Spatial processes in macroecology

Brian McGill, U of Arizona
mail@brianmcgill.org

Superimposing individual species' ranges creates local community structure

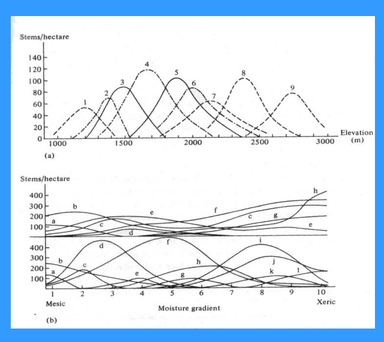
In collaboration w/ Cathy Collins

Abstract

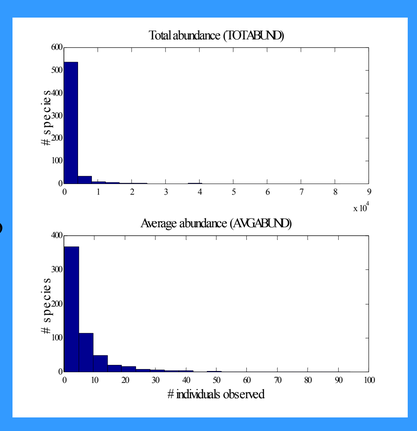
1) **I start with three empirically well documented patterns**
   - A. Species ranges are placed independently of each other
   - B. Global abundance varies between species according to a hollow curve
   - C. Intraspecific abundance varies roughly "normally" across the range
2) **These three patterns serve as proximate mechanisms to produce four well-known, qualitative macroecological patterns:**
   - I. Species abundance distributions (SAD) which follow a hollow curve shape in arithmetic space
   - II. Species area relationship (SPAR), $S=cA^Z$
   - III. Similarity between sites decays in a roughly hyperbolic fashion with distance
   - IV. Range size increases with global abundance
3) **The mathematical model also produces three novel, quantitative predictions, which I test empirically:**
   - i. The SPAR should have particular c (intercept) and z (slope) values
   - ii. The shape of the local SAD should be the same as the shape of the distribution of global abundances
   - iii. Local abundance should be correlated with and predicted by distance from peak abundance and global abundance of the species
4) **The empirically robust inputs and the empirically tested outputs (over and above producing curves of the right shape) make for a strongly empirically grounded unified theory of macroecology**

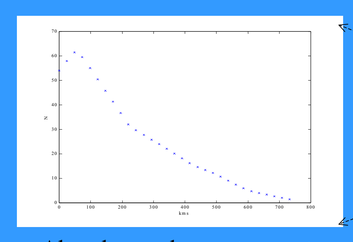
Three empirically well-documented macroecological patterns …

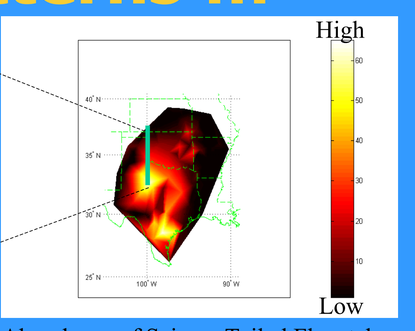
Mechanism A - Ranges are located independently of each other
(Whittaker 1969, see also Hoagland & Collins 1997)



Mechanism B - Global abundances are distributed according to a hollow curve. Various measures of global abundance correlate strongly. Data here from the BBS. See also Gaston 1994




Abundance along a transect



Abundance of Scissor-Tailed Flycatcher

Mechanism C - Abundance is distributed roughly "normally" across a species' range (Brown 1984, Hengeveld & Haeck 1982). Real world data often has multiple peaks and is asymmetric. The key for our model is that the data is strongly spatially autocorrelated and has small peaks relative to the tails. We often model this as Gaussian, but in some computer simulations use more complex representations.

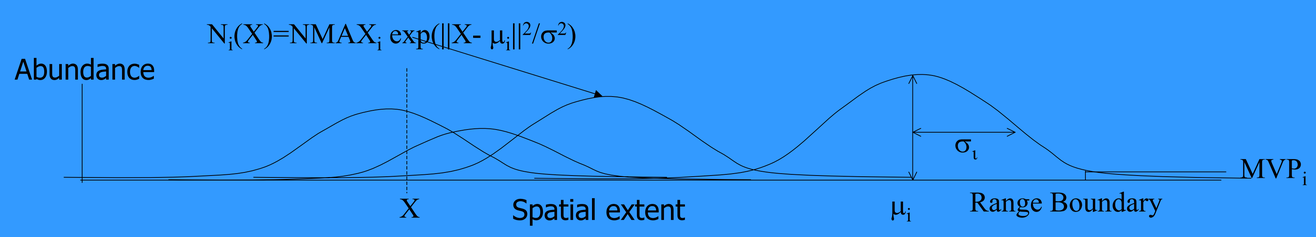
… Translate mathematically and in computer simulations as …

Mechanism A - Range centers (μ_i) are distributed according to a Poisson process. The width of the range is given by σ_i.

