Universidade de São Paulo

**Ajuste e seleção de modelos na descrição de comunidades arbóreas:
estrutura, diversidade e padrões espaciais**

(versão corrigida)

RENATO AUGUSTO FERREIRA DE LIMA

Tese apresentada para obtenção do título de
Doutor em Ecologia, Área de Concentração:
Ecossistemas Terrestres e Aquáticos

(A versão original desta tese encontra-se disponível no Instituto de Biociências da USP)

**São Paulo, SP
2013**

RENATO AUGUSTO FERREIRA DE LIMA
Ecólogo

**Ajuste e seleção de modelos na descrição de comunidades arbóreas:
estrutura, diversidade e padrões espaciais**

(versão corrigida)

Orientador:
Prof. Dr. **PAULO INÁCIO K. L. DE PRADO**

Tese apresentada para obtenção do título de
Doutor em Ecologia, Área de Concentração:
Ecossistemas Terrestres e Aquáticos

São Paulo, SP
2013

FICHA CATALOGRÁFICA

Lima, Renato Augusto Ferreira de

Ajuste e seleção de modelos na descrição de comunidades arbóreas :
Estrutura, diversidade e padrões espaciais / Renato Augusto Ferreira de
Lima ; orientador Paulo Inácio de Knecht López de Prado. --. São Paulo,
2013.

196 f.

Tese (Doutorado) – Instituto de Biociências da Universidade de
São Paulo. Departamento de Ecologia.

1. Distribuição diamétrica. 2. Distribuição espacial. 3. Distribuição de
abundância de espécies. 5. Ecologia de comunidades. 6. Floresta tropical.
I. Universidade de São Paulo. Instituto de Biociências. Departamento de
Ecologia.

Comissão Julgadora:

Prof. Dr. Fernando Roberto Martins

Profa. Dra. Luciana Ferreira Alves

Prof. Dr. Tadeu Siqueira de Barros

Prof. Dr. Alexandre Adalardo de Oliveira

Prof. Dr. Paulo Inácio K. L. Prado
(Orientador)

*Aos meus pais, Mara e
Raimundo, e à minha
amada companheira,
Isabel, dedico.*

AGRADECIMENTOS

Ao Prof. Dr. Paulo Inácio K.L. Prado por toda sua paciência, dedicação e tranquilidade na difícil tarefa de ensinar pessoas a pensar de maneira crítica sobre ecologia e sobre ciência em geral. Com certeza estou muito mais preparado hoje e devo isso em grande parte a sua capacidade de formar pesquisadores com competência e entusiasmo dignos de um mestre. Obrigado e espero poder apreender ainda mais com você no futuro!

Aos professores Alexandre Adalardo de Oliveira (IB/USP), Adriana Maria Z. Martini (IB/USP) e João Luís F. Batista (ESALQ/USP) pelas valiosas contribuições e idéias dadas durante meu exame de qualificação. Agradeço ainda ao João Luís F. Batista pela participação e sugestões que resultaram no Capítulo 4.1 e à colaboração com Paula Condé (IBAMA/AC) que resultou no Capítulo 3.

Ao Smithsonian Tropical Research Institution (STRI), em especial aos pesquisadores Helene C. Muller-Landau e Richard S. Condit por todo o apoio nas análises dos dados e discussão dos resultados apresentados no Capítulo 4.2. Agradeço também ao diretor do Center for Tropical Forest Science (CTFS/STRI), Stuart J. Davies, pela oportunidade de participar do ‘2012 CTFS/SIGEO Dimensions on biodiversity workshop’.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado concedida e ao projeto temático “Diversidade, dinâmica e conservação em Florestas do Estado de São Paulo: 40,96 ha de parcelas permanentes” (Processo no. 99/09635-0) do programa BIOTA-FAPESP (www.biota.org.br), que disponibilizou a base de dados utilizada durante grande parte desta tese. Agradeço ainda ao Instituto de Biociências da Universidade de São Paulo e ao Programa de Pós-graduação em Ecologia por toda a estrutura relacionada ao curso de doutorado e a execução desta tese. Agradeço ainda à FAPESP (Processo no. 09/53413-5) pelo apoio às atividades do Capítulo 3.

Ao pessoal do Laboratório de Ecologia Teórica (LET) da USP, em especial à Camila Mandai, Rodolpho Rodrigues, Camila Castanho, Paula Condé, Sara Mortara e Marcelo Awade pelas discussões, comentários e convivência. Ao pessoal da LAGE pela companhia, discussões e risadas.

À minha família, Mara, Raimundo e Ive, por todo amor e apoio à minha formação pessoal e profissional. À Pina e principalmente à Isabel, por todo amor, paciência e companheirismo nos momentos gostosos e difíceis do caminho que me trouxe até aqui.

“Essentially, all models are wrong, but some are useful.”

– George E.P. Box, 1987.

SUMÁRIO

RESUMO.....	2
ABSTRACT	4
CAPÍTULO 1: INTRODUÇÃO GERAL	6
1.1 Descrevendo padrões em comunidades	6
1.2 Ajuste e seleção de modelos	10
1.3 Abordagens de inferência estatística.....	12
1.4 Objetivos gerais e plano da tese.....	15
1.5 Metodologia geral	17
1.5.1 Protocolo analítico.....	17
1.5.2 Áreas de estudo.....	23
1.6 Referências bibliográficas.....	25
CAPÍTULO 2: COMPARANDO DESCRITORES FLORESTAIS BÁSICOS.....	32
CAPÍTULO 3: DISTRIBUIÇÃO DE ABUNDÂNCIA DE ESPÉCIES	58
CAPÍTULO 4: DISTRIBUIÇÃO DIAMÉTRICA DOS FUSTES.....	80
4.1 Modeling tree diameter distributions in tropical forests	80
4.2 How well do demographic rates predict tree size distributions?	108
CAPÍTULO 5: DISTRIBUIÇÃO ESPACIAL DAS ESPÉCIES.....	152
CAPÍTULO 6: CONSIDERAÇÕES FINAIS.....	185
6.1 Descrevendo padrões de comunidades arbóreas em parcelas permanentes	185
6.2 Ajuste e seleção de modelos por máxima verossimilhança	186
ANEXOS	189

RESUMO

Ajuste e seleção de modelos na descrição de comunidades arbóreas: estrutura, diversidade e padrões espaciais

A descrição de padrões, *i.e.*, tendências ou arranjos não aleatórios em comunidades, possui um longo histórico em ecologia vegetal. Comumente, a estrutura e diversidade de comunidades vegetais são descritas a partir de sua distribuição em classes de tamanho (SDD), distribuição espacial (SSD) e de sua distribuição abundância de espécies (SAD). Isto porque há um pressuposto de que padrões existentes nestes descritores de comunidades são assinaturas de processos fundamentais na sua organização e funcionamento. Assim, a descrição de padrões é com frequência o primeiro passo para gerar ou testar hipóteses sobre esses processos que regulam a estrutura e diversidade de comunidades.

Organizada em diferentes capítulos, esta tese teve como objetivo central descrever e comparar padrões em diferentes comunidades arbóreas Neotropicais, buscando gerar hipóteses sobre os processos que regulam sua organização e funcionamento. Para tanto, buscou-se utilizar uma abordagem de inferência baseada no ajuste e seleção de modelos, que foi realizado usando máxima verossimilhança estatística. Em todos os capítulos, os dados sobre as comunidades arbóreas são oriundos de diferentes parcelas florestais permanentes, quatro delas com 10,24 hectares, localizadas no Brasil, e outra com 50 hectares, localizada no Panamá. Além da introdução geral sobre os conceitos e técnicas utilizadas nesta tese (Capítulo 1), foram avaliados e comparados: (i) os descritores básicos da estrutura florestal (*i.e.* abundância, área basal e riqueza de espécies por sub-parcela – Capítulo 2); (ii) a SAD e como ela varia com o aumento da escala e tamanho amostral (Capítulo 3); (iii) a SDD e como ela se relaciona com a demografia das espécies (Capítulo 4.1 e 4.2); e, por fim, (iv) a SSD e como ela varia entre parcelas permanentes (Capítulo 5). Apenas a relação entre a SDD e a demografia das espécies foi realizada como os dados da parcela panamenha, enquanto que os demais capítulos se referem as quatro parcelas brasileiras. Cada capítulo utilizou métodos e modelos probabilísticos distintos para a descrição e comparação das variáveis de interesse.

O Capítulo 2 mostrou que os descritores básicos florestais foram muito diferentes entre as quatro parcelas brasileiras. Estes descritores foram raramente normais nas escalas estudadas e as diferenças entre as parcelas foram mais evidentes quando a variância é considerada, fornecendo informações extras sobre os processos geradores de variabilidade dentro das parcelas. O Capítulo 3 mostrou que a mudança no formato da SAD com o aumento da escala é predominantemente um efeito indireto do tamanho da amostra. Assim, pode haver um efeito de escala, mas esse efeito é pequeno e parece depender do grau de similaridade de espécies entre amostras. No capítulo 4.1, foi necessária uma combinação de quatro diferentes distribuições de probabilidade para descrever a ampla gama de SDD, visto que os modelos candidatos raramente foram adequados para a maioria das espécies. No capítulo 4.2, verificou-se que o crescimento e recrutamento determinam o formato da SDD, o que não aconteceu com a mortalidade. No geral, curvas decrescentes de crescimento por diâmetro (*i.e.*, maior crescimento juvenil) levaram à SDD menos íngremes, enquanto que taxas altas de recrutamento estiveram relacionadas à SDD mais íngremes. Apesar das previsões da teoria de equilíbrio demográfico terem apresentado relações positivas com as SDD observadas, houve muita variação, fazendo com que as previsões fossem pouco confiáveis. No capítulo 5, confirmou-se que a grande maioria das espécies se distribui de maneira agregada no espaço. No entanto, as parcelas apresentaram diferentes padrões de intensidade e tamanho de agregação. As diferenças nestes padrões entre parcelas foram, em geral, similares àquelas encontradas ao comparar populações de uma mesma espécie entre parcelas.

Assim, as parcelas permanentes brasileiras apresentaram padrões bem distintos umas das outras, tanto em termos de estrutura quanto de diversidade de espécies, padrões estes que provavelmente foram determinados pelas condições ambientais as quais estas comunidades estão sujeitas. Diferentes padrões também foram encontrados em relação à distribuição espacial das espécies (*i.e.* frequência, intensidade e tamanho de agregação). Contudo, os resultados sugeriram que estes padrões estiveram mais ligados a heterogeneidade ambiental interna das parcelas do que com condições de clima e solo as quais estas comunidades estão sujeitas. Por outro lado, os padrões de distribuição espacial das espécies parecem ter influenciado os padrões de diversidade das comunidades. Não houve, entretanto, um número suficiente de parcelas sob as diferentes combinações de condições ambientais para testar estas sugestões ou para fazer generalizações para cada formação florestal, sendo necessário o confronto destas sugestões com outros estudos realizados em condições similares. Já na parcela panamenha, foi confirmada a expectativa teórica de que distribuições diamétricas refletem a demografia das espécies, em especial os padrões de crescimento e recrutamento. Mas, houve grande variação entre as espécies, dificultando a inferência precisa de padrões demográficos passados das espécies a partir de SDD atuais. Por fim, a abordagem analítica baseada no ajuste e seleção de modelos por máxima foi uma alternativa viável, flexível e apropriada, principalmente em relação à comparação simultânea de diferentes modelos e à busca de processos por trás dos padrões encontrados. Apesar de algumas limitações de cunho operacional, a abordagem baseado em modelos é uma alternativa adequada para a descrição de comunidades arbóreas, podendo ser utilizada de maneira consorciada com outras abordagens (*e.g.* testes de hipóteses) para descrever padrões e para gerar ou testar hipóteses sobre esses processos fundamentais que regulam a estrutura e diversidade destas comunidades.

Palavras-chave: distribuição diamétrica, distribuição espacial, distribuição de abundância de espécies, diversidade de espécies, ecologia de comunidades, estrutura florestal, floresta tropical, máxima verossimilhança

ABSTRACT

Model fit and selection in the description of tree communities: structure, diversity and spatial patterns

Pattern description – search for trends or non-random arrangements in communities, has a long history in plant ecology. Commonly, the structure and diversity of plant communities are described based on their size class distribution (SDD), spatial distribution (SSD) and species abundance distribution (SAD). This is because there is an underlying assumption that the existing patterns in these community descriptors are signatures of key processes determining their organization and functioning. Thus, pattern description is often the first step to generate or test hypotheses about the processes governing community structure and diversity.

Organized in different chapters, the main goal of this thesis was to describe and compare different patterns in Neotropical tree communities and to generate hypotheses about the processes that regulate them. To do so we used an approach based on model selection, which was performed using maximum likelihood. In all chapters the data on tree communities came from different permanent forest plots, four of them of 10.24 ha located in Brazil and another 50 ha located in Panama. In addition to the general introduction of key concepts and techniques used along the thesis (Chapter 1), it was evaluated and compared: (i) the basic forest descriptors (*i.e.* abundance, basal area and species richness per subplot - Chapter 2), (ii) the SAD and how it varies with increasing sample size and scale (Chapter 3), (iii) the SDD and how it relates to species demography (Chapter 4.1 and 4.2) and, finally, (iv) the SSD and how it varies between plots (Chapter 5). Only the relationship between SDD and species demography was performed using the Panama plot data, while the remaining chapters relate to the four Brazilian plots. Each chapter used different methods and probabilistic models for the description and comparison of the variables of interest.

In Chapter 2, it was found that basic forest descriptors were very different between the four Brazilian plots. These descriptors were rarely normal at the studied scales and differences between plots were more evident when variance is accounted for, which seems to provide information on processes generating within-plot variability. Chapter 3 showed that the change in shape of the SAD due to increasing scale is predominantly an indirect effect of sample size. Thus, there may be an effect of the scale, but this effect is minor and seems to depend on the degree of species turnover between samples. In Chapter 4.1, it was shown that the combination of four different probability distributions was necessary to describe the wide range of SDD, since models were rarely appropriate for the majority of tree species. In Chapter 4.2, it was found that growth and recruitment, but not mortality, shape the SDD. On average, decreasing growth-diameter curves (*i.e.* higher juvenile growth) were associated to less steep SDD, whereas high recruitment rates were related to steeper SDD. Although the predictions of demographic equilibrium theory were positively related to the observed SDD, there was lots of variation, making predictions quite unreliable. In Chapter 5, it was confirmed that the great majority of species had clumped spatial distributions. However, the results of intensity and size of clumps showed that the patterns of aggregation were different among plots. Species shared between two plots generally showed patterns of spatial distribution that matched the patterns found for individual plots.

Therefore, the Brazilian plots presented very distinct patterns, both in terms of structure and species diversity, which were most probably determined by the environmental conditions to which these communities are subjected. Different patterns among plots were also found in respect to species spatial distribution (*i.e.* frequency, intensity and size of

aggregation). However, the results suggested that these patterns were more connected to within-plot environmental heterogeneity than with climate and soil conditions. On the other hand, the spatial distribution of the species seems to have influenced the patterns of diversity of communities. There was not, however, a sufficient number of plots under different combinations of environmental conditions to test these suggestions or to make generalizations for each forest type, being necessary to confront these suggestions with other studies conducted in similar conditions. In the Panamanian plot, it was confirmed the theoretical expectation that diameter distributions reflect the demographics of the species, in particular the patterns of growth and recruitment. But there was great variability among species, making it difficult to infer past demographic patterns from current SDD. Finally, the analytical approach based on model fit and selection by maximum likelihood was a viable, flexible and appropriate approach, particularly in respect to the simultaneous comparison of different models and to the search for mechanisms underlying patterns. Despite some more operational limitations, the model-based approach is an appropriate alternative for the description of tree communities and can be jointly used with other approaches (*e.g.* hypothesis testing) for pattern description and to generate and test hypotheses on the fundamental processes that determine the structure and diversity of these communities.

Keywords: community ecology, diameter distribution, forest structure, maximum likelihood, spatial distribution, species abundance distribution, species diversity, tropical forest

CAPÍTULO 1: INTRODUÇÃO GERAL

Introdução

1.1 Descrevendo padrões em comunidades

Desde o início da ecologia como ciência até os dias de hoje, a descrição de padrões é uma das abordagens mais tradicionais (MacArthur 1972, Lawton 1996). Em ecologia de comunidades, padrões são entendidos como tendências ou arranjos não aleatórios nos descritores da comunidade ou de suas espécies (Greig-Smith 1961). Em comunidades com alta complexidade estrutural, taxonômica ou funcional, nas quais esses arranjos ou tendência podem não ser tão aparentes (Hill 1973), descrições analíticas e quantitativas da comunidade são passos essenciais na busca por padrões, que tradicionalmente se inicia pela quantificação da sua diversidade e estrutura (MacArthur 1969, Losos *et al.* 2004). Quanto à diversidade, os descritores básicos são o número de espécies, a equidade de abundâncias, ambos (*e.g.* distribuição de abundância de espécies – SAD) ou como a riqueza de espécies varia com a área amostral (*i.e.* relação espécie-área – SAR). Alguns estudos incluem também a maneira com a qual o número de espécies amostradas se acumula ao longo do processo de amostragem, descrita pela curva de acúmulo de espécies ou curva do coletor (SAC). Quanto à estrutura, é possível quantificar a distribuição dos indivíduos no espaço (SSD) ou em classes de tamanho (SDD). A descrição desta estrutura pode ser feita para a comunidade como um todo (identidade dos indivíduos não importa) ou espécie a espécie.

Existe um motivo pelo qual estes descritores são frequentemente usados na detecção de padrões em comunidades. Há um pressuposto básico de que padrões na SAD, SAR, SSD ou SDD são assinaturas de processos fundamentais na organização e funcionamento de comunidades. Este pressuposto, portanto, faz com que a descrição destes padrões seja uma maneira de gerar ou testar hipóteses sobre esses processos fundamentais (Lepš 1990, Gilbert & Lechowicz 2004, Hubbell 2004). Ou seja, a descrição de padrões precede a busca pelas causas dos padrões encontrados, uma abordagem que tem sido considerado essencial para o progresso da ecologia (Swihart *et al.* 2002). Apesar de um mesmo padrão poder ser gerado a partir de processos diferentes ou a partir de contribuições diferentes destes processos (Lawton 1996), inúmeras hipóteses sobre processos reguladores da estrutura e diversidade de comunidades surgiram a partir da descrição da SAD (Tokeshi 1999, Hubbell 2001), SSD (Hubbell 1979, Condit *et al.* 2000), SDD (Enquist & Niklas 2001, Muller-Landau *et al.* 2006)

e SAR (Condit *et al.* 1996, Lomolino 2000). Vale destacar o valor histórico das comunidades vegetais, em especial das florestas, como provedoras de evidências empíricas para o reconhecimento de importantes processos em ecologia. Alguns exemplos clássicos são: sucessão (Clements 1936, Connell & Slatyer 1977), partição de nicho (Denslow 1980), denso-dependência (Connell *et al.* 1984), interações inter-específicas (Callaway & Walker 1997) e neutralidade (Ågren & Fagerström 1984, Hubbell 2001).

Apesar de simples, a SAD é um dos descritores de comunidades mais úteis e difundidos em ecologia (Pielou 1975, Hubbell 1979, Magurran 2004). Isto porque ela é capaz de descrever, ao mesmo tempo, quantas espécies a comunidade possui e como suas abundâncias estão distribuídas. A SAD é quase sempre caracterizada pela abundância de espécies raras e pela raridade de espécies abundantes, um padrão que é considerado como um dos mais gerais e intrigantes em ecologia (Preston 1948, McGill *et al.* 2007). A generalidade do formato da SAD, tanto em comunidades vegetais como animais, fez com que muitos ecólogos propusessem processos por trás deste padrão tão recorrente. Assumindo a abundância como bons indicadores do desempenho das espécies na comunidade, alguns ecólogos sugeriram que a SAD resulta da partição dos recursos disponíveis entre espécies (MacArthur 1960, Tokeshi 1999), de modo que as espécies mais abundantes seriam aquelas com maior capacidade de assimilação dos recursos. Outros têm sugerido que padrões na SAD são resultantes de uma dinâmica aleatória nas taxas de nascimento, morte e dispersão das espécies (Hughes 1986, Engen & Lande 1996, Hubbell 2001). Um terceiro grupo propõe que o formato da SAD é que um efeito amostral relacionado ao tamanho (Preston 1948, Dewdney 1998, McGill 2003) ou à escala da amostragem (Connolly *et al.* 2005, Šizling *et al.* 2009). Tantas explicações para um mesmo padrão levou à proliferação de modelos para descrever a SAD (McGill *et al.* 2007). Por outro lado, várias outras aplicações da SAD permanecem inexploradas. Dentre elas, por exemplo, a aplicação de SAD para entender distribuições de variáveis além do o número de indivíduos, como a biomassa (McGill *et al.* 2007).

Enquanto a SAD descreve as abundâncias por espécie, a SDD descreve as abundâncias por classe de tamanho, fornecendo uma descrição sintética da estrutura vertical da comunidade. Em florestas, o tamanho das árvores é tradicionalmente estimado usando o diâmetro, mas a área basal, volume ou biomassa também podem ser usados. Assim como a SAD, a SDD possui um padrão que se repete na maioria das florestas: há muitas árvores pequenas e poucas grandes. A forma da SDD depende fundamentalmente das taxas de crescimento e mortalidade da floresta ou de suas espécies (Muller-Landau *et al.* 2006) e por isso a SDD tem sido usada para inferir sobre a estabilidade da floresta, estratégias específicas

de regeneração e tendências populacionais (Knight 1975, Lorimer & Frelich 1984, Condit *et al.* 1998, Wright *et al.* 2003). Nesse contexto, fatores capazes de afetar as taxas vitais das espécies (*e.g.* competição ou distúrbios – Goff & West 1975, Coomes & Allen 2007) afetarão a forma da SDD. Outros autores propuseram que o formato da SDD seja uma consequência da relação entre a taxa fotossintética e o tamanho de plantas vasculares e como essa relação determina a densidade por classe de tamanho (densidade $\propto$ diâmetro⁻² – Enquist *et al.* 1998, Enquist & Niklas 2001). Essas relações metabólicas, contudo, não tem encontrado grande suporte de dados empíricas, especialmente no que se refere ao excesso de árvores grandes previstas por essas relações (Coomes *et al.* 2003, Muller-Landau *et al.* 2006).

Além da SDD, outro importante descritor da estrutura de comunidades é a maneira com a qual os indivíduos se distribuem no espaço (*i.e.* estrutura horizontal - SSD), que pode ser classificada como agregada, aleatória ou uniforme/regular. Existem muitos métodos para detectar padrões na SSD, como aqueles baseados em ocorrência/abundância por amostra e aqueles na localização dos indivíduos (ver Perry *et al.* 2006 e McGill 2011 para detalhes sobre esses métodos). Antes de Hubbell (1979), que propôs a agregação na SSD da maioria das espécies em uma floresta decidual, havia uma crença de que espécies florestais se distribuíam aleatoriamente pela floresta. Desde então, estudos em florestas tropicais e temperadas tem confirmado o padrão predominantemente agregado das espécies (Armesto *et al.* 1986, He *et al.* 1997, Condit *et al.* 2000). Alguns processos têm sido sugeridos para explicar a agregação espacial das espécies, dentre eles a associação com habitat, disponibilidade de recursos, limitação na dispersão de sementes e facilitação ou competição entre espécies (Valencia *et al.* 2004, Li *et al.* 2009). Apesar da maioria destes estudos apenas descrever padrões de distribuição específicos, alguns trabalhos buscaram separar a variação explicada por estes diferentes processos ou tentaram associar padrões espaciais entre espécies (Wiegand *et al.* 2007, Wang *et al.* 2010). Certamente, estas abordagens são muito promissoras para examinar os mecanismos capazes de manter a grande diversidade de espécies em florestas tropicais.

