

Um Protocolo para Ajuste e Seleção de Modelos de Abundância de Espécies

Paulo Inácio de Knecht López de Prado

Depto. de Ecologia, Instituto de Biociências, Universidade de São Paulo

4 de setembro de 2009

Resumo

Este projeto faz parte da linha de pesquisa que iniciei ao ingressar no Departamento de Ecologia da USP, em 2006, e que já está em andamento. Seu foco é o desenvolvimento e aplicação de um protocolo analítico inédito no Brasil, para descrever um padrão básico em ecologia, que é a predominância de espécies raras em comunidades naturais. Diversas distribuições teóricas de probabilidade foram propostas para descrever este padrão, o que permite a construção de modelos para estimar a probabilidade de ocorrência de uma espécie com uma certa abundância, em função de variáveis preditoras. Por meio de parceria com outros grupos de pesquisa, serão analisados dados de inventários padronizados de diferentes táxons ao longo de diferentes gradientes ambientais. Abordagens baseadas no conceito de verossimilhança estatística serão usadas para estimar os parâmetros de modelos que relacionem as distribuições de abundância com gradientes ambientais, bem como para compará-los. A expectativa é que tal refinamento analítico resulte em protocolos que possam ser usados por outros grupos de pesquisa, mas também que o conjunto de estudos de casos permita construir descrições gerais e realistas das mudanças da diversidade biológica em resposta aos fatores investigados.

Abstract

This proposal is part of an ongoing research program started when I moved to the Ecology Department of the University of São Paulo, in 2006. It aims to develop and apply a novel analytical framework to describe a basic pattern in ecology, namely, the prevalence of rare

species in biological communities. Many theoretical probability distributions had been proposed to describe this pattern, which can be used to model the probabilities of different abundance values of the species in response to explanatory variables. In collaboration with other research groups I will analyze datasets of standardized inventories of many *taxa* across different kinds of environmental gradients. Likelihood-based approaches will be used to estimate parameters of competing models of species abundance distributions (SADs) in response to environmental gradients, and also to compare them. My goals are to improve some analytical tools that can be further used by other groups, and to use the case studies to make general and realistic descriptions of the changes in biological diversity in response to the environmental factors investigated.

Equipe Executora :

- Paulo Inácio K. L. Prado, IB - USP (responsável)
- Camila Yumi Mandai, mestrado em Ecologia, USP
- Paula Alves Condé, mestrado em Ecologia, USP
- Renato Augusto Ferreira de Lima, doutorado em Ecologia, USP
- Rodolpho Credo Rodrigues, mestrado em Ecologia, USP

Pesquisadores Associados :

- João Luis Ferreira Batista, ESALQ - USP
- Paulo De Marco Jr., UFG
- Zolacir Trindade de Oliveira Jr., UESC

Colaboradores :

- Alexandre Adalardo de Oiveira, IB - USP
- Érica Hasui, UFAL
- Jean Paul Metzger, IB - USP
- Malva Isabel Medina Hernandez, UFSC
- Rafael Guerra Pimentel, mestrado em Ecologia, USP
- Renata Pardini, IB - USP
- Ricardo Ribeiro Rodrigues, ESALQ - USP

1 Fundamentação

1.1 Distribuição das Abundâncias das Espécies em uma Comunidade

A dominância numérica de poucas espécies nas comunidades biológicas é uma das poucas leis gerais da ecologia (McGill et al., 2007). A maioria das espécies nas comunidades são representadas por poucos indivíduos, e poucas são as espécies abundantes, resultando em um histograma de distribuição de abundâncias tipicamente côncavo, conhecido na literatura como “*hollow curve*”. Há pouquíssimas exceções a este padrão, o que o torna relevante para a análise de quaisquer outros parâmetro das comunidades. Uma de suas consequências mais diretas é que a riqueza de espécies é determinada pelas espécies raras (Whittaker, 1965), mas há outras menos óbvias. A concavidade da distribuição de abundância das espécies (**DAE**¹) pode explicar outros padrões importantes nas comunidades, como a inclinação da relação espécie-área (Preston, 1960; Pielou, 1977), o aninhamento (Fischer and Lindenmayer, 2002), e a relação positiva entre abundâncias e frequência de ocupação de manchas (Gaston, 1996). Além disso, as DAE têm enorme potencial aplicado, pois permitem uma análise muito mais informativa de padrões de diversidade do que índices sintéticos de riqueza e diversidade (McGill et al., 2007). Entre estes padrões está a raridade de espécies, cuja compreensão é uma das principais metas da biologia da conservação (Hubbell, 2001).

