Density dependence and a colonization rule predict spatial patterning

Micah Brush[1] **and John Harte**[2]

[1]micbru@berkeley.edu, Department of Physics, 366 LeConte Hall MC 7300, University of California, Berkeley, CA, 94720-7300, USA
[2]Energy and Resources Group, 310 Barrows Hall, University of California, Berkeley, CA, 94720-3050, USA

ABSTRACT

We present a model of intraspecific spatial aggregation that relates static spatial patterning to density dependence. Individuals are placed according to the colonization rule consistent with the Maximum Entropy Theory of Ecology (METE), and die with probability proportional to their abundance raised to a power α. $\alpha = 1$ recovers the strongly aggregated METE distribution, and as $\alpha \to 2$ our prediction approaches the binomial distribution. In between our model predicts less aggregation than METE, but more than random placement. Thus, more density dependence randomizes spatial patterning. A tropical forest has non-zero density dependence with $\alpha = 1.12$ at a 50 ha scale, and $\alpha = 1.0003$ in a 64 m^2 serpentine grassland plot. α increases at smaller scales in both datasets and the degree of density dependence varies across individual species and a 2^8 range of spatial scales. We infer this underlying density dependence from empirical spatial patterns.

Keywords: Theoretical ecology, density dependence, METE, spatial, aggregation, macroecology, colonization, scale, abundance, BCI

INTRODUCTION

Quantitative understanding of spatial patterns in ecosystems can illuminate underlying processes (Levin, 1992; Rosenzweig, 1995) and allow us to better predict ecosystem response to natural and anthropogenic disturbances (Thomas *et al.*, 2004; Newman *et al.*, 2018). It is also critical in understanding the species-area relationship, a well studied (Arrhenius, 1921; Plotkin *et al.*, 2000; Drakare *et al.*, 2006; Harte & Kitzes, 2015) macroecological pattern, and has applications in reserve designs and conservation (Kitzes & Shirley, 2016).

In this paper we consider the spatially explicit species-level abundance distribution $\Pi(n|A,A_0,n_0)$, defined as the probability that if a species has n_0 individuals in a plot of area A_0, then it has n individuals in a randomly selected subplot of area A.

A prediction of this function comes from the Maximum Entropy Theory of Ecology (METE), which simultaneously predicts many macroecological patterns (Harte, 2011; Harte & Newman, 2014) across a wide range of spatial scales, taxa, and habitats (Xiao *et al.*, 2015; White *et al.*, 2012). For the $\Pi(n|A,A_0,n_0)$ function, METE predicts very strong spatial aggregation. This is in contrast to the random placement model (Coleman, 1981), which predicts no spatial aggregation.

Empirically, we see that observations are often in between these two extremes (Conlisk *et al.*, 2012). Conlisk *et al.* (2007) added an extra parameter to the colonization rule that allows $\Pi(n)$ to vary, but it has no mechanistic interpretation and is used only as a fit parameter. Alternatively, a negative binomial is sometimes used to fit the spatial point process (Bliss & Fisher, 1953; He & Gaston, 2000, 2003) but again this leaves us without any ecological intuition about the degree of aggregation to expect in different ecosystems.

We derive a new model that uses the colonization rule consistent with METE and adds a density dependent death rule. This model predicts a more random spatial arrangement with greater density dependence and more spatial aggregation with less density dependence, allowing us to infer underlying density dependence from observed spatial patterning.

First, we briefly review the METE and random placement predictions. In Methods, we derive our new probability distribution and explain how we will compare it to data, as well as introduce the datasets that we will be using for comparison. In Results, we look at the shape of this probability distribution and then present aggregate and species-level statistics for the data. Finally, in Discussion, we place our results for the analysed datasets in a broader context, and describe possible applications.

METE

In METE, the $\Pi(n)$ function is given by (Harte *et al.*, 2008)

$$\Pi(n|A,A_0,n_0) = \frac{e^{-\lambda_\Pi n}}{Z_\Pi} \tag{1}$$

where Z_Π is a normalization factor, and λ_Π is the Lagrange multiplier determined by the constraint condition

$$\sum_n n\Pi(n|A,A_0,n_0) = \frac{n_0 A}{A_0}. \tag{2}$$

In the case of a bisection, $A = A_0/2$ and the Π function simplifies to

$$\Pi(n|n_0) = \frac{1}{n_0+1}, \tag{3}$$

which is independent of n. This means that given n_0 individuals, any arrangement of them on the two sides of a bisected plot or quadrat is just as likely as any other. Ecologically, this is equivalent to very strong spatial aggregation. This is in agreement with some datasets but fails in others, where the theory over-predicts aggregation (Conlisk *et al.*, 2007).

The prediction from METE is equivalent to the distribution obtained from using the Laplace rule of succession as a colonization rule (Harte, 2011). This rule states that in a colonization process, the probability of placing an individual on one side of the bisected area is roughly proportional to the fraction of individuals already there. This "rich get richer" effect results in strong spatial aggregation. More precisely, the probability for placing an individual on the left half of a bisected plot with n_L individuals on the left and n_R individuals on the right is

$$p_L = \frac{n_L+1}{n_L+n_R+2}.$$

To make our notation consistent with that above, let the number on the left be n and the total number to be n_0. The probabilities of a new individual arriving on the left or on the right are then:

$$\begin{aligned} p_L(n|n_0) &= \frac{n+1}{n_0+2}, \\ p_R(n|n_0) &= \frac{n_0-n+1}{n_0+2}. \end{aligned} \tag{4}$$

If we place n_0 individuals using this rule, the resulting probability distribution is given by Eq. 3.

