

# A multi-fractal model for the species-area relationship

Henri Laurie<sup>1</sup> & Edith Perrier<sup>2,1</sup>

<sup>1</sup>*Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa.*

<sup>2</sup>*IRD, UR 079 GEODES, 32 Ave. H. Varagnat, Bondy 93143 cedex, France.*

The species-area relationship (SAR) of an ecologically defined region quantifies the way species number  $S$  in a survey increases with area  $A$  surveyed. It has long been fundamental to discussions of species richness<sup>1,7,16,25,28</sup>, which is one of the central problems of theoretical ecology, and it has also been used in applied ecology<sup>6,21,32</sup>. SAR's are routinely presented as real-valued functions  $S = f(A)$ , often

$$S = cA^z, \tag{1}$$

though competing forms of  $f$  exist<sup>5,16</sup>. Such an SAR does not model spatial variability; in applications the variability in  $S$  at a given size  $A$  is smoothed out by taking an average. Here we view  $S$  as instead intrinsically variable at all scales of  $A$ —in other words, as multifractal. We propose a two-parameter multiplicative cascade to model this phenomenon. We derive a graphical characterisation of multi-fractality and density variation across scales. Our model generalises Equation 1, of which it retains the scaling properties recently emphasised in the literature<sup>9,11,13,14,17,18,22,23,26</sup>.

As an application, we show that data on spatial variability in richness of Proteaceae in the Cape Floristic Region.

## Model description

Consider a rectangle with area  $A_0$  that contains  $S_0$  species (the geometric assumptions can be relaxed, as we do when considering real data). Bisect the rectangle into two congruent rectangles of area  $A_1 = 2^{-1}A_0$ , then bisect each of these with a line at right angles to the previous bisector to reach four rectangles of area  $A_2 = 2^{-2}A_0$ . Continue this process, reaching  $2^k$  rectangles of area  $A_k = 2^{-k}A_0$  after  $k$  bisections. This is the basis of the Hartean derivation <sup>11</sup>, from which we now depart.

As each parent rectangle is split in two, spatial heterogeneity in species richness implies that its offspring inherit unequal amounts of species. Denote by  $p_1$  the fraction of species inherited by the richer half and by  $p_2$  the fraction inherited by the poorer half. Since every species in the larger rectangle must occur in at least one of the smaller rectangles, we have  $p_1 + p_2 \geq 1$ . At each stage, the position of the richer half is randomly assigned.

The simplest case is when  $p_1$  and  $p_2$  are constant; this assumption implies self-similarity. After  $k$  bisections the heterogeneity is captured in the binomial  $(p_1 + p_2)^k = \sum_{j=0}^k \binom{k}{j} p_1^{k-j} p_2^j$ . Thus species richness takes the  $(k + 1)$  values  $p_1^{k-j} p_2^j S_0$  for  $j = 0$  to  $k$ , and each level of richness occurs in  $\binom{k}{j}$  of the  $2^k$  rectangles of area  $2^{-k}A_0$ . If one denotes by  $P_k(S = n)$  the probability that a rectangle of area  $2^{-k}A_0$  contains  $n$  species, then we can restate this result as the species richness distribution

$$P_k(S = S_0 p_1^{k-j} p_2^j) = \binom{k}{j} / 2^k. \quad (2)$$

Equation 2 applies to any self-similar density which is variable at every scale, and is perhaps the simplest of all models with this property. Typical realisations are shown in Figure . Moreover, the support of the density may be any geometric object with finite area, because for a given slope such an area has a unique bisector, and hence the subdivision process described above is shape-independent.

One can visualise the variability of  $S$  versus  $A$  in a simple diagram: plot  $S = p_1^{k-j} p^j S_0$  versus  $A = 2^{-k}$ , where  $j = 0$  to  $k$  for  $k = 0$  to  $k_{\max}$ . This gives  $(k + 1)$  vertical lines, the leftmost of which has  $k + 1$  points and the rightmost a single point. On a loglog plot (see Figure ) these vertical lines are bounded by a triangle with its vertex at  $(\log(A_0), \log(S_0))$  on the right and two lines sloping up to it. The line through the upper bounds of richness has slope  $\log(p_1)/\log(2)$ , and the line through the lower bounds of richness has slope  $\log(p_2)/\log(2)$ . It is obvious that a model which produces a diagram very like Figure must essentially be the same as Equation 2. In fact, any multifractal density will have a roughly triangular footprint in a loglog plot of density versus sample area. As we show below, this is also true for at least some data. We suggest that the triangular footprint is a simple way to characterise a multifractal density.