Os padrões existentes nos descritores das comunidades não são totalmente independentes, de modo que um mesmo processo poderá gerar os padrões encontrados em dois ou mais descritores da comunidade. Por exemplo, o grau de agregação das espécies na comunidade (*i.e.* SSD) é capaz de influenciar o formato da SAD e da SAR (May 1975, Plotkin *et al.* 2000, He & Legendre 2002, Green & Plotkin 2007), que por sua vez influenciam o formato da SAC (Condit *et al.* 1996, Colwell *et al.* 2004). Padrões de beta diversidade (*distance-decay of similarity*) também dependem da SSD e da SAD (Plotkin &

Muller-Landau 2002). Essas inter-relações entre padrões colocam os processos geradores da SAD e SSD como os processos mais fundamentais na organização das comunidades. Tal constatação levou McGill (2010) a propor uma teoria que unifica padrões de distribuição espacial dentre e entre espécies (*i.e.* SSD) com padrões de abundância em grandes escalas (*i.e.* SAD da metacomunidade) como os únicos determinantes da SAD em escala local, da SAR, dos padrões de beta-diversidade e, conseqüentemente, nos padrões da SAC. Assim, parece existir uma hierarquia entre os descritores das comunidades (Harte *et al.* 2005), de modo que o padrão existente em um descritor pode ser o “processo” que determina os padrões em outros.

Com o intuito de investigar esses padrões em comunidades florestais, grandes parcelas permanentes (*large, long-term plots* - Condit 1995) têm sido instaladas no mundo todo desde meados dos anos 70. Apesar de terem tido seu início nos trópicos (Hubbell 1979), hoje existe um grande número de parcelas sob variadas condições ambientais e com tamanhos que variam entre 10 e 150 ha. Através de inventários florestais detalhados, estas parcelas têm fornecido descrições mais fiéis de comunidades florestais, permitindo avaliar padrões em escalas temporais e espaciais compatíveis aos processos estudados. Além disso, o grande volume de dados gerados permite descrever a estrutura e dinâmica de populações de muitas de espécies, incluindo as mais raras. Por estes motivos, os dados obtidos nestas parcelas oferecem valiosas oportunidades para descrever padrões e para inferir sobre processos fundamentais na estrutura e diversidade de comunidades florestais. Isto porque a combinação entre grandes bases de dados e métodos analíticos apropriados tem revelado novos padrões e como eles variam no tempo e no espaço (Condit 1995, Hubbell 2004). Atualmente, muitas destas parcelas estão sendo reunidas em redes de monitoramento (*e.g.* CTFS - <http://www.ctfs.si.edu>). Essas redes estão possibilitando a comparação de padrões em escala global, capazes de responder desde perguntas mais teóricas sobre o funcionamento de comunidades até prover conhecimento básico para solucionar questões práticas sobre manejo florestal ou impactos relacionados às mudanças climáticas. Assim, não é uma coincidência que muitos dos avanços modernos em ecologia teórica tenham sido desenvolvidos ou testados com dados coletados nestas grandes parcelas (*e.g.* Hubbell 1979, 2001; Hubbell *et al.* 1999, Condit *et al.* 2000, 2004, Plotkin *et al.* 2000, Harms *et al.* 2001; Muller-Landau *et al.* 2006).

1.2 Ajuste e seleção de modelos

A análise de dados coletados em comunidades ecológicas, assim como em outros sistemas naturais complexos, não é uma tarefa trivial. Ao contrário dos experimentos realizados em condições controladas, estudos observacionais em ecologia (e evolução) geralmente se deparam com fenômenos não-replicáveis que se expressam em ambientes difíceis de serem controlados ou manipulados (Hilborn & Mangel 1997). Comunidades naturais resultam de muitos efeitos e interações causadas por variações individuais, ambientais e/ou históricas, geralmente desconhecidas ou desconsideradas (Burnham & Anderson 2002). Isto faz com que as variações causadas por tais efeitos e interações não possam ser consideradas apenas como erros sem implicações biológicas. Assim, um ecólogo que busca desvendar padrões e inferir sobre processos em comunidades deverá lançar mão de uma abordagem analítica capaz de lidar com esse grau de complexidade e com a natureza observacional da maioria dos dados.

A existência de padrões produzidos por muitos efeitos sugere que poderá haver mais de uma explicação para um mesmo padrão, o que vai de encontro com a abordagem dos testes de hipótese nula usada na análise de experimentos. A idéia de uma abordagem capaz de oferecer e comparar múltiplas explicações para um mesmo fenômeno é, portanto, indicada para o avanço teórico em ecologia e evolução (Platt 1964, Johnson & Omland 2004). Apesar de essas idéias terem sido concebidas em 1890 por um naturalista que buscava explicar a origem da Terra (Chamberlin 1890), apenas um século depois ecólogos e biólogos passaram a perceber a força desta abordagem em estudos observacionais (Hilborn & Mangel 1997). Hoje, essa abordagem foi aprofundada e incorporada a um contexto estatístico sólido e está se tornando cada vez mais difundida em ecologia e evolução (Johnson & Omland 2004). O termo ‘hipóteses múltiplas de trabalho’, cunhado por Chamberlin (1890), hoje foi operacionalizado pelo procedimento estatístico conhecido como ‘seleção de modelos’. Entende-se como modelo uma descrição conceitual ou matemática que é útil no entendimento de um fenômeno. Modelos “verdadeiros” em sistemas naturais geralmente são indescritíveis ou inacessíveis (Burnham & Anderson 2002). Assim, modelos são simplificações ou generalizações que descreverão melhor ou pior um fenômeno dependendo das evidências empíricas disponíveis.

A abordagem de seleção de modelos possui algumas vantagens sobre as demais abordagens de inferência estatística (Hilborn & Mangel 1997, Burnham & Anderson 2002, Lindsey 2004, Johnson & Omland 2004). Em primeiro lugar, ela permite a comparação simultânea de dois ou mais modelos concorrentes. Apesar de mais de um modelo poder ser selecionado, essa abordagem geralmente ajuda a eliminar modelos que definitivamente não

são boas descrições, um processo fundamental para o avanço científico em qualquer disciplina (Platt 1964). Segundo, ela requer um exercício teórico e criativo prévio para buscar modelos que representem explicações que façam sentido. Além de ser um exercício válido por si só, ele tende a evitar o que Lohle (1987) chamou de ‘viés de confirmação’, que é uma tendência em pensar em estudos que tentam confirmar e não falsificar uma hipótese (*i.e.* efeito “hipótese de pelúcia”¹). Em terceiro lugar, a comparação entre modelos candidato é baseada em critério baseados no suporte que os dados observados oferecem. Nos testes de hipóteses nulas, comum na abordagem frequentista de inferência estatística (detalhes no próximo item), os valores de probabilidade são obtidos baseados nos dados observados e nos dados que poderiam ter sido observados, assumindo um determinado modelo amostral, o que não é muito intuitivo e tem gerado muitas críticas (Anderson *et al.* 2000).

Existem outros benefícios mais práticos ligados ao uso da seleção de modelos como abordagem de inferência sobre fenômenos quaisquer. Quando aliada à teoria da informação e aos métodos de máxima verossimilhança (ver item 1.5.1 Protocolo analítico), a abordagem de seleção de modelos é bastante flexível. Ela não exige a definição *a priori* de uma hipótese nula e permite a comparação entre modelos não aninhados (Lindsey 2004). A abordagem de inferência por seleção de modelos também permite lidar com diferentes pressupostos quanto à distribuição dos dados, quanto à forma da relação entre variáveis dependentes e independentes, presença de co-variáveis, etc. Vale lembrar, contudo, que o uso dos testes de hipótese nula ainda são apropriados para experimentos ou em situações que envolvem a tomada de decisão (*i.e.* há ou não há efeito? – Lindsey 2004). Além disso, o uso consorciado da seleção de modelos (*e.g.* sugestão de processos plausíveis) seguido de experimentação (*e.g.* testes dos processos sugeridos) pode ser bastante frutífero (Johnson & Omland 2004).

Boa parte dos procedimentos necessários para a seleção de modelos depende de computadores. Como veremos abaixo, os métodos baseados em máxima verossimilhança, assim como os métodos Bayesianos, comumente dependem de procedimentos numéricos intensivos para gerar estimativas dos parâmetros e, conseqüentemente, avaliar a aderência dos modelos aos dados. Por isso, boa parte da responsabilidade pela difusão da seleção de modelo em ecologia e outras disciplinas está relacionada aos avanços computacionais das últimas décadas (Hilborn & Mangel 1997). A velocidade de processamento que computadores têm hoje permite que o ajuste e comparação de modelos complexos a bases de dados grandes sejam feitos em segundos. Antes da era dos computadores, as possibilidades de análises eram

¹ Termo usado pelo Prof. Dr. Glauco Machado do Instituto de Biociências da Universidade de São Paulo.

limitadas e o teste de hipótese nula assumindo a distribuição normal (Gaussiana) como modelo de erros era uma das únicas saídas viáveis (Lindsey 2004). Junto com a seleção de modelos, outros métodos também se beneficiaram dos avanços computacionais, dentre eles pode-se citar os métodos baseados em randomizações ou em análise de modelos nulos (Gotelli 2001, Manly 2006). Contudo, a abordagem de seleção de modelos emerge como uma alternativa pujante em ecologia de comunidades, pois ela permite quantificar e comparar padrões usando protocolos analíticos unificados, que são baseados em modelos probabilísticos e não apenas em tendência centrais ou índices de diversidade ou de agregação.

1.3 Abordagens de inferência estatística

Em estatística, inferência significa tirar conclusões sobre um todo conhecendo apenas uma parte dele (*i.e.* amostras ou conjuntos de dados), sendo esse todo governado por variações aleatórias. Quanto mais fortes os padrões existentes nas amostras, maior é nosso poder de inferir sobre o todo (Edwards 1992). Todas as diferentes abordagens de inferência estatística possuem premissas e procedimentos para que possamos dar este “pulo” para além dos dados. Apesar de não serem totalmente exclusivas, estas abordagens diferem quanto a essas premissas e procedimentos, podendo ser divididas em: (i) frequentista, (ii) verossimilhança e (iii) Bayesiana (Bolker 2008). A escolha entre algumas destas ‘escolas’ de inferência estatística é uma decisão que cabe ao pesquisador e depende de aspectos como coerência entre as premissas de cada abordagem e o contexto científico do pesquisador, assim como a facilidade de compreensão, implementação e comunicação da abordagem escolhida. A primeira abordagem é a mais tradicional, tendo sido desenvolvida particularmente para a análise de dados experimentais e de outros procedimentos replicáveis, como o controle de qualidade de processos industriais. A segunda e terceira abordagens, apesar de criadas nos meados do século passado e no século XVIII, respectivamente, têm sido mais amplamente utilizadas nas últimas décadas devido aos avanços computacionais (Royle & Dorazio 2008) e principalmente em estudos com caráter não-experimentais. A abordagem baseada na verossimilhança foi aquela mais usada durante esta tese. A terceira abordagem, ao contrário, foi incluída para fins didáticos e não é discutida em igual profundidade por não ter sido usada em nenhum momento aqui. Todas as definições e conceitos apresentados abaixo são apresentados de maneira bastante simplificada, sob a perspectiva de um pesquisador que se preocupa em compreender criticamente as ferramentas que utiliza.

A abordagem frequentista, também conhecida como teoria ou paradigmas Fisherianos e de Neyman-Pearson, tem como objetivo central escolher uma hipótese em detrimento de outra (Royall 1997), um procedimento conhecido como teste de hipótese. Para tanto, essa abordagem tem como premissa que uma amostra é apenas uma das infinitas repetições possíveis para um mesmo experimento, de modo que cada amostra será uma reprodução independente do processo amostral em estudo. Essa repetição geraria uma distribuição amostral, que é um modelo teórico sobre o comportamento das estimativas calculadas a partir da amostra (Batista 2009). Essas distribuições amostrais são definidas a partir de alguns pressupostos básicos (*e.g.* normalidade dos dados e igualdade de variâncias) e sua forma depende de uma hipótese nula escolhida *a priori*. Geralmente, assume-se que tais distribuições têm parâmetros fixos e desconhecidos (Royle & Dorazio 2008), cujo valor não é afetado pelos dados observados. A estimativa dos parâmetros, do intervalo de confiança e da probabilidade de erro associada à escolha de uma das hipóteses é feita baseado nestas distribuições amostrais. Contudo, a idéia de inferência baseada em ‘amostras que poderiam ter sido realizadas’ e em distribuições amostrais que são fixas e assumidas *a priori* (*i.e.* dados não importam) tem recebido muitas críticas nas últimas décadas (Edwards 1992, Anderson *et al.* 2000, Bolker 2008). Essas idéias fariam mais sentido em condições experimentais, contexto no qual há a possibilidade de réplicas e controle amostral. Apesar de pouco intuitiva e problemática em determinadas circunstâncias, esta abordagem tornou-se muito difundida devido a sua objetividade e de sua atratividade no processo de tomada de decisão (Royall 1997, 2004).

Passo a passo, a abordagem frequentista é feita da seguinte forma. Primeiro, define-se a estatística de interesse, que é uma métrica calculada a partir da amostra (*e.g.* diferenças entre médias ou razões entre variâncias). Em seguida, assumindo a re-amostragem indefinida da população, é definida a distribuição amostral sob a ótica da hipótese nula (*e.g.* não há diferença entre médias ou razões entre variâncias) que pode ter sido deduzida analiticamente ou obtidas por randomização das amostras. Um exemplo de distribuição amostral analítica é a distribuição *t* de *Student* que sob a ótica da hipótese de médias iguais é uma distribuição com média zero e variância que depende do número de graus de liberdade (*i.e.* tamanho amostral). Uma vez definida essa distribuição amostral, o valor observado da estatística de interesse é comparado ao valor predito pela distribuição amostral, visando determinar qual a probabilidade de ser obter um valor igual ou mais extremo. Essa probabilidade é, então, usada para decidir se a hipótese nula deve ou não ser rejeitada. Ou seja, a pergunta central na abordagem frequentista é: dado a amostra e o modelo de amostragem, a hipótese nula é

válida? Assim, o foco principal desta abordagem é qual a chance da amostra gerar um certo valor da estatística de interesse, dado um processo de amostragem pré-definido (hipótese nula) e não qual o processo que gerou os dados observados. O principal método de estimação de parâmetros usados nessa abordagem é o método dos mínimos quadrados (em inglês, *Least Squares*) que busca minimizar a soma dos quadrados das diferenças entre os valores estimados pela distribuição amostral hipotética e os dados observados.

Na abordagem de verossimilhança, há uma mudança de postura em relação à importância dos dados observados na inferência estatística. Agora, são os dados quem irão definir qual hipótese possui maior plausibilidade de ser verdadeira (Edwards 1992). Ou seja, a inferência não se baseia mais em usar os dados para aceitar ou rejeitar uma hipótese nula, mas sim em verificar qual o valor de evidência dos dados para cada hipótese, o que é a própria definição de verossimilhança, também conhecida por suporte. Assumindo os dados como realizações de um processo aleatório, as hipóteses são agora as variáveis enquanto os dados são fixos, o que é a diferença fundamental da abordagem frequentista (Edwards 1992). Seguindo a Lei de Verossimilhança (Edwards 1992, Royall 1997), aceita por todas as abordagens de inferência estatística, uma hipótese terá maior suporte que outra quando ela atribuir uma probabilidade maior aos dados, de modo que a força de evidência a favor de uma destas hipóteses será dada pela razão entre essas probabilidades, conhecida como Razão de Verossimilhança. Ou seja, nesta abordagem a força de evidência que os dados atribuem a uma hipótese só faz sentido quando comparada a outra hipótese (Royle & Dorazio 2008). Assim, verossimilhança é sempre condicionada aos dados, enquanto que a força de evidência expressa a relação entre hipóteses concorrentes.

Apesar da diferença na postura em relação aos dados, a abordagem de verossimilhança também é capaz de gerar estimativas de parâmetros e intervalos de confiança, bem como fornecer meios para a comparação de hipóteses. Contudo, nesta abordagem a pergunta central é: qual hipótese possui maior força de evidência frente aos dados observados? Ou seja, seu foco principal está nos possíveis processos que geraram o conjunto de dados observados. Essa abordagem de inferência estatística tem sido considerada a mais apropriada para a análise de dados de natureza observacional (Edwards 1992), sendo especialmente recomendada em ecologia e no contexto de seleção de modelos (Johnson & Omland 2004, Bolker 2008). Por ser mais intuitiva, a inferência por máxima verossimilhança pode ser mais facilmente entendida, aplicada e visualizada graficamente (Edwards 1992). Além disso, essa abordagem é capaz de reproduzir a maioria dos testes de significância tradicionais (Bolker 2008). Como veremos com mais detalhes abaixo, o valor de verossimilhança atribuído pelos dados varia de

acordo com os valores dos parâmetros de funções de probabilidade, o que é definido pela função de verossimilhança. Assim, a estimativa dos parâmetros na abordagem de verossimilhança é feita pelo método de máxima verossimilhança, que consiste em encontrar os parâmetros que possuem maior verossimilhança para um determinado conjunto de dados, os chamados estimadores de máxima verossimilhança. O passo a passo da abordagem de verossimilhança no contexto de seleção de modelos é fornecido no item 1.5.1 Protocolo analítico.

Por fim, a abordagem Bayesiana também foi desenvolvida também para tentar acessar o suporte de hipóteses fornecido pelos dados. Mas nesta abordagem os dados empíricos são combinados com informações prévias sobre o fenômeno em estudo, de modo que essas informações são atualizadas e refinadas. Na prática, esta abordagem combina a função de probabilidade anterior (conhecimento prévio) com a função de verossimilhança (dados empíricos) para produzir uma função de probabilidade posterior (conhecimento atualizado), de modo que quanto maior a amostra, menor a importância relativa da probabilidade anterior. Como a abordagem Bayesiana também se baseia na função de verossimilhança para obter a função posterior, a abordagem Bayesiana e de verossimilhança estão conectadas pela Lei da Verossimilhança (Edwards 1992, Royle & Dorazio 2008). Edwards (1992) foi além ao dizer que “assim como verossimilhança é a abordagem Bayesiana sem as probabilidades anteriores, ela é a abordagem frequentista sem os erros”. Apesar de embutir certa subjetividade no processo de inferência relacionado à escolha da probabilidade anterior, a abordagem Bayesiana é recomendada para pesquisas onde já há um bom conhecimento prévio do fenômeno, para pesquisas realizadas de maneira sequencial ou em análises hierárquicas (Bolker 2008). Assim, nesta abordagem a pergunta principal é: como os dados alteram meu conhecimento prévio sobre um fenômeno? De maneira geral, a obtenção da função de probabilidade posterior não é feita analiticamente, mas sim com o uso de algoritmos conhecidos como *Markov Chain Monte Carlo*, que amostram e aproximam a função posterior sem calculá-la diretamente (Royle & Dorazio 2008).

1.4 Objetivos gerais e plano da tese

Esta tese foi organizada em capítulos que, apesar de abordarem aspectos distintos dentro da ecologia de comunidades, tiveram como objetivo comum descrever e comparar padrões coletivos de comunidades ecológicas usando uma abordagem de inferência estatística baseada no ajuste e seleção de modelos por máxima verossimilhança. Assim, apesar de cada capítulo

conter seus objetivos e protocolos analíticos específicos, todos têm em comum a filosofia de que os dados são as melhores evidências que temos da realidade e, por isso, os próprios dados são aqueles que irão selecionar os modelos que melhor descrevem os padrões observados (Hilborn & Mangel 1997, Burnham & Anderson 2002). Essa filosofia não foi usada de maneira ‘ortodoxa’, visto que em alguns capítulos foram usadas técnicas frequentistas de inferência estática (*e.g.* teste de hipótese nula e intervalos de confiança) aliadas ao ajuste e/ou seleção de modelos por máxima verossimilhança. Em alguns capítulos foi possível ir além da simples descrição e comparação de padrões e testar hipóteses sobre os mecanismos geradores dos padrões observados nestas comunidades. Isto porque, em alguns casos, os modelos concorrentes eram expressões matemáticas de hipóteses alternativas para explicar um mesmo padrão (Burnham & Anderson 2002).

Além deste objetivo geral comum, todos os capítulos abordaram padrões emergentes de comunidades arbóreas inventariadas usando parcelas florestais permanentes. Em sua maioria, os dados são provenientes de parcelas florestais com 10,24 ha montadas em quatro formações florestas do Estado de São Paulo, Brasil. No Capítulo 2, é feita a comparação de descritores básicos da estrutura florestal destas quatro áreas, incluindo a comparação da densidade, área basal e riqueza de espécies, visando reconhecer semelhanças e diferenças entre as áreas estudadas. No terceiro capítulo, avaliou-se a distribuição de abundância das espécies nestas quatro áreas e como a forma desta distribuição varia com o aumento da escala e tamanho amostral. O capítulo 4 abordou as distribuições de diâmetro das espécies em dois momentos distintos. Primeiro, fez-se uma avaliação de quais modelos estatísticos são mais apropriados para descrever distribuições diamétricas de *stands* e espécies destas quatro florestas (Capítulo 4.1). Em seguida, foi escolhido um destes modelos para descrever distribuições diamétricas de espécies e examinar se estas distribuições podem ser previstas a partir de taxas demográficas passadas (Capítulo 4.2). A exceção dos demais capítulos, este capítulo foi avaliado usando dados provenientes de uma parcela florestal permanente instalada no Panamá que possui 30 anos de monitoramento das taxas de recrutamento, crescimento e mortalidade das espécies. Por fim, no Capítulo 5 foram avaliados, de maneira preliminar, os padrões de distribuição espacial das espécies presentes nas mesmas quatro formações florestais e como estes se relacionam com alguns atributos básicos das espécies.

Ao longo do desenvolvimento desta tese e das demais atividades relacionadas ao programa de doutorado, algumas das idéias contidas no plano original tiveram que ser deixadas de lado por questões estratégicas ou de tempo. Alguns capítulos consumiram muito mais tempo do que o esperado e outros mostraram-se mais interessantes nesse momento. O

capítulo 2 era originalmente dividido em dois sub-capítulos, o primeiro que é apresentado aqui, comparando os descritores básicos por sub-parcela entre as quatro parcelas permanentes do Estado de São Paulo, e um segundo que faria a comparação dos descritores das comunidades (*i.e.* SAD, SAR, SAC e SDD) por parcela de 10,24 ha. O mesmo ocorreu com o capítulo sobre SADs (Capítulo 3), que conteria um sub-capítulo avaliando a previsibilidade da SAR baseada nos ajustes da SAD, visto que há uma relação direta entre esses dois descritores. O capítulo sobre SDD conteria um terceiro sub-capítulo comparando os ajustes às distribuições em nível de stand e de espécie usando um dos modelos estatísticos examinado no Capítulo 4.1, a Odd Weibull. Por fim, havia ainda um último capítulo que trataria da composição florística e comparação dos padrões de beta-diversidade entre as quatro parcelas, tema esse que nem sequer começou a ser explorado aqui.

1.5 Metodologia geral

1.5.1 Protocolo analítico

Como apresentado acima, a abordagem de inferência estatística utilizada na grande maioria dos capítulos desta tese foi baseada no ajuste e seleção de modelos (Johnson & Omland 2004, Bolker 2008). Na prática, cada capítulo utilizou esta abordagem de diferentes formas ou com diferentes intensidades. Contudo, o uso desta abordagem durante as análises seguiu um protocolo geral comum que é apresentado abaixo. Esse protocolo procurou usar uma mesma abordagem analítica ao longo de suas diferentes etapas, abordagem esta baseada essencialmente no princípio de verossimilhança, otimização numérica e na teoria da informação. Sua construção se baseou fortemente nos conceitos e práticas sugeridas por Hilborn & Mangel (1997), Burnham & Anderson (2002), Johnson & Omland (2004), Lindsey (2004), Royall (1997, 2004) e Bolker (2008), que são referências fundamentais, mas não as únicas, no que diz respeito à aplicação do ajuste e seleção de modelos por máxima verossimilhança em biologia e ecologia. Um roteiro prático bastante útil que ilustra boa parte dos conceitos e técnicas apresentadas a seguir pode ser encontrado em Batista (2009). As explicações que seguem são bastante simplificadas de modo que se recomenda uma consulta a estes trabalhos para a completa compreensão dos conceitos tratados.

