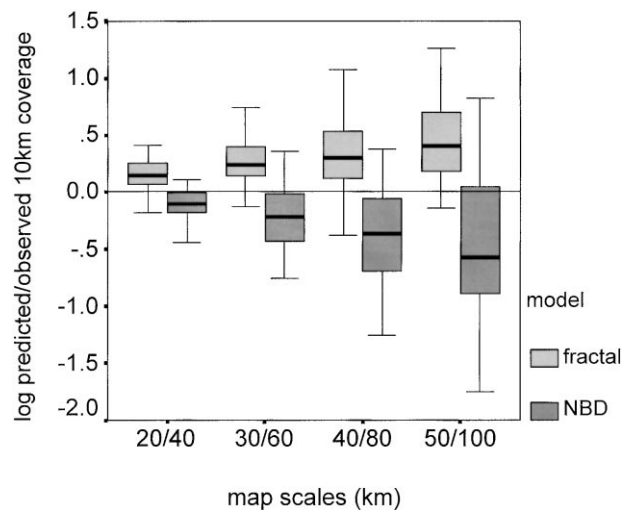
A second weakness of the box-counting fractal method used in Kunin (1998) is its tendency to overestimate the area occupied by a species. This effect was noted both in Kunin's original analysis and, more markedly, in He and Gaston (2000). The phenomenon appears to be caused, at least in part, by "grid saturation" (the converse of the "dilution effect" discussed by Gautestad and Myrsetrud 1994). The box-counting method is based on extrapolating the rate at which the area occupied by a species rises as one moves to coarser scales. For a truly fractal distribution, this rate of change should be constant (on logarithmic axes) at all scales. Yet the maximum area that can be occupied by a species is constrained by the total area of the map, and so this rise cannot continue at a constant pace indefinitely. Instead, the slope of the scale-area curve inevitably becomes shallower at coarser scales, especially in fairly common species (which approach map saturation soonest). Extrapolating down to fine scales from these flattened curves produces inappropriately high estimates for fine-scale occupancy.

Map saturation creates errors for a box-counting fractal analysis even when the underlying distribution is genuinely

fractal. We have confirmed this by generating synthetic fractal distributional patterns and then extrapolating down from the area occupied at coarse scales using the box-counting technique. The result was a systematic overestimation of fine-scale occupancy, exactly as found in real data. This property is probably a weakness of the box-counting approach to fractal analysis, rather than of fractal methods in general, and other techniques (e.g., correlation fractals, Hastings and Sugihara 1993; information fractals, Li, in press) deserve further attention in this regard.

### Difficulties with the NBD Model

If the fractal model does not seem ideally suited to the task set it by He and Gaston (2000), it is equally true that the NBD model does not seem to perform well when applied to national scale data of the sort used in Kunin (1998). We have applied He and Gaston's model to Kunin's original data set of "scarce" British plants and find that it consistently underestimates fine-scale occupancy, often by a considerable degree (fig. 2). As noted above, Kunin (1998) found that the fractal model tended to overestimate systematically the area occupied at fine scale in this data



**Figure 3:** Prediction of area occupied at the finest available scale (10 km) for 92 species of British grasses. Fractal and negative binomial distribution (NBD) models were fit to pairs of coarser scale maps, as indicated. The accuracy of prediction is expressed as the ratio of predicted to observed number of 10-km cells occupied ( $\log_{10}$  transformed). The NBD is a better fit at 20 km/40 km scale ( $t = 2.43$ ,  $P = .02$ ), the models are equally good at 30 km/60 km ( $t = 0.14$ ,  $P = .88$ ), but the fractal model is a better fit for the two remaining coarser scales ( $t = 5.54$ ,  $P < .001$  and  $t = 5.93$ ,  $P < .001$ ). For all map scales in both models, the bias away from predicted = observed is significant ( $P < .001$  throughout). Data are expressed as box and whisker plots, showing median values, quartiles, and trimmed ranges (following Tukey 1977).

**Table 1:** Effect of a scale-dependent  $k$  on the accuracy of NBD population estimates

$k$ ratio ( $k_c/k_f$ )	25,000/100,000 grid cells	2,500/10,000 grid cells	250/1,000 grid cells	25/100 grid cells
.9	1.034 (.19)	1.318 (.18)	4.125 (.15)	128.8 (.12)
1	1.000 (.20)	1.000 (.20)	1.000 (.20)	1.000 (.20)
2	.768 (.35)	.313 (.47)	.067 (.64)	.011 (.84)
4	.549 ( $5.6 \times 10^5$ )	.152 ( $4.1 \times 10^5$ )	.027 ( $1.5 \times 10^5$ )	.004 (23,019)

Note: Ratio of estimated to true population size ( $\hat{N}/N^*$ ) is followed by the estimated  $k$  ( $\hat{k}$ ) in parentheses. Both are derived by fitting He and Gaston's (2000) common- $k$  NBD model to pairs of maps at different scales. The shift in  $k$  across scales is expressed as a ratio:  $k_c/k_f$  where  $k_c$  and  $k_f$  represent the true value of  $k$  at coarse and fine scales, respectively. In all cases,  $N^* = 10,000$  and  $k_c = 0.2$ . Note that the effect of a given shift in  $k$  on parameter estimates grows larger as the map is analyzed at progressively coarser scales.