Random placement

Another model for spatial ecology, perhaps the simplest, is the random placement model (Coleman, 1981). Instead of the placement rules in Eq. 4, each individual is placed randomly. In a bisected plot this means each individual has a 50 percent chance of being placed on either side, $p_L = p_R = 0.5$. Placing n_0 individuals this way gives the binomial distribution

$$\Pi_{\mathrm{RP}}(n|n_0) = \binom{n_0}{n}\left(\frac{1}{2}\right)^{n_0} \tag{5}$$

which, if n_0 is large, means we are very likely to have roughly half the individuals on each side. This is equivalent to having no spatial aggregation; there is no preference for any new individual to stay close to any previous individual as each placement is a random coin flip.

METHODS

Introducing a density dependent death rule

We now introduce a density dependent death rule in addition to the METE colonization rule in Eq. 4. In the logistic model, for example, death depends quadratically on population size. To allow for more general density dependence we introduce a death rate proportional to n^α. Density dependence may result from resource limitation or some other mechanism (e.g. the Janzen-Connell effect (Janzen, 1970; Connell, 1971)).

In the case of a bisected plot, each death must be on the left or right and the probability of a single death on the left, $p_{D,L}$, or right, $p_{D,R}$, is

$$\begin{aligned} p_{D,L}(n|n_0) &= \frac{n^\alpha}{n^\alpha + (n_0-n)^\alpha} \\ p_{D,R}(n|n_0) &= \frac{(n_0-n)^\alpha}{n^\alpha + (n_0-n)^\alpha}. \end{aligned} \tag{6}$$

Deriving the Π distribution

Now that we have the colonization (Eq. 4) and death rules (Eq. 6), we can derive the general $\Pi_\alpha(n|n_0)$ for bisections. At each step in the model, we will have one death followed by the placement of one new individual. Each placement can be interpreted ecologically as a birth, or as the immigration of an individual from the same species. We can solve for the resulting steady state distribution where we reach an equilibrium in the spatial pattern.

There are several approaches for deriving the steady state solution for such a system. Here, we equate the rates leaving and entering any individual state $\Pi_\alpha(n|n_0)$. We take

the probability that we start with n individuals on the left, one on the right dies and then one is placed on the left resulting in $n+1$ individuals on the left, and equate that to the probability that we have $n+1$ individuals on the left, one on the left dies and then one is placed on the right resulting in n individuals on the left. We could have equivalently done the same thing with n and $n-1$. Equating these rates using the probabilities in Eq. 4 and Eq. 6 leads to

$$\frac{(n+1)^{\alpha-1}}{(n_0-n-1)^{\alpha}+(n+1)^{\alpha}}\Pi_{\alpha}(n+1) = \frac{(n_0-n)^{\alpha-1}}{(n_0-n)^{\alpha}+n^{\alpha}}\Pi_{\alpha}(n). \tag{7}$$

We can solve this recursion relation to obtain a general stationary solution for Π_{α} with a given n_0 and α:

$$\Pi_{\alpha}(n|n_0) = \frac{n^{\alpha}+(n_0-n)^{\alpha}}{C(n_0,\alpha)n_0^{\alpha}}\binom{n_0}{n}^{\alpha-1} \tag{8}$$

where $C(n_0,\alpha)$ is the overall normalization that has to be calculated numerically. See Supporting Information S1 for the details of this derivation.

If $\alpha = 2$, we can solve for the normalization explicitly to get

$$\Pi_{\alpha=2}(n|n_0) = \frac{n^2+(n_0-n)^2}{2^{n_0-1}n_0(n_0+1)}\binom{n_0}{n}, \tag{9}$$

and if $\alpha = 1$, we recover the METE prediction $\Pi_{\alpha=1}(n|n_0) = \frac{1}{n_0+1}$.

Beyond the first bisection

Bisecting the plot more than once (into quadrants, then into 8 cells, etc.) when comparing the model to data allows us to consider how aggregation changes with scale, as well as look at α at the level of an individual species. We use the following method when we discuss bisecting a single plot more than once.

We begin by bisecting the plot in half in one direction, then bisecting each of those plots in the opposite direction. We alternate this bisection pattern until we have 2^b cells, where b is the number of bisections. We can then compare adjacent cells as if they were their own bisection to obtain 2^{b-1} points.

When we go beyond the first bisection in the subsequent analysis, we only consider species that could have at least one individual per bisection ($n_0 > 2^{b-1}$). The smallest scale we consider in our datasets is $b = 8$, so when we bisect the plot more than once we will only consider species with more than 128 individuals.

Fitting α

Bisection predictions can be compared to data by rank ordering the fraction of individuals present in one half of the plot for each species (e.g. Harte (2011)). This method, however, obscures trends in species abundance in the data and the predictions, and can lead to incorrect conclusions about which model is preferred (see Supporting Information S2).

We instead find the maximum likelihood α given the data. This method gives us values for α that are the most consistent with the data, even if they may not look like they fit the rank ordered fractions (see Fig. S1 and Table S1).

For a single bisection, we will have a number of points p equal to the number of species, $p = S_0$. For multiple bisections where we consider the species on aggregate, the number of points will be roughly equal to the number of species multiplied by the number of cells, $p \approx S_0 2^{b-1}$. The equality is not exact as not all cell pairings beyond the first bisection will have all of the species present at the single bisection level. For multiple bisections where we consider individual species, the number of points will be roughly the number of cell pairings, $p \approx 2^{b-1}$, where again there will be fewer points in practice as some cell pairings will be empty. The statistical error in estimating α this way goes as $1/\sqrt{p}$ (see Supporting Information S3).

Data used

We will compare our results to two datasets. First, we will use data from a sparse Californian serpentine grassland site (Green *et al.*, 2003, 2019) at the McLaughlin University of California Natural Reserve censused in 1998. This is a 64 m^2 plot divided into 256 cells with 24 species and 37 182 individuals. There are 10 species with abundance greater than 128 individuals that constitute 36 783 individuals.