Como foi necessária a construção de modelos e funções específicas para responder as perguntas de cada capítulo, foi necessário usar uma ferramenta capaz de prover essa flexibilidade. Existem vários programas estatísticos e de programação disponíveis. Nessa tese, optou-se pelo uso pela linguagem e ambiente de programação R (<http://www.r-project.org>),

no qual todos os procedimentos analíticos descritos no protocolo abaixo foram programados. Há algumas razões para isso. Além de ser um *software* livre, o R oferece um ótimo suporte de procedimentos estatísticos e computacionais além de grande flexibilidade na construção de gráficos e figuras. Por outro lado, há vários pacotes internos e contribuídos que executam muitas das funções e cálculos descritos neste protocolo. Nesta tese, os pacotes mais utilizados foram: *bbmle* (Bolker 2012), *MASS* (Venables & Ripley 2002), *vegan* (Oksanen *et al.* 2012), *sads* (Prado & Miranda, 2013), *spatstat* (Baddeley & Turner 2005) e *nlme* (Pinheiro *et al.* 2013). Apesar do R ser mais lento e consumir mais memória que outras linguagens de programação e pacotes estatísticos (Bolker 2008), para os fins desta tese e para o tamanho dos conjuntos de dados, o R se mostrou suficientemente rápido e flexível.

Passo 1 – *Definição dos modelos concorrentes*

Na abordagem de seleção de modelos por máxima verossimilhança, assim como nas demais abordagens de inferência estatística, a escolha dos modelos é fundamental. Isso porque, além dos dados, não se pode fazer qualquer tipo de inferência estatística sem um modelo. Ao contrário da obtenção dos dados, a escolha de modelos ou hipóteses apropriadas tem grau variável de subjetividade, pois ela depende em grande parte da experiência prévia do pesquisador (Burnham & Anderson 2002) e independe da abordagem de inferência adotada. Idealmente, cada modelo concorrente deve representar uma hipótese científica que quando confrontado com dados empíricos possam auxiliar na busca pelos mecanismos biológicos por trás dos dados. Frequentemente, modelos são inclusos entre os concorrentes apenas por que eles descrevem bem um dado fenômeno (*i.e.* abordagem fenomenológica), mas mesmo nesta abordagem é necessário ter critérios de escolha. Além da natureza dos dados em si, teorias propostas na literatura, experiência prévia com o sistema que se pretende modelar e o número e facilidade de interpretação dos parâmetros do modelo são alguns destes critérios (Burnham & Anderson 2002, Bolker 2008). Um bom conjunto de modelos concorrentes deve incluir desde os mais simples (poucos parâmetros) até modelos mais complexos (muitos parâmetros), porém a escolha de todos eles deve estar bem fundamentada para que modelos que ofereçam aproximações melhores não fiquem de fora. Existe um balanço entre a complexidade e a capacidade de generalização de um modelo que também deve ser considerado na definição dos modelos candidatos. O ideal é ter modelos capazes de fornecer bons ajustes, mas com simplicidade (Lindsey 2004).

Na prática, modelos são traduções matemáticas de hipóteses, que podem ser expressas por diferentes funções matemáticas, distribuições de probabilidade ou por distribuições de

probabilidade cujos parâmetros são influenciados por co-variáveis, como por exemplos a identidade da espécie ou o tipo de ambiente onde a observação foi coletada (Hilborn & Mangel 1997, Bolker 2008). Diferentes combinações de co-variáveis (*i.e.* diferentes estruturas do modelo) darão origem a modelos concorrentes diferentes, que serão então comparados quanto à importância das co-variáveis na explicação do conjunto de dados. Este é o princípio básico dos modelos de regressão múltiplas tradicionais que podem facilmente ser comparadas num contexto de seleção múltipla de modelos concorrentes. Nesta tese, em alguns capítulos utilizou-se as próprias distribuições de probabilidade (*e.g.* Poisson, Binomial Negativa, Normal ou Log-normal) como modelos concorrentes. Exemplos e aplicações de distribuições usadas nesta tese como modelos concorrentes são apresentados no Anexo 1. Em outros capítulos, testou-se o efeito de diferentes co-variáveis sobre a distribuição das variáveis de interesse.

Passo 2 – *Ajuste dos modelos aos dados*

Uma vez definido o conjunto de modelos concorrentes, faz-se o ajuste dos modelos aos dados. O ajuste aos dados tem como principal objetivo verificar através de uma métrica de ajuste qual o suporte que os dados oferecem a cada um dos modelos concorrentes, lembrando que no princípio de verossimilhança estatística são os dados quem decidem quais modelos são mais apropriados e não o inverso (Hilborn & Mangel 1997). Em outras palavras, avalia-se qual é a plausibilidade de se obter o conjunto de dados observado assumindo um determinado modelo como explicação da realidade. Como veremos em maior detalhe a seguir, um dos componentes básicos para o ajuste aos dados é a função de verossimilhança. De maneira geral, a função de verossimilhança pressupõe que as variáveis sejam variáveis aleatórias e independentes, para que as probabilidades atribuídas a cada dado possam ser multiplicadas. Para os fins desta tese a aleatoriedade das variáveis de interesse (*e.g.* densidade de indivíduos ou espécies, abundância por espécie ou medidas de diâmetros) é uma pressuposição bastante plausível. Para algumas variáveis, contudo, a independência entre as observações não pode ser assumida, ao menos em determinadas escalas, como descrito em maior detalhe no Capítulo 2.

Nesta tese, o ajuste aos dados foi feito usando o método de máxima verossimilhança que foi aplicado da seguinte forma. Primeiro, definiu-se a função de verossimilhança $L(\theta|X)$, que é uma função onde os parâmetros do modelo concorrente são variáveis e os dados são fixos (*i.e.* uma função dos parâmetros condicionada aos dados e não uma função dos dados condicionados aos parâmetros). Nesta função, o modelo concorrente é representado por uma função de densidade probabilística (PDF) que possui o conjunto de parâmetros θ . A PDF atribui um valor de probabilidade a cada observação para cada valor de θ . No caso de

observações múltiplas independentes, a função de verossimilhança torna-se uma multiplicação das probabilidades atribuídas a cada uma das observações:

$$L(\theta|X = x_i) = \prod f(x_i, \theta),$$

onde x_i são os dados observados e $f(x_i, \theta)$ é a PDF. Como essa multiplicação de probabilidades rapidamente tende a zero com o aumento da amostra utilizou-se o logaritmo da função de verossimilhança, o que resulta em valores negativos. Por questões práticas, faz-se a transformação do sinal resultando na função de log-verossimilhança negativa, que em linguagem matemática é dada por:

$$L(\theta|X) = - \sum_{i=1}^n \log [L(\theta|X)],$$

onde X representa um conjunto de dados com n valores e $L(\theta|X)$ é a função de verossimilhança de um modelo com parâmetros θ (note o *itálico* na grafia da função de verossimilhança). Assim, o valor de verossimilhança se torna um número positivo com uma escala mais fácil de trabalhar. Esse é a notação geral das funções utilizada nas análises desta tese.

Em seguida, é necessário encontrar os valores dos parâmetros θ que resultam no melhor ajuste aos dados (*i.e.* maior valor de verossimilhança), chamado de estimador de máxima verossimilhança (MLE). Para tanto é necessário encontrar o valor de θ que maximiza a função de verossimilhança ou, como foi feito nesta tese, encontrar o valor de θ que minimiza a função de log-verossimilhança negativa. Para funções de log-verossimilhança negativa mais simples é possível encontrar o MLE de maneira analítica, ao igualar a derivada da função à zero. Mas, para outras funções não há solução analítica e se faz necessário o uso de otimização numérica ou computacional. Esse foi o procedimento padrão utilizado aqui para encontrar o MLE nos diferentes capítulos desta tese, pois ele possui propriedades consideradas desejáveis para a inferência estatística (Royle & Dorazio 2008). De maneira geral, utilizou se os algoritmos de Nelder-Mead e de Broyden-Fletcher-Goldfarb-Shanno (BFGS) para encontrar as soluções numéricas das funções de log-verossimilhança negativa. No caso de problemas de convergência do otimizador numérico, o que geralmente ocorre quando os perfis ou superfícies de verossimilhança têm vários mínimos locais (Bolker 2008), foi utilizado o algoritmo estocástico *simulated annealing* (SANN), que possui menor precisão que os demais e é mais lento. Sempre que disponível para uma determinada distribuição de probabilidade, o método dos momentos foi usado para fornecer valores iniciais dos parâmetros ao otimizador. Em alguns casos, o resultado do SANN foi usado como valores iniciais dos demais algoritmos, visando maior precisão nas estimativas dos MLEs.

Uma vez encontrado o MLE, é possível determinar o intervalo ou combinações de θ para qual a diferença de log-verossimilhança negativa em relação à log-verossimilhança no MLE não seja maior que um dado limiar. Para modelos com apenas um parâmetro este intervalo é chamado de intervalo de verossimilhança, enquanto que para modelos com dois ou mais parâmetros usa-se o termo superfície de verossimilhança. Vale destacar que intervalos ou superfícies de verossimilhança podem não ser simétricos em torno do MLE e sua interpretação é sempre relativa a ele, que varia dependendo do conjunto de dados. Obviamente, quanto maior o conjunto de dados, maior a precisão da estimativa, o que resulta em intervalos menores. Durante esta tese, intervalos e superfícies de verossimilhança de modelos com mais complexos foram aproximados examinando a segunda derivada ou as segundas derivadas parciais da função de log-verossimilhança negativa no MLE. Para modelos com mais de dois parâmetros têm-se uma matriz de segundas derivadas, chamada de matriz de *Hessian*, e para obter os intervalos ou superfícies é necessário inverter numericamente essa matriz (Bolker 2008).

Passo 3 – *Comparação e seleção dos modelos concorrentes*

No caso de dois ou mais modelos concorrentes é necessário que eles sejam comparados usando algum critério objetivo, capaz de auxiliar na seleção de quais modelos são plausíveis na explicação dos dados observados. Existem diferentes maneiras de realizar essa comparação e nesta tese optou-se pelo uso de uma abordagem baseada na teoria da informação, que permite a comparação simultânea de dois ou mais modelos concorrentes. Optou-se pelo critério de proposto por Akaike (1973) que estabelece uma relação assintótica entre a distância relativa de Kullback–Leibler e a máxima verossimilhança de um modelo. A distância ou discrepância de Kullback–Leibler, de maneira muito resumida, expressa a distância entre um modelo qualquer e um modelo “verdadeiro” que gerou os dados. Como esse modelo não é conhecido na análise de dados empíricos, mas é o mesmo para todos os modelos concorrentes, conhecê-lo torna-se irrelevante quando o intuito é comparar modelos concorrentes, onde o que importa é apenas a distância relativa entre eles e o modelo “verdadeiro” (Burnham & Anderson 2002). Por isso, o Critério de Informação de Akaike (AIC) possui importantes implicações na abordagem de ajuste e seleção de modelos, sendo definido por:

$$AIC = -2 \log(L) + 2k,$$

onde L é o valor de máxima verossimilhança de um modelo (*i.e.* verossimilhança quando $\theta = \text{MLE}$) e k é o número de parâmetros do modelo. O valor de AIC, portanto, é uma medida da

distância relativa de um modelo a um modelo verdadeiro desconhecido, ou seja, uma estimativa da distância relativa esperada de Kullback–Leibler. Quanto menor o AIC, mais próximo o modelo está do modelo verdadeiro. Contudo, um valor absoluto de AIC por si só não importa, o que importa é a relação desse valor com os valores e AIC dos demais modelos concorrentes. Assim, a diferença de AIC do modelo concorrente com menor valor de AIC, ΔAIC , foi usada para comparar e ordenar os modelos quanto a sua plausibilidade de descrição dos dados. Há um consenso entre pesquisadores de que valores de ΔAIC menores que dois são indicadores de modelos ou combinações de parâmetros plausíveis e que valores de ΔAIC maiores que 10 indicam modelos não plausíveis na explicação de um conjunto de dados (Burnham & Anderson 2002). Mas, esses limiares de ΔAIC valor são arbitrários e dependendo das circunstâncias pode levar o pesquisador a descartar modelos plausíveis (Richards 2005).

Vale lembrar que, ao contrário do teste de hipótese tradicional, o resultado final do processo de seleção de modelos pode ser (e em ecologia comumente é) a obtenção de mais de um modelo plausível. Isso pode ser um indicativo de que a coleta de dados não foi apropriada para diferenciar hipóteses alternativas (baixo esforço amostral ou desenhos experimentais ruins) ou que o conjunto de modelos concorrentes não foi o melhor possível (Burnham & Anderson 2002). Mas, se não for esse o caso, a contribuição do pesquisador ao decidir quais modelos realmente fazem sentido. Como dito acima, a teoria e a facilidade de interpretação dos parâmetros de um modelo podem ser usados como critérios de desempate. Por outro lado, como os dados são apenas uma amostra de uma população maior, o bom ajuste de um modelo à amostra não necessariamente se repetirá para todas as amostras dessa população. Assim, no caso de empate, o pesquisador poderá escolher o modelo mais simples (Lindsey 2004). Adicionalmente, a quantidade de dados disponíveis é proporcional o nível de explicação que se pode dar a um fenômeno. Em conjuntos de dados menores apenas os padrões mais fortes serão detectado, enquanto que tamanhos amostrais grandes permitem detectar nuances do processo ou co-variáveis por trás da variação existente nos dados.

Passo 4 – *Interpretação dos modelos selecionados*

Após a seleção dos modelos plausíveis é feita a interpretação do significado biológico do modelo eleito. Seja pelas propriedades do modelo em si, seja pela combinação dos MLEs obtidos, é possível fazer as seguintes perguntas: qual o padrão predominante existente nos dados? Quais variáveis influenciam a geração do padrão observado e qual a importância relativa de cada uma delas? Existe algum mecanismo assumido pelo modelo selecionado que

faz sentido na explicação do fenômeno em estudo? Caso o conjunto de modelos concorrentes tenha sido bem definido dentro da teoria disponível e após percorrer os possíveis processos por trás dos dados, a interpretação do resultado da seleção de modelos poderá ser bastante direta. Além disso, o diagnóstico sempre estará condicionado ao conjunto de modelos concorrentes. Caso boas explicações para o fenômeno estudado tenham sido deixadas de fora, a seleção de modelos ainda escolherá o(s) melhor(es) modelo(s) concorrente(s). Ou seja, a seleção não indicará a existência de explicações melhores, ressaltando que boas inferências sempre dependem da definição de hipóteses bem estabelecidas, o que é uma tarefa exclusiva do pesquisador.

No caso de empates entre dois ou mais modelos plausíveis, o resultado do processo de seleção pode parecer frustrante. Mas considerando que a ciência caminha mais pela exclusão de hipóteses pouco plausíveis do que a seleção da melhor hipótese (Platt 1964), conhecer quais modelos possuem pouco suporte empírico é um importante resultado por si só. Esse conhecimento é a base para o contínuo aperfeiçoamento das hipóteses, que poderão ser testadas em novos estudos ou experimentos. Além disso, não se pode excluir nesses casos a possibilidade de que o empate signifique de fato que os dados tenham sido gerados simultaneamente por processos previstos em ambos os modelos.

1.5.2 Áreas de estudo

As evidências empíricas utilizados durante esta tese são provenientes de diferentes parcelas florestais permanentes. A maior parte dos capítulos utilizou dados do projeto “Diversidade, dinâmica e conservação em Florestas do Estado de São Paulo: 40,96 ha de parcelas permanentes” (Processo no: 99/09635-0 – programa BIOTA/FAPESP), que implantou parcelas permanentes nas quatro principais formações florestais do Estado de São Paulo, Brasil (*sensu* Veloso 1992): Savana Florestada (*i.e.* Cerradão), Floresta Estacional Semidecidual Montana (FESM), Floresta Ombrófila Densa Submontana (FODS) e Floresta Ombrófila Densa de Terras Baixas sobre depósitos marinhos (*i.e.* Floresta Alta de Restinga - FAR). Atualmente, tratam-se dos inventários mais completos já realizados nestas formações florestais no Estado de São Paulo e no Brasil para algumas delas. Apesar da variação em distância e em latitude serem relativamente pequenas, as quatro áreas apresentam ambientes físicos e, conseqüentemente, formações florestais distintas (ver Tabela 1 do Capítulo 2). Estas parcelas foram instaladas em quatro unidades de conservação: Estação Ecológica de Assis (EEA – área total: 1.760 ha), Estação Ecológica de Caetetus (EEC – 2.178 ha), Parque

Estadual da Ilha do Cardoso (PEIC – 13.500 ha), e Parque Estadual Carlos Botelho (PECB – 37.644 ha), respectivamente.

Em cada uma das áreas foi instalada uma parcela permanente de 10,24 ha (320×320 m), subdividida em 256 sub-parcelas de 20×20 m (400 m²). Estas sub-parcelas foram demarcadas com estacas de madeira, enquanto que os vértices da parcela foram demarcados usando tubos de PVC preenchidos com cimento. Dentro destas sub-parcelas, todos os arbustos e árvores com perímetro do fuste à altura do peito (1,3 m) – PAP ≥ 15 cm foram marcados, mapeados, identificados e tiveram seu perímetro anotado. O perímetro de todos os fustes foi medido com precisão de 5 mm e a fórmula da área do círculo foi usada para o cálculo da área basal dos fustes. Foram coletadas mais 5.500 exsicatas para auxiliar a identificação das espécies e ao menos um *voucher* de cada espécie foi incorporado ao herbário ESA. As parcelas foram inventariadas em 2002 e novamente entre 2004 e 2005. Os dados usados nesta tese se referem apenas aos indivíduos encontrados vivos no segundo censo das parcelas. Contudo, o refinamento das identificações botânicas têm sido contínuo desde então, tendo sido realizada a conferência de todos os espécimes coletados e algumas excursões a campo para a conferência de inconsistências e de espécies de identificação duvidosa. Mais informações sobre o ambiente e sobre os procedimentos de escolha e instalação das parcelas podem ser encontrados em Rodrigues (2005).

No capítulo 3.2 foram usados dados provenientes da parcela permanente da Ilha de Barro Colorado (BCI), um trecho de Floresta Estacional Semidecidual medindo 1000×500 m (50 ha), localizada em no Lago Gatun no Panamá (coordenadas: 9° 9'N, 79° 51'W). A temperatura e precipitação médias na ilha são de 31 °C e 2.551 mm y⁻¹, respectivamente, com uma estação seca que se estende entre Dezembro e Abril (Leigh *et al.* 2004). A área da parcela é relativamente plana com altitudes variando entre 120 e 160 acima do nível do mar. A maior parte da parcela é recoberta por uma floresta que não sofre perturbações humanas a pelo menos 400 anos, exceto por área de 2 ha que foi perturbada a 100 anos atrás. A parcela foi montada entre 1981 e 1983 com procedimentos muito parecidos àqueles descritos para as parcelas permanentes do Estado de São Paulo (ver detalhes em Condit 1998). Contudo, em BCI foram incluídos todos os arbustos e árvores com diâmetro à altura do peito – DAP ≥ 10 mm. Além disso, o diâmetro dos fustes nesta parcela foi mensurado com uma precisão de 1 mm. Após 1983, parcela foi re-inventariada em 1985 e depois a cada cinco anos, completando sete avaliações ao longo de 30 anos de monitoramento. A cada novo inventário o número de recrutas e mortes foi anotado e todos os recrutas e sobreviventes tiveram seu DAP mensurado. A parcela tem sido intensivamente estudada e descrita desde os anos 80 com inúmeras

publicações sobre seu ambiente físico, estrutura, dinâmica e composição (para mais detalhes e estudos sobre a parcela de BCI ver: <http://www.ctfs.si.edu/site/Barro+Colorado+Island>).

1.6 Referências bibliográficas

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N. & Csaki, F. (eds) Second International Symposium on information theory, pp. 267-281. Akademiai Kiado, Budapest, HU.
- Anderson, D.R., Burnham, K.P. & Thompson, W.L. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* 64: 912-923.
- Armesto, J.J., Mitchell, J.D. & Villagran, C. 1986. A comparison of spatial patterns of trees in some tropical and temperate forests. *Biotropica* 18, 1-11.
- Ågren, G.I. & Fagerström, T. 1984. Limiting dissimilarity in plants: Randomness prevents exclusion of species with similar competitive abilities. *Oikos* 43, 369-375.
- Batista, J.L.F. 2009. Verossimilhança e máxima verossimilhança. (Apostila de estudos) ESALQ/USP, Piracicaba.
- Baddeley, A & Turner, R. 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12, 1-42.
- Bolker, B.M. 2008. Ecological models and data in R. Princeton University Press, Princeton.
- Bolker, B. 2012. *bbmle*: Tools for general maximum likelihood estimation. Pacote R, versão 1.0.5.2. <http://CRAN.R-project.org/package=bbmle>.
- Burnham, K.P. & Anderson, D.R. 2002. Model selection and multi-model inference: a practical information-theoretic approach. 2^a ed. Springer, New York.
- Callaway, R.M. & Walker, L.R., 1997. Competition and facilitation: A synthetic approach to interactions in plant communities. *Ecology* 78, 1958-1965.
- Chamberlin, T.C., 1890. The method of multiple working hypotheses. *Science* 148, 759.
- Clements, F.E. 1936. Nature and structure of the climax. *Journal of Ecology* 24, 252-284.
- Colwell, R.K., Mao, C.X. & Chang, J. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85: 2717-2727.
- Condit, R. 1995. Research in large, long-term tropical forest plots. *Trends in Ecology and Evolution* 10, 18-22.
- Condit, R. 1998. Tropical Forest Census Plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots. Springer-Verlag, Georgetown.

- Condit, R., Hubbell, S.P., LaFrankie, J.V., Sukumar, R., Manokaran, N., Foster, R.B. & Ashton, P.S. 1996. Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology* 84, 549-562.
- Condit, R., Sukumar, R., Hubbell, S.P. & Foster, R.B. 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. *American Naturalist* 152, 495-509.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Hubbell, S.P., Foster, R.B., Itoh, A., LaFrankie, J.V., Lee, H.S., Losos, E., Manokaran, N., Sukumar, R. & Yamakura, T. 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288, 1414-1418.
- Condit, R. & CTFS Work Group. 2004. Species-area relationships and diversity measures in forest dynamics plots. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp. 79-89. University of Chicago Press, Chicago.
- Connell, J.H. & Slatyer, R.O. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. *American Naturalist* 111, 1119-1144.
- Connell, J.H., Tracey, J.G. & Webb, L.J., 1984. Compensatory recruitment, growth, and mortality as factors maintaining rain forest tree diversity. *Ecological Monographs* 54, 141-164.
- Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. 2005. Community structure of corals and reef fishes at multiple scales. *Science* 309, 1363-1365.
- Coomes, D.A. & Allen, R.B. 2007. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology* 95, 27-40.
- Coomes, D.A., Duncan, R.P., Allen, R.B. & Truscott, J. 2003. Disturbances prevent stem size-density distributions in natural forests from following scaling relationships. *Ecology Letters* 6, 980-989.
- Denslow, J.S., 1980. Patterns of plant species diversity during succession under different disturbance regimes. *Oecologia* 46, 18-21.
- 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. 1992. *Likelihood*. The John Hopkins University Press, Baltimore.
- Engen, S. & Lande, R., 1996. Population dynamic models generating the Lognormal species abundance distribution. *Mathematical Biosciences* 132, 169-183.