In the limit as  $k \rightarrow \infty$ , each rectangle is a scaled copy of the any of the larger ones. The scaling applies not only to size but also to density: as we rescale the smaller to the size of the larger, we must also rescale all densities in proportion so that the average densities are equal. In this sense, the density defined by Equation 2 is exactly self-similar. It therefore has all the scaling properties that derive from self-similarity <sup>9,11</sup>. In fact, it reduces to Equation 1 as follows. Define

$a = (p_1 + p_2)/2$ . One can show that the average number of species in a rectangle of area  $2^{-k}A_0$  is  $a^k S_0$ . Then Equation 1 is easily derived <sup>11</sup>, with  $a = 2^{-z}$ . Because of self-similarity as  $k \rightarrow \infty$ , the same value of  $z$  holds for any rectangle irrespective of its average density.

### **Case study: spatial variability in species richness of Cape Proteaceae**

The multifractal density of Equation 2 should be widely applicable. Indeed, the only requirement is a representative sample of masses at each of a wide range of scales. Nevertheless, species richness data are particularly suitable, both for wide range of scales at which they can be found and for ease of interpretation.

We consider all species in the family Proteaceae that occur in the Cape Floristic Region<sup>15,29,31</sup> (CFR henceforth). This is the smallest of the world's six floristic regions, among which it has the highest endemism. The CFR ranges in latitude from 30°S to 34°S and contains almost the entire fynbos biome. By permission of Tony Rebelo of SANBI we obtained a map of the CFR rasterised at 1' resolution, and the counts of Proteaceae species richness in the 9 426 1'*times*1' cells that contain fynbos; for a map see Figure . Larger areas were also counted; our full data set contains nearly 100 000 ( $A, S$ ) data pairs. See Figure for a loglog plot. By least squares regression on the logarithms of  $S$  and  $A$  we find that  $z \approx 0.47$ , which implies that  $p_1 + p_2 \approx 1.44$ .

We calculated the approximate Renyi dimensions (see Methods) of the data, and compared them to the exact Renyi dimensions of the model, subject to  $p_1 + p_2 = 1.44$ . By varying  $p_1$ , it is evident that there must be a best fit, and that it must occur near  $p_1 = 0.83$ ; see Figure . In other

words, we claim that the spatial pattern of species richness predicted by the model for the square in Figure (a) is equivalent to the observed richness at a comparable scale as seen on the map in Figure (b).

We can test the estimate  $p_1 = 0.83$ ,  $p_2 = 0.61$  by looking at the data in other ways. First we compare scatterplot of the data in loglog space with the triangular footprint of the model. The data do fill a roughly triangular region, and we see from Figure that again the best fit is near  $p_1 = 0.83$ . We use  $k = 14$  because the model then has a larger number of finest scale units than the data, so that one would expect the data to correspond to a subset of the model.

A more exacting test is to compare histograms of the data to histograms derived from Equation 2 for various levels of  $k$ ; see Figure . We see by the approximate match between the largest predicted richness and the largest observed richness that the model is in qualitative agreement with the data. Note that the map in Figure (b) predicts  $0 < S < 31$  as against observed  $1 < S < 38$ , also qualitatively in good agreement.

## Methods

**Exact Renyi dimensions for densities known on tessellations of arbitrary mesh size** The Renyi dimensions of an object with possibly varying density are given by <sup>8</sup>

$$D(q) = \begin{cases} \lim_{r \rightarrow 0} \frac{1}{q-1} \frac{\log \sum_{i=1}^{N(r)} p_i^q(r)}{\log r} & \text{if } q \neq 1 \\ \lim_{r \rightarrow 0} \frac{\sum_{i=1}^{N(r)} p_i(r) \log p_i(r)}{\log r} & \text{if } q = 1, \end{cases} \quad (3)$$

where  $r$  is a characteristic length,  $p_i$  is the proportion of the density in box  $i$ , and it takes  $N$  identical boxes with length  $r$  to cover the object. For a fractal object  $D(0)$  is a non-integer; for a standard fractal like the Sierpinski gasket  $D$  is a constant; for a multifractal  $D$  may be a curve.

Substituting the probabilities from Equation 2 into Equation 3 gives

$$D(q) = \begin{cases} \frac{\log(p_1^q + p_2^q) - q \log(p_1 + p_2)}{(1-q) \log(\sqrt{2})} & \text{if } q \neq 1 \\ \frac{(p_1 + p_2) \log(p_1 + p_2) - p_1 \log(p_1) - p_2 \log(p_2)}{(p_1 + p_2) \log(\sqrt{2})} & \text{if } q = 1. \end{cases} \quad (4)$$

This formula yields the exact Renyi dimensions in Figure .

Note that spatially homogeneous species richness is not fractal in the geometric sense, even when it satisfies Equation 1 (in which case  $p_1 = p_2 = a = 2^{-z}$ ).