Second, we will use data from Barro Colorado Island (BCI) in Panama (Condit *et al.*, 2019; Hubbell *et al.*, 2005; Condit, 1998; Hubbell *et al.*, 1999), a 50 ha plot in a moist tropical forest. We will work with the 2005 census and consider plants with a diameter at breast height (dbh) greater than 10 cm. This dataset has 229 species and 20 852 individuals, and 40 species with abundance greater than 128 individuals that constitude 15 960 individuals.

RESULTS

Comparison to METE and random placement

Figure 1 compares the bisection predictions for $\Pi(n)$ from METE, random placement, and our density dependent model for various α, at $n_0 = 10$ and 50. We can relate our

distribution to both the METE and random placement distributions for different values of α. $\alpha = 1$ corresponds exactly to the METE solution, which makes sense given that the placement and death rules are both linear in n. As $\alpha \to 2$, our distribution approaches the random placement prediction if n_0 is large enough (Supporting Information S4 shows this result analytically). For $1 < \alpha < 2$, we vary continuously between METE and random placement. We can make the distribution even more spatially aggregated than METE with $\alpha < 1$ and even less than random placement (overdispersed) with $\alpha > 2$.

We can also relate this distribution to the conditional negative binomial distribution in the limit of large n_0, assuming that matching the peak of the distributions is a good approximation for the entire distribution. In that limit, the aggregation parameter k is approximately related to the density dependent parameter α by $k \approx \frac{n_0}{2}\left(\frac{\alpha-1}{2-\alpha}\right) + 1$. See Supporting Information S5 for the derivation.

Comparing to data

Figure 2 shows the direct comparison between our model and the serpentine and BCI datasets. Each data point is the observed fraction of individuals in one half of the plot versus the species abundance. The curves in this figure show the 95% contour intervals for $\Pi(n)$ distributions predicted by METE, random placement, and our density dependent model with the maximum likelihood α value. The contours narrow at larger n_0 as the distributions narrow.

The maximum likelihood result for the serpentine dataset is indistinguishable from $\alpha = 1$, so the confidence interval curves on the plot for METE and the density dependent model overlap for most n_0. For the BCI data, the maximum likelihood value is $\alpha = 1.12$, slightly larger than 1. We can see the contours in this case are clearly in between those for METE and random placement. The likelihoods for each of the models are shown in Table 1.

Scale dependence

Going beyond the first bisection allows us to see how α varies depending on the scale of our plot. Figure 3 shows how α scales with fractional area for both the serpentine and BCI plots. Density dependence increases at smaller scales in both datasets.

Individual species

Since the Π function is defined on the species level, we can consider each species separately and find the maximum likelihood α for each, rather than looking at the aggregate α across the community. To do this we have to go beyond the first bisection to

get multiple data points for the same species at smaller scales.

For the serpentine data, we exclude *Eriogonum nudum* from the following figures as an outlier (see Discussion). This leaves 9 species with abundance greater than 128 individuals.

Figure 4 shows the distribution of α values among the species at each scale. The median α increases at smaller scales for both datasets, consistent with what happens for the aggregate α across species in Fig. 3.

Figure 5 traces out the corresponding α value for each species as we change scale. The darker lines correspond to more abundant species. Individual species have a slight scale dependence that typically follows the aggregate α, where α increases at smaller scale. However, not all species have a monotonically decreasing α.

DISCUSSION

We first directly compare our results between our two datasets, serpentine and BCI, in Figs. 2 and 3. Because the serpentine site was very sparse, whereas the BCI forest is very tightly packed, we expect a higher α and greater density dependence at BCI than at the serpentine site. This is consistent with our maximum likelihood values for α. Additionally, the serpentine data agrees well with other METE predictions, such as the species-abundance distribution, whereas the BCI data does not agree as well (Harte, 2011). It is therefore not surprising that there is more deviation from METE at the BCI site.

Beyond the single bisection results, our scaling results in Fig. 3 make ecological sense. We expect that at smaller scales, the density dependence would be larger as individuals compete more for resources at that scale. At large scales, we expect α to be close to 1 as the individuals do not compete over large distances. This means that the spatial distributions look more aggregated on large scales than on small scales as the individuals within species broadly group together, but repel each other at small scales.

Condit *et al.* (2000) find that at small scales individuals are aggregated, but that they are distributed randomly at large scales. This seemingly contradictory result is explained by the different metrics of aggregation. They use Ω_x, which is the total number individuals found within a radius x of a single individual, normalized by the average density of individuals. $\Omega_x = 1$ corresponds to random placement, and $\Omega_x > 1$ indicates aggregation. At small scales at BCI, they find $\Omega_{0-10\mathrm{m}} >> 1$, and at large scales $\Omega_{>100\mathrm{m}} \approx 1$. On the other hand, we find $\alpha > 1$ at small scales and $\alpha \approx 1$ at large scales (Fig. 3). To see that these results are consistent, consider a plot with a cluster of

individuals on one side where the spatial structure within the cluster is random. At the single bisection level this would be consistent with $\alpha = 1$, but within the cluster $\alpha > 1$. That same arrangement of individuals would give an Ω_x that shows strong aggregation at small scales and random placement at large scales. This is what we find empirically in the BCI data. This spatial pattern is also consistent with the Poisson cluster model outlined in Plotkin *et al.* (2000), where clusters of individuals are placed uniformly with random internal spatial structure.

Our repeated bisection analysis also indicates at which scale density dependence becomes important. We see this as a shoulder for both the serpentine and BCI datasets in Fig. 3 where α moves away from ≈ 1. In absolute scale, this corresponds to $< 0.5\text{m}^2$ for the serpentine plot and < 1.6 ha for the BCI dataset. This again makes sense given that the serpentine grassland is much more sparse than the BCI forest.