- Enquist, B.J., Brown, J.H. & West, G.B., 1998. Allometric scaling of plant energetics and population density. *Nature* 395, 163-165.
- Enquist, B.J. & Niklas, K.J. 2001. Invariant scaling relations across tree-dominated communities. *Nature* 410, 655-660.
- Gilbert, B. & Lechowicz, M.J., 2004. Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences* 101, 7651-7656.
- Goff, F.G. & West, D. 1975. Canopy-understory interaction effects on forest population structure. *Forest Science* 21, 98-108.
- Gotelli, N.J. 2001. Research frontier in null model analysis. *Global Ecology and Biogeography* 10, 337-343.
- Green, J.L. & Plotkin, J.B. 2007. A statistical theory for sampling species abundances. *Ecology Letters* 10, 1037-1045.
- Greig-Smith, P. 1961. Data on pattern within plant communities: I. The analysis of pattern. *Journal of Ecology* 49, 695-702.
- Harms, K.E., Condit, R., Hubbell, S.P. & Foster, R.B. 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* 89, 947-959.
- Harte, J., Conlisk, E., Ostling, A., Green, J.L. & Smith, A.B. 2005. A theory of spatial structure in ecological communities at multiple spatial scales. *Ecological Monographs* 75, 179-197.
- He, F. & Legendre, P. 2002. Species diversity patterns derived from species-area models. *Ecology* 83, 1185-1198.
- He, F., Legendre, P. & LaFrankie, J.V. 1997. Distribution patterns of tree species in a Malaysian tropical rain forest. *Journal of Vegetation Science* 8, 105-114.
- Hilborn, R. & Mangel, M. 1997. *The ecological detective: Confronting models with data* (Monographs in Population Biology 28). Princeton University Press, Princeton.
- Hill, M.O., 1973. The intensity of spatial pattern in plant communities. *Journal of Ecology* 61, 225-235.
- Hubbell, S.P., 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* 203, 1299-1309.
- Hubbell, S.P. 2001. *The unified neutral theory of biodiversity and biogeography*. (Princeton Monographs in Population biology). Princeton University Press, Princeton.
- Hubbell, S.P. 2004. Two decades of research on the BCI Forest Dynamics Plot: where we have been and where we are going. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest*

- diversity and dynamism: Findings from a large-scale plot network, pp.8-30. University of Chicago Press, Chicago.
- Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. & Loo de Lao, S. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a Neotropical forest. *Science* 283, 554-557.
- Hughes, R.G. 1986. Theories and models of species abundance. *American Naturalist* 128, 879-899.
- Johnson, J.B. & Omland, K.S. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19, 101-108.
- Knight, D.H. 1975. A phytosociological analysis of species-rich tropical forest on Barro Colorado Island, Panama. *Ecological Monographs* 45, 259-284.
- Lawton, J.H. 1996. Patterns in ecology. *Oikos* 75: 145-147.
- Leigh Jr., E.G., Lao, S.L., Condit, R., Hubbell, S.P., Foster, R.B. & Perez, R. 2004. Barro Colorado Island Forest Dynamics Plot, Panama. In: E.C. Losos and Leigh Jr., E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp.451-463. University of Chicago Press, Chicago.
- Lepš, J., 1990. Can underlying mechanisms be deduced from observed patterns? In: Krahulec, F., Agnew, A.D.Q., Agnew, S., Willems, J.H. (Eds.), *Spatial processes in plant communities*. Academia, Praga, pp. 1-11.
- Li, L., Huang, Z., Ye, W., Cao, H., Wei, S., Wang, Z., Lian, J., Sun, I.-F., Ma, K. & He, F. 2009. Spatial distributions of tree species in a subtropical forest of China. *Oikos* 118, 495-502.
- Lindsey, J.K. 2004. *Introduction to applied statistics: A modelling approach*. Oxford University Press, Oxford.
- Loehle, C., 1987. Hypothesis testing in ecology: psychological aspects and the importance of theory maturation. *Quarterly Review of Biology* 62, 397-409.
- Lomolino, M.V. 2000. Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography* 27, 17-26.
- Lorimer, C.G. & Frelich, L.E. 1984. A simulation of equilibrium diameter distributions of Sugar Maple (*Acer saccharum*). *Bulletin of the Torrey Botanical Club* 111, 193-199.
- Losos, E.C. & CTFS Work Group. 2004. The structure of tropical forests. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp. 69-78. University of Chicago Press, Chicago.
- MacArthur, R.H. 1960. On the relative abundance of species. *American Naturalist* 94, 25-36.

- MacArthur, R.H. 1969. Patterns of communities in the tropics. *Biological Journal of the Linnean Society* 1, 19-30.
- MacArthur, R.H. 1972. *Geographical ecology: patterns in the distribution of species*. Princeton University Press, Princeton.
- Magurran, A.E. 2004. *Measuring biological diversity*. Blackwell Publishing, Oxford.
- Manly, B.F.J. 2006. *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman & Hall/CRC Press, London.
- May, R.M. 1975. Patterns of species abundance and diversity. In: Cody, M.L., Diamond, J.M. (Eds.), *Ecology and Evolution of Communities*. Harvard University Press, Cambridge, pp. 81-120.
- McGill, B.J. 2003. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecology Letters* 6, 766-773.
- McGill, B.J., 2010. Towards a unification of unified theories of biodiversity. *Ecology Letters* 13, 627-642.
- McGill, B.J. 2011. Measuring the spatial structure of biodiversity. In: Magurran, A.E., McGill, B.J. (Eds.), *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp. 152-171.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10, 995-1015.
- Muller-Landau, H.C., Condit, R., Harms, K.E., Marks, C.O., Thomas, S.C., Bunyavejchewin, S., Chuyong, G., Co, L., Davies, S., Foster, R.B., Gunatilleke, S., Gunatilleke, N., Hart, T.B., Hubbell, S.P., Itoh, A., Kassim, A.R., Kenfack, D., LaFrankie, J.V., Lagunzard, D., Lee, H.S., Losos, E., Makana, J., Ohkubo, T., Samper, C., Sakumar, R., Sun, I.-F., Supardi, N., Tan, S., Thomas, D., Thompson, J., Valencia, R., Vallejo, M.I., Muñoz, G.V., Yamakura, T., Zimmerman, J.K., Dattaraja, H.S., Esufali, S., Hall, P., He, F., Hernández, C., Kiratiprayoon, S., Suresh, H.S., Wills, C. & Ashton, P. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology Letters* 9, 589-602.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. 2012. *vegan: Community Ecology* Pacote R, versão 2.0-4. <http://CRAN.R-project.org/package=vegan>.

- Perry, G.L.W., Miller, B.P. & Enright, N.J., 2006. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecology* 187, 59-82.
- Pielou, E.C. 1975. *Ecological Diversity*. John Wiley and Sons, New York.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Development Core Team 2013. *nlme: Linear and Nonlinear Mixed Effects Models*. Pacote R, versão 3.1-108.
- Platt, J.R.. Strong inference. *Science* 146[3642], 347-353. 1964.
- Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R.B., Hubbell, S.P., LaFrankie, J.V., Manokaran, N., Lee, H.-S., Nowak, M.A., Ashton, P.M. 2000. Predicting species diversity in tropical forests. *Proceedings of the National Academy of Science* 97, 10850-10854.
- Plotkin, J.B. & Muller-Landau, H.C. 2002. Sampling the species composition of a landscape. *Ecology* 83, 3344-3356.
- Prado, P.I. & Miranda, M. 2013. *sads: Modeling Species Abundance Distributions*. Pacote R em desenvolvimento, versão 0.2.06. <https://r-forge.r-project.org/projects/sads>
- Preston, F.W. 1948. The commonness, and rarity, of species. *Ecology* 29, 254-283.
- Richards, S.A. 2005. Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* 86, 2805-2814.
- Rodrigues, R.R. 2005. Parcelas Permanentes em 40ha de florestas do Estado de São Paulo: uma experiência interdisciplinar. URL: <http://www.lerf.esalq.usp.br/parrel2005.php>. ESALQ/USP, Piracicaba.
- Royall, R.M. 1997. *Statistical evidence: a likelihood paradigm*. (Monographs on Statistics and Applied Probability 71). Chapman & Hall, London and New York.
- Royall, R.M. 2004. The likelihood paradigm for statistical evidence. In: Taper, M.L., Lele, R.S. (Eds.), *The nature of scientific evidence: statistical, philosophical, and empirical considerations*. University of Chicago Press, Chicago, pp. 119-152.
- Royle, J.A. & Dorazio, R.M. 2008. *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Academic Press, London.
- Swihart, R.K. Dunning Jr., J.B. & Waser, P.M. Gray matters in ecology: dynamics of pattern, process, and scientific progress *Bulletin of the Ecological Society of America* 83, 149-155.
- Šizling, A.L., Storch, D., Šizlingová, E., Reif, J. & Gaston, K.J., 2009. Species abundance distribution results from a spatial analogy of central limit theorem. *Proceedings of the National Academy of Science* 106, 6691-6695.

- Tokeshi, M. 1999. Species coexistence: ecological and evolutionary perspectives. Blackwell Science, Malden. 239p.
- Valencia, R., Foster, R.B., Villa, G., Condit, R., Svenning, J.C., Hernández, C., Romoleroux, K., Losos, E., Magård, E. & Balslev, H. 2004. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *Journal of Ecology* 92, 214-229.
- Veloso, H.P. 1992. Manual técnico da vegetação brasileira. (Manuais técnicos de geociências n°1). IBGE, Rio de Janeiro. 92p.
- Venables, W.N. & Ripley, B.D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. <http://www.stats.ox.ac.uk/pub/MASS4>.
- Wang, X.G., Wiegand, T., Hao, Z.Q., Li, B.H., Ye, J. & Lin, F. 2010. Species associations in an old-growth temperate forest in north-eastern China. *Journal of Ecology* 98, 674-686.
- Wiegand, T., Gunatilleke, S. & Gunatilleke, N. 2007. Species associations in a heterogeneous Sri Lankan Dipterocarp Forest. *American Naturalist* 170, E77-E95.
- Wright, S.J., Muller-Landau, H.C., Condit, R. & Hubbell, S.P. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84, 3174-3185.

CAPÍTULO 2: COMPARANDO DESCRITORES FLORESTAIS BÁSICOS

Comparing community descriptors in four Neotropical forest plots: when variability also matters

Renato A. Ferreira de Lima, Paulo Inácio K.L. Prado & Ricardo R. Rodrigues

Lima, R.A.F. (Corresponding author, raflima@usp.br): Programa de pós-graduação em Ecologia, Instituto de Biociências – Universidade de São Paulo (USP)

Lima, R.A.F. & Prado, P.I.K.L. (prado@ib.usp.br; <http://ecologia.ib.usp.br/let>): Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, Universidade de São Paulo (USP), Rua do Matão, trav. 14, nº 321, São Paulo, CEP 05508-090, Brazil

Rodrigues, R.R. (rrr@esalq.usp.br; <http://www.lerf.esalq.usp.br>): Departamento de Ciências Biológicas, Universidade de São Paulo (USP), Av. Pádua Dias, nº 11, Caixa Postal 9, Piracicaba, CEP 13018-900, Brazil

Abstract

Aims: account for variability in forest descriptors (i.e. density, basal area, species density, richness, and diversity) to compare four 10.24-ha permanent forest plots along a gradient of climate and soil conditions.

Location: Cerrado and Atlantic Forests, state of São Paulo, southeastern Brazil

Methods: Based on trees >4.8 cm in diameter at breast height surveyed in subplots of 20×20 m, we used model selection by maximum likelihood to select the statistical distributions and mathematical functions that best described the observed distribution of the above-mentioned descriptors and the relationship between them. The distribution of each descriptor was compared among plots based on the likelihood surface of their maximum likelihood estimates of expected mean and variance.

Results: Descriptions of variables showed that the same theoretical distribution was never appropriate for all plots, which usually presented non-normal distributions at the studied scale. The comparison between plots revealed strong differences in their community structure, with the woodland savanna plot having almost twice as many trees than other plots but having almost half of the species found in the rainforest plot. There was evidence that soil and climate conditions may act together to produce these differences. As expected, comparisons of species density and diversity between plots were scale-dependent. For all variables, differences between plots became more evident when we account for their variance.

Conclusions: We conclude that the use of proper statistical description has implications other than purely statistical ones, providing valuable information on meaningful processes generating within-site variability. Besides means, the inclusion of variance in plot comparisons was useful to weight the influence of local and regional process, suggesting that site-specific history as well as regional abiotic factors might act together to shape the structure of these communities.

Keywords: Brazil; environmental gradient; forest structure; maximum likelihood; species richness; variability

Abbreviations: EEA= Estação Ecológica de Assis (Savanna forest); EEC= Estação Ecológica de Caetetus (Seasonal forest); PEIC= Parque Estadual da Ilha do Cardoso (White-sand forest); PECB= Parque Estadual Carlos Botelho (Rainforest)

Running head: Comparing patterns in four Neotropical forests

Introduction

Many community ecologists have dedicated themselves to detect patterns, that is, non-random arrangements in the community (Greig-Smith 1961). Some patterns may be readily distinguishable, such as changes in species distributions along succession or across ecological gradients. But less obvious patterns can only be detected by detailed descriptions (Hill 1973). Quantitative descriptions of forest communities are often used to detect patterns, which are then attributed to biological, environmental, or historical causal factors (He & Legendre 2002; Losos et al. 2004). Even though the same pattern may emerge from different processes, such pattern description generally is the basic step into gaining an empirical understanding of natural communities, being used as indirect tests of ecological hypothesis or even as ground-evidence to formulate new ones (Hubbell 2004). Therefore, the search for repeated patterns to infer on processes is pervasive in the ecological literature (MacArthur 1972; Gilbert & Lechowicz 2004; Freestone & Inouye 2006) and many of the existing hypotheses on forest structure and diversity have arisen from patterns in seed dispersal, seedling distribution, species interactions (e.g. predation), populations and canopy dynamics (see Table 1 in Wright 2002 for several examples).

Early studies in plant ecology have focused on the description of the ‘scale and intensity’ of patterns in basic forest descriptors, such as abundance, biomass and diversity (Greig-Smith 1961; Anderson et al. 1969; Hill 1973). These early studies were gradually replaced by studies carried in large permanent forest plots, which have provided important contributions to the better understanding of forest communities in the past decades. Although there were some critics at first (cf. Hulbert 1984), it is known today that the combination of such large datasets with proper analytical tools allows precise descriptions of communities and how they vary across time (Condit 1995; Herben 1996; Hubbell 2004). So, it is not a coincidence that many of the advances in modern plant ecology come from detailed community descriptions of large permanent forest plots (Hubbell et al. 1999; Harms et al. 2001; Condit et al. 2000, 2004; Plotkin et al. 2000). Besides the simple description of communities using central tendencies and/or diversity indexes (e.g. Losos et al. 2004), data from large permanent plots can also be used to assess patterns of variability in forest descriptors across space. Variability itself may not be an irrelevant source of noise and can be useful to reveal the influence of possible causal factors of community organization (Taylor et al. 1978; Dutilleul & Legendre 1993). In addition to non-normal patterns of variability in community descriptors (Chave et al. 2003), the ways in which means and variances are

related may also have implications other than purely statistical ones (Taylor 1961; Gaston et al. 2000).

Forests occur over a wide range of environmental conditions, so the establishment of forest plots using standardized methods in sites with contrasting environments offers excellent opportunities to assess the role of abiotic factors on community organization. We can ask, for instance, if there are similarities among forests in spite of contrasts in rainfall regime, soil types, among others. If so, we would expect little abiotic control (Greig-Smith 1961). So far, however, there is certain agreement that forest structure and diversity are likely to be determined by both abiotic and historical factors (Ricklefs et al. 1999). Few generalities may be taken from recent studies. Forests growing in wetter and less seasonal climates tend to support higher tree densities and basal area values than forests submitted to drier or more seasonal climates (Phillips et al. 1994; Malhi et al. 2002; Losos et al. 2004). Species richness also tends to be higher in wetter and hotter sites, which generally coincides with low latitude forests (Gentry 1988; Phillips et al. 1994; Gaston 2000). Soil fertility may also influence forest structure and richness (Gentry 1988; Paoli et al. 2008). Large-scale historical events may also play a foremost role in local community diversity (Ricklefs et al. 1999) and composition (Gentry 1988). Finally, life-history traits of the most abundant species may also play a role in community pattern formation (cf. Peh et al. 2011).

The state of São Paulo, Brazil, harbors a remarkable diversity of biome types, driven by a clear climatic gradient from coastal to the inland areas. Along this gradient, different forest occur over various soil conditions. These conditions may thus be used to assess the roles of climate and soil on community patterns, and this is the goal of the present study. We compare tree community patterns in four 10.24-ha permanent forest plots set up in contrasting abiotic conditions. Differently from other studies (e.g. Condit et al. 1996, 2004), these forests are geographically close (max. distance ~350 km), lie in a small latitudinal range (22–25°S), and share many biogeographical affinities. Plots also encompass a wider range of forest types, including well-studied seasonal and rain forests, and poorly known and highly threatened savanna and white-sand forests. As climate and soil vary between plots, we ask the following questions: (i) How similar are the four plots in terms of the shape and expected values of the distribution of variables such as abundance, basal area, species density and diversity per quadrat? (ii) The relationship between number of species and abundance can be described by the same mathematical functions? (iii) Are these similarities maintained when we also take into account the variability in these descriptors? We expect that climate and soil act together to generate differences between sites, so that communities will have higher density, lower

basal area and diversity, including lower inclination in the species-abundance curves, in areas where climate and/or soil are more restrictive. In addition, since we used methods that are similar to other plot-based studies, results can be compared to assess if the suggestions of soil and climate control on tree communities can be generalized extended to other forests in Brazil and abroad.

Material and Methods

Study sites and tree census

We used surveys of the four main forest types of São Paulo State, southeastern Brazil: one type of savanna forest (locally known as ‘*Cerradão*’), semi-deciduous seasonal forest, white-sand forest (locally known as ‘*Restinga*’ forest), and the evergreen rain forest. Four 10.24-ha forest plots were placed in protected areas, namely, ‘Estação Ecológica de Assis’ (EEA – total area: 1,760 ha), ‘Estação Ecológica de Caetetus’ (EEC – 2,178 ha), ‘Parque Estadual da Ilha do Cardoso’ (PEIC – 13,500 ha), and ‘Parque Estadual Carlos Botelho’ (PECB – 37,644 ha). We will refer to these plots as the savanna, seasonal, white-sand and rainforest plots, respectively. Location and site characteristics of each plot are provided in Table 1 (see also Lima et al. 2008, 2011, 2012; Lima & Gandolfi 2009). Each plot has 320×320 m and is subdivided into 256 subplots of 20×20 m (400 m²). All trees, treelets, and tree ferns with stem girth at breast height ≥ 15 cm (dbh ≥ 4.8 cm) were tagged and identified to species. Over 5,500 specimens were collected to help species identification, and at least one voucher specimen was prepared for each species. The girth of all stems was measured to the nearest 5 mm and the sum of their circular cross-sectional area ($A = \text{girth}^2/4\pi$) was used to measure individual basal area. Plots were first surveyed in 2002 and then again in 2004/2005. The results presented here refer to the second census, with several identifications corrections made after that. Plots had different species composition with only five species were common to all plots (Lima et al. unpubl.).

Data analysis

Each plot was described based on tree abundance, basal area, number of species (i.e. species density) and Fisher’s α per subplot. These variables were described using standard statistical distributions whose probability density functions (PDF) are easily found in statistical textbooks (e.g. Bolker 2008), used here as candidate distributions. These distributions were selected based on the nature of the data (type of data) and patterns of symmetry. Data inspection suggested possible bimodal distributions for some variables, so finite mixtures of

two PDFs (e.g. Normal + Lognormal) were added to the set of candidate models (Burnham & Anderson 2002). We used Fisher's α instead of Simpson index as a measure of diversity because the former is directly related to a well-known species abundance distribution model, the Log-series. The Simpson index (D), although being more sample-size independent (Lande et al. 2000), was not appropriate to detect differences in richness between sites (not shown). And since α is strongly biased towards rare species (Magurran 2004), Simpson evenness ($E_{1/D} = (1/D)/\text{species}$) is also presented as a complementary measure of species relative abundances. Preliminary analysis on singletons confirmed its close relation to Fisher's α (Magurran 2004), thus results for singletons were omitted. We also present the results averaged for all 144 possible square hectares (5×5 subplots) that could be fitted inside the 10.24-ha plots.

Since the number of species depends on the number of individuals sampled, species density was also modeled holding tree density as a covariate. In this case, the expected mean was modeled as a function of density, that is, Species $\sim$ Normal ($\mu = f(N)$, σ^2). Linear ($f(N) = a + bN$), logarithmic ($f(N) = a + b \ln(N)$), power-law ($f(N) = aN^b$), negative exponential ($f(N) = a(1 - e^{-bN})$), and Michaelis-Menten/Monod ($f(N) = aN / (b + N)$) functions were used as candidate link functions to model this relationship (Bolker 2008). The first three are always-increasing functions while the last two are saturating. The parameter b can be used as a proxy to species richness, except for the Michaelis-Menten in which a/b may be used to assess richness. No function was forced through the origin so that the parameters could be properly estimated. Whenever needed, variance was also modeled as a power-law function of the number of individuals (i.e. $\sigma^2 = cN^d$). The same approach was used to model basal area and Fisher's α as function of tree density, aiming to detect whether these variables were related at subplot scale. We also used rarefaction curves to compare plot-scale species richness for the same number of individual (Magurran 2004).

Model selection by maximum likelihood, a flexible and intuitive way to describe ecological data (Hilborn & Mangel 1997; Burnham & Anderson 2002; Johnson & Omland 2004), was used to select the distributions that best described the data. The PDFs of the candidate distributions were used to construct -log-likelihood functions (Bolker 2008) and numeric optimization was used to maximize them, that is, to find the combination of parameters that best fit the dataset (i.e. maximum likelihood estimates - MLE). Since candidate distributions have different number of parameters, comparison between them were based on the Akaike's Information Criterion (AIC) that gauges the relative distance of each model to an unknown true model, so that the model closest to this true model will be the one with lowest AIC (Burnham & Anderson 2002). Differences in AIC values (ΔAIC) > 2 were

regarded as indicators of differences between two models fit. Since plots were not always best described by the same PDF, comparisons between plots had to be made based on their two first moments, namely the expected values of mean ($E[x]$) and variance ($\text{var}[x]$). Because mean and variance are monotonic functions of the PDF parameters, the likelihood functions could be parameterized to provide the MLEs for $E[x]$ and $\text{var}[x]$, which are comparable among all candidate distributions. The likelihood profiles of $E[x]$ and $\text{var}[x]$ of the best distribution were used to produce likelihood surfaces and compare plots. If likelihood surfaces overlap, there is no evidence of difference between the MLEs of two different plots.

Although the 20×20 m subplot is a standard and appropriate scale to compare the 10.24-ha plots (He & Legendre 2002), subplot observations are clearly dependent. Spatial correlation matrixes could be added to likelihood functions in order to account for such dependence. As described by Bolker (2008), this can be done by using multivariate distributions to build the likelihood functions (e.g. $f(x) \sim \text{MVNormal}(\mu, C)$, where C is the correlation matrix). However, this procedure proved to be difficult to implement, especially for non-normal multivariate distributions and non-linear relationships (Bolker 2008). Therefore, we assessed the effect of spatial dependence in parameter estimation by using generalized least squares models (Pinheiro & Bates 2000) that also provide likelihood values for model comparison. Models with and without spatial dependence were compared based on their AIC value and whenever the spatial model provided different MLEs, the results from the former were presented. Although for some variables the approximation to a normal distribution is inappropriate, this approach provided a rough assessment of the effect spatial dependence on overall patterns. All analyses were performed in R (www.R-project.org), using the packages MASS (Venables & Ripley 2002), nlme (Pinheiro et al. 2008), bbmle (Bolker 2009), and vegan (Oksanen 2010). The R codes are available upon request from the authors.

Results

Describing patterns

The search for adequate descriptions for the distribution of variables showed that a single theoretical distribution was never appropriate for all plots (Appendix S1). Apart from species density, the normal distribution was a poor description for most variables that presented varying degrees of asymmetry. In the rainforest plot tree density, basal area, and species density were fitted much better by mixture than single models, with differences between single and mixture models being more pronounced for tree and species density (ΔAIC = 24.5, 5.7, and 18.5 respectively; see Appendix S1 for details on these mixtures). Except for basal

area distribution in the seasonal forest, mixtures did not provide better fits for other plots. The selection of functions to relate tree density with other subplot variables can be seen as an exception, since power-laws almost always provided plausible fits in all plots (Appendix S2). The linear function also performed well except for the relationship between species and tree density. It was remarkable how other functions performed well at some plots and poorly at others. Modeling variance as a function of tree density improved model fit for at least one of the relationships at all plots.