One additional, technical assumption is needed. Range boundaries occur when the abundance drops below some population size, $\mathbf{MVP_i}$

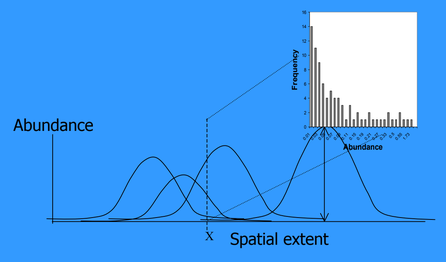
Mechanism B - Peak abundances ($\mathbf{NMAX_i}$) vary and are distributed according to a power distribution $f(x)=ax^c$

Mechanism C - Abundance ($\mathbf{N_i}$) is Gaussian across space ($\mathbf{X}$)



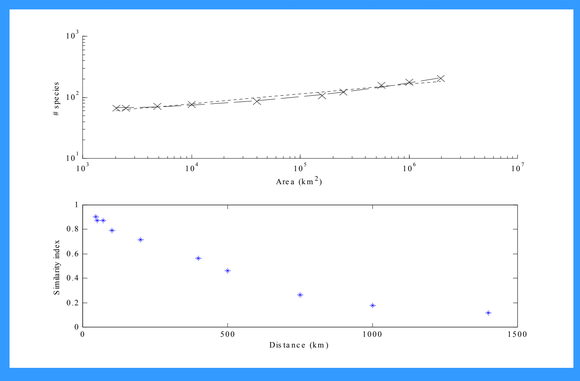
… To produce four qualitative, well-known patterns …

Pattern I – Species abundance distributions follow a hollow curve



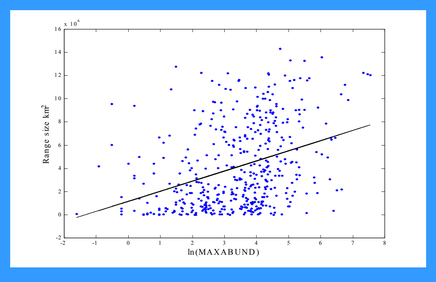
The above figure shows intuitively why this works. At a given point, x, we sample many species in their tail and a few species in their peaks. An analytical model makes this more rigorous. Computer simulations show that this pattern is robust to a variety of changes in assumptions including the use of the normal curve (if each species has three asymmetric peaks the pattern still holds).

Pattern II – Species Area Relationship (SPAR, $S=cA^Z$)
Pattern III – Species similarity decays hyperbolically



The above figure is based on Monte Carlo simulations. Note the curvilinearity in the SPAR — something that is also observed in empirical data. The similarity index used here is Jaccard's.

Pattern IV – Range size increases with global abundance



A simple rearrangement of the formula for the Gaussian curve gives:

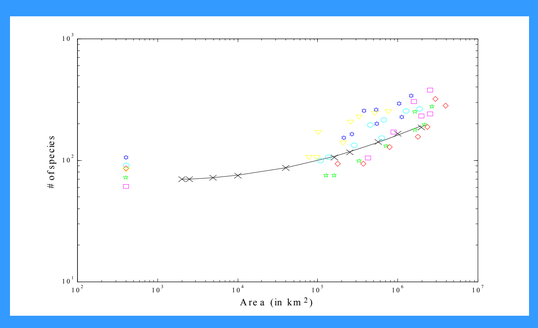
$$RangeSize = \pi \sigma^2 \ln\left(\frac{NMAX_i}{MVP}\right)$$

This produces a positive correlation. If we allow σ and MVP to vary by species we reproduce the scatter observed in real data (this figure based on Breeding Bird Survey data.

… And three novel, quantitative predictions, which I test

Prediction i) A priori estimates of the SPAR model parameters produce the correct slope, intercept and curvilinearity.

We estimated the 5 parameters of the SPAR model using NA Breeding Bird Survey data. These simulations give z-values of 0.16±0.03, almost identical to Rosenzweigs range of empirically observed z's (0.15±0.03). The diagram to the right contains data from 5 different empirical SPARs based on the BBS data and different staring points. You can see that the simulation (black line and X's) predicted the correct slope, intercept, and curvilinearity.