The 2-dimensional multifractal constructed by Stanly and Meakin<sup>30</sup> takes  $k/2$  steps to reach  $2^k$  rectangles of area  $2^{-k} A_0$  and its parameters are the same  $p_1$  and  $p_2$  that we use. Denoting by  $D_{SM}$  the Renyi dimensions of their multifractal<sup>24</sup>, one has  $D_{SM} = 2D + 2$ , where  $D$  is given by Equation 4.

**Approximate Renyi dimensions for densities known only from data on irregular domains** We noted above that a cascade of bisections is possible on any domain with finite area. However, for real data the calculation of the limit in Equation 3 is infeasible. We used the following approximation. Assume that the data are available at several scales  $A_i$ , where  $A_0$  is the area of the whole region. At each such scale, randomly select from the data  $m$  replicates such that  $mA_i \approx A_0$ . We found that good approximation was important here, and kept the relative errors below about 5%. The slope of linear regression of  $S^q$  against  $A$  then yields  $D(q)$  in the standard way<sup>4,24,33</sup>. Obviously this requires that the  $m$  replicates at each scale represent the true variability and that many scales over a large range are available; the Protea Atlas CFR data that we used met both requirements. We chose only those  $A_i$  with  $m$  or more data in the sample, and easily had enough scales for good estimates of  $D(q)$  (all fits had  $R^2 > 0.99$ ).

1. Arrhenius, O. 1921. Species and area. *Journal of Ecology* 9, 95–99.
2. Arita, H. and Rodríguez, P. 2002. Geographic range, turnover rate and the scaling of species diversity. *Ecography* 25, 541–550.
3. Banavar, J., Green, J., Harte, J., and Maritan, A. 1999. Finite size scaling in ecology. *Physical Review Letters* 83, 4212–4214.
4. Borda-de Água, L., Hubbell, S., and McAllister, M. 2002. Species-area curves, diversity indices, and species abundance distributions: a multi-fractal analysis. *The American Naturalist* 159, 138–155.

5. Connor, E. and McCoy, E. 1979. The statistics and biology of the species-area relationship. *The American Naturalist* 113, 791–833.
6. Driver, A., Desmet, P., Rouget, M., Cowling, R., and Maze, K. 2003. Succulent Karoo Ecosystem Plan. Technical Report CCU 1/03, Cape Conservation Unit of the Botanical Society of South Africa, Cape Town.
7. Gleason, H. A. 1922. On the relation between species and area. *Ecology* 3, 158–162.
8. Harte, D. 2001. *Multifractals: Theory and Applications* Chapman & Hall/CRC, Boca Raton.
9. Harte, J., Blackburn, T., and Ostling, A. 2001. Self-similarity and the relationship between abundance and range size. *The American Naturalist* 157, 374–386.
10. Harte, J. and Kinzig, A. 1997. On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. *Oikos* 80, 417–427.
11. Harte, J., Kinzig, A., and Green, J. 1999. Self-similarity in the distribution and abundance of species. *Science* 284, 334–336.
12. Harte, J., McCarthy, S., Taylor, K., Kinzig, A., and Fischer, M. 1999. Estimating species-area relationships from plot to landscape scale using species turnover data. *Oikos* 86, 45–54.
13. He, F. and Legendre, P. 2002. Species diversity patterns derived from species-area models. *Ecology* 83, 1185–1198.
14. Lennon, J. J., Kunin, W. E. and Hartley, S. 2002. Fractal species distributions do not produce power-law species-area relationships. *Oikos* 97, 378–386.

15. Laurie, H., Rebelo, A. G., Silander Jr, J. A. and Smit, W. *submitted*. Spatial variation in power-law species-area curves for a single clade in a single biome. *EER submitted*.
16. Lomolino, M. 2001. The species-area relationship: new challenges for an old pattern. *Progress in Physical Geography* 25, 1–21.
17. Maddux, R. D. 2004. Self-similarity and the species-area relationship. *The American Naturalist* 103, 616–626.
18. Martin, H. G. and Goldenfeld, N. 2001. On the origin and robustness of power-law species-area relationships in ecology. *Proceedings of the National Academy of Sciences of the USA* 103, 10310–10315.
19. Nichols, J. D., Boulinier, T., Hines, J. E., Pollock, K. H. and Sauer, J. R. 2001. Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology* 12, 1390–1398.
20. Matter, S., Hanski, I., and Gyllenberg, M. 2002. A test for the metapopulation model of the species-area relationship. *Journal of Biogeography* 29, 977–983.
21. Ney-Nifle, M. and Mangel, M. 2000. Habitat loss and changes in the species-area relationship. *Conservation Biology* 14, 893–898.
22. Ostling, A. and Harte, J. 2003. A community-level fractal property produces power-law species-area relationships. *Oikos* 103, 218–224.