Except for basal area at the seasonal and white-sand plots, the inclusion of spatial dependence improved model fit for all variables at all sites. Thus, the variance:mean ratio ≈ 1 of species density in the white-sand plot was not a good indication of spatial randomness. According to the spatial models, most variables had range estimates below 20 m (maximum of 35.4 m), suggesting that the use of 40×40 m subplots would break up spatial dependence. At the savanna and the rainforest plots the inclusion of spatial dependence had an effect in expectation and variance estimation, being this effect more pronounced regarding forest structure (Appendix S3). In general, this spatial dependence tended to widen the intervals around parameter estimates, altering considerably their likelihood surfaces. The interpretation of the results, however, changed only for species density, in which the inclusion of spatial dependence resulted in the overlap in $E[x]$ between the savanna and rainforest. Regardless of the different strengths of spatial dependence between plots (Oda-Souza et al. 2010), the use of spatial models has not changed the general patterns. For instance, even when we account for spatial dependence in the data, there were greater variance estimates in the rainforest plot for virtually all variables.

Comparing patterns

Overall results indicated substantial differences between plots with the savanna plot having more than one third of total abundance, but also presenting the smaller total basal area. The rainforest plot showed the opposite trend (Table 2). The comparison of variables per subplot confirmed this suggestion, but it also revealed tendencies different than expect based solely on overall results. Although having the smallest and highest density estimates of $E[x]$, the savanna and rainforest plots were at least two fold more variable than others. Differences in $E[x]$ for basal area were expressed in a different fashion. The great variability of this descriptor in the rainforest (range: 1.7–93.4 m² ha⁻¹) resulted in $E[x]$ estimates not different from the seasonal and white-sand forests (Figure 1), although overlap with the seasonal forest

was little. As for tree density, even when expected values of basal area per subplot were similar between plots, they were all different in terms of variance.

The rainforest plot had nearly twice as many species as the savanna and white-sand plots. But its variability in the number of species per subplot generated similar mean values between these plots at this scale (Figure 2). The seasonal forest, owner of the second largest overall richness, had the lowest species density at this scale. Overall results for Fishers' α confirmed the greater diversity of the rainforest. The confidence intervals around Fishers' α overlapped between the three other plots, although overlap between the savanna and seasonal plots was small. At subplot scale, the inland plots had mean Fishers' α estimates similar to each other, while the white-sand plot had relatively higher estimates. Estimates of mean Fishers' α per subplot were much more variable in coastal plots as well. Differently of α , Simpson evenness was greater at the white-sand plot and smaller at the seasonal plot where a single species represented almost 40% of all trees (Table 2). Plots with higher richness tended to have lower evenness, while inland plots tended to have higher proportions of rare species.

Species density modeled as functions of tree density resulted in the rainforest having the highest richness (i.e. power-law parameter b), followed by the seasonal, white-sand and savanna forests. At this scale, species richness in the seasonal was not distinguishable from the white-sand forest. Species richness comparison based on much larger samples (i.e. rarefied samples of 11,618, 12,043, and 15,248 trees), however, confirmed the overall pattern in species density: Savanna= White-sand< Seasonal< Rainforest. In general, the relation to tree density was much stronger for species density than for basal area (Figure 3). The species-tree density relationship in the rainforest was almost linear ($b \approx 1$), while there was no evidence of any relationship for basal area in the seasonal plot ($b \approx 0$). For other plots, the relationship was very weak although definitely positive. There was a tendency of smaller samples having smaller estimates of Fisher's σ in the rainforest. This dependence is broken when samples with $N < 42$ are taken out of the analysis. For other plots this dependence was non-existent or negligible.

Discussion

Describing patterns

We found that the forest descriptors tested here are usually better described by asymmetrical distributions, and that these distributions may change among sites. This has implications other than providing proper statistical descriptions. If we assume, for instance, that variation across samples (in our case, across space) is linked to resource or niche heterogeneity (Gaston et al.

2000), the existence of asymmetrical distributions of forest descriptors are therefore signatures of non-random process. Therefore, the decision to transform data to obtain normality, a common approach to remove asymmetry before performing statistical tests, may have the side effect of losing useful information on the relationship between forest descriptors and possible causal factors (Taylor et al. 1978). As found by Chave et al. (2003) for above ground biomass, the use of subplots larger than 20×20 m probably would lead to more symmetrical distributions. As predicted by the Central Limit Theorem, this tendency is related to the sum of many random variables. But the merge of subplots also leads to a loss of resolution (Dutilleul & Legendre 1993), limiting the designation of possible interactions of community descriptors with small-scale biotic or abiotic heterogeneity. This would be particularly true for the search of biotic-related mechanisms of community organization, since the interaction between trees occur predominantly at scales smaller than 20 m.

We found that the joint analysis of mean and variance highlighted the differences between plots, meaning that their differences go beyond central tendencies. So, looking also at variance in forest descriptors may help to tell apart processes generating differences between forests. In the rainforest plot, contour maps (not shown) suggested that variability may be related to topographic heterogeneity, and to bamboo-dominated stands. A 1962 aerial photo of the savanna plot suggested a link between forest structure variability and within-plot differences in the disturbance history. The rainfall regimes and major soil types of plots, on the other hand, may act at broader scales generating tendencies that are common to forests of the same type (Walter 1986; Gentry 1988; Gaston 2000). Therefore, we hypothesize that variability is more linked with local processes while central tendencies can be related to more regional ones. This may be a more general rule with examples in the studies plots. Hence, under the assumption that mean and variance reflect processes at different scales, the explicit inclusion of variability in forest description may provide insights into the interplay between local and regional processes shaping the structure of communities (Hunter & Price 1992; Ricklefs et al. 1999; Crawley & Harral 2001).

The better fit of mixture models for some variables also provided useful indications of within-plot patterns. In the rainforest plot these mixtures efficiently capture the influence of bamboo-dominated stands in tree density, and the estimated mixing proportion (0.20) was not different from the real proportion of this stands inside the plot (0.19, R.A.F. Lima et al., unpubl.). For basal area and species density such estimated proportions were smaller (both 0.10), suggesting that the influence of bamboo-stands is less pronounced for these variables. In the seasonal plot, although data inspection did not suggest any evident bimodal pattern,

basal area was fitted better by a mixture model as well. Interestingly, the estimated mixing proportion (0.31) was close to the plot proportion covered by canopy gaps (0.36, Lima et al. 2008). However, the results indicated higher basal area for gap-dominated plots, a result that is not intuitive since the presence of gaps denotes the absence of canopy trees, which contribute most to forest basal area (Chave et al. 2003). Maybe soil moisture or nutrient conditions may offer an alternative variation in such pattern in basal area. Although both suggestions have to be tested directly, what is important is that the mixture model can reveal within-plot differences that could not be detected by simple inspection of the data or that would not be properly described by finite models.

Comparing patterns

The four forests studied differed markedly in tree density, basal area, species density, and diversity. Even when we consider only the expected means, the differences between plots was the rule. At both scales studied here (i.e. plot and subplot scales), the savanna plot had the highest tree density and lowest basal area that probably resulted from a larger proportion of small trees. In fact, there was no individual > 60 cm dbh in this plot. Forest structure estimates for the savanna plot were different than other 13 forest plots around the world (Losos et al. 2004). Since these and other researchers (Malhi et al. 2002) found that less seasonal climates tend to have higher tree densities, it seems that there is an abiotic threshold beyond which this trend is reversed, which may be related to the joint effect of seasonality and low water retention capacity in the savanna plot soil. An alternative or additional explanation would be the successional status of the plot, which was partially disturbed in the beginning of the last century. The other three plots had comparable values to these 13 plots, except for basal area at the seasonal forest that was relatively smaller. Life-history traits of the dominant species may have influenced the results in this plot (Peh et al. 2011). The elevated dominance of the small-sized *Metrodorea nigra* (38% of all trees) probably generated such smaller basal area estimates. In addition, its dominance certainly caused the low diversity in this plot. Another possible explanation would be that the higher rate of wind-related disturbance at this plot. Besides the wind-throw of canopy trees that leads to considerable reductions basal area, such higher disturbance rate may have favored some species, including *M. nigra*, in detriment of others (Martini et al. 2008), changing plot evenness. In addition, there are evidences of allelopathic effects going on in this plot (Maluf et al. 2010) that may have some extra effect on species diversity at small scales.

As plots are at similar latitudes, the higher basal area of coastal plots confirmed that the rainfall regime does influence basal area (Losos et al. 2004) and most probably biomass stocks, assuming that basal area and biomass are positively related (Chave et al. 2003). And this influence seems strong because there is more photosynthetically active radiation reaching inland than coastal plots (Table 1). This suggestion was supported by the higher basal area of white-sand compared to the savanna plots. Both forests occur over restrictive soils, but it seems that higher rainfall can be compensating for soil limitations (cf. Malhi et al. 2002). At a lower level, within region trends in basal area indicated that soils can also influence mean estimates and how basal area it is distributed among size classes (Uhl & Murphy 1981; Clark & Clark 2000; Paoli et al. 2008). Preliminary inspection of plots dbh distribution indicated that plots over less favorable soils had higher density of smaller individuals. If so, the dbh cutoff adopted here may be excluding a greater proportion of the trees in such plots. But since small trees generally make small contributions to total basal area (Chave et al. 2003), underestimation were probably slight. Underestimation might have occurred at PECB and EEC since woody-bamboos and woody lianas were not included in the tree census (Vieira et al. 2008).

The discrepancy between the results at subplot and plot scales revealed that species density and diversity comparison were both scale dependent. At subplot scale, similarities in species density between the savanna, white-sand and rainforests was due to the packing of trees (Denslow 1995) that generally occurs in communities where there are less large-sized trees (Crawley & Haral 2001). The ‘smaller than expected’ number of species per subplot in the seasonal and rainforest plots suggests that they probably have higher beta-diversities than plots over more restrictive soils. These differences in sample-size dependence between plots suggest that plots’ species accumulation curves would probably cross (Lande et al. 2000; Magurran 2004). Differences in sample-size dependence can also be triggered by differences in species relative abundances and aggregation between plots (He & Legendre 2002; Colwell et al. 2004). Differences in species aggregation may be the cause why of Fisher’s α ranked plots differently at different scales. This reinforce that α is not always sample independent (Condit et al. 2004) and suggest that plots may not be well described by Fisher’s log-series. Therefore, the diversity patterns found would have been different if sample size was smaller and most probably if it was larger.

Acknowledgements

We would like to thank the ‘Instituto Florestal de São Paulo’ and all the people responsible for the data collection at the four sites: A.A. Oliveira, A.R. Duarte, A.P. Savassi-Coutinho, M.T.Z. Toniato, D. Sampaio, G.A.D.C. Franco, G. Durigan, N.M. Ivanauskas, T.E. Barreto, and V.C. Souza. Financial support was provided by FAPESP (process no. 99/09635-0 and 09/53413-5). R.A.F.L. was supported by CAPES (Demanda Social program), and P.I.K.L.P. by CNPq (Bolsa Produtividade 303878/2008-8). We also thank A.A. Oliveira, A.M.Z. Martini, J.L.F. Batista, and J. Chave for their valuable comments on earlier version of the manuscript.

References

- Anderson, D.J., Jacobs, S.W.L. & Malik, A.R. 1969. Studies on structure in plant communities. VI. The significance of pattern evaluation in some Australian Dry-land vegetation types. *Australian Journal of Botany* 17: 315-322.
- Bolker, B.M. 2008. *Ecological models and data in R*. Princeton University Press, Princeton, USA.
- Bolker, B.M. 2009. *bbmle*: Tools for general maximum likelihood estimation. R package version 0.9.3. URL: <http://CRAN.R-project.org/package=bbmle>.
- Burnham, K.P. & Anderson, D.R. 2002. *Model selection and multi-model inference: a practical information-theoretic approach*. 2nd ed. Springer, New York, USA.
- Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B. & Hubbell, S. P. 2003. Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *Journal of Ecology* 91: 240-252.
- Clark, D.B. & Clark, D.A. 2000. Landscape-scale variation in forest structure and biomass in a tropical rain forest. *Forest Ecology and Management* 137: 185-198.
- Colwell, R.K., Mao, C.X. & Chang, J. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85: 2717-2727.
- Condit, R. 1995. Research in large, long-term tropical forest plots. *Trends in Ecology and Evolution* 10: 18-22.
- Condit, R., Hubbell, S.P., LaFrankie, J.V., Sukumar, R., Manokaran, N., Foster, R.B. & Ashton, P.S. 1996. Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology* 84: 549-562.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Hubbell, S.P., Foster, R.B., Itoh, A., LaFrankie, J.V., Lee, H.S., Losos, E.,

- Manokaran, N., Sukumar, R. & Yamakura, T. 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288: 1414-1418.
- Condit, R. & CTFS Work Group. 2004. Species-area relationships and diversity measures in forest dynamics plots. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp. 79-89. University of Chicago Press, Chicago, USA.
- Crawley, M.J. & Harral, J.E. 2001. Scale dependence in plant biodiversity. *Science* 291: 864-868.
- Denslow, J.S. 1995. Disturbance and diversity in tropical rain forests: The density effect. *Ecological Applications* 5: 962-968.
- Dutilleul, P. & Legendre, P. 1993. Spatial heterogeneity against heteroscedasticity: an ecological paradigm versus a statistical concept. *Oikos* 66: 152-171.
- Freestone, A.L. & Inouye, B.D. 2006. Dispersal limitation and environmental heterogeneity shape scale-dependent diversity patterns in plant communities. *Ecology* 87: 2425-2432.
- Gaston, K.J. 2000. Global patterns in biodiversity. *Nature* 405: 220-227.
- Gaston, K.J., Blackburn, T. M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. 2000. Abundance-occupancy relationships. *Journal of Applied Ecology* 37: 39-59.
- Gentry, A.H. 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Annals of the Missouri Botanical Garden* 75: 1-34.
- Gilbert, B. & Lechowicz, M.J. 2004. Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences* 101: 7651-7656.
- Greig-Smith, P. 1961. Data on pattern within plant communities: I. The analysis of pattern. *Journal of Ecology* 49: 695-702.
- Harms, K.E., Condit, R., Hubbell, S.P. & Foster, R.B. 2001. Habitat associations of trees and shrubs in a 50-ha Neotropical forest plot. *Journal of Ecology* 89: 947-959.
- He, F. & Legendre, P. 2002. Species diversity patterns derived from species-area models. *Ecology* 83: 1185-1198.
- Herben, T. 1996. Permanent plots as tools for plant community ecology. *Journal of Vegetation Science*: 195-202.
- Hilborn, R. & Mangel, M. 1997. *The ecological detective: Confronting models with data*. Princeton University Press, Princeton, USA.
- Hill, M.O. 1973. The intensity of spatial pattern in plant communities. *Journal of Ecology* 61: 225-235.

- Hubbell, S.P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, USA.
- Hubbell, S.P. 2004 Two decades of research on the BCI Forest Dynamics Plot: where we have been and where we are going. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp.8-30. University of Chicago Press, Chicago, USA.
- Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. & Loo de Lao, S. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a Neotropical forest. *Science* 283: 554-557.
- Hunter, M.D. & Price, P.W. 1992. Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology* 73: 724-732.
- Hurlbert, S.H. 1984 Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187-211.
- Johnson, J.B. & Omland, K.S. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19: 101-108.
- Lande, R., DeVries, P.J. & Walla, T.R. 2000. When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos* 89: 601-605.
- Lima, R.A.F. & Gandolfi, S. 2009. Structure of the herb stratum under different light regimes in the Submontane Atlantic Rain Forest. *Brazilian Journal of Biology* 69: 631-637.
- Lima, R.A.F., Martini, A.M.Z., Gandolfi, S. & Rodrigues, R.R. 2008. Repeated disturbances and canopy disturbance regime in a tropical semi-deciduous forest. *Journal of Tropical Ecology* 24: 85-93.
- Lima, R.A.F., Oliveira, A.A., Martini, A.M.Z., Sampaio, D., Souza, V.C. & Rodrigues, R.R. 2011. Structure, diversity, and spatial patterns in a permanent plot of high Restinga forest in Southeastern Brazil. *Acta Botanica Brasilica* 25(3): 633-645.
- Lima, R.A.F., Rother, D.C., Muler, A.E. Lepsch, I.F & Rodrigues, R.R. 2012. Bamboo overabundance alters forest structure and dynamics in the Atlantic Forest hotspot. *Biological Conservation* 147: 32-29.
- Losos, E.C. & CTFS Work Group. 2004. The structure of tropical forests. In: Losos, E.C. & Leigh Jr. E.G. (eds) *Tropical forest diversity and dynamism: Findings from a large-scale plot network*, pp. 69-78. University of Chicago Press, Chicago, USA.
- MacArthur, R.H. 1972. *Geographical ecology: patterns in the distribution of species*. Princeton University Press, Princeton, USA.
- Magurran, A.E. 2004. *Measuring Biological Diversity*. Blackwell, Malden, USA.

- Malhi, Y., Phillips, O.L., Lloyd, J., Baker, T.R., Wright, J., Almeida, S., Arroyo, L., Frederickson, T.S., Grace, J., Higuchi, N., Killeen, T., Laurance, W.F., Leão, C., Lewis, S., Meir, P., Monteagudo, A., Neill, D.A., Núñez-Vargas, P., Panfil, S. N., Patiño, S., Pitman, N., Quseada, C. A., Rudas-Ll, A., Salomão, R., Saleska, S., Silva, N., Silveira, M., Sombroek, W.G., Valencia, R., Vásquez Martínez, R., Vieira, I.C.G. & Vinceti, B. 2002. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *Journal of Vegetation Science* 13: 439-450.
- Martini, A.M.Z., Lima, R.A.F., Franco, G.A.D.C. & Rodrigues, R.R. 2008. The need for full inventories of tree modes of disturbance to improve forest dynamics comprehension: An example from a Semideciduous forest in Brazil. *Forest Ecology and Management* 255: 1479-1488.
- Oda-Souza, M., Batista, J.L.F., Ribeiro Jr., P.J., & Rodrigues, R.R. 2010. Comparação das estruturas de continuidade espacial em quatro formações florestais do estado de São Paulo. *Floresta* 40: 515-522.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. 2010. *vegan*: Community Ecology Package. R package version 1.17-2. <http://CRAN.R-project.org/package=vegan>
- Paoli, G.D., Curran, L.M. & Slik, J.W.F. 2008. Soil nutrients affect spatial patterns of aboveground biomass and emergent tree density in southwestern Borneo. *Oecologia* 155: 287-299.
- Peh, K. S.-H., Lewis, S. L. & Lloyd, J. 2011. Mechanisms of monodominance in diverse tropical tree-dominated systems. *Journal of Ecology* 99: 891-898.
- Phillips, O. L., Hall, P., Gentry, A. H., Sawyer, S. A. & Vazquez, R. 1994. Dynamics and species richness of tropical rain forests. *Proceedings of the National Academy of Science* 91: 2805-2809.
- Pinheiro, J., Bates, D., 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York, USA.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core team (2008). *nlme*: Linear and nonlinear mixed effects models. R package version 3.1-90. URL: <http://cran.r-project.org/web/packages/nlme/index.html>.
- Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R.B., Hubbell, S.P., LaFrankie, J.V., Manokaran, N., Seng, L.H., Nowak, M.A. & Ashton, P.M. 2000. Predicting species diversity in tropical forests. *Proceedings of the National Academy of Science* 97: 10850-10854.

- Ricklefs, R.E., Latham, R.E. & Qian, H. 1999. Global patterns of tree species richness in moist forests: distinguishing ecological influences and historical contingency. *Oikos* 86: 369-373.
- Rodrigues, R.R., 2005. *Parcelas Permanentes em 40ha de florestas do Estado de São Paulo: uma experiência interdisciplinar*. URL: <http://www.lerf.esalq.usp.br/parrel2005.php>. ESALQ/USP, Piracicaba.
- Souza, F.M., Gandolfi, S., Perez, S.C.J.G.A. & Rodrigues, R.R. 2010. Allelopathic potential of bark and leaves of *Esenbeckia leiocarpa* Engl. (Rutaceae). *Acta Botanica Brasilica* 24: 169-174.
- Taylor, L.R. 1961. Aggregation, variance and the mean. *Nature* 189:732–735.
- Taylor, L.R., Woiwod, I.P. & Perry, J.N. 1978. The Density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology* 47: 383-406.
- Uhl, C. & Murphy, P.G. 1981. Composition, structure, and regeneration of a tierra firme forest in the Amazon basin of Venezuela. *Tropical Ecology* 22: 219-237.
- Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Fourth Edition. Springer, New York.
- Vieira, S.A., Alves, L.F., Aidar, M., Araújo, L.S., Baker, T., Batista, J.L.F., Campos, M.C., Camargo, P.B., Chave, J., Delitti, W.B.C., Higuchi, N., Honório, E., Joly, C.A., Keller, M., Martinelli, L.A., Mattos, E.A., Metzker, T., Phillips, O.L., Santos, F.A.M., Shimabukuro, M.T., Silveira, M. & Trumbore, S.E. 2008. Estimation of biomass and carbon stocks: the case of the Atlantic Forest. *Biota Neotropica* 8: 21-29.
- Walter, H. 1986. Vegetação e zonas climáticas: tratado de ecologia global. EPU, São Paulo.
- Wright, S.J. 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia* 130: 1-14.

Supporting Information

Appendix S1. Summary of the model-selection of tree density, basal area, species density and Fisher's α .

Appendix S2. Summary of the model-selection of the relationships between basal area, species density and Fisher's α with tree density at each plot.

Appendix S3. Results of the comparison between models with and without spatial dependence between subplot observations.

Tables

Table 1. Location and description of the four plots studied here. Values of temperature, rainfall, air humidity are annual averages obtained near the plots between 2003 and 2005. Rainfall in parentheses, number of dry months and water surplus refer to long-term averages obtained at the nearest cities, and may not reflect real climate at each plot, specially the rainforest plot. Soil types with minor expression inside plots are not shown. PAR= photosynthetically active radiation. Source: Rodrigues (2005).

	Savanna (EEA plot)	Seasonal (EEC plot)	White-sand (PEIC plot)	Rainforest (PECB plot)
Large region	Inland	Inland	Coastal	Coastal
Plot coordinates	22°35'S 50°22'W	22°24'S 49°42'W	25°04'S 47°55'W	24°10'S 47°55'W
Altitude (m)	540 – 560	570 – 610	3 – 8	350 – 450
Temperature (°C)	21.0	20.9	21.4	20.9
Rain (mm)	865 (1255)	1169 (1303)	2521 (2261)	3082 (1582)
Dry months*	5	5	1	3
Rain 3 driest months	145	122	260	203
Air humidity (%)	72.2	74.1	84.4	89.1
Water surplus (mm)	165	278	1123	518
PAR (MJ m ⁻² d ⁻¹)	6.5	6.5	4.4	2.5
Main soil types	Distrophic Oxisols	Eutrophic Ultisols	Distrophic Psamments, and Spodosols	Distrophic Inceptisols, and Neosols
Soil hydraulic conductivity (Kfs)	High/medium; (10 ⁻⁶ –10 ⁻⁴)	Medium/low (10 ⁻⁸ –0)	High/medium (10 ⁻⁷ –10 ⁻⁵)	Medium/low (10 ⁻⁸ –0)

* Months whose 95% confidence interval of monthly rainfall is lower than 100 mm.

Table 2. Summary of the forest descriptors evaluated in four 10.24-ha forest plots. The averages were calculated for all 144 possible square hectares that could be fitted inside plots. Dominances per hectare were the same as the overall value and thus are not presented. For overall Fisher's alpha, the analytical standard deviation was used to estimate the 95% confidence intervals (CI). Legend: $E_{1/D}$ = Simpson evenness; Dominance= maximum relative abundance; Rare species are defined as those with less than one tree per hectare.