We can compare the scale at which the shoulder occurs that we find here to other studies of these sites. Conlisk *et al.* (2007) introduce a parameter ϕ that modifies the colonization rule Eq. 4 and allows the Π distribution to vary continuously between random placement and METE. For the serpentine data, they find that at scales larger than around 0.5 m^2 (the 8th bisection), ϕ approaches 0.5, which corresponds to the METE prediction. At scales around 0.5 m^2 or smaller, $\phi \approx 0.2$, where $\phi = 0$ corresponds to random placement. This is consistent with our scaling results in Fig. 3.

At the BCI site, Volkov *et al.* (2005) showed that density dependence can explain the species-abundance distribution. Their added parameter $c = 1.80$ corresponds to density dependence entering at scales smaller than a fractional area of $1/27.2 = 1/2^{4.77}$, which is close to where we see α increase in Fig. 3.

Individual species largely follow the aggregate scaling, but the overall range is large (Figs. 4 and 5). This variation is expected considering the small number of data points, especially for rarer species. The values of the median α in Fig. 4 are largely consistent with the aggregate α in Fig. 3.

In the serpentine dataset, we excluded *Eriogonum nudum* as an outlier for part of our analysis. The maximum likelihood α was > 6 at the smallest scale and the maximum itself was very shallow. This species has a large canopy compared to the other grassland plants, and tends to be found far from other individuals. It makes sense that it would be overdispersed with $\alpha > 2$.

For individual species at the BCI dataset, *Gustavia superba* stood out with an average α of 1.001 across scales in Fig. 5. This species is largely limited to 2 hectares of young secondary forest along the edge of the plot (J. Wright, personal communication, 2019),

making it look especially aggregated and resulting in a maximum likelihood α close to 1.

At the individual species level more broadly, we might expect to find trends in α with species abundance or size. More abundant species may be competing more for the same resources, or larger species may compete over larger distances. For example, Condit *et al.* (2000) find that both rarer species and smaller individuals tend to be more aggregated, however at a much smaller scale (within a 10 m radius) and with a different metric of aggregation.

In the BCI data, we did not find any species with high α and high abundance (Fig. S4), but otherwise we do not see a trend in abundance. We also find no evidence of a trend with species' mean dbh (Fig. S5), though it is possible this trend is obscured by variance in individual size within a species, or that the range of mean dbh we considered (about $100-500$ mm) is too small to see its effect.

Finally, we looked for an overall energy effect. Considering that the most abundant species tend to be smaller, it may be that density dependence depends on the total metabolic rate of a species. Plotting this relationship (Fig. S6) again does not reveal a scaling relationship. That none of these trends are found in the data might indicate that the operating density dependent mechanism is not resource limitation, in that it doesn't depend on the abundance, size, or metabolic rate of the species. See Supporting Information S6 for more information on these trends.

A plausible mechanism for the observed density dependence at BCI is the Janzen-Connell effect (Janzen, 1970; Connell, 1971), whereby areas near parent trees are inhospitable for offspring, resulting in density dependence. Various studies (Harms *et al.*, 2000; Carson *et al.*, 2008; Comita *et al.*, 2014) have observed this effect at BCI, which is consistent with our result that $\alpha > 1$ there.

In general, our result that $\alpha < 2$ at large scales makes ecological sense. Choosing a logistic death rate is common in modelling, but it corresponds to the case where each individual interacts with all other individuals. In larger plots, we would expect a general exponent to be smaller as not all individuals are competing with all other individuals. Any one individual may only be interacting with some of its neighbors, which would result in $\alpha < 2$.

The connection we have derived between spatial pattern and density dependence offers a way to infer density dependence at species and community levels. This result stands in contrast to arguments that density dependence is inherently unmeasurable (e.g. Ehrlich & Birch (1967)).

Our model considers only bisections, and it would be useful to extend it to be more

general. There are many spatial arrangements that can not be accurately captured by dividing plots into bisection. For example, if we divide our plot into an m by m grid, and have one individual per cell, we would see exactly 0.5 as the fraction for each bisection. This result would be consistent with random placement with a large number of individuals, which doesn't well describe this exceptionally uniform arrangement. There could also be different degrees of spatial aggregation within a cell that we will not accurately capture with a bisection. Despite these limitations, bisections are useful for understanding commonly observed macroscopic spatial patterns. We leave the derivation of the full $\Pi_\alpha(n|n_0, A, A_0)$ to future work.

A conceptually simple extension to our model is to divide plots into quadrisections rather than bisections. The colonization and death rules then have three unknowns rather than one (the number of individuals in each quadrant, where the fourth is determined by constraining the sum to be n_0). This makes it hard to solve analytically, however we can simulate the birth-death process until it reaches steady state. We find no significant difference in our simulation compared to our prediction from two bisections, and find that $\alpha = 1.12$ is still consistent with the BCI data.

Because we consider the steady state solution in our model, we are assuming that the density dependence time scale is longer than the time scale of individual births or deaths. That is, α must not change too rapidly in time. This assumption is justified for many systems roughly in steady state with their environment and not undergoing rapid change (Newman *et al.*, 2018).

Solving for the steady state solution also assumes that births and deaths are in balance. We assume here that there is a single death followed by a placement, however simulating two deaths followed by two placements gives a probability distribution consistent with our analytic prediction. We expect our result to hold with other numbers of deaths and placements.

In summary, our model predicts that increased density dependence leads to more random spatial patterning. We predict that this result is general across ecosystems and taxonomic groups, although here we have only analysed a serpentine grassland and a moist tropical forest. We find that density dependence is important in the tropical forest site, although considerably weaker (α closer to 1 than 2) than in the commonly used logistic model. The serpentine grassland site is consistent with the absence of a density dependent correction with the strong spatial aggregation predicted by METE. Our model allows us to infer this underlying density dependence from static spatial patterning.