23. Ostling, A., Harte, J., Green, J. L. and Kinzig, A. P. 2004. Self-similarity, the power-law form of the species-area relationship and a probability rule: a reply to Maddux. *The American Naturalist* 163, 627–633.
24. Perrier, E., Tarquis, A. and Dathe, A. 2006. A program for fractal and multifractal analysis of two-dimensional binary images: computer algorithms versus mathematical theory. *Geoderma* 134, 284–294.
25. Preston, F. 1960. Time and space in the variation of species. *Ecology* 41, 611–627.
26. Pueyo, S. 2006. Self-similarity in species-area relationship and in species abundance distribution. *Oikos* 112, 156–162.
27. Preston, F. 1962. The canonical distribution of commonness and rarity. *Ecology* 43, 185–215 and 410–432.
28. Rosenzweig, M. 1995. *Species diversity in space and time*. Cambridge: Cambridge University Press.
29. Rebelo, A. 1991. *Protea Atlas Manual: instruction booklet to the Protea Atlas Project*. Cape Town: Protea Atlas Project.
30. Stanley, H. E. and Meakin, P. 1988. Multifractal phenomena in physics and chemistry. *Nature* 335, 405–409.
31. Takhtajan, A. 1986. *Floristic Regions of the World*. Berkeley, CA: University of California Press.

32. Ulrich, W. 2005. Predicting species numbers using species-area and endemics-area relations  
*Biodiversity and Conservation* 14, 3351–3362.
33. Zhang, Y. X., Ma, K. M., Anand, M. and Fu, B. J. 2006. Do generalised scaling laws exist for  
species distribution in mountains? *Oikos* 115, 81–88.

**Figure 1** Multiplicative cascade, stages  $k = 0, 1, 2$  and  $3$ . At each stage, there are  $2^k$  bisections.  $p_1$  is the fraction of richness inherited by the richer half,  $p_2$  is the fraction of richness inherited by the poorer half. Position of richer half is randomly chosen among the two possibilities at each bisection

**Figure 2** Realisations of the multifractal defined by Equation 2 with  $p_1 = 1, p_2 = 0.6$ , and (a)  $k = 2$ , (b)  $k = 3$ , (c)  $k = 6$  and  $k = 14$ . Density scale is the same in all four cases: the maximum richness is  $p_1^2$ , which appears as black squares in the  $k = 2$  example, and the minimum richness is  $p_2^{14}$ , which appears only in the  $k = 14$  example.

**Figure 3** The footprint of Equation 2 in loglog space. Here,  $S_0 = 400, A_0 = 2^{10}, p_1 = 0.8, p_2 = 0.6$  and  $k$  ranges from 0 to 10.

**Figure 4** Observed richness as a function of area for Cape Proteaceae (blue). The black line shows the best least squares fit of  $\log(S) = \log(c) + z \log(A)$  to the 94 260  $(\log(A), \log(S))$  data pairs. Only about 5% of the data appear in this diagram.

**Figure 5** Exact Renyi spectra of various model parameterisations (red) compared to approximate Renyi spectrum of the data (blue). Model parameters satisfy  $p_1 + p_2 = 1.44$  in each case; see Methods of how the spectra were calculated. (a)  $p_1 = 0.72$  (b)  $p_1 = 0.80$  (c)  $p_1 = 0.83$  (d)  $p_1 = 0.86$  and (e)  $p_1 = 0.94$

**Figure 6** Visual comparison of model and data. (a) The square shows spatial variation predicted by the model using  $k = 14$  and the best parameter estimates  $p_1 = 0.83$ ,  $p_2 = 0.61$ .

(b) The map shows Proteaceae richness in the Cape Floristic Region at  $1' \times 1'$  resolution.

(c) Colour map of density ranges; the same map is used in both images.

**Figure 7** Loglog scatterplots of data (blue) with model predictions (red). Model uses  $k = 14$  and  $S_0 = 421$ , the extrapolation using  $z = 0.47$  in Equation 1 from the observed  $S = 372$  at  $A = 9426$  to  $S_0$  at  $A_0 = 2^{14}$ . In each case  $p_1 + p_2 = 1.44$ . (a)  $p_1 = 0.72$  (b)  $p_1 = 0.83$  (c)  $p_1 = 0.94$

**Figure 8** Histograms of model prediction (red) versus data (blue). In all cases, the model uses  $p_1 = 0.83$ ,  $p_2 = 0.61$ ,  $S_0 = 421$  and  $k = 2, 4, 6, 8, 10$  and  $12$ . For a given  $k$  the area  $A = 2^{14-k}$  is used; one can visualise such a sample as taking a vertical slice of the data in Figure 4.