Descriptor	Savanna (EEA plot)	Seasonal (EEC plot)	White-sand (PEIC plot)	Rainforest (PECB plot)
Trees ha ⁻¹	2147 ± 252	1174 ± 50	1484 ± 123	1206 ± 211
Basal area ha ⁻¹	21.8 ± 1.5	26.1 ± 1.9	28.3 ± 1.9	31.3 ± 7.0
Species ha ⁻¹	61.1 ± 4.9	78.3 ± 5.7	79.6 ± 6.2	145.7 ± 14.4
Fisher's α ha ⁻¹	11.8 ± 1.3	18.9 ± 1.8	18.0 ± 1.7	43.6 ± 3.6
$E_{1/D}$ ha ⁻¹	0.22 ± 0.04	0.08 ± 0.02	0.19 ± 0.04	0.11 ± 0.02
Dominance*	0.183	0.384	0.199	0.224
Fisher's α	16.4	23.6	17.2	39.2
Fisher's α CI	13.4–19.8	19.7–28.0	14.1–20.9	33.8–45.1
Rare species (%)	45.8	50.3	27.4	36.8

* Refers to the relative abundance of *Copaifera langsdorfii* (Fabaceae), *Metrodorea nigra* (Rutaceae), and *Euterpe edulis* (×2 – Arecaceae) at EEA, EEC, PEIC, and PECB plots, respectively.

Figures

Figure 1. Probability distributions that best fitted tree density, and basal area per 20×20 m subplots (left) and their respective likelihood surfaces of mean and variance (right). In likelihood surface plots, the cross represents the maximum likelihood estimates (MLE) of the expected mean and its variance for each plot, while the ellipse around MLEs delimit the combinations of mean and variance that are equally plausible, that is, result in $\Delta\text{AIC} \leq \ln(8)$. Since confidence regions were delimited based on spatial model estimates, they may look different (generally wider) than what would be expected based solely on models fit without spatial structure. For clarity, variance estimates are presented in terms of their square roots. See Supporting Information for the results on models selection for each variable.

Figure 2. Probability distributions that best fitted species density, and Fisher's α per 20×20 m subplots (left) and their respective likelihood surfaces of mean and variance (right). See caption of Figure 1 for details.

Figure 3. Basal area (top), species (middle) and Fisher's α (bottom) as functions of abundance per 20×20-m subplot. Lines are the predicted mean estimated by maximum likelihood using Linear (basal area and Fisher's α) and Power-law functions (species). Legends stand for the maximum likelihood estimates (MLE) and likelihood intervals of the

inclination (Linear) and exponent (Power-law) parameter b . See Supporting Information for the results of the model selection on these relationships. Note the log-scale for Fisher's α .

Supporting Information

Appendix S1. Summary of the model-selection to describe the distribution of tree density (N), basal area (BA), species density and Fisher's α at each plot. Legend: NB= Negative binomial; k = number of parameters; Δ_i = difference in AIC from the best model; w_i = Akaike weights. Various mixtures were placed among the set of candidate models but only the mixture with best support from the data is shown. The models used to build these mixtures are presented only when they were plausible descriptions. When mixture results are not presented, optimization fail to converge. In bold, models that were equally plausible ($\Delta_i \leq \ln(8)$) to describe variables at each plot.

Variable	PDF	k	Savanna		Seasonal		White-sand		Rainforest	
			Δ_i	w_i	Δ_i	w_i	Δ_i	w_i	Δ_i	w_i
N	Poisson	1	443	0.000	45.1	0.000	69.2	0.000	1469	0.000
	NB	2	10.7	0.003	2.8	0.127	0.0	0.404	60.7	0.000
	Normal	2	0.0	0.515	11.2	0.002	3.0	0.090	24.5	0.000
	Lognormal	2	27.7	0.000	0.0	0.505	2.7	0.105	118.4	0.000
	Gamma	2	13.6	0.000	0.8	0.339	0.2	0.368	66.0	0.000
	Weibull	2	0.3	0.438	34.9	0.000	19.2	0.000	27.2	0.000
	Mixture	5	4.9	0.044	5.9	0.027	5.0	0.033	0.0*	1
BA	Normal	2	0.6	0.270	123	0.000	21.5	0.000	53.9	0.000
	Lognormal	2	4.5	0.037	2.3	0.242	4.4	0.059	63.0	0.000
	Gamma	2	0.0	0.357	23.7	0.000	0.5	0.400	14.8	0.000
	Weibull	2	20.4	0.000	84.4	0.000	39.7	0.000	5.7	0.055
	Hyperbolic	3	0.4	0.286	72.9	0.000	0.0	0.521	23.3	0.000
	Mixture	5	3.9	0.050	0.0§	0.758	6.5	0.020	0.0†	0.945
Species	Poisson	1	49.8	0.000	638	0.000	0.0	0.568	605	0.000
	NB	2	61.0	0.000	569	0.000	4.6	0.058	44.4	0.000
	Normal	2	0.0	0.860	0.0	0.789	2.9	0.131	20.6	0.000
	Lognormal	2	9.8	0.007	13.1	0.001	8.8	0.007	94.8	0.000
	Gamma	2	4.5	0.092	4.7	0.075	3.8	0.086	51.2	0.000
	Weibull	2	10.0	0.006	5.0	0.065	9.4	0.005	18.5	0.000
	Mixture	5	6.4	0.035	4.9	0.070	2.7	0.145	0.0‡	1
α	Normal	2	19.5	0.000	14.6	0.000	8.3	0.012	52.0	0.000
	Lognormal	2	0.0	0.752	8.0	0.016	9.7	0.006	29.0	0.000
	Gamma	2	2.2	0.248	0.0	0.089	0.0	0.747	0.9	0.375
	Weibull	2	35.3	0.000	11.3	0.003	4.9	0.065	5.3	0.041
	Mixture	5	-	-	4.7	0.086	3.0	0.170	0.0§	0.584

* Lognormal + Negative binomial (mixing proportion, $p = 0.20$ [0.14–0.26])

§ Lognormal + Lognormal ($p_{\text{ecc}} = 0.69$ [0.50–0.84]; $p_{\text{pcb}} = 0.50$ [0.37–0.62])

† Lognormal + Gamma ($p = 0.10$ [0.06–0.15])

‡ Normal + Normal ($p = 0.10$ [0.06–0.15])

Appendix S2. Summary of the model-selection to describe the relationship of basal area (BA), species density and Fisher's α with tree density (N) at each plot. Legend: Neg. exp.= Negative exponential; MM= Michaelis-Menten; k = number of parameters (a , b , σ^2); Δ_i = difference in AIC from the best model; w_i = Akaike weights. In cases when modeling variance as a power-law function of density ($\sigma^2 = cN^d$) resulted in a better fit ($\Delta_i > \ln(8)$), the results of this better fit are presented instead, followed by '*'. In this case, $k = 4$. In bold, functions equally plausible to describe the relationship at each plot. See text for functions details.

Variables	Function	k	Savanna		Seasonal		White-sand		Rainforest	
			Δ_i	w_i	Δ_i	w_i	Δ_i	w_i	Δ_i	w_i
$BA \sim N$	Linear	3	0.0	0.566	0.3	0.192*	2.3	0.121	0.0	0.694*
	Logarithmic	3	3.4	0.103	0.2	0.202*	0.5	0.285	36.6	0.000*
	Power-law	3	1.5	0.272	0.5	0.173*	1.1	0.219	1.7	0.301*
	Neg. exp.	3	14.5	0.000	0.0	0.228*	12.3	0.000	10.2	0.005*
	MM	3	4.5	0.059	0.2	0.205*	0.0	0.374	14.8	0.000*
Species $\sim N$	Linear	3	4.9	0.028	3.5	0.051	0.0	0.295*	0.0	0.753*
	Logarithmic	3	0.9	0.202	0.0	0.294	0.2	0.263*	151	0.000*
	Power-law	3	1.9	0.126	1.6	0.129	1.0	0.183*	2.3	0.237*
	Neg. exp.	3	0.0	0.323	0.1	0.274	1.5	0.137*	17.2	0.000*
	MM	3	0.0	0.321	0.3	0.252	1.8	0.122*	8.7	0.010*
$\alpha \sim N$	Linear	3	0.1	0.300*	1.1	0.175	0.0	0.240	0.0	0.448
	Logarithmic	3	0.0	0.315*	0.3	0.261	0.6	0.182	9.5	0.004
	Power-law	3	0.0	0.316*	0.5	0.238	0.1	0.227	0.8	0.306
	Neg. exp.	3	14.8	0.000*	5.6	0.018	1.0	0.142	2.7	0.116
	MM	3	3.0	0.069*	0.0	0.307	0.3	0.208	2.5	0.126

Appendix S3. Maximum likelihood estimates (MLE) of mean ($E[x]$) and variance ($\text{var}[x]$), and comparison between models with and without spatial dependence between subplot observations. The results of mean and variance for the non-spatial models were obtained based on the theoretical distributions that best fitted the observations at each plot (see Appendix S1), except for the cases when the mixture model was the best fit. In these cases, we presented the MLE for the second best model. Ranges right of the MLEs are their likelihood intervals.

Variable	Plot	AIC	AIC _{sp}	$E[x]$	$E[x]_{\text{sp}}$	$\sqrt{\text{var}[x]}$	$\sqrt{\text{var}[x]_{\text{sp}}}$	range (m)
N	EEA	2225.7	2079.2	85.7 [83.4–88.0]	81.9 [72.9–90.9]	18.6 [17.1–20.3]	19.4 [15.8–23.9]	28 [23–62]
	EEC	1855.3	1849.5	47.1 [45.9–48.2]	47.0 [45.5–48.5]	8.9 [8.2–9.8]	9.0 [8.2–9.8]	10 [7–14]
	PEIC	1948.1	1909.7	59.6 [58.3–60.9]	59.6 [57.1–62.1]	10.8 [9.9–11.8]	10.7 [9.7–11.9]	16 [12–21]
	PECB	2234.8	2086.4	45.4 [43.1–4.7]	43.1 [35.0–51.3]	18.8 [17.4–20.6]	18.9 [15.9–22.4]	35 [23–52]
Basal area	EEA	-205.1	-219.3	.883 [.863–.903]	.883 [.852–.912]	.161 [.148–.176]	.161 [.146–.176]	12 [8–16]
	EEC	350.4	351.8	1.03 [.974–1.08]	1.030 [.966–1.093]	.423 [.377–.481]	.476 [.436–.519]	6 [3–13]
	PEIC	150.5	150.0	1.15 [1.11–1.19]	1.147 [1.102–1.192]	.316 [.290–.351]	.322 [.295–.351]	7 [4–12]
	PECB	536.7	540.6	1.19 [1.09–1.29]	1.159 [0.994–1.324]	.819 [.536–.857]	.734 [.664–.811]	15 [11–20]
Species	EEA	1339.8	1311.7	21.9 [21.5–22.3]	21.7 [21.0–22.5]	3.3 [3.0–3.6]	3.3 [3.0–3.6]	15 [11–20]
	EEC	1375.3	1350.2	16.3 [15.8– 16.7]	16.2 [15.4–16.9]	3.52 [3.2–3.9]	3.5 [3.2–3.9]	14 [10–19]
	PEIC	1538.5	1458.2	24.1 [23.5–24.8]	24.4 [22.9–25.8]	4.9 [4.8–5.0]	4.8 [4.2–5.4]	22 [16–30]
	PECB	1936.9	1802.0	25.1 [23.8– 26.4]	23.82 [19.7–28.0]	10.5 [9.7–11.5]	10.44 [8.9–12.2]	31 [21–45]
Fisher's α	EEA	1156.2	1119.5	9.9 [9.6–10.2]	9.8 [9.2–10.4]	2.3 [2.1–2.6]	2.3;2.1–2.6]	17 [13–23]
	EEC	1304.0	1291.6	9.3 [8.9–9.7]	9.2 [8.5–9.8]	3.2 [2.9–3.5]	3.2 [2.9–3.5]	14 [10–19]
	PEIC	1578.0	1534.4	16.0 [15.3–16.7]	16.2 [14.8–17.6]	5.5 [5.0–6.0]	5.3 [4.7–5.9]	18 [13–24]
	PECB	2040.2	2058.5	25.7 [24.0–27.5]	21.2 [22.1–28.3]	14.5 [13.0–16.2]	14.1 [12.8–15.6]	15 [11–19]

CAPÍTULO 3: DISTRIBUIÇÃO DE ABUNDÂNCIA DE ESPÉCIES

Disentangling the effect of sampling scale and size in species abundance distributions

Renato A. Ferreira de Lima,^{1,2*} Ricardo R. Rodrigues³ and Paulo I. Prado²

¹ *Programa de pós-graduação em Ecologia, Instituto de Biociências – Universidade de São Paulo (USP), São Paulo, SP, Brazil*

² *Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, Instituto de Biociências – Universidade de São Paulo (USP), Rua do Matão, trav. 14, 321, São Paulo, SP, CEP 05508-090, Brazil*

³ *Departamento de Ciências Biológicas, Escola Superior de Agricultura ‘Luiz de Queiroz’ – Universidade de São Paulo (USP), Av. Pádua Dias, 11, Piracicaba, SP, CEP 13018-900 P.O. Box 9, Brazil*

E-mail: raflima@usp.br (RAFL), rrresalq@usp.br (RRR), prado@ib.usp.br (PIP)

Running title: Sample scale and size in species abundance distributions

Statement of authorship: RRR was responsible for the data collection; RAFL and PIP wrote the R codes; RAFL analysed the data; RAFL and PIP wrote the draft of the manuscript, and RRR contributed substantially to revisions.

Type of article: Letters (with 149 words in the abstract, 4025 in the main text, two figures and no table).

Correspondence: Rua do Matão, trav. 14, 321, São Paulo, SP, CEP 05508-090, Brazil.
Telephone: +55 11 30910969. Fax +55 11 30918096

(Draft of the manuscript to be submitted to Ecology Letters)

ABSTRACT

Species abundance distributions (SAD) are characterized by the abundance of rare and scarcity of abundant species. Recently, it has been suggested that scale-related mechanisms are important in shaping SADs. However, these suggestions did not consider that the scale effect on the SAD has an indirect effect of sample size, which increases as samples of different scales are combined. Here, we tried to separate the effect of scale and sample size on tree SAD, by comparing observed with rarefied SADs generated using the same species relative abundances but controlling for community size. We found that the scale effect on SADs is predominantly an indirect effect of sample size. There were, however, significant deviations between observed and rarefied SADs, especially in the proportion of rare species. Thus, although there may be a scale effect, this effect is minor and dependable on the species turnover between samples.

Keywords: Atlantic Forest, biomass, Brazil, community ecology, rarefaction, sample size effects, species abundance distributions, trees, veil line.

INTRODUCTION

The species abundance distribution (SAD) is one of the simplest yet powerful ways to describe and compare samples of ecological communities (Pielou 1975; Alonso *et al.* 2008). Because species abundances are generally regarded as good proxies of species performance (Preston 1962; Tokeshi 1999), the analysis of SADs is long considered as a useful tool to assess community organization and diversity (May 1975; Hubbell 2001; Engen & Lande 1996; McGill *et al.* 2007). The repeated description of such distributions in tropical and temperate communities revealed one of the most intriguing and general patterns in ecology: communities have few common and many rare species (Preston 1948; MacArthur 1960; McGill *et al.* 2007; Kempton & Taylor 1974). Since species in the community are sampled based on their frequency and not on their sizes, this pattern is more conspicuous when using counts of individuals to describe the SAD (Williamson 2010). However, patterns are still asymmetric when using biomass as measures of abundance (Connolly *et al.* 2005; Morlon *et al.* 2009).

Because of its remarkable generality, many ecologists tried to explain the almost universal steeply declining shape of SADs and what would be the main processes generating it. These explanations can be separated into three groups. One group relies on how species share niches or resources, assuming that the share received by each species is proportional to

their abundance in the community (MacArthur 1960; Tokeshi 1999). A second group relies on stochastic population dynamics which would be created by random differences in the balance between density-dependent vital rates of species through time (Routledge 1980; Hughes 1986; Hubbell 2001; Engen & Lande 1996). A third group provides explanations that have more statistical rather than biological grounds (Preston 1948, 1962; McGill 2003). These authors have shown that the shape of SAD is fundamentally affected by sample size, although species aggregation and detectability may also play a role in shaping the SAD in samples (Green & Plotkin 2007; Alonso *et al.* 2008). It was shown that increasing sample sizes would lead to the disclosure of a mode in the SAD, meaning that singletons are no longer the largest abundance class in the community, a process known as the "veil line" effect (Preston 1948). Despite of the long-standing debate of how well this effect defines the shape of SAD (Dewdney 1998; Chisholm 2007 and citations therein), it is known for a long time now that sample size affects the shape of the SAD.

More recently, scale-related mechanisms have been proposed to explain the shape of SADs (Connolly *et al.* 2005; Morlon *et al.* 2009; Šizling *et al.* 2009). In particular, Connolly *et al.* (2005) working with corals and reef fishes suggested that the shape of SADs depends on the scale of study considered: more asymmetrical and amodal shapes would be found at smaller scales of sampling while more symmetrical and modal shapes would be found at larger ones. Years later, Morlon *et al.* (2009) claimed that this same phenomenon also applied for birds, trees and island fishes. Using different arguments, Šizling *et al.* (2009) suggested that the steeply declining shape of SADs arises from a statistical process associated with increasing sampling scale, a process that would be defined mainly by the degree of spatial aggregation and turnover of species in the community. However, these attempts to explain the shape of SADs based on sample scale have not taken into account sample size effects. In these studies, larger sampling scales are obtained by pooling samples from smaller scales together, which also leads to an increase in sample size. So, the scale effect claimed by these authors may simply be an indirect effect of increasing sample size, predicted much earlier by Preston (1948, 1962).

The goal of this study is to disentangle the role of sampling scale and size in shaping species abundance distributions. Based on Preston (1948, 1962), our basic and more parsimonious prediction is that sample size has a greater influence on the shape of SADs than scale. Based on similar principles of existing frameworks that take into account sampling intensity in the shape of SADs (Bulmer 1974; Pielou 1975; Dewdney 1998; Green & Plotkin 2007), we generated random draws that preserve the properties of observed communities at

each scale but have progressively smaller sample sizes, in a process similar to rarefaction. These re-sampling procedure controls for possible effects of sample size while increasing the scale of observation of the community (Gotelli & Colwell 2001). More specifically, our prediction would be supported if the shape of observed SAD at lower scales matches the shape of the SAD from the random “rarefied” samples of higher-scale communities. Aiming to assess the generality of our prediction, we assess these predictions using tree species, and SADs constructed using counts of individuals and biomass as abundance metrics. As a result, we found that the change in shape of SAD due to increasing scale is predominantly an indirect effect of sample size, a result that was quite similar to both measures of abundance. There were only minor deviations between observed and rarefied SADs which may be attributed to a sample scale effect especially in respect to the proportion of rare species.

MATERIALS AND METHODS

Study sites and sampling

Trees were surveyed using 10.24-ha plots placed in four tropical forest types of São Paulo State, southeastern Brazil, namely the savanna, seasonal, white-sand, and rain forests. The two first sites are inland and 75 km apart from each other. These two inland sites are ca. 300 km apart from the other two coastal plots that are 100 km apart from each other. Placed inside different conservation units, each 10.24-ha plot is subdivided into 256 subplots of 20×20 m. However, we used 40×40-m subplots as sample units (64 samples per 10.24-ha plot) to avoid possible spatial correlation between samples, which ranged up to 30 m for basic descriptors such as number of trees, species, and basal area (R.A.F. Lima, unpublished data) and to avoid samples with low number of species. Inside each sample unit all trees, palms, and tree ferns with stem girth at breast height (1.3 m) $\geq$ 15 cm were tagged, measured and identified to species. The girth of all tree stems was measured to the nearest 5 mm and the sum of their biomass was used as a measure individual tree aboveground biomass, that was calculated using the allometric equations based on diameter at breast height (dbh) and wood specific gravity provided in Chave *et al.* (2005) for moist forests. For species other than trees, we used the formula presented by Wendling (1998) for *Euterpe edulis* and by Tiepolo *et al.* (2002) for tree ferns. The results presented here refer to the second census of the plots conducted in the summer of 2004/2005. At the end of this census there was 21 933, 121 05, 15 276, and 11 618 alive individuals at the savanna, seasonal, white-sand, and rain forest plots belonging to 118, 147, 117 and 224 species, respectively. Plots have low species overlap with only five species were common to all plots (Lima *et al.* unpublished data).

Study design and data analysis

SADs were produced using two measures of abundance – number of individuals and biomass, at four different spatial scales: sample, habitat (within-site), site, and regional scales. To be specific, at sample scale one SAD was obtained for each sample unit (40×40-m subplots) while at site scale one SAD was obtained by pooling all sample units from a single site. Habitat scale was defined by the existence of marked environmental differences within sites, such as topographic positions and soil types. Although each site had different types and number of habitats and that there was environmental variability within them, these habitats are in fact clusters of sample units representing type of environments that can be unequivocally identified in the field. At each spatial scale, SADs were represented using histograms of the number of species in abundances classes in log base 2, the Preston octaves. However, differently than Preston (1948) we used true octaves that double in abundance between successive abundance classes and are closed on the right. To allow the comparison of patterns between sites with different number of species we used the fraction of observed species instead of the total number of species in each octave. Whenever a Preston plot is presented graphically, the SAD represents the mean of the observed distributions at each scale (e.g. the SAD at sample scale represents the mean of 256 abundance distributions).

To test whether the unveiling process of the modal SAD is a matter of sample size and not of sample scale, we used a rarefaction approach (Gotelli & Colwell 2001). From the observed SAD at scales higher than sample scale (i.e. habitat, site and regional scales) we generated 10,000 random draws which had the same properties of the observed SADs (i.e. size of the species pool and relative abundances) but smaller sample sizes (Appendix 1). These sample sizes were equivalent to the median sample sizes observed at site, habitat, and sample scales. For instance, the observed median sample sizes were 250 individuals and this was the sample sizes used to generate the random draws.

Although random draws from observed SAD at regional scale could be generated right away, the draws from SADs at site and habitat scales were generated in a nested way. First, we generated separate draws based on the community properties of each individual site (four sites) or habitat (15 habitats). These separate draws were generated preserving the relative abundances of each site or habitat, but using the same sample size for all sites or habitats. Again, this sample sizes were defined based on the median sample sizes observed at site and habitat scales, which corresponded to 600 and 3000. At a second moment, we averaged each separate draw to produce a single mean draw at each scale, which was the same procedure

used to obtain the mean observed SAD at each scale (see above). Thus, each of the 10,000 random draw actually represents the mean of separate draws generated from individual sites or habitats. For both approaches (non-nested and nested draws) we generated Preston plots in the same way as described for the observed samples except that the plots for the rarefied samples had to be truncated at zero. For the 10,000 random draws we were able to obtain the 95% sample quantiles of the proportion of species at each octave and these quartiles were used for comparison between observed and rarefied SADs. If the mean octaves of the observed SAD were within the range defined by these 95% quantiles, we considered that there was no difference between observed and rarefied SADs.

Although independent sampling is reasonable for some groups of organisms, the same is not true for trees sampled using plots. Since the great majority of tree species are spatially aggregated in the scales considered here (Condit *et al.* 2000), trees sampled using a plot-based design will have more individuals of a given species than what would be expected at spatially random sampling (Hughes 1986). Thus, to generate the random community draws described in the last paragraph we used the Negative Binomial (NB) distribution. This distribution has a long tradition in ecological literature to simulate data under the assumption of spatial aggregation of conspecific individuals (Pielou 1975; Dewdney 1998; Green & Plotkin 2007; Brown *et al.* 1995; Alonso *et al.* 2008). The use of the NB adds complexity to the analysis because of the additional aggregation parameter κ . The estimation of this parameter is scale dependent and is generally related to population size –species with larger populations tend to have more random distributions (higher κ parameter) and vice-versa (Taylor *et al.* 1978). Moreover, there were not enough sites to confidently estimate the parameter κ at regional scale, and for rare species this was commonly the case even at lower scales. Therefore, we assigned an arbitrary value of κ for rare species, which was equal to the community average of the parameter at the given scale. At regional scale, we assumed that the available abundance values for each species followed a NB distribution and thus we randomly sampled abundance values from the empirical data to create a random negative binomial regional SAD. More details on the estimation of κ are given in Appendix 1.