O uso de um histograma para representar as DAE evoca uma distribuição teórica subjacente. Não por acaso, essas duas idéias foram apresentadas juntas por Fisher et al. (1943), que propuseram a distribuição de série logarítmica de probabilidades para explicar a forma das DAE de amostras de comunidades biológicas com muitas espécies, e em seguida por Preston (1948) que defendeu o uso de uma distribuição log-normal como um modelo mais adequado. A falta de consenso persiste, com o agravante de que pelo menos outros vinte modelos foram propostos (Tokeshi, 1993; McGill et al., 2007). A proliferação de tantos modelos teóricos para explicar um mesmo padrão empírico é um reconhecido embaraço para os ecólogos (Whittaker, 1975; Magurran, 2004) e foi definido em uma revisão recente sobre o assunto como “*fracasso científico coletivo*” (McGill et al., 2007).

Uma das razões desse fracasso é que a abundância de modelos teóricos

¹a literatura recente em inglês usa a sigla SAD, de *Species Abundance Distribution*. Neste projeto usamos a forma em português "DAE"

contrasta com uma notável escassez de testes empíricos. Comparações sistemáticas de vários modelos por meio de conjuntos de dados representativos de vários grupos taxonômicos ou diferentes situações ecológicas são extremamente raros na literatura (McGill, 2003). Não é descabido arriscar que a maioria dos modelos só foi avaliada no artigo em que foram propostos, e com poucos conjuntos de dados. Quando diferentes autores avaliam os mesmos modelos, tendem a usar os mesmos conjuntos de dados. O exemplo atual é a distribuição multinomial de soma zero, prevista pela teoria neutra de Hubbell (2001), cujos critérios de avaliação têm sido intensamente discutidos mas invariavelmente com referência às abundâncias de árvores nas parcelas permanentes de Barro Colorado e de outras poucas florestas tropicais (e.g., Etienne and Olff 2005; McGill et al. 2006; Chave et al. 2006). Quando muitos conjuntos de dados são analisados, normalmente são poucos os modelos avaliados. As duas maiores compilações publicadas são de 222 comunidades por Hughes (1986) e de 70 comunidades por Dewdney (1998, 2000) e em ambas os autores compararam apenas os seus próprios modelos com o previsto pelas distribuições série logarítmica e log-normal.

Uma segunda dificuldade é que os métodos estatísticos tradicionalmente usados não são adequados. Em geral, infere-se que abundâncias observadas ajustam-se aos valores previstos por diferentes modelos se testes de aderência, como Qui-quadrado ou Kolmogorov-Smirnov, não indicam desvios significativos (e.g. Magurran 2004; Hughes 1986). É um uso equivocado da lógica de teste de significância, pois coloca o modelo de distribuição testado como a hipótese nula, o que faz com que sua aceitação dependa mais da força do teste do que da qualidade do ajuste (Routledge, 1980). Além disso, esses testes não são adequados para comparação de diferentes modelos, pois avaliam o ajuste a uma única distribuição por vez. Testes múltiplos trazem outros problemas (Manly, 1992) e muitas vezes são inconclusivos, pois é comum que os ajustes de diferentes modelos pareçam iguais.

Uma solução promissora para estes problemas metodológicos é o uso do princípio de verossimilhança estatística (Edwards, 1972; Royall, 2000) para criar protocolos de comparação simultânea de várias hipóteses estatísticas concorrentes. Um dos mais simples é a seleção de modelos baseada em índices de informação (Burnham and Anderson, 2002), que ordena um conjunto de modelos de acordo com sua plausibilidade frente aos dados, levando em conta sua parcimônia. Na escala proposta recentemente por McGill (2003), essa seria a categoria mais rigorosa de testes de ajustes dos dados a modelos teóricos. O uso crescente destes conceitos é considerada uma mudança no

paradigma de análise de dados, em resposta às limitações que as abordagens tradicionais impõem a muitas áreas da biologia (Johnson and Omland, 2004). O desenvolvimento de um tal protocolo depende, essencialmente, da correta especificação dos modelos concorrentes o que, no caso, envolve uma teoria de amostragem específica.

1.2 Teoria da Amostragem das DAE

Um teoria probabilística para descrever as abundâncias das espécies em uma amostra deve combinar dois modelos: da distribuição das abundâncias das espécies na comunidade e do procedimento de amostragem. Esta foi a abordagem adotada por Fisher et al. (1943) para deduzir a distribuição de série logarítmica, e que foi posteriormente generalizada e formalizada por Pielou (1977) e Green and Plotkin (2007).