ACKNOWLEDGEMENTS

This material is based upon work is supported by the National Science Foundation under Grant No. DEB-1751380. MB acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [PGSD2-517114-2018]. JH thanks the Santa Fe Institute for their hospitality. We thank Kaito Umemura for valuable discussion and feedback, and Egbert Leigh and Joseph Wright for their help with the BCI dataset.

We thank Jessica Green for the serpentine data. The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed.

Data accessibility statement

No new data were used. The serpentine data are available from the Dryad Digital Repository at `https://doi.org/10.6078/D1MQ2V`. The BCI data can be found at `https://forestgeo.si.edu/explore-data` and are available from the Dryad Digital Repository at `https://doi.org/10.15146/5xcp-0d46`. Relevant code is available at `https://github.com/micbru/density_dependence_public`.

Statement of authorship

Both authors were involved in conceptualization. MB conducted the formal analysis and led the writing of the manuscript.

REFERENCES

Arrhenius, O. (1921). Species and Area. *Journal of Ecology*, 9, 95–99.

Bliss, C. I. & Fisher, R. A. (1953). Fitting the Negative Binomial Distribution to Biological Data. *Biometrics*, 9, 176–200.

Carson, W. P., Anderson, J., Leigh, E. & Schnitzer, S. A. (2008). Challenges Associated with Testing and Falsifying the Janzen–Connell Hypothesis: A Review and Critique. In: *Tropical Forest Community Ecology* (eds. Carson, W. P. & Schnitzer, S. A.). Wiley-Blackwell, pp. 210–241.

Coleman, B. D. (1981). On random placement and species-area relations. *Mathematical Biosciences*, 54, 191–215.

Comita, L. S., Queenborough, S. A., Murphy, S. J., Eck, J. L., Xu, K., Krishnadas, M.

et al. (2014). Testing predictions of the Janzen-Connell hypothesis: a meta-analysis of experimental evidence for distance- and density-dependent seed and seedling survival. *Journal of Ecology*, 102, 845–856.

Condit, R. (1998). *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots*. Springer-Verlag, Berlin Heidelberg.

Condit, R., Ashton, P. S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N. *et al.* (2000). Spatial Patterns in the Distribution of Tropical Tree Species. *Science*, 288, 1414–1418.

Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R. & Hubbell, S. (2019). Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. URL `https://doi.org/10.15146/5xcp-0d46`.

Conlisk, E., Bloxham, M., Conlisk, J., Enquist, B. & Harte, J. (2007). A New Class of Models of Spatial Distribution. *Ecological Monographs*, 77, 269–284.

Conlisk, J., Conlisk, E., Kassim, A. R., Billick, I. & Harte, J. (2012). The shape of a species' spatial abundance distribution. *Global Ecology and Biogeography*, 21, 1167–1178.

Connell, J. (1971). On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: *Dynamics of populations: Proceedings of the Advanced Study Institute on Dynamics of numbers in populations, Oosterbeek, the Netherlands, 7-18 September 1970* (eds. Boer, P. J. d. & Gradwell, G. R.). Pudoc, Wageningen, pp. 298–312.

Drakare, S., Lennon, J. J. & Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecology Letters*, 9, 215–227.

Ehrlich, P. R. & Birch, L. C. (1967). The "Balance of Nature" and "Population Control". *The American Naturalist*, 101, 97–107.

Green, J., Harte, J. & Ostling, A. (2019). Data from: Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland. URL `https://doi.org/10.6078/D1MQ2V`.

Green, J. L., Harte, J. & Ostling, A. (2003). Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland. *Ecology Letters*, 6, 919–928.

Harms, K. E., Wright, S. J., Calderón, O., Hernández, A. & Herre, E. A. (2000). Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest. *Nature*,

404, 493–495.

Harte, J. (2011). *Maximum entropy and ecology : a theory of abundance, distribution, and energetics*. Oxford University Press, Oxford.

Harte, J. & Kitzes, J. (2015). Inferring Regional-Scale Species Diversity from Small-Plot Censuses. *PLOS ONE*, 10, e0117527–e0117527.

Harte, J. & Newman, E. A. (2014). Maximum information entropy: a foundation for ecological theory. *Trends in Ecology & Evolution*, 29, 384–389.

Harte, J., Zillio, T., Conlisk, E. & Smith, A. B. (2008). Maximum Entropy and the State-Variable Approach to Macroecology. *Ecology*, 89, 2700–2711.

He, F. & Gaston, K. (2003). Occupancy, Spatial Variance, and the Abundance of Species. *The American Naturalist*, 162, 366–375.

He, F. & Gaston, K. J. (2000). Estimating Species Abundance from Occurrence. *The American Naturalist*, 156, 553–559.

Hubbell, S., Condit, R. & Foster, R. (2005). BCI 50-ha plot data. URL `http://ctfs.si.edu/webatlas/datasets/bci/`.

Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B. *et al.* (1999). Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science*, 283, 554–557.

Janzen, D. H. (1970). Herbivores and the Number of Tree Species in Tropical Forests. *The American Naturalist*, 104, 501–528.

Kitzes, J. & Shirley, R. (2016). Estimating biodiversity impacts without field surveys: A case study in northern Borneo. *Ambio*, 45, 110–119.

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73, 1943–1967.

Newman, E. A., Wilber, M. Q., Kopper, K. E., Moritz, M. A., Falk, D. A., McKenzie, D. *et al.* (2018). Disturbance macroecology: integrating disturbance ecology and macroecology in different-age post-fire stands of a closed-cone pine forest as a case study. *bioRxiv*, 309419.

Plotkin, J. B., Potts, M. D., Leslie, N., Manokaran, N., Lafrankie, J. & Ashton, P. S. (2000). Species-area Curves, Spatial Aggregation, and Habitat Specialization in Tropical Forests. *Journal of Theoretical Biology*, 207, 81–99.

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press, Cambridge.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C. *et al.* (2004). Extinction risk from climate change. *Nature*, 427, 145–148.