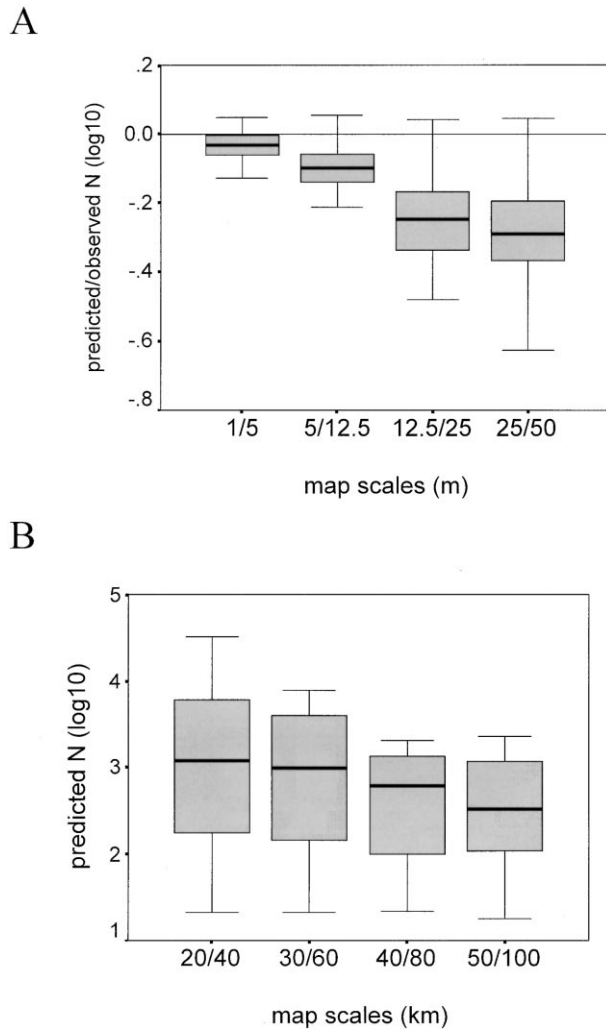
set. Applying both the NBD and the fractal model to national distributional data on 92 species of British grasses (data from Perring and Walters 1962), confirms this pattern: while the NBD model does as well or better than the fractal model in predicting cell occupancy over very short (e.g., fourfold area) ranges of scales, it predicts considerably less well as the range of extrapolation increases (fig. 3). The pattern of errors is consistent with that shown in the scarce plant data set: while the fractal model overestimates fine-scale abundance, the NBD model consistently underestimates it. This difficulty, along with some additional problems outlined below, raise doubts as to how well the NBD model can perform when applied to regional and national scale data.

We have two related concerns regarding the appropriateness of the NBD to modeling spatially continuous patterns and, in particular, to relatively coarse-scale data. Strictly, the NBD should not be applied to contiguous quadrats (Ripley 1981), as spatial autocorrelation renders quadrat counts nonindependent. This may not be a serious problem when the NBD is applied at a single map scale, but it can have serious consequences when maps at more than one scale are used together in parameter estimation. The degree of spatial autocorrelation will determine whether the value of  $k$  grows, shrinks, or remains constant as one moves to coarser scales of analysis (appendix; see also Taylor et al. 1979). In practice  $k$  generally seems to grow progressively larger at coarse scales. For example, of the 434 species in the Pasoh data set with positive  $k$  values at both scales, 420 have larger  $k$ 's when mapped at 50-m resolution than at a 1-m scale, with a 27.5-fold increase in the median value over a 50-fold change in scales ( $k$  estimated at single scales from observed  $N$  and cell occupancy; binomial  $P \ll .001$ ). Looking at a very different system (temperate herbaceous plants rather than tropical trees) at a very different range of scales (tens of kilometers rather than tens of meters), we find the same pattern in the 92 British grass species referred to above: 90 of 92 species show an increase in  $k$  at coarser scales of analysis (here using He and Gaston's two-scale method for esti-

ating  $k$ ; binomial  $P \ll .001$ ), with a 2.4-fold increase in median  $k$  over the fivefold range of scales considered (10–20-km to 50–100-km scales).