Modelo da DAE

Com a premissa de que as abundâncias de cada espécie são variáveis aleatórias independentes e que seguem a mesma distribuição de probabilidades (uma *i.i.d.*², Pielou 1977), podemos descrevê-las com uma única função de distribuição probabilística (Engen, 1977). As abundâncias das espécies são variáveis discretas, mas se aceitamos que na comunidade há uma grande quantidade de indivíduos, uma aproximação para uma variável contínua é razoável. Acrescentando esta premissa, nosso modelo para a distribuição de abundâncias das espécies na comunidade será uma a distribuição de probabilidades de uma variável aleatória contínua, uma função de densidade probabilística (*pdf*³), que chamaremos de $f(n)$, sendo n o número de indivíduos de uma espécie tomada ao acaso da comunidade.

Modelo de Amostragem

Se a abundância de uma espécie na comunidade é n e uma fração a da comunidade é amostrada ao acaso, o número esperado de indivíduos na amostra é $E[X] = a \times n$. No entanto, a variação ao acaso imposta pela amostragem faz com que algumas amostras tenham menos ou mais indivíduos do que este valor, mas números muito diferentes do esperado são menos prováveis dos que os valores próximos.

²*independently and identically distributed random variable*

³*probability density function*

Logo, o número de indivíduos de cada espécie em uma amostra é também uma variável aleatória, que podemos descrever com uma função de distribuição de probabilidades. Como o número de indivíduos na amostra pode ser pequeno, modelamos esta variável aleatória como uma variável discreta, que descreve a probabilidade de que a abundância X na amostra seja um certo valor x , dada uma certa abundância n na comunidade:

$$g(x|n) = P(X = x | n) \quad (1)$$

Combinando os Dois Modelos

A probabilidade de que uma espécie tenha uma certa abundância n na comunidade e uma certa abundância x na amostra, dado n , é o produto das probabilidades destes dois eventos, $f(n) \times g(x | n)$. Como cada abundância n na comunidade é um evento excludente das demais, a probabilidade de uma espécie ser representada por x indivíduos na amostra é dada por :

$$h(x) = c \int_0^{\infty} f(n) g(x|n) dn \quad (2)$$

Portanto, $h(x)$ é a *pdf* das abundâncias de cada espécie na amostra, dados os modelos que descrevem as abundâncias na comunidade e o processo amostral (Green and Plotkin, 2007). Esta função pode ser definida de modo a preservar os parâmetros da função $f(n)$, que descreve a DAE da comunidade. Com isso, temos um modelo que, uma vez ajustado a uma amostra, recupera parâmetros da comunidade de onde foi tomada esta amostra.

1.3 Confrontando Modelos com Dados: A Abordagem da Verossimilhança

Se dois modelos atribuem probabilidades diferentes a um resultado, a Lei da Verossimilhança é um critério objetivo e simples para a escolha do modelo mais plausível: será aquele que atribuir a maior probabilidade ao resultado obtido. A rigor, o que usamos nesta comparação não são as probabilidades atribuídas pelos modelos, mas uma função delas, chamadas função de verossimilhança. A Lei também nos dá uma medida da plausibilidade de um modelo em relação a outro, que é a razão entre as suas verossimilhanças (*likelihood ratio*), o que fornece um critério objetivo de comparação (Edwards, 1972). Também é possível levar em consideração a complexidade dos modelos concorrentes, com funções de verossimilhança que penalizam a adição de

parâmetros, como o Critério de Informação de Akaike (AIC), ou o Critério de Informação Bayesiano (BIC).

A comparação simultânea de múltiplos modelos - que podem representar hipóteses com graus crescentes de complexidade - é mais adequada para estudos mensurativos (ou observacionais *sensu* Krebs 1999) tão comuns em ecologia, especialmente em escalas espaciais e temporais amplas (e.g. Brown 1995). Nestes casos, abordagens baseadas em verossimilhança são uma excelente alternativa aos testes de significância tradicionais, limitados a avaliar um só modelo (a hipótese nula), e deduzir a validade de outro (a hipótese alternativa) pela sua rejeição.

Além disso, a verossimilhança é uma função de probabilidades de um modelo condicionadas aos dados, e não é afetada pelo conjunto teórico dos resultados alternativos que não ocorreram. Nesta abordagem, uma vez feito o estudo, o espaço amostral é uma abstração irrelevante (Royall, 2000). Logo, tomam-se os dados como um fato, cujo valor de evidência a favor de um modelo contra outros não depende de se postular o que ocorreria se o estudo fosse repetido outras vezes. Testes de significância, ao contrário, têm a premissa de que os dados provêm de experimentos controlados que podem ser repetidos (Royall, 2000), que é inadequada para a maioria dos estudos mensurativos de ampla escala em ecologia e evolução (Johnson and Omland, 2004).