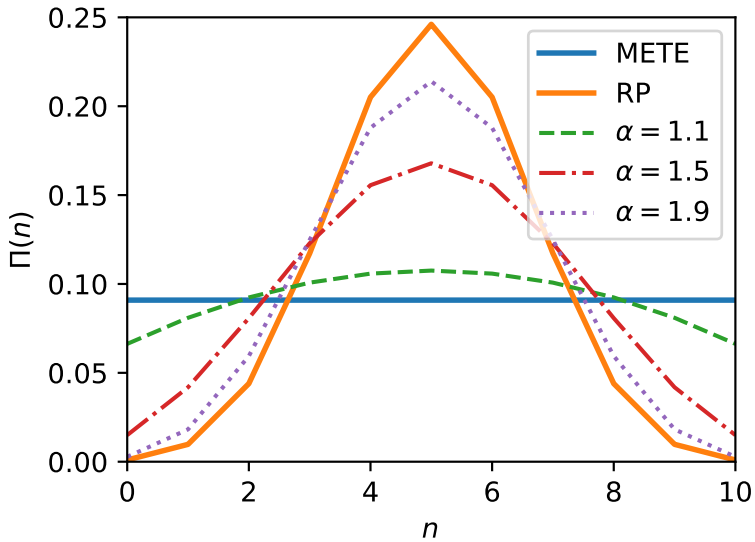
Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. (2005). Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, 438, 658–661.

White, E. P., Thibault, K. M. & Xiao, X. (2012). Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, 93, 1772–1778.

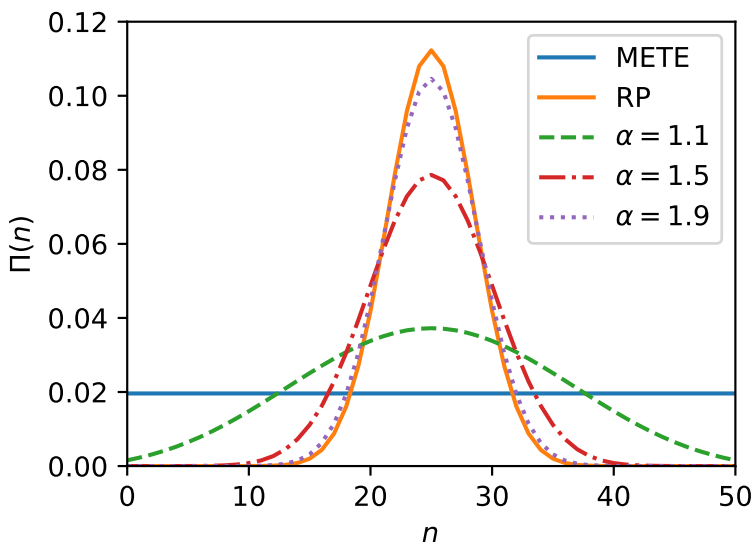
Xiao, X., Mcglinn, D. J. & White, E. P. (2015). A Strong Test of the Maximum Entropy Theory of Ecology. *The American Naturalist*, 185, E70–E80.

Serpentine		BCI	
Model	Log-likelihood	Model	Log-likelihood
METE	-114.8	METE	-729
RP	-5188.6	RP	-963
$\alpha = 1.0003$	-114.6	$\alpha = 1.12$	-660

Table 1. Log-likelihood values for the three different models. We can compare our model to METE using the deviance in log-likelihood and obtain a p-value. The deviance is defined as twice the difference in log-likelihood. For the serpentine dataset, the deviance is 0.6 which corresponds to a p-value of 0.45. For the BCI dataset, the deviance is 138 which corresponds to a p-value of $< 10^{-30}$.