Since the NB is a discrete distribution, it cannot be used directly to assess changes in the species biomass distributions. Thus, the generation of random draws of species biomass distributions was carried in two steps. The first step was identical to the NB draws described above used to obtain random number of individuals per species. Then, for each species with n individuals we sampled with replacement n values of biomass from the empirical distribution of biomass of the same species at the same scale of observation. This procedure was easy to

implement but it may have generated some unrealistic situations for rare species – since these species have few biomass observations, sampling with replacement may have generated unrealistic draws with identical biomass values. However, we believe that these unrealistic draws have not changed the overall outcomes of the simulation. We could have tried to fit continuous distribution to empirical biomass distributions and sample values from this fitted distributions but this would add more complexity to the simulation related to the selection of a particular statistical distribution and to parameter estimation using such few observations at some scales or from some species. All analyses were performed in R (R Core Team 2012), using the contributed R package *sads* (Prado & Miranda 2013).

Another way to test if the change in the shape of SAD is only a matter of sample size would be to use the analytical framework that takes into account the effect of sampling intensity to model SADs (Bulmer 1974; Green & Plotkin 2007). Based on sampling theory, this framework has explicit predictions of how the shape of sampled SAD should change according to the sampling intensity ($N_{\text{sample}}/N_{\text{metacommunity}}$), also referred to as sampling ratio (Dewdney 1998), which has a direct influence in the shape of the SAD. However, we decided for our approach for the following reasons: (i) there is no need to assume a given underlying SAD model, such as a Lognormal or Gamma distributions, since all the information needed for generating the draws comes from the observed SADs themselves; (ii) for the same reason, there is no need to assume any size of this unknown metacommunity; (iii) although the use compound models of SADs was recently facilitated by implementation of specific software solutions (e.g. Poisson Lognormal – Engen *et al.* 2002), there is no such ease for the NB sampling and trying to fit it using numerical integration usually leads to an impracticable amount of optimization errors (but see Foster & Warton 2007); and (iv) both approaches suffer from the same limitations when dealing with species biomass instead of counts of individuals. It should be notice, however, that both approaches are quite similar in their basic principles, in particular the existence of a sampling intensity ‘control’ on the shape of the distribution at each scale.

RESULTS

We first generated the mean observed SAD at the four scales considered here (regional, site, habitat and sample scales). There was a clear trend of change in the shape of the distribution when we use counts of individuals as a measure of abundance: averaged SADs became more symmetrical when we increase the scale of observation, with the emergence of one or more modes in the distributions at higher scales (Fig. 1, leftmost column). For SADs constructed

using biomass instead of counts of individuals, this trend was less clear but there was still a tendency of more symmetrical shapes at higher scales (Fig. 2, leftmost column). The observed biomass distributions were already modal even at the smallest scale considered (40×40 m subplots) and there was a slight trend of a change from a positive to a negative skewness with increasing scale.

We then generated random draws based on the community properties at each scale but with sample sizes that matched the median sample sizes at lower scales. The result was basically the same for both measures of abundance and revealed that random draws with smaller sample sizes were able to reproduce SADs with shapes very similar to the observed (Figs. 1 and 2, second to fourth columns). Although the overall shape was recovered with the control of sample size, a closer look at the overlaid empirical and rarefied SADs revealed some deviations (Fig. S1). Rarefied SADs showed a small trend of an underestimation in the octaves with greater species abundances and in particular for counts of individuals. In addition, for rarefied SADs obtained by using the smallest sample size there was a clear overestimation of the number of singletons. Thus, the greatest deviations of rarefied SADs were found when using counts of individuals and the smallest sample size. The tendencies found for count of individuals were in general the same as found for biomass; however the overestimation of tree species in the first octave did not happen when the measure of abundance was biomass. Moreover, although we found deviations between observed and rarefied SADs, we were not able to detect any clear pattern of deviation in respect to sample size or scale of origin (regional, site or habitat) of the simulated communities (Appendix 2).

DISCUSSION

The process of pooling samples to generate SADs at progressively higher scales did change the shape of the SAD, which became more symmetrical and modal at larger scales, as previously found for corals, fishes, trees and birds (Connolly *et al.* 2005; Morlon *et al.* 2009). However, when we decouple the effect of sample size, the effect of scale is much less evident or nonexistent. Although there were differences between SADs constructed using counts of individuals and biomass, when accounting for sample size we were able to reconstruct shapes very similar to the ones observed at lower scales for both measures of abundance. Even when using the regional community as a reference – the scale at which the distributions were more modal, the rarefied SADs built using the sample scale community size resulted in markedly asymmetric (inverse-J shaped) distributions. Thus, the change in shape of SADs with increasing scale should be properly decomposed in terms of sampling scale and sample size

(see example below). Indeed, Preston (1948) once showed that a modal SAD can arise when samples with increasing sizes are taken from the same community, what he called the ‘veil line’ effect. Although his proposal was later proved to be an approximation (Dewdney 1998; see Chisholm 2007 for a reconciliation between Preston and Dewdney views on the ‘veil line’), it is known for decades that the sample size greatly affects the shape of the SAD (Green & Plotkin 2007).

Although the sample size effect was preponderant in this study to unveil the SAD, there were few deviations in the shape of the SAD that could not be attributed to sample size or sampling scheme (random or aggregated – Green & Plotkin 2007) and suggest a true scale effect. This effect seems to produce less asymmetrical SADs, that is, when we enhance sample scale and keep sample size fixed the tendency is having less rare and more common species in the community than the expected based solely on sample size. This was more evident for the case of counts of tree species. The amount of sample scale effect will probably depend on the amount of similarity in species composition and relative abundances between sample units (Šizling *et al.* 2009), which are both beta diversity components. If the samples are taken from communities having similar species composition and abundances (low beta diversity), when we pool samples together the shape of the SAD will tend to be more symmetric even if we control for sample size. In contrast, one can predict a smaller scale effect if samples are taken from areas with higher beta diversity. In Appendix 3 we provide a simple simulation to exemplify this suggestion. In this study there was a low degree tree species overlap between sites, which may explain the small scale effect observed. Moreover, we may expect different scale effects when species shared are common in both communities or if they are common in one and rare in the other. Future studies comparing scenarios with different degrees of overlap in species composition and abundance are welcome to refine the understanding of the spatial scale effect and how it interacts with the existent spatially-implicit hypothesis that try to explain the shape of SADs based on species-range distributions (McGill & Collins 2003; Brown *et al.* 1995) or on the existence of species with different strategies (e.g. core vs. occasional species – Magurran & Henderson 2003).

In addition, Connolly *et al.* (2005), Morlon *et al.* (2009), and Ulrich *et al.* (2010) found that Preston plots built using biomass instead of individuals were much more modal, a phenomenon termed ‘differential veiling’ by Williamson (2010). We found a similar result, which we attribute to the fact that biomass distributions are essentially the distribution of sums of biomass values of each species. If we take these values as random variables, the existence of a modal shape of biomass SAD can be explained by the Central Limit Theorem

(CLT) which predicts that the sum of independent, identically distributed random variables will tend to a normal distribution (Johnson *et al.* 1994). Assuming that species distributions of biomass are independent seems reasonable, but since these distributions are probably not identical among all species (nor have the same number of observations per species) this may be more the case of a weak variant of the CLT that does not require variables to be identically distributed (Johnson *et al.* 1994). If this is the case, the biomass SAD will also tend to a more modal distribution (in $\log_2$ scale) but perhaps more slowly or with some deviations from normality. Note that this is a different explanation than the one given by Williamson (2010), which relies on differential consequences of sampling by individuals or by biomass. In practice, a modal biomass SAD means that species have lower chances of falling in the lower or higher octaves. For instance, the chance of observing a species in the lowest octave of biomass will be the chance of observing only one individual times the chance that this individual has a biomass value equal or very close to the minimum observed (i.e., singletons will not necessarily fall in the lowest octave of biomass). Thus, the predominant modal shape of the biomass SAD is probably due more to a statistical property than to a sample size or scale effect.

Although the modality of biomass SADs already at sample scale can be attributed to the CLT, the same application of the theorem does not hold for counts of individuals. In fact, SADs at sample scale were markedly asymmetric and even at regional scale they were still not normal. Some authors have made use of the CLT to explain why SADs based on counts of individuals on log scale would also tend to a normal distribution (May 1975; Šizling *et al.* 2009), an explanation that is commonly based on stochastic variation in population size over time or space (Hughes 1986; Engen & Lande 1996, Foster & Warton 2007; Connolly *et al.* 2005). Other authors criticized this view (Pielou 1975; Routledge 1980; Williamson & Gaston 2005) mainly because very large sample sizes still present SADs with deviations from normality, generally a log-left-skew (McGill 2003; Nee *et al.* 1991). While we want to avoid this long-standing debate on which is the best underlying model of SAD, it seems that the change in its shape with increasing scale is more a statistical than biological issue, to which the CLT may offer a good but not exclusive explanation (Connolly *et al.* 2005; McGill *et al.* 2007). For instance, it has been proposed that when we combine samples from a range of environments and communities a modal distribution is expected (Routledge 1980 and citations therein). Therefore, only when using complete enumerations on large-scale sampling schemes we will be able to assess the true underlying shape of the SAD (McGill 2003; Ulrich *et al.* 2010), if there is one.

ACKNOWLEDGMENTS

Tree data collection was supported by FAPESP (process no. 99/09635-0). RAFL was supported by a scholarship from CAPES (Demanda Social program), and PIP was supported by CNPq (no. 303878/2008-8). This manuscript was conceived and partially written during a workshop funded by FAPESP (process no. 09/53413-5).

REFERENCES

- Alonso, D., Ostling, A. & Etienne, R.S. (2008). The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.*, 11, 93-105.
- Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995). Spatial variation in abundance. *Ecology*, 76, 2028-2043.
- Bulmer, M.G. (1974). On fitting the Poisson Lognormal distribution to species-abundance data. *Biometrics*, 30, 101-110.
- Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D. *et al.* (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145, 87-99.
- Chisholm, R.A. (2007). Sampling species abundance distributions: Resolving the veil-line debate. *J. Theor. Biol.*, 247, 600-607.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N. *et al.* (2000). Spatial patterns in the distribution of tropical tree species. *Science*, 288, 1414-1418.
- Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005). Community structure of corals and reef fishes at multiple scales. *Science*, 309, 1363-1365.
- Dewdney, A.K. (1998). A general theory of the sampling process with applications to the "veil line". *Theor. Popul. Biol.*, 54, 294-302.
- Engen, S. & Lande, R. (1996). Population dynamic models generating the Lognormal species abundance distribution. *Math. Biosci.*, 132, 169-183.
- Engen, S., Lande, R., Walla, T.R. & DeVries, P.J. (2002). Analyzing spatial structure of communities using the two-dimensional Poisson Lognormal species abundance model. *Am. Nat.*, 160, 60-73.

- Foster, M.A. & Warton, D.I. (2007). A metacommunity-scale comparison of species-abundance distribution models for plant communities of eastern Australia. *Ecography*, 30, 449-458.
- Gotelli, N.J. & Colwell, R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.*, 4, 379-391
- Green, J.L. & Plotkin, J.B. (2007). A statistical theory for sampling species abundances. *Ecol. Lett.*, 10, 1037-1045.
- Hubbell, S.P. (2001). The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton.
- Hughes, R.G. (1986). Theories and models of species abundance. *Am. Nat.*, 128, 879-899.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions* (vol. 1). John Wiley & Sons, New York.
- Kempton, R.A. & Taylor, L.R. (1974). Log-series and Log-normal parameters as diversity discriminants for the Lepidoptera. *J. Anim. Ecol.*, 43, 381-399.
- MacArthur, R.H. (1960). On the relative abundance of species. *Am. Nat.*, 94, 25-36.
- Magurran, A.E. & Henderson, P.A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422, 714-716.
- May, R.M. (1975). Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities* (eds. Cody, M.L. & Diamond, J.M.). Harvard University Press, Cambridge, pp. 81-120.
- McGill, B.J. (2003). Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecol. Lett.*, 6, 766-773.
- McGill, B.J. & Collins, C. (2003). A unified theory for macroecology based on spatial patterns of abundance. *Evol. Ecol. Res.*, 5, 469-492.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995-1015.
- Morlon, H., White, E.P., Etienne, R.S., Green, J.L., Ostling, A., Alonso, D. *et al.* (2009). Taking species abundance distributions beyond individuals. *Ecol. Lett.*, 12, 488-501.
- Nee, S., Harvey, P.H. & May, R.M. (1991) Lifting the veil on abundance patterns. *P. Roy. Soc. Lond. B. Bio.*, 243, 161-163.
- Pielou, E.C. (1975). *Ecological Diversity*. John Wiley and Sons, New York.
- Prado, P.I. & Miranda, M.D. (2013). *sads*: Modeling species abundance distributions. R package version 0.2.06. Available at: <https://r-forge.r-project.org/projects/sads>

- Preston, F.W. (1948). The commonness, and rarity, of species. *Ecology*, 29, 254-283.
- Preston, F.W. (1962). The canonical distribution of commonness and rarity: part I. *Ecology*, 43, 185-215.
- R Core Team. (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>.
- Routledge, R.D. (1980). The form of species-abundance distributions. *J. Theor. Biol.*, 82, 547-558.
- Šizling, A.L., Storch, D., Šizlingová, E., Reif, J. & Gaston, K.J. (2009). Species abundance distribution results from a spatial analogy of central limit theorem. *P. Natl. Acad. Sci. USA*, 106, 6691-6695.
- Taylor, L.R., Woivod, I.P. & Perry, J.N. (1978). The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.*, 47, 383-406.
- Tiepolo, G., Calmon, M. & Feretti, A.R. (2002). Measuring and monitoring carbon stocks at the Guaraqueçaba Climate Action Project, Paraná, Brazil. In: *International Symposium on Forest Carbon Sequestration and Monitoring*. Taiwan Forestry Research Institute, Taiwan. pp. 98–115.
- Tokeshi, M. (1999). *Species coexistence: ecological and evolutionary perspectives*. Blackwell, Oxford.
- Ulrich, W., Olrik, M. & Ugland, K.I. (2010). A meta-analysis of species–abundance distributions. *Oikos*, 119, 1149-115.
- Wendling, J.L.G. (1998). *Modelos matemáticos de crescimento e de produção em biomassa para árvores de Euterpe edulis Mart. plantadas a céu aberto*. M.Sc dissertation, Universidade Federal do Paraná, Curitiba.
- Williamson, M. (2010). Why do species abundance distributions of individuals and of biomass behave differently under sampling? *Oikos*, 119, 1697-1699.
- Williamson, M. & Gaston, K.J. (2005). The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J. Anim. Ecol.*, 74, 409-422.

Figures

Figure 1. Preston plots of tree communities in different scales and sample sizes using counts of individuals. The leftmost column is the mean observed SAD at each scale of observation. The other columns are the mean distributions obtained by generating random draws with decreasing samples sizes but based on the same community properties (rarefied SADs) for each reference scale of observation (rows). In the figure, scale decreases downwards, whereas sample size decreases rightwards. Sample sizes N_{site} , N_{habitat} and N_{sample} were defined based on the mean observed sample size of the correspondent scale. The x-axis represents the abundance octaves in log base 2 which are closed to the right – first octave contains species with less than 2^1 individuals, second octave contains species with $\geq 2^1$ and $< 2^2$ individuals, and so on. Vertical bars represent the 95% sample quantiles of the proportion of species in each octave.

Figure 2. Preston plots of tree communities in different scales and sample sizes using biomass. See caption of Fig. 1 for details. The x -axis represents the abundance octaves in log base 2 which are closed to the right – first octave contains species with biomass less than $<2^{-6}$ kg, second octave contains species with $\geq 2^{-6}$ and $<2^{-5}$ kg, and so on.

SUPPORTING INFORMATION

Appendix 1. Methods used to generate the random draws based using the observed community properties from the Poisson and the Negative Binomial distributions. Codes are given in language R (<http://www.r-project.org>).

Under the assumption of spatial randomness of species, generating draws to simulate community samples with decreasing sample sizes is quite straightforward, especially when dealing with counts of individuals. Suppose that we have an original community based on which we want to generate random samples and let this community be represented by a vector of species abundances V . Since each element of V is simply the abundance of each species present in this community, the length of V is the community richness while its sum is the community size. The Poisson distribution is often used to mimic a sampling process where there is an assumption of spatial randomness. Therefore, the generation of random draws with sample intensity a based on an arbitrary community can be performed using the following R codes:

```
# Creating an arbitrary vector of abundances V, with 20 species and 300
#individuals:
V <- c(76,5,30,3,1,16,9,1,2,42,13,14,1,1,8,53,3,18,3,1)
S <- length(V)
N <- sum(V)
# Setting sample size to one third of V, i.e., 100 individuals:
a <- 100/N
# Poisson draws with the same community properties but smaller size:
v <- rpois (n= S, lambda=V*a)
```

Note that v has on average the same relative abundances of V but the sum of the vector v is on average a times smaller than N . In addition, v has species with abundance zero, meaning that the species is present in the pool of the original community but was not ‘sampled’. Thus, the species richness of v will be smaller than V (14.5 ± 1.5 species in the example). If we replicate this procedure many times we can establish what would be the mean shape of the species abundance distribution (SAD) if the original community had the same community properties (i.e. species richness and relative abundances) but a sample size a times smaller.

If species are aggregated, we can use the Negative Binomial (NB) distribution to assess changes in the shape of the SAD as a function of sample size. Differently than the Poisson distribution, which has only one parameter, the NB has an additional overdispersion parameter κ that measures the amount of species aggregation (Bolker 2008). Thus, to generate

random draws from the BN distribution we use the same vector of species abundances V but we have to provide values of species aggregation as well:

```
# Creating arbitrary values of k for each species
k <- c(0.5,0.15,1,0.1,0.1,0.15,0.1,0.1,0.2,1.5,0.3,0.1,0.05,
       0.05,0.8,0.5,0.1,0.9,0.1,0.05)
# NB draws with the same community properties but smaller size:
v <- rnbinom (n= S, mu=V*a, size=k)
```

Once again, note that sum of the vector v is on average a times smaller than N and it also has species with abundance zero. But since species aggregation decreases the expected number of species in samples (Hughes 1986, Colwell *et al.* 2004), the mean species richness of v generated using the NB distribution (8.5 ± 1.7 species) is smaller if compared to the Poisson distribution.

In the analytical framework used in the manuscript, the vector V was defined based on the observed communities of each reference scale (regional, site, and habitat scales) and a was defined based on the median sample size n of each target scale (site, habitat, and sample scales). For instance, when the regional-scale was used as a reference V was defined as a vector containing species from all sites and their overall abundances, and the values of a for the target scales were defined as $a_{\text{site}} = n_{\text{site}}/N$, $a_{\text{habitat}} = n_{\text{habitat}}/N$, and $a_{\text{sample}} = n_{\text{sample}}/N$. Since the higher scale of observation is the regional one, than N in the manuscript represent the total of individuals sampled at this scale (60932 trees).

To perform the NB draws, the values of the overdispersion parameter κ for each species and each scale were estimated based on the empirical data using maximum likelihood techniques. Estimates of κ for species with few individuals generally fail to converge and for those species we assigned an arbitrary value of κ that was equal to the community average of the parameter. The only exception was the estimation of κ at site scale that had only four observations. In this special case, although the number of observations per species was low, we assumed that they followed a BN distribution and thus we randomly sampled abundance values from the empirical data to create a random negative binomial regional SAD. This procedure guarantees that the scale of observation is preserved (the abundance of each species can come from any site) and that on average the size of the community is n_{site}/N times smaller than N .

This same procedure was used to generate random draws of species biomass distribution but with an additional step. After generating the draw v , the abundance of species was used to conduct a sample with replacement of the same number of observation from the

empirical distribution of biomass of the same species at the same scale. If the species abundance was zero, a biomass value equals zero was assigned to the species. This additional step was easy to implement but it certainly generated some unrealistic situation for rare species – since these species have few observed biomass values, sampling may have generated unrealistic draws of populations with identical biomass values. However, we believe that these unrealistic draws have not changed the overall outcomes of the simulation. Therefore, as illustrated above, the species biomass distributions were simulated using the general codes:

```
# Defining the empirical biomass distributions of each species
data1 <- tapply (data$biomass, list(data$sp), function(x) x)
# Sampling with replacement biomass values from the empirical distribution
sim.data = NULL
for (i in 1:length(data1)) {
  if (v[i]==0) { value= 0 } else {
    value= sum(sample(x= data1[[i]], size= v[i], rep=TRUE)) }
  if (is.null(sim.data)) { sim.data= value } else {
    sim.data= c(sim.data,value) }
}
```

In this example above, the sampled individuals of each species are represented by the rows of a data frame arbitrarily named 'data' that are specified by columns 'biomass' (empirical biomass observations) and 'sp' (name of the species to which the individual belongs). The loop guarantees that the number of biomass values sampled for each species matches the abundance of the same species in the vector of abundance v . It's noteworthy that this procedure accepts v sampled from both Poisson and BN distributions.

References

- Bolker B.M. (2008) *Ecological Models and Data in R*. Princeton University Press, Princeton.
- Colwell R.K., Mao C.X. & Chang J. (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85, 2717–2727.
- Hughes R.G. (1986) Theories and models of species abundance. *Am. Nat.* 128, 879-899.

Appendix 2. Comparison between observed and rarefied Preston plots of tree communities at different scales and sample sizes. These are exactly the same curves presented in Figs. 1 and 2 of the main text but arranged in a different way. Observed SADs are in solid black while the 95% quantiles of the 10 000 random draws are in colored dashed lines. Different colors represent different reference scales (regional, site and habitat scales) from which the community properties were used to parameterize the random draws. These draws were generated using different sample sizes (columns: N_{site} , N_{habitat} , N_{sample}) that matches the mean observed sample size at each scale.

Figure S1. Comparison between observed and rarefied Preston plots using count of individuals (top row) and biomass (bottom row) at different scales and sample sizes. The mean observed SAD (solid black line) and the 95% confidence intervals of the rarefied SADs (dashed lines) are presented. The x-axis represents the abundance octaves in log base 2 which are closed to the right. For counts, the first octave contains species with less than 2^1 individuals, second octave contains species with $\geq 2^1$ and $< 2^2$ individuals, and so on. For biomass, the first octave contains species with biomass less than $< 2^{-6}$ kg, second octave contains species with $\geq 2^{-6}$ and $< 2^{-5}$ kg, and so on.

Appendix 3. Example of the scale effect in the shape of SADs under different scenarios of beta diversity. Here we present the results of a preliminary simulation that tried to mimic as far as possible the procedures found in the manuscript. For the sake of simplicity we assume the lognormal as the underlying SAD model and a random sampling scheme. We defined the lognormal parameter of the simulations based on the empirical findings of the regional tree community. As in the Appendix 1, codes are given in language R.