No caso das DAE, os modelos teóricos de distribuição de abundâncias são funções que atribuem probabilidades à ocorrência de uma espécie com abundância de um a infinitos indivíduos numa amostra de uma comunidade. Dado um vetor de abundâncias de espécies desta amostra, o modelo teórico terá a forma expressa pela equação 2. A verossimilhança é uma função das probabilidades atribuídas pelo modelo para cada abundância observada (Alonso et al., 2008). A função de verossimilhança é usada para encontrar os valores mais plausíveis para os parâmetros de cada modelo e, em seguida, para comparar diferentes modelos. As funções de verossimilhança de muitos dos modelos de distribuição de probabilidade propostos para as DAE já foram deduzidas (e.g. Bulmer 1974; Kempton and Taylor 1974; Newman 2005), alguns recentemente (Etienne and Olf, 2005; Alonso et al., 2008). Mesmo para os modelos ainda sem funções de verossimilhança definidas, há hoje recursos computacionais para a aproximação de estimativas de máxima verossimilhança, por meio de algoritmos de otimização (Bolker, 2008b,a). Nos dois casos, a análise do comportamento da função de verossimilhança na vizinhança de seu máximo permite também definir o intervalo de valores

igualmente plausíveis dos parâmetros de cada modelos, além de comparar diferentes modelos (Hilborn and Mangel, 1997; Royall, 2000).

O ajuste de um modelo de DAE para cada conjunto de dados pode ainda ser generalizada em um modelo em que os parâmetros da distribuição não são constantes. A lógica é a mesma de modelos mais conhecidos, como o de regressão linear simples, que descreve uma variável dependente com uma distribuição normal, cujo parâmetro média não é constante, e sim uma função linear de uma variável preditora. Para os modelos aqui considerados, a idéia é que mudanças na diversidade possam ser descritas com funções que expressem a resposta dos parâmetros das DAE a um conjunto de variáveis preditoras. Podemos vislumbrar, por exemplo, um modelo que ajuste a distribuição teórica de série logarítmica a amostras ao longo de um gradiente de fertilidade do solo, sendo que o parâmetro alfa (índice de diversidade de Fisher) da distribuição seja uma função linear das variáveis de fertilidade. Esta abordagem de construção de modelos probabilísticos é baseada no conceito de verossimilhança, e apenas generaliza procedimentos bem conhecidos como os de regressão linear com erros normais (Hilborn and Mangel, 1997; Bolker, 2008b).

2 Justificativa

Esta proposta faz parte de um projeto intelectual de longo prazo, que é promover mudanças nos protocolos de análise estatística usados na ecologia de comunidades, a meu ver necessárias para que ocorram avanços teóricos nesse campo de conhecimento. Minha principal estratégia tem sido o trabalho em parceria com outros grupos de pesquisa, pois acredito que o entendimento e aplicação de novos protocolos analíticos é facilitada se tais procedimentos são construídos em colaboração com colegas interessados em desenvolvê-los e aplicá-los. Os conjuntos de dados escolhidos foram produzidos por grupos que partilham o interesse pelas questões teóricas aqui propostas, o que se refletiu em delineamentos amostrais adequados para responder tais questões. Todas as colaborações já estão em andamento, por meio de teses de mestrado ou doutorado sob minha orientação, com co-orientação de pesquisadores dos grupos parceiros.

A colaboração entre pesquisadores e a divisão de tarefas me parece uma condição necessária para respondermos a questões ecológicas complexas, que demandam especialidades variadas articuladas em redes de pesquisa. O uso de conjuntos de dados já disponíveis me permite alocar tempo e recursos

para maior aprofundamento dos aspectos analíticos, que em si já é tarefa que também demanda parcerias. Neste sentido, venho mantendo colaboração com outros colegas que também focam sua pesquisa na área analítica, como o Dr. João L. F. Batista (ESALQ-USP), com quem criei e ministro uma disciplina de modelagem estatística baseada em verossimilhança⁴, e ministro uma disciplina de introdução à linguagem R⁵, juntamente com o Dr. Alexandre Adalardo. Além do Drs. João Batista e Alexandre, esta proposta prevê também a colaboração com a equipe do Dr. Paulo de Marco Jr. (UFG) que vem pesquisando as DAE (De Marco, 2006), e com o Dr. Zolacir Trindade Oliveira Jr. (UESC), com quem pretendo explorar aplicações de ferramentas analíticas de física estatística para os protocolos aqui propostos.