Prediction ii – The shape of the global abundance and *every* local abundance distribution should be the same.

This prediction comes from the analytical model for the SAD (pattern I above). In fact the c for global abundance in the BBS is c=0.293, while the c for 1000 routes averages c=0.289 w/ 95% of all values in (0.204,0.380). Contrast with c=0.23 in the BCI tree SAD or an average of 0.39 in the Gentry data set.

Prediction iii – Local abundance of a species should be correlated w/ its distance from peak abundance and w/ its global abundance.

In fact, Pearson correlation gives r=0.42 & r=0.57 respectively. Regression on these two variables for 1000 routes gives an average R^2=45% (14.6%,67.86%) w/ high significance. These values are conservatively low for two reasons.

Sidebar #1: Testing macroecology processes

1) A major goal of macroecology is to identify processes that underlie the patterns identified. The most common test is to build the hypothesized mechanisms into a model and then see if the model produces a curve that can match the observed patterns. This approach is favored in part due to the difficulty of experimentation in macroecology
2) The Central Limit Theorem (CLT) of statistics predicts the shape of many of the curves without any biological mechanisms
3) Thus if one is testing a theory by producing a curve of the correct shape, it is important to do this rigorously. The following is a sequence of increasingly rigorous tests of a model:
   a) Curve appears to match the data by eyeball
   b) Curve is close to the data by an objective measure (e.g. R^2)
   c) Curve is closer to the data than the null hypothesis (lognormal) by some objective measure
   d) Curve is statistically significantly closer (i.e. 95% confident) than the null hypothesis (lognormal)
   e) Curve is statistically significantly better than the null with all parameters predicted independently (i.e. without curve fitting)
4) It is even more desirable to examine those aspects of the pattern which are not predicted by the CLT:
   a) The parameters These can be meaningfully studied in two ways
      i. Constancy of parameters at various spatial and temporal scales (e.g. Rosenzweig 1995 found three different spatial scales/structures with three different corresponding constant z-values).
      ii. A priori, independent prediction of the parameters (i.e. not chosen to maximize the fit of the curve). An example, albeit now disproved, would be Preston's Canonical Lognormal. See also predictions i) & ii) above.
   b) The spatial, temporal, and taxonomic scales of "exchangeability" The CLT treats all entities (usually species) as identical or exchangeable. In practice however, unlike gas molecules, species exchange roles only rarely. The spatial and temporal scales over which species are unexchangeable inform us about the scale of the processes involved.. For example, pollen and fossil records tell us that species rank abundances remain constant over long periods of time (until a revolution occurs). Condit et al (2002) show that rank similarity of tropical trees decays slowly over large spatial scales. McGowan & Walker (1985) show that marine copepod rank abundance remains constant over 16 years and 800 km.
   c) Numerous other predictions (other than curve shape), such as correlations. See prediction iii) above for an example.
5) To test some of these ideas about the CLT, we tried fitting various distributions to data from the North American Breeding Bird Survey (BBS). The numbers below represent the mean and 95% range for 1000 routes from the BBS (100 for the zero sum multinomial distribution). We look at six different measures of goodness of fit between the predicted and observed distribution. We also look at seven different distributions. We note the following
   a) The lognormal fits the data best. Several other distributions fit the data very well.
   b) The Zero Sum Multinomial (neutral biogeography) distribution fails to perform better than the the lognormal (rigor level c above) and fails to reject it as a null hypothesis (rigor level d above).
   c) Similar results were obtained with the BCI 50 hectare tree dataset (Condit et al 2002).

	R2	R2MeanCorrect	R2Correlation	K (Kolmogorov)	X2 (10 bins)
MLE Power	0.93 (0.88,0.97)	0.48 (0.04,0.77)	0.89 (0.77,0.96)	0.26 (0.18,0.36)	51.4 (17.9,87.2)
Moment Power	0.95 (0.87,0.99)	0.62 (0.02,0.94)	0.82 (0.51,0.97)	0.50 (0.27,0.70)	93.8 (16.0,207.2)
Lognormal	1.00 (0.99,1.00)	0.98 (0.93,0.99)	0.98 (0.96,1.00)	0.10 (0.06,0.15)	19.7 (5.2,63.7)
Truncated LN	1.00 (0.99,1.00)	0.97 (0.94,0.99)	0.97 (0.94,0.99)	0.13 (0.08,0.22)	15.9 (4.6,40.2)
LogSeries	0.99 (0.97,1.00)	0.94 (0.82,0.99)	0.98 (0.94,1.00)	0.19 (0.15,0.24)	28.1 (7.5,46.6)
Neg Binomial	0.98 (0.93,1.00)	0.88 (0.38,0.98)	0.94 (0.83,0.99)	0.24 (0.09,0.50)	Inf (Inf,NaN)
ZSMultinomial	0.99 (0.97,1.00)	0.93 (0.76,0.99)	0.97 (0.92,0.99)	0.25 (0.19,0.33)	Inf (Inf,NaN)
reject LN via MC	6%	14%	19%	17%	1%
ZSM beats LN	20%	20%	32%	0%	NA