He and Gaston (2000) acknowledge these difficulties, citing the potential effects of spatial autocorrelation on parameter estimation and, indeed, illustrating a shift in  $k$  for one species (*Dacryodes rubiginosa*, their table 2). They conclude, however, that the effect of such shifts on numerical estimates is relatively minor, at least in the Pasoh data set. We are less certain that the issue can be ignored when coarser scale data are considered. We have used He and Gaston's method to estimate  $N$  and  $k$  from distributions in which the actual  $k$  values at two map scales are allowed to differ. If  $k$  is kept constant across scales, the method consistently finds the correct  $k$  and  $N$  values. When two maps representing the same population of  $N^*$  individuals at different resolutions differ in their  $k$  values, however, then the estimated  $k$  and  $N$  ( $\hat{k}$  and  $\hat{N}$ , respectively) are systematically in error. Where the coarser-scale  $k$  is the larger (by far the commonest situation), the estimated  $\hat{k}$  is greater than either of the true  $k$  values, and the estimated  $\hat{N}$  is consistently lower than the true  $N^*$ . The relative size of this error appears to depend on the number of individuals per cell and thus grows progressively greater at coarser map scales (table 1; see also Anson 1950).

This predicted downward shift in abundance estimates at coarser scales is reflected both in the Pasoh data set and in our analyses of British grasses. In each case, there is a progressive decline in the estimated population size of species as the scale of analysis grows coarser (fig. 4). This bias should accumulate as one moves from local to regional and then national or even global scales of analysis. As a result, population estimates based on national-scale maps become unrealistically low: the median population size predicted by the NBD model for the 73 scarce British plant species considered in Kunin (1998) is 34.7 (range 13.0–113.9), while the commonest of the 92 species of grass in Britain could be predicted to have a population



**Figure 4:** Effects of scale on estimates of population size ( $N$ ) using He and Gaston's negative binomial distribution model. *A*, Prediction of  $N$  for each of the 824 tree species in the Pasoh data set, using pairs of maps at successively coarser scales. The  $\log_{10}$  ratio of predicted to observed abundance declines with the coarseness of the maps used (regression slope  $t = -22.6$ ,  $P < .001$ ). *B*, Estimated  $N$  for 92 species of British grasses. Again, estimated  $N$  declines with coarsening scale of the maps used (regression slope  $t = -5.60$ ,  $P < .001$ ). Data are expressed as box and whisker plots, following Tukey (1977).

of only 2,256 individuals (projecting the NBD model from 50- and 100-km scales; median = 328).

This difficulty in making long-range projections is disappointing; ultimately, the goal of projecting abundance across scales is not just to make accurate predictions of populations within a local site (as He and Gaston 2000 do), nor to make mesoscale predictions at a national scale (as Kunin 1998 does), but to make local-scale population predictions from fairly coarse-resolution national data-

bases—that is, to span the range of scales considered in both papers combined. The NBD model seems unlikely to be successful over this range of scales; perhaps no model can be.

#### Future Directions

Despite the limitations of current methodologies, results in both Kunin (1998) and He and Gaston (2000) suggest that it is possible to deduce useful information about fine-scale population patterns from coarse-scale data. Just how accurate such extrapolations can become over how great a range of scales remains an open question and a key focus for future work.

In the meantime, which model is to be preferred? The NBD model seems much the better suited to dealing with scales close to the size of an individual, scales where the current area-based fractal model breaks down, and it provides much more direct estimates of  $N$  (although these estimates become unreliable when made from coarse scales). At relatively coarse scales, the fractal model seems to perform best, at least for species well below map saturation. As expected, the accuracy of both models deteriorates as extrapolations are made across progressively greater ranges of scales. It may well turn out that there are intrinsic limits to the range of such extrapolations, if, for instance, distributional patterns are determined by completely different processes at local and regional scales.

On the other hand, the complementary success of two very different methods, one working down from national distributional geometry, the other building up conceptually from individual point patterns, is encouraging. It is also encouraging to note that, over most of the data sets considered, reality seems to lie somewhere between the strict linearity of the fractal extrapolation and the specific curvature of the NBD model. Perhaps an intermediate model can be devised with just the right amount of curvature across scales to capture successfully the strengths of both models. Alternatively, post hoc models might be adapted to fit the specific curvature evident in different species or groups. Ultimately, perhaps the goal should be to find ways to avoid the rather arbitrary and inefficient “blocking-up” process completely and to concentrate instead on making better use of the information contained in the autocorrelation structure of the single most detailed map. The question is no longer whether abundance can be estimated from occurrence but how best it should be done.

#### Acknowledgments

We thank K. J. Gaston and F. He for discussions on this topic and access to their data set and the Biological Records

Centre at the Institute of Terrestrial Ecology, Monks Wood, for access to distributional data on British plants. The work was made possible by a grant from the Natural Environment Research Council (GR3/11916) and by a joint venture agreement with the Pacific Northwest Research Station of the U.S. Forest Service (PNW 99-1006-1-JVA).