A escolha das distribuições de abundâncias relativas das espécies, as DAE, deve-se à sua condição de atributo básico das comunidades que segue um padrão universal, que são as curvas côncavas. Se aceitamos este padrão como lei empírica que pode ser expressa por meio de modelos quantitativos, podemos usá-lo para deduzir novos enunciados, como regularidades e diferenças entre as comunidades. Este uso de leis empíricas tem justificativas epistemológicas e, na prática, já promoveu importantes avanços teóricos em diversas ciências, incluindo a biologia (El-Hani, 2006). Foi esta a motivação para desenvolvimento e aplicação de modelos de DAE, que é muito bem resumida por Pielou (1977, pp.269):

If it should turn out that one single form of probability distribution with a small number of parameters (say two or three) fitted the data from the majority of observed communities, with only the parameter values varying from one community to the another, interesting relationships might be discovered between the values of the parameters and the types of community they describe.

As abordagens de ajuste e comparação de modelos ampliaram as possibilidades de implementação desta idéia simples, trazendo-a de volta ao centro das atenções na ecologia de comunidades (McGill et al., 2007). Contribuições neste campo têm alto potencial de impacto nesta área de conhecimento básico, bem como para quaisquer de suas aplicações que envolvam monitoramento da diversidade biológica.

⁴<http://cmq.esalq.usp.br/wiki/doku.php?id=biometria:verossim:start>

⁵<http://ecologia.ib.usp.br/bie5782>

3 Objetivos

1. Desenvolver protocolos baseados em verossimilhança estatística para ajustar e avaliar modelos probabilísticos de abundância de espécies que permitam:
 - (a) incorporar aos modelos efeitos amostrais;
 - (b) incorporar aos modelos efeitos de co-variáveis ambientais;
 - (c) identificar o(s) modelo(s) mais plausível(is) para um conjunto de dados;
 - (d) comparar diferentes conjuntos de dados, em termos dos parâmetros dos modelos mais plausíveis.
2. Aplicar esses protocolos para analisar mudanças na diversidade biológica em quatro conjuntos de dados, a saber:
 - (a) Inventários padronizados de aves, mamíferos e anfíbios em fragmentos florestais de diferentes tamanhos e áreas-controle de floresta contínua no Planalto Atlântico;
 - (b) Inventário padronizado de besouros Scarabaeidae em fragmentos florestais ao longo de gradientes de perturbação antrópica;
 - (c) Inventários de aves em paisagens do Sudeste Brasileiro com diferentes graus de fragmentação, obtidos da literatura;
 - (d) Parcelas permanentes de vegetação arborea em quatro formações florestais do estado de São Paulo.

4 Procedimentos

4.1 Desenvolvimento dos Protocolos de Análise

Implementação em R

Para ajustar e comparar os modelos usaremos a linguagem estatística R (R Development Core Team, 2009), que já inclui funções para a maioria das distribuições teóricas de probabilidades propostas para as DAE das comunidades, rotinas para definir novas distribuições, bem como para estimar (ou aproximar por otimização) e analisar o comportamento de funções de verossimilhança, e comparar modelos por critérios verossimilhança (Venables and Ripley, 2002; Bolker, 2008b). Embora haja estes recursos básicos para

o ajuste e comparação de diferentes modelos, é preciso usá-los para programar um protocolo específico para as DAE, o que inexistente em R, ou em qualquer outra plataforma. Isto inclui não só o uso de funções já disponíveis, como a criação de novas, notadamente as que incorporam o efeito de amostragem nos modelos.

A linguagem R é de código aberto e permite a criação de pacotes com novas funções, que podem ser publicados em um repositório público (<http://www.r-project.org/cran>). Desenvolveremos um protocolo de ajuste, diagnóstico e comparação de modelos de DAE a dados de abundância obtidos de inventários biológicos, que será reunido e documentado em um pacote na linguagem R. As funções básicas deste pacote incluirão:

- Densidade probabilísticas correspondentes aos principais modelos teóricos de DAE para as comunidades: série logarítmica, log-normal, exponencial, *broken-stick*, função de potência e função de potência atenuada.
- Densidades probabilísticas das combinações de DAE das comunidades com processos amostrais Poisson e Binomial Negativo, para as quais haja solução analítica;
- Rotinas de solução numérica para as combinações de DAE das comunidades e processos amostrais que não tenham solução analítica;
- Rotinas de ajuste e comparação dos modelos acima por máxima verossimilhança;
- Funções para diagnóstico numérico e visual dos modelos selecionados.