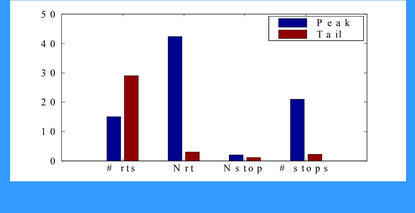
Sidebar #2: What causes abundances to vary across a range?

Abstract

1) Little is known about what causes the structure of abundance across a range.
2) There are two important changes in abundance in a transect across a range:
   a) The edge or boundary where abundance drops to zero
   b) The interior rapid dropoff where abundances drop from very high levels to the very low abundances of the tail over a fairly short distance
3) Three mechanisms are often proposed to explain the edge (boundary) of a range (climatic limitations, Gaussian competitive exclusion, and metapopulation dynamics). These three mechanisms all lead to very abrupt transitions to zero fitness. This makes it difficult to explain the tails which we observe. The idea of spatial refugia might provide a partial explanation for these tails, but they would usually be small.
4) Moreover, in the NA Breeding Bird Survey, both the number of occupied patches and the abundance per patch drop. See Figure 1. An explanation for the drop in abundance is needed.
5) I hypothesize a model based on a tradeoff between abiotic tolerance and competitive dominance (in the context of diffuse competition). This model uses empirically established physiological and behavioral responses to conditions.

   An example of such a tradeoff is found in the Saguaro Cactus. It is more cold tolerant than other large columnar cacti (e.g. Cardon) due to a greater number of spines and deeper fluting which lead to a thicker blanket of air. The Saguaro is out competed in the southern portion of its range despite growing well there when there is no competition. This is presumably due to the energetic and water costs of the traits that allow it to tolerate colder temperatures.
6) The proposed model produces a bell-shaped distribution of equilibrium abundances across space.

Figure 1



Model

1) Let x stand for a position in space along 1 dimension
2) Let dominance rank (0=dominates all others) be $\mathbf{r(x)=exp(-cx)}$ (Fig 2)
3) Let abiotic/stress-related mortality be $\mathbf{m(x)=exp(-exp(-x))}$ (Fig 2)
4) Let energy intake in the face of diffuse competition depend on density and rank: $\mathbf{I(x)=N^{-r(x)}}$ (Goss-Custard et al, various; Schwinning & Fox 1995)
5) Let total population density (all species) be constant across space (i.e. $\mathbf{N(x)=K}$)
6) Let fecundity be $\mathbf{F(x)=c(I(x)-E_{maint})}$ (Lawton, Hassell & Beddington 1975; Rees & Crawley 1989)
7) Let number of offspring reproducing be $\mathbf{R(x)=F(x)(1-m(x))}$
8) Let fitness/population dynamics be according to $\mathbf{1/N\ dN/dt = R(x)-uN}$ (here N is just 1 species) (Holt et al 1997)
9) We can now solve for N*(x), the equilibrium population size (Figure 3)

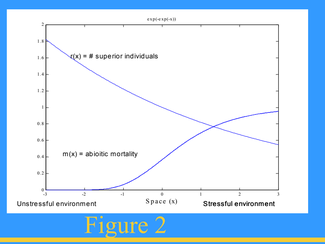


Figure 2

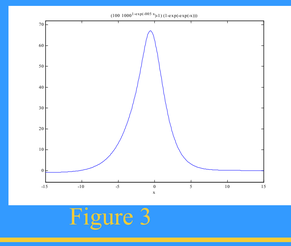


Figure 3

Acknowledgements

In partial collaboration: with Cathy Collins

Mike Rosenzweig
Leticia Aviles
Jim Cushing
Brian Enquist
Wayne Maddison
Larry Venable

NSF
Biodiversity RTG
Flinn Biomath
Bruce Walsh
Functional Ecology Group
Will Turner