## APPENDIX

### Spatial Autocorrelation and Its Effects on the Scaling Behavior of $k$

The aggregation parameter  $k$  of the NBD (here at the finer of two scales), is given by the mean and variance of counts of individuals between sample units (quadrats), such that

$$k_f = \frac{\mu_f^2}{\sigma_f^2 - \mu_f}.$$

We generally expect  $k$  to change with scale when quadrats are amalgamated. To illustrate this, we consider two extreme cases. At one extreme (no autocorrelation),  $k$  tends to increase at coarser scales, while at the other (strong autocorrelation), it tends to decrease.

#### Case 1: No Spatial Autocorrelation in Quadrat Counts

We define the coarser-scale map as the amalgamation of  $2 \times 2$  quadrats of the finer-scale map, so the mean count within coarse quadrats is four times greater. If we have no autocorrelation in counts between adjacent fine-scale quadrats, the variance in counts between the coarser quadrats is four times larger, so  $k$  at the coarser scale is

$$k_c = \frac{(4\mu_f)^2}{4\sigma_f^2 - 4\mu_f} = \frac{16\mu_f^2}{4\sigma_f^2 - 4\mu_f} = 4k_f.$$

Therefore,  $k$  is expected to increase with amalgamation of uncorrelated quadrats.

#### Case 2: Strong Spatial Autocorrelation in Quadrat Counts

Spatial autocorrelation will tend to make counts in adjacent quadrats more similar. As the limiting case of extreme autocorrelation, we can take the case where the  $2 \times 2$  quadrats amalgamated together have identical counts. Different coarse-scale quadrats, however, may still have different counts; indeed the variance in counts between the coarser quadrats should now be 16 times that of the fine-scale counts, and

$$k_c \approx \frac{(4\mu_f)^2}{16\sigma_f^2 - 4\mu_f} \approx \frac{\mu_f^2}{\sigma_f^2 - \mu_f/4}.$$

This is smaller than the fine-scale  $k$ , so we expect  $k$  to decrease with amalgamation of strongly autocorrelated quadrats.

Reality will almost certainly lie somewhere between these two extreme cases of no autocorrelation and extreme autocorrelation. Indeed, between these extremes, there should be an intermediate level of spatial autocorrelation where  $k$  remains constant as cells are amalgamated. As we tend to find  $k$  increasing at coarse scales (see text), most species distributions examined here seem to lie between this intermediate case and the extreme case with no autocorrelation.

### Literature Cited

- Anscombe, F. J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* 37:358–382.
- Bliss, C. I., and R. A. Fisher. 1953. Fitting the negative binomial distribution to biological data: note on the efficient fitting of the negative binomial. *Biometrics* 9: 176–200.
- Gautestad, A. O., and I. Mysterud. 1994. Fractal analysis of population ranges—methodological problems and challenges. *Oikos* 69:154–157.
- Hartley, S. 1998. A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same). *Journal of Animal Ecology* 67:992–994.
- Hastings, H. M., and G. Sugihara. 1993. *Fractals: a user's guide for the natural sciences*. Oxford University Press, Oxford.
- He, F., and K. J. Gaston. 2000. Estimating species abundance from occurrence. *American Naturalist* 156: 000–000.
- Jalas, J., and J. Suominen, eds. 1987. *Atlas florae Europaeae: distribution of vascular plants in Europe*. Vol. 1. Cambridge University Press, Cambridge.
- Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. *Science* (Washington, D.C.) 281:1513–1515.
- Li, B. In press. Applications of fractal geometry and percolation theory to landscape analysis and assessments. Chapter 14 in P. S. Bourgeron, M. Jensen, and G. Lesard, eds. *An integrated ecological assessment protocols guidebook*. Springer, New York.
- Perring, F. H., and S. M. Walters, eds. 1962. *Atlas of the British flora*. Nelson, London.
- Pfeifer, P., and M. Obert. 1989. Fractals: basic concepts and terminology. Pages 11–43 in D. Avnir, ed. *The fractal approach to heterogeneous chemistry: surfaces, colloids, polymers*. Wiley, Chichester.

- Ripley, B. D. 1981. *Spatial statistics*. Wiley, New York.
- Stewart, A., D. A. Pearman, and C. D. Preston, eds. 1994. *Scarce plants in Britain*. Joint Nature Conservancy Council, Peterborough.
- Taylor, L. R., I. P. Woiwod, and J. N. Perry. 1979. The negative binomial as a dynamic ecological model for aggregation, and the density dependence of  $k$ . *Journal of Animal Ecology* 48:289–304.
- Tukey, J. W. 1977. *Exploratory data analysis*. Addison-Wesley, Reading, Mass.
- Wright, D. H. 1991. Correlations between incidence and abundance are expected by chance. *Journal of Biogeography* 1:463–466.

Associate Editor: Lenore Fahrig