Testes do Protocolo

As funções de ajuste e seleção de modelos serão testadas com amostras simuladas de modelos teóricos de DAE. Como os modelos corretos e os valores de seus parâmetros serão conhecidos, este procedimento pode ser usado para avaliar a precisão e consistência dos estimadores usados, bem como sua resposta ao aumento do tamanho amostral. As etapas destes testes são:

1. Definir o tamanho da comunidade e sortear abundância de suas espécies de um modelo teórico de distribuição ($f(n)$ na equação 2);
2. Sortear uma amostra desta comunidade de acordo com um modelo de amostragem (equação 1);

3. Ajustar o modelo de distribuição de abundâncias a esta amostra (equação 2), por máxima verossimilhança;
4. Repetir as etapas acima várias vezes e guardar as estimativas dos parâmetros;
5. Avaliar a precisão pela amplitude de variação das estimativas dos parâmetros nas simulações;
6. Avaliar a consistência comparando a média das estimativas com os valores dos parâmetros.

Em teoria, estimadores de máxima verossimilhança têm consistência e eficiência assintóticas (Edwards, 1972). Os procedimentos acima serão usados para avaliar estas propriedades dos estimadores obtidos, em simulações de comunidades e amostras similares a de dados reais.

4.2 Aplicação do Protocolo

Uma vez implementado o protocolo de análise, ele será aplicado no ajuste e seleção de modelos para todos os conjuntos de dados. Cada análise buscará expressar mudanças na diversidade entre amostras como respostas dos parâmetros de um modelo de DAE a variáveis preditoras.

Estrutura das Análises

As análises seguirão as mesmas etapas para todos os conjuntos de dados:

Escolha dos Modelos: Os dados de cada amostra serão ajustados por máxima verossimilhança a diferentes distribuições teóricas propostas para as DAE, e os modelos mais adequados identificados com critérios de seleção de modelos baseados no AIC.

Avaliação da Variação dos Parâmetros: uma vez escolhidas as distribuições teóricas mais adequadas, avaliaremos o comportamento dos parâmetros estimados para cada amostra em função das variáveis preditoras, por meio de gráficos exploratórios.

Ajuste e comparação dos modelos finais: as duas etapas anteriores guiarão a criação de modelos sintéticos que expressem as diferenças dos modelos ajustados a cada amostra como mudanças nos parâmetros da distribuição teórica em função de variáveis preditoras. Os coeficientes destes modelos serão estimados por máxima verossimilhança, e os

modelos resultantes comparados, por métodos baseados em verossimilhança. Entre os modelos concorrentes será sempre incluído o modelo nulo de que não há efeito de nenhuma variável preditora sobre os parâmetros.

Análise de sensibilidade: a precisão das estimativas dos parâmetros de cada modelo selecionado será avaliada por meio dos seus perfis de verossimilhança (Bolker 2008b). A sensibilidade das previsões será avaliada examinando-se os valores previstos pelo modelo sob diferentes combinações dos valores dos parâmetros, tomados dentro do intervalo de plausibilidade obtido pelo perfil de verossimilhança (Bolker 2008a).

Conjuntos de Dados e Questões

a) PLANALTO ATLÂNTICO

Origem dos dados: Projeto “Conservação da biodiversidade em paisagens fragmentadas”⁶, coordenado pelos Drs. Jean Paul Metzger e Renata Pardini (IB-USP), financiado atualmente pelo Programa Mata Atlântica (CNPq/Alemanha), Auxílio Jovem Pesquisador FAPESP 05/56555-4, e anteriormente pelo Auxílio Temático FAPESP 99/05123-4 .

Caracterização dos dados: inventários padronizados de anfíbios, aves e mamíferos em três paisagens com graus crescentes de perda de habitat, no Planalto Atlântico Paulista. Os inventários foram feitos em 15 a 20 fragmentos com características equivalentes, e em áreas controle de floresta contínua.

Questões a investigar: mudanças nas DAE em função de tamanho e isolamento dos fragmentos, e comparação das DAE entre fragmentos e áreas controle.

b) AVES DO SUDESTE

Origem dos dados: Projeto “Avaliação de fatores biogeográficos associados com a vulnerabilidade das espécies de aves de Mata Atlântica à fragmentação”⁷, coordenado pela Dra. Erica Hasui, UFAL (pós-doutorado CNPq 151677/2005-0), com supervisão do Dr. Jean Paul Metzger.