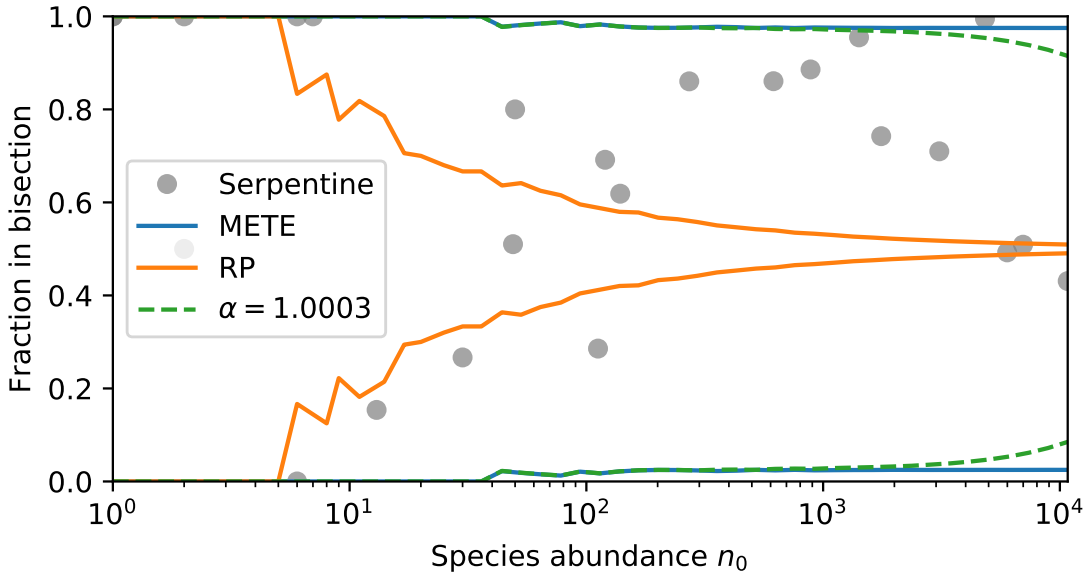


(a) $n_0 = 10$

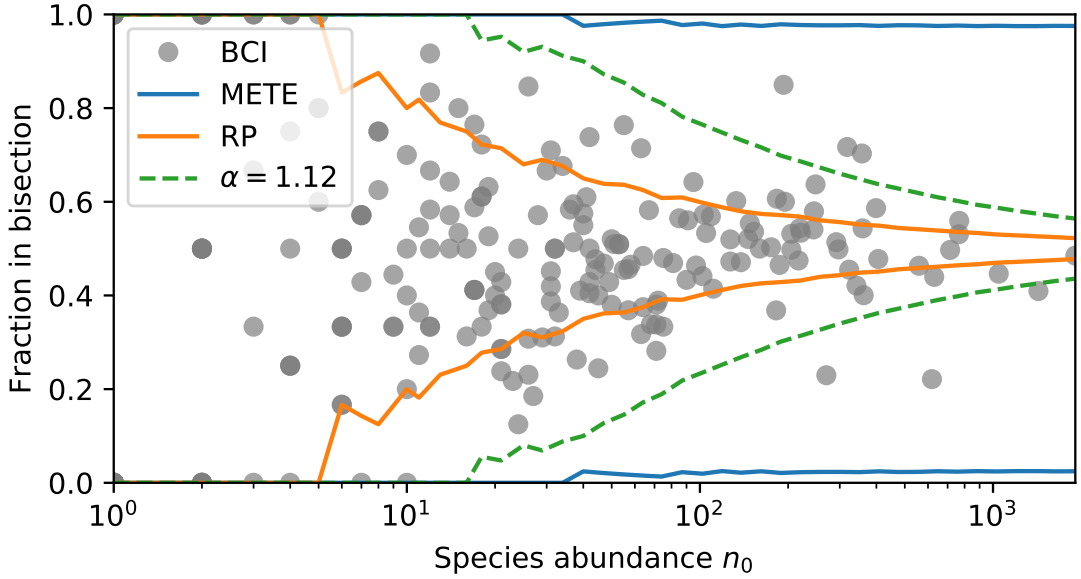


(b) $n_0 = 50$

Figure 1. Comparison of the bisection probability distributions $\Pi(n)$ from METE, random placement (RP), and our density dependent model with varying α. At $\alpha = 1$, our model corresponds exactly to METE. At larger n_0, $\alpha \to 2$ approaches the random placement distribution. Our model varies continuously between METE and random placement for $1 < \alpha < 2$.



(a) Serpentine



(b) BCI

Figure 2. 95% contour intervals for the different predicted probability distributions overlayed with (2a) the serpentine dataset, and (2b) the BCI dataset. For our density dependent model, $\alpha = 1.0003$ maximizes the log-likelihood for the serpentine dataset, and $\alpha = 1.12$ maximizes log-likelihood for BCI dataset.

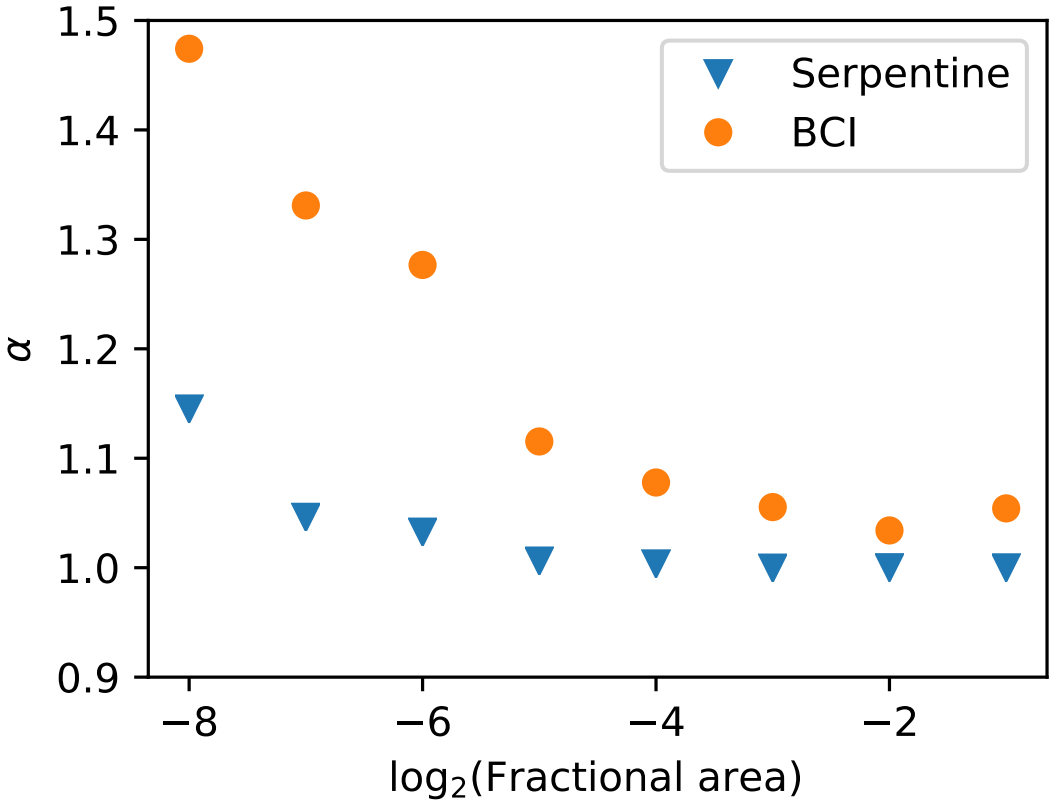
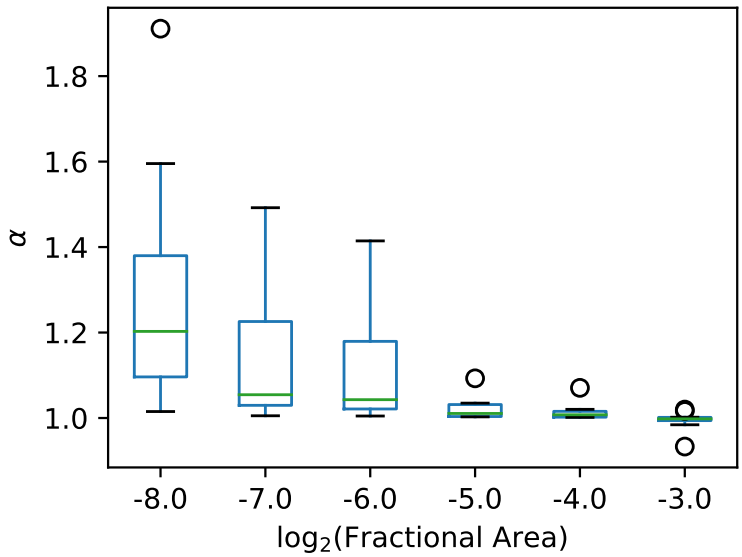
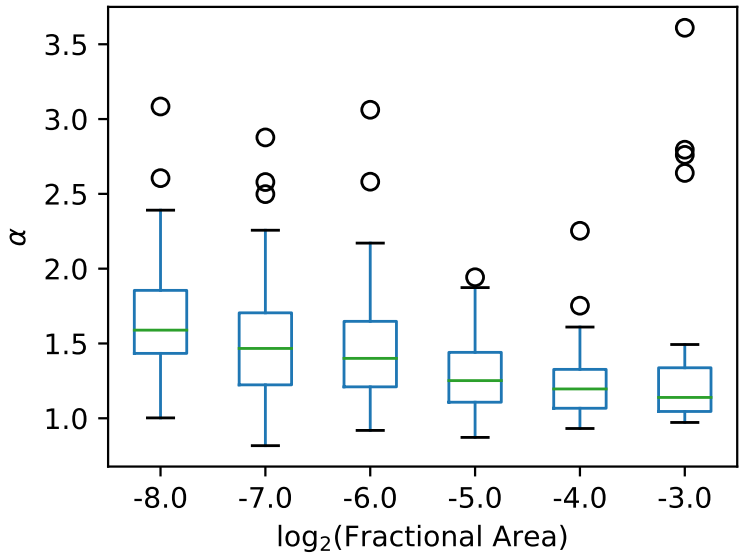