⁶http://eco.ib.usp.br/lepac/paisagem_fragmentada.htm

⁷http://eco.ib.usp.br/lepac/paisagem_fragmentada.htm

Caracterização dos dados: Compilação de registros de ocorrência e abundância de espécies de aves da literatura, coleções e bases de dados *on-line*, nos estados de Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná e Santa Catarina. Foram selecionadas 32 áreas com inventários da comunidade de aves com redes de neblina, em quatro fitofisionomias (florestas ombrófila densa, semidecidual e mista; e cerrado). Para cada uma destas áreas estão calculadas métricas de fragmentação da paisagem, e evapotranspiração potencial, a partir de dados de sensoriamento remoto.

Questões a investigar: mudanças nas DAE em função da fitofisionomia, métricas de fragmentação e evapotranspiração potencial.

c) ESCARAVELHOS NA MATA ATLÂNTICA

Origem dos dados: Projeto “Insetos como indicadores ecológicos em áreas degradadas e em restauração”, coordenado pela Dra. Malva I. M. Hernández (UFSC), financiado pela Fundação de Apoio a Pesquisa Científica e Tecnológica do Estado de SC e CNPq.

Caracterização dos dados: inventários de besouros Scarabaeidae com armadilhas de queda com iscas em áreas litorâneas com diferentes graus e tipos de perturbação ambiental.

Questões a investigar: respostas das DAE à gradientes de perturbação ambiental em pequena escala, e potencial indicador dos Scarabaeidae.

d) ÁRVORES EM FORMAÇÕES FLORESTAIS DE SÃO PAULO

Origem dos dados: Parcelas permanentes de vegetação em São Paulo, coordenação geral do Dr. Ricardo Ribeiro Rodrigues (ESALQ-USP), por meio do Projeto Temático “Diversidade, dinâmica e conservação em Florestas do Estado de São Paulo: 40,96 ha de parcelas permanentes”⁸, projeto temático FAPESP (Biota, 99/09635-0). Parcela da Ilha do Cardoso financiada também pela Petrobras e Center for Tropical Forest Studies, por meio do projeto “Recuperação e Conservação dos Ecossistemas de Restingas do Litoral Sul de São Paulo”⁹, coordenado pelo Dr. Alexandre Adalardo de Oliveira, IB-USP.

⁸<http://www.lerf.esalq.usp.br/parapre.php>

⁹<http://ecologia.ib.usp.br/labtrop>

Caracterização dos dados: inventários de plantas lenhosas em parcelas permanentes de 10,24 ha, instaladas em unidades de conservação representativas das quatro principais formações florestais do Estado de São Paulo: cerradão, floresta semidecídua, floresta ombrófila densa, e floresta alta de restinga. Todas as plantas lenhosas com perímetro à altura do peito maior ou igual a 15 cm foram plaquetados, identificados e medidos. Na parcela de restinga (Ilha do Cardoso), um novo censo das plantas com DAP maior que 1 cm (PAP = 3,14 cm) está andamento.

Questões a investigar: comparação das DAE entre formações florestais, efeitos da mudança do critério de inclusão sobre os modelos de DAE, avaliação dos protocolos de ajuste por comparação entre o estimado para subparcelas e os valores conhecidos para a parcela.

5 Cronograma de Execução

ETAPA	TRIMESTRE							
	1	2	3	4	5	6	7	8
Reuniões parciais da equipe	X	X	X	X	X	X	X	X
Simpósios com toda a equipe				X			X	
Dedução dos modelos com soluções analíticas	X	X						
Desenvolvimento das soluções numéricas	X	X	X					
Programação das funções em R		X	X	X	X			
Teste do protocolo				X	X			
Relatório Parcial					X			
Ajuste aos dados: escolha dos modelos					X			
Ajuste aos dados: avaliação da variação dos parâmetros					X	X		
Ajuste e comparação dos modelos finais						X	X	X
Análises de sensibilidade							X	X
Documentação e publicação do pacote em R							X	X
Relatório Final								X
Redação de manuscritos							X	X

Planejei um projeto com duração de dois anos, sendo o primeiro dedicado ao desenvolvimento, implementação e teste dos protocolos, e o segundo ano para a sua aplicação. Reuniões de trabalho com parte da equipe e colaboradores ocorrerão ao longo de todo o projeto, conforme as necessidades.

Além disto, estão previstas duas reuniões (“simpósios”) com toda a equipe, incluindo colaboradores.

Referências

- Alonso, D., A. Ostling, and R. S. Etienne, 2008. The implicit assumption of symmetry and the species abundance distribution. *Ecology Letters* **11**:93–105.
- Bolker, B., 2008a. *bbmle*: Tools for general maximum likelihood estimation, based on *stats4* by the R Development Core Team. R package version 0.8.9.
- Bolker, B., 2008b. *Ecological Models and Data in R*. Princeton University Press, Princeton.
- Brown, J. H., 1995. *Macroecology*. University of Chicago Press, Chicago.
- Bulmer, M. G., 1974. On fitting the poisson lognormal distribution to species abundance data. *Biometrics* **30**:651–660.
- Burnham, K. P. and D. R. Anderson, 2002. *Model Selection and Multimodel Inference - A Practical-Theoretic Approach*. Springer-Verlag.
- Chave, J., D. Alonso, and R. S. Etienne, 2006. Comparing models of species abundance. *Nature* **441**:E1–E2.
- De Marco, P. J., 2006. Um longo caminho até uma teoria unificada para a ecologia. *Oecologia Brasiliensis* **10**:120–126.
- Dewdney, A., 2000. A dynamical model of communities and a new species-abundance distribution. *Biological Bulletin* **198**:152–165.
- Dewdney, A. K., 1998. A general theory of the sampling process with applications to the veil line. *Theoretical Population Biology* **54**:294–302.
- Edwards, A. W. F., 1972. *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*. Cambridge University Press.
- El-Hani, C. N., 2006. Generalizações ecológicas. *Oecologia Brasiliensis* **10**:17–68.

- Engen, S., 1977. Comments on two different approaches to the analysis of species frequency data. *Biometrics* pages 205–213.
- Etienne, R. S. and H. Olf, 2005. Confronting different models of community structure to species-abundance data: a bayesian approach. *Ecology Letters* **8**:493–504.
- Fischer, J. and D. Lindenmayer, 2002. Treating the nestdness temperature calculator as a 'black box' can lead to false conclusions. *Oikos* **99**:193–199.
- Fisher, R., A. Corbet, and C. Williams, 1943. The relation between the number of the species and the number of individuals in a random sample from animal population. *Journal of Animal Ecology* **12**:42–58.
- Gaston, K., 1996. The multiple forms of the interspecific abundance-distribution relationships. *Oikos* **76**:211–220.
- Green, J. L. and J. B. Plotkin, 2007. A statistical theory for sampling species abundances. *Ecology Letters* **10**:1037 – 1045.
- Hilborn, R. and M. Mangel, 1997. The Ecological Detective - Confronting Models with Data. Princeton University Press, Princeton.
- Hubbell, S. P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.
- Hughes, G., 1986. Theories and models of species abundance. *The American Naturalist* **128**:879–899.
- Johnson, J. and K. Omland, 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* **19**:101–108.
- Kempton, R. and L. Taylor, 1974. Log-series and log-normal parameters as diversity discriminants for the lepidoptera. *Journal of Animal Ecology* **43**:381–399.
- Krebs, C. J., 1999. Ecological Methodology. Addison-Wesley, New York, second edition.
- Magurran, A., 2004. Measuring Biological Diversity. Blackwell, London.
- Manly, B. F. J., 1992. The design and analysis of research studies. Cambridge University Press, Cambridge.

- McGill, B., 2003. Strong and weak tests of macroecological theory. *Oikos* **102**:679–685.
- McGill, B., R. Etienne, J. Gray, D. Alonso, M. Anderson, H. Benecha, M. Dornelas, B. Enquist, J. Green, F. He, A. Hurlbert, A. E. Magurran, P. Marquet, B. Maurer, A. Ostling, C. Soykan, K. Ugland, and E. White, 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* **10**:995–1015.
- McGill, B., B. Maurer, and M. Weiser, 2006. Empirical evaluation of neutral theory. *Ecology* **87**:1411–1423.
- Newman, M., 2005. Power laws, pareto distributions and zipf’s law. *Contemporary Physics* **46**:323–351.
- Pielou, E. C., 1977. *Mathematical Ecology*. John Wiley and Sons, New York.
- Preston, F. W., 1948. The commonness and rarity of species. *Ecology* **29**:254–283.
- Preston, F. W., 1960. Time and space and the variation of species. *Ecology* **41**:611–627.
- R Development Core Team, 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Routledge, R., 1980. The form of species-abundance distributions. *Journal of Theoretical Biology* **82**:547–558.
- Royall, R., 2000. *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, London.
- Tokeshi, M., 1993. Species abundance patterns and community structure. *Advances in Ecological Research* **24**:111–186.
- Venables, W. N. and B. D. Ripley, 2002. *Modern Applied Statistics with S*. Springer, New York, fourth edition.
- Whittaker, R., 1965. Dominance and diversity in land plant communities. *Science* **147**:250–260.
- Whittaker, R., 1975. *Communities and Ecosystems*. MacMillan Publishers, New York, second edition.