Figure 3. Aggregate α scaling with area for species with abundance $n_0 > 128$. The density dependence increases at smaller scales. The serpentine dataset has 36 783 individuals and the BCI dataset has 15 960 individuals.

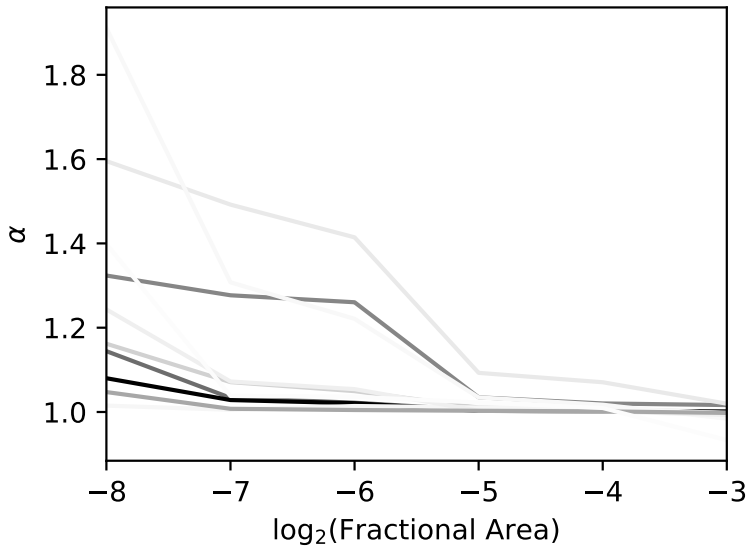


(a) Serpentine, 10 species

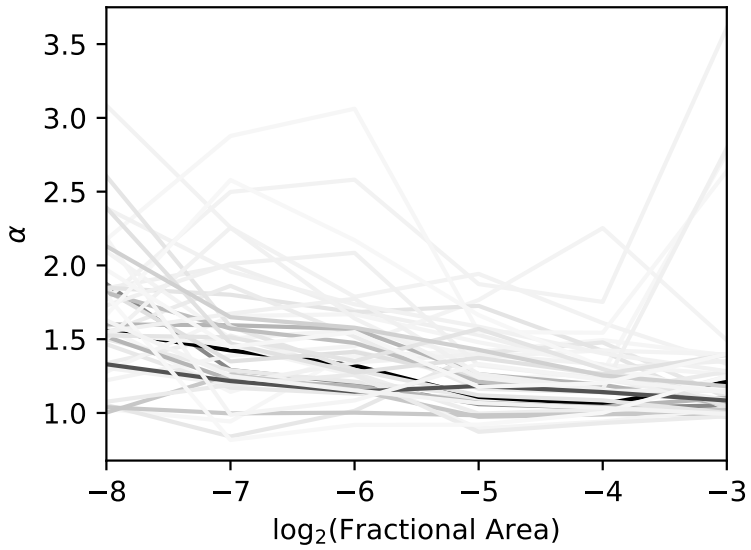


(b) BCI, 40 species

Figure 4. Boxplots for α among the species at different scales. At smaller scales α is larger, as in the aggregate α case, however here we can also see the spread in α between species. The boxplots show boxes from quartile 1 (Q1) to quartile 3 (Q3) with a line at the median. The whiskers extend to 1.5×(Q3-Q1). The remaining points are plotted as individual circles.



(a) Serpentine, 10 species



(b) BCI, 40 species

Figure 5. α values for individual species across scales, from $b = 8$ to 3 bisections. Darker lines are more abundant species.