To illustrate the effect of beta diversity on the shape of the SAD, we generated three different regional communities (or meta-communities – M) but with the same total number of species S and same lognormal parameters, that were fixed based on the observed regional tree community SAD ($S= 486$, $\mu= 2.9$ and $\sigma= 2.1$). Although having the same number of species, the first metacommunity $M1$ shares only half of its species with $M2$ and no species with $M3$. This try to mimic a distance decay effect on species similarity that would reflect the spatial distance between communities. Next, we simulated a random sampling process in which n samples are taken from each meta-community with the same sample intensity a :

```
# setting the total species pool, sample size and sampling intensity:
S <- 486
n <- 100
a <- 1/4

# generating the lognormal meta-communities with  $\mu= 2.9$  and  $\sigma= 2.1$ :
M1 <- rlnorm (S, meanlog = 2.9, sdlog = 2.1)
M2 <- rlnorm (S, meanlog = 2.9, sdlog = 2.1)
M3 <- rlnorm (S, meanlog = 2.9, sdlog = 2.1)

# simulating  $n$  Poisson samples that are a times smaller from M1, M2 and M3:
data1 = NULL; data2 = NULL; data3 = NULL
for (i in 1:n) {
  x = rpois (S, lambda= M1*a)
  names(x) = paste("sp", 1:S, sep = '')
  if(is.null(data1)) data1 = x else data1 = rbind (data1, x)
}

for (i in 1:n) {
  x = rpois (S,lambda= M2*a)
  names(x) = paste("sp", (S/2+1):(S + S/2), sep='')
  if(is.null(data2)) data2 = x else data2 = rbind (data2,x)
}

for (i in 1:n) {
  x = rpois (S, lambda= M3*a)
  names(x) = paste("sp", (S+1):(2*S), sep='')
  if(is.null(data3)) data3 = x else data3 = rbind (data3,x)
}

# creating exclusive sample names for each communities. These allows the
#use of function 'merge' (see below)
```

```
rownames(data1) = paste("dat1", 1:n, sep='_')
rownames(data2) = paste("dat2", 1:n, sep='_')
rownames(data3) = paste("dat3", 1:n, sep='_')
```

Then, different scenarios of beta diversity were created by merging $n/2$ samples from $M1$ to $n/2$ samples from $M2$, so that the resulting community $M12$ covers a larger spatial scale than $M1$ and $M2$, but with the same sample size n . The same procedure was carried using $M1$ and $M3$ to create the community $M13$. When we compare $M1$ with $M12$ and $M13$ we are comparing sampling scenarios from meta-communities with low beta diversity ($M1$: all samples taken from the same species pool), medium ($M12$: samples taken from meta-communities that share half of the species pool), and high beta diversity ($M13$: samples taken from meta-communities that share none of the species). Thus, when we compare the shape of the resulting SAD from $M1$, $M12$ and $M13$ we are simulating sampling processes from meta-communities with markedly different beta diversity but with the exactly same sample size. Note that the low beta scenario is the $M1$ itself, since it represents the scenario with the lowest beta-diversity possible.

```
# medium beta diversity scenario (M1 vs. M2)
M12 = merge(t(data1[1:(n/2),]), t(data2[1:(n/2),]), by="row.names",
all= T)
M12[is.na(M12)] = 0
# high beta diversity scenario (M1 vs. M3)
M13 = merge(t(data1[1:(n/2),]), t(data3[1:(n/2),]), by="row.names",
all=T)
M13[is.na(M13)] = 0
```

As a result we found that the SAD generated in a high diversity scenario has on average more rare and less abundant species when compared to the low diversity scenario. The Preston plots reveal this pattern through a more asymmetric and steeply declining curve (Figure S3a) while on Whittaker rank-abundance plots this pattern is characterized by a less steep curve and by a much greater number of rare species, i.e., a larger number of species on the right tail of the curve (Figure S3b). Since we simulated the two most extreme scenarios of beta diversity, we believe that they cover all possible amounts of deviation in SAD due to beta diversity. Although we have arbitrarily fixed a value for a , the general patterns between the three beta diversity scenarios remain the same when we use different values (increases in the values of a only generate less asymmetrical curves on Preston plots and vice-versa). In this preliminary study, we did not evaluate the effect of sampling scheme (random vs. aggregated), the effect of meta-communities with different sizes of the species pool or

lognormal parameters and the possible effect of the species abundance overlap, that is, the degree of similarity in the abundance of shared species.

Figure S3. Preston (a) and Whittaker plots (b) constructed for the simulated samples of communities with low, medium and high beta diversity. These curves were obtained by averaging samples from scenarios of low, medium and high beta diversity, meaning that samples were taken from fictitious meta-communities with exactly the same species pool, with half of shared species, and with no species in common, respectively.

CAPÍTULO 4: DISTRIBUIÇÃO DIAMÉTRICA DOS FUSTES

4.1 Modeling tree diameter distributions in tropical forests

Renato Augusto F. de Lima^{a,b}, João Luís F. Batista^c, Paulo Inácio Prado^b

^a Programa de pós-graduação em Ecologia, Instituto de Biociências – Universidade de São Paulo (IB–USP)

^b Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, Instituto de Biociências – Universidade de São Paulo (IB–USP), Rua do Matão, trav. 14, nº 321, Cidade Universitária, CEP 05508–090, São Paulo, Brazil

^c Departamento de Ciências Florestais, Escola Superior de Agricultura ‘Luiz de Queiroz’ – Universidade de São Paulo (ESALQ–USP), Av. Pádua Dias, nº 11, CEP: 13418–900 Caixa Postal 9, Piracicaba, Brazil

Corresponding author: Renato Augusto F. de Lima (raflima@usp.br)

Running title: Modeling diameter distributions

Abstract

The description of tree size distribution is useful for various purposes in forest ecology and management. This description is traditionally performed using diameter at breast height (dbh) as a proxy of tree size and by fitting statistical distributions (i.e. models) to empirical dbh distributions. However, there is no agreement of what models are more appropriate to describe this type of data and what criteria should be used to assemble a reasonable set of candidate models. Here we provide a simultaneous comparison of the performance of ten models to fit empirical dbh distributions obtained in four 10.24-ha plots placed on different types of forests of south-eastern Brazil. These comparisons were carried at species and stand (40×40 m plots) levels and aimed to detect which models provide best fit to data and under which samples properties (size, median, variance, skewness, and kurtosis). We show that, in terms of fit to data, the combination of the Logit-Logistic (LL), Odd Weibull (OW), Weibull, and Johnson's Special Bounded (JSB) models provided plausible descriptions to nearly all species (92%) and stands (99%) in the forests studied here. However, there was relatively low overlap in their performance, meaning that models were rarely appropriate for the majority of cases. This complementarity was evident for the two models with best performance in terms of fit. The OW had a better performance for distributions that were more symmetrical or with trends toward bimodal or rotated sigmoid shapes, while the LL performed better for typical right-skewed patterns, particularly for heavy-tailed distributions. The performance of all models was significantly related to sample properties and type of forest, but in different directions for each model. Besides fit, the models with more than three parameters had more technical problems related to optimization convergence (sensitivity to bad starting values), unreliable confidence intervals and unrealistic fits, particularly for bounded models. Finally, we discuss more theoretical issues that may help foresters to choose appropriate candidate models.

Keywords: forest structure, lifetime distributions, likelihood, model selection, size structure

Abbreviations

LN: Log-normal; EXP: Exponential; JSB: Johnson's Special Bounded; Beta: Generalized Beta; BS: Birnbaum-Saunders, LL: Logit-Logistic; MW: New Modified Weibull; OW: Odd Weibull.

Introduction

The description of tree size distribution is a simple yet effective tool in forest ecology and management. In management, size distributions are used to value forests, plan harvest activities, predict forest growth, and thus to increase forest productivity (Bailey and Dell, 1973; Borders and Patterson, 1990; Hyink and Moser Jr., 1983; Newton et al., 2005). In ecology, besides the comparison between forest stands, size distributions can be used to infer on past disturbance events, successional status, estimate above-ground biomass and carbon stocks, and test predictions of different ecological theories (Coomes and Allen, 2007; Lorimer and Frelich, 1984; Muller-Landau et al., 2006; Schmelz and Lindsey, 1965). When assessed at species level, the number of trees in each given dbh can also provide information on species-specific regeneration strategies, demographic rates and population trends (Condit et al., 1998; Knight, 1975; Van Sickle, 1977; Wright et al., 2003).

Traditionally tree size is assessed through stem diameter at breast height – dbh, and its frequency distribution is described using probability distributions (but see Goff and West, 1965). At least from the 50s on, foresters started to search for appropriate distributions to mathematically described dbh distributions, which are typically skewed to the right (Leak, 1965; Meyer, 1952). Classic examples of distributions used to this end are the Exponential (a.k.a. inverse J-shaped distribution), Log-normal, Weibull, and Gamma distributions (Bailey, 1980; Bailey and Dell, 1973; Bliss and Reinker, 1964; Hafley and Schreuder, 1977). The use of such distributions, known as lifetime distributions, was initially phenomenological, that is, because they described dbh distributions well (Bailey and Dell, 1973; Leak, 1965). But as the name suggests, this class of distributions is used in engineering and biology to model the survival of machines, processes, or organisms. A non-phenomenological use of lifetime distributions to model dbh distributions would rely on the assumption that tree size and age are proportional. Although it seems quite reasonable to assume that larger trees are older than smaller ones, different growth rates at different sizes or sites can generate non-linear size-age relationships (Lorimer et al., 2001; Smith et al., 1998). But if there is compelling evidence that size and age are proportional, then lifetime distributions can be used to infer on the mechanisms responsible for generating patterns of tree size distribution. This may be the case of even-aged stands where trees share more similar conditions and thus would have similar growth and mortality rates (Bailey and Dell, 1973).

Since the 60s new statistical distributions were derived, with some of them being only recently applied to model tree dbh distributions (e.g. Podlaski, 2008; Wang and Rennolls, 2005). Amidst the numerous distributions available foresters need reasonable criteria to

choose the most appropriate set of distributions for their own purposes. Most likely, the simplest criterion is the assessment of distribution fit to data. This assessment has been carried out for decades, using different numbers and combinations of distributions, different parameter estimation methods and using empirical dbh data from natural or planted, managed or unmanaged forest stands (Hafley and Schreuder, 1977; Li et al., 2002; Nanang, 1998; Palahí et al., 2007; Podlaski, 2008; Wang and Rennolls, 2005; Westphal et al., 2006). But the selection of candidate distributions should not be based solely on fit to data. Other important criteria are the number and interpretability of the parameters and the existence of proper methods for parameter estimation and model comparison. And if the assumption of proportionality between tree size and age holds then the choice of candidate distributions should also take into account theoretical arguments or the biological assumptions of each lifetime distribution. One should also take into account the hazard function of each distribution that is a competing hypothesis about tree mortality patterns (Leiva et al., 2009).

Aiming to help foresters to choose appropriate probability distributions to describe dbh distributions, we assess the performance of traditional and recently derived distributions to model empirical dbh distributions on a representative set of natural tropical forests of southeastern Brazil. Aiming to cover a wide range of possible situations and applications, we perform this evaluation at species and stand levels using dbh distributions obtained from sites with markedly different forest structures. Additionally, this study aims to correct a bias in the evaluation of dbh data description that until now was largely confined to temperate forests (but see Nanang, 1998). We use maximum likelihood techniques and information-based model selection (Burnham and Anderson, 2002) to estimate parameters and to compare distribution performances, methods which have desirable properties over moments and percentile methods or goodness-of-fit tests (Nanang, 1998; Podlaski, 2008; Zhang et al., 2003; Zhou and McTague, 1996). Thus, besides the description of general patterns of dbh distribution in four representative tropical forest sites, we ask the following questions: which distribution provides best fit to the observed dbh distributions? What properties of the samples (sample size, median, variance, skewness, and kurtosis) influence the fit of distributions? Which distributions have less analytical drawbacks during parameter estimation? Finally, we discuss other practical or theoretical properties of distributions that may help foresters to choose proper candidate distributions to model dbh data.

Material and methods

Study sites and data collection

Diameter measurements were conducted in four Brazilian types of forests that are very contrasting with respect to its composition, structure and diversity: the savanna, seasonal, white-sand and rain forests. In each of these forests a 10.24-ha permanent forest plot were placed in different protected areas of São Paulo State, southeastern Brazil (plot coordinates vary between 22°24'–25°04' S and 47°55'–50°22' W). Each plot has 320×320 m, which results in a total sampled area of 40.96 ha in the four forests. Within each plot, all trees, palms and tree ferns with stem girth at breast height ≥ 15 cm (i.e. diameter at breast height – dbh ≥ 4.78 cm) were measured and identified to species level. In cases of trees with multiple stems (i.e. ramifications below 1.3 m above ground height), if at least one stem met the dbh cutoff criterion, measurements were taken for all stems. In case of buttresses and bent trees, measurements were taken following the recommendations of Condit (1998). Dbh data used here refer to the second survey (2004/2005) that resulted in an overall tree density of 2142, 1492, 1176, and 1135 ha⁻¹ for the savanna, white-sand, seasonal, and rainforest plots, respectively. Correspondingly, maximum dbh and total plot richness were 45.2, 85.4, 216.5, and 151.2 cm, and 118, 117, 146, and 223 species.

Data analysis

The evaluation of different candidate distributions was carried at species and stand levels. Species-wise, we modeled the dbh distribution of individual species with ≥ 30 individuals at each 10.24-ha plot. We excluded from this analysis tree ferns and palm species that have limited or no secondary stem growth after initial development. Thus, species-level analysis involved a total of 225 species and ca. 92% of all tree individuals sampled (range: 31–4605 individuals per species; mean $\pm$ sd: 248 $\pm$ 525). When the same species occurred with ≥ 30 individuals at two or more plots, distributions from each plot were modeled separately. Stand-wise, we modeled the dbh distribution of all trees inside 40×40 m subplots (0.16 ha), hereafter referred to as stands, resulting in 64 contiguous stands per 10.24-ha plot (total of 256 stands; range: 37–470 individuals per stand; mean $\pm$ sd: 238 $\pm$ 79). This stand size was chosen aiming to break possible spatial autocorrelation between stands, which had an average range of ca. 20 m (R.A.F. Lima, unpublished results). To account for species-specific strategies of growth we used the equivalent dbh of the multi-stemmed trees during all analysis. That is, instead of using maximum dbh, we calculated the single dbh which is equivalent to the total basal area of all stems together. We also calculated general descriptive statistics for all dbh distributions including sample size, maximum dbh, variance, skewness, and kurtosis. Maximum dbh was

positively correlated to sample variance and skewness, thus the results of analyses including this variable were omitted.

Evaluation was based on ten candidate continuous distributions: Log-normal (LN), Exponential (EXP), Weibull, Gamma, Johnson's Special Bounded (JSB), Generalized Beta (Beta), Birnbaum-Saunders (BS), Logit-Logistic (LL), New Modified Weibull (MW) and Odd Weibull (OW) distributions. The first five are commonly used to model empirical dbh distribution (Bailey, 1980; Hafley and Schreuder, 1977) while the Generalized Beta, BS and the LL distributions were only recently applied to this end (Li et al., 2002; Podlaski, 2008; Wang and Rennolls, 2005). As far as we know, this is the first time that the MW (Lai et al., 2003) and OW (Cooray, 2006), two Weibull modifications that have the Exponential and Weibull distributions as special cases, are used to model dbh distributions. General properties of the ten distributions are given in Table 1. We did not attempt to fit distributions with more than four parameters, such as the generalized modified Weibull (Carrasco et al., 2008), which often have parameters that are difficult to estimate, to interpret biologically and to represent graphically. We also have not tried to evaluate distributions which had poor performance in one or more previous studies, such as the Gaussian and Burr II distribution (Nanang, 1998; Wang and Rennolls, 2005).

Parameter estimation was carried using maximum likelihood techniques (Burnham and Anderson, 2002) using the probability density functions (PDF) of the ten above-cited distributions (Appendix A). The negative log likelihood function constructed for each PDF was minimized using numeric optimization, aiming to find the combination of parameters that best fitted the data (i.e. maximum likelihood estimates – MLE). MLEs have many desirable statistical properties (Royle and Dorazio, 2008). In addition, several studies using different distribution have reported MLE as superior or comparable to other parameter estimation methods (Nanang, 1998; Podlaski, 2008; Zhang et al., 2003; Zhou and McTague, 1996; but see Cousineau, 2008). Since the dbh data is left truncated, we used location-free versions of the PDFs to build the likelihood functions and we fixed the location parameter at minimum dbh - c , a constant which simply is half the dbh measurement precision in each plot. Exception to this rule was the BS distribution that was very sensitive to the truncating point and had the location parameter fixed at values slightly lower (minimum dbh - $2 \times c$). Although arbitrary, fixing the location parameter diminishes the chances of convergence problems and biased estimates (e.g. Cousineau, 2008). Ideally, PDFs should have been truncated using the accumulated probability at the point of location, but the cumulative distribution functions

were not available for all models or did not have a closed form, which greatly complicates the process of estimating the parameters numerically.

For the bounded distributions (i.e. Beta, JSB, and LL) the range parameter was kept free during optimizations and maximum - minimum observed dbh was used as the starting value of the optimization routine. For other parameters, the starting values were defined using the method of moments, whenever available for a given distribution. All fits were visually inspected by superimposing them to density histograms. In cases of bad fits we varied the start parameters, the maximum number of iterations, or the optimization method (e.g. Nelder-Mead, quasi-Newton or Simulated annealing – Bolker, 2008) to check for a better fit or convergence. The bounded distributions occasionally resulted in unreasonable fits, such as U-shaped fits when the data was not U-shaped. To avoid such fits, we changed the start parameters (generally the range parameter) to find more realistic fits.

The comparison of candidate distributions was based on the Akaike's Information Criterion ($AIC = -2 \ln(L) + 2k$, where L = likelihood, and k = number of parameters) that measures the distance of a candidate distribution relative to an unknown true model. The smaller the AIC value, closer the candidate will be to this true model. Additionally, AIC penalizes distributions according to the number of parameters. So, the difference in AIC from the distribution with lowest AIC (i.e. ΔAIC) was used here as a measure of "goodness-of-fit", allowing to compare distributions with different numbers of parameters simultaneously considering model fit and parsimony (Burnham and Anderson, 2002). Thus, the performance of the ten distributions was evaluated using model ranking, that is, distributions were classified into plausible or non-plausible fits according to their AIC. We used three different ΔAIC thresholds to rank distributions: two, four, and seven. The use of larger thresholds avoids that plausible models are discarded due to small sample sizes (when N is small models with fewer parameters tend to perform better – Richards, 2005).

To assess the performance of the ten candidate distributions under different sample properties, we modeled the performance of each candidate distribution using sample size, median dbh, variance, skewness, kurtosis, and forest type as covariates. As the performance of the distributions was classified into plausible and not plausible, analysis was performed using mixed-effects logistic regression (Pinheiro and Bates, 2000) where the covariates were the fixed effects and the level (i.e. species or stands) were the random effects. Thus, although the performance of each distribution at species and stand levels were lumped together ($N=481$), we controlled for the variation in model performance between the two levels of evaluation. The importance of the covariate effect on the logistic models was assessed by

comparing the ΔAIC of models with and without the covariate. We used the logit link and covariates were log-transformed whenever needed. The difference of coefficient estimates between forest types was assessed based on their likelihood profiles. All analyses were performed using R (www.R-project.org) and the packages *bbmle* (Bolker, 2010) and *lme4* (Bates et al., 2012).

Results

As expected, species and stand dbh distributions were markedly right-skewed with average values of sample skewness and kurtosis of 2.1 ± 1.1 and 6.8 ± 8.5 (mean $\pm$ standard deviation). But the observed skewness and kurtosis were highly variable (range of skewness and kurtosis: -0.1–6.9, and -1.1–64.0) meaning that observed dbh distributions covered a wide range of shapes. The distribution of sample skewness and kurtosis was different between species and stands distributions (Student *t*-test on log-transformed data: *t*-value= 17.1 and 16.0, respectively; d.f.= 479; $p < 0.0001$), with both variables having greater mean values for stands. Skewness and kurtosis of stand dbh distributions were also different between forest types (Analysis of Variance on log-transformed data: *F*-value= 58.2 and 34.9, respectively; d.f.= 252; $p < 0.0001$), with mean values of both variables arranged as follows: savanna < white-sand < seasonal forest < rainforest (for kurtosis only there was no statistical difference between values of the two last plots). Differences in the sample asymmetry between forest types were probably due to differences in mean maximum dbh which presented differences in the same way as for sample asymmetry (ANOVA on log-transformed data: *F*-value= 117.7; d.f.= 252; $p < 0.0001$).

The evaluation of the ten candidate distributions in terms of their fit to dbh data resulted in the following order of general performance, from best to worst: LL, OW, Weibull, Gamma, Beta, MW, JSB, EXP, LN and BS. This outcome varied little between species and stand levels assessments and between ranking using different ΔAIC thresholds (Table 2). It should be noted, however, that candidate distributions were rarely appropriate to describe the majority of dbh distributions. The distribution with best fit to dbh data, the LL, provided plausible fits for 52% and 60% of the species and stands distributions, respectively, at ΔAIC threshold of 2. The top three distributions provided plausible fits for 89% of species and 92% of stands at the same threshold. The JSB was the model with least overlap with the top three distributions (Figure 1), raising these proportions to 92 and 99%. The Gamma, however, had a great overlap with the Weibull (88% at ΔAIC threshold of 2). There was great overlap between the Weibull and MW with the OW (62.4 and 89.9%), and between the JSB with the

LL (69.3%). The one-parameter EXP provided good fits for only 16% of the distributions, and out of this proportion only 7% did not overlap with the Weibull or OW, which have the EXP as a special case (shape parameter(s) ≈ 1).

The performance of distributions was significantly related to sample properties and type of forest. In general, their performance worsened as size, total basal area, variance, skewness and kurtosis of the sample increased (Table 3). Exceptions were the negative relationships of LN, OW, and BS with sample skewness and kurtosis, and the negative relationship between LL with sample variance and skewness. The absence of relationship between the LL with skewness and kurtosis, and between OW with sample size, basal area and forest type were remarkable as well. The direct comparison between the LL and OW performances revealed that the OW was more appropriate for dbh distributions with larger medians and smaller variances (i.e. more symmetric distributions), and that both models performed equally well when dbh distributions had high skewness and kurtosis (Appendix B). The visual inspection of models fits revealed this and other trends (Figures 2 and 3). The OW performed better for fitting unimodal distributions and distributions with tendencies to bimodality, while distributions with long tails were more efficiently fitted by the LL. The one-parameter EXP could not fit well unimodal and bell-shaped distributions. The Beta performed better to fit more symmetric distributions (skewness and kurtosis close to 0). The BS was only appropriate for distributions with a well-defined mode at values away from the location; that is probably the reason why the distribution performed worst at stand-level. Similar outcomes were found for the LN distribution.

The two and three-parameter distributions presented no problems in the convergence of numerical optimization and fits were very robust to changes in the start parameters, indicating that the optimization was not trapped in local minima. The Log-normal and BS distributions, however, were very sensitive to the minimum location parameter; when fixed at values close to minimum dbh these two distributions performed much worst (results not shown). For other distributions, namely OW, MW, Beta, JSB, and LL, parameter estimation generated optimization problems (non-convergences using the initial starts was 15, three, nine, nine, and 45, respectively). Most of convergence problems occurred at species-level and were easily solved by changing the start parameters. For LL distribution, however, convergence was not attained in 15 cases. In general, the OW and MW were much less sensitive than the Beta, JSB, and LL to bad start parameters. The flexibility of the Beta, JSB, and especially the LL resulted in some U-shaped fits, i.e. fits that predicted increasing probabilities of having larger dbh values. Such unrealistic fits were more frequent for more

uniform distributions and for those with ‘higher-than-expected’ proportion of larger trees. Very large range estimates such as values > 500 were also common for Beta distribution. The inversion of the Hessian matrix to find the confidence intervals (Bolker, 2008) of the parameters frequently failed for the four-parameter distributions, mainly for bounded ones. And when these intervals were attainable they often were so broad that they became meaningless. This was particularly true for range parameter estimates.

Discussion

The great number of probability distributions currently available provides foresters descriptions to virtually all types of empirical frequency distributions of tree dbh. However, excessive number of options can hinder rather than help the selection appropriate models. In terms of fit to dbh data, the overall result of the evaluation of the ten models was very straightforward: the joint use of LL, OW, Weibull, and JSB provided plausible descriptions to nearly all species and stands dbh distributions, at least in the natural tropical forests studied here. Good performances of the LL, Weibull, and JSB have been demonstrated before in natural or managed forests (Nanang, 1998; Palahí et al., 2007; Podlaski, 2006; Wang and Rennolls, 2005) and our results placed the OW among models with better performances in the description of dbh data. The Beta distribution, although being very flexible in terms of shape (Table 1), performed worse than the Weibull and Gamma, and we found no differences in overall fit with JSB at both species or stand levels, somewhat contradicting the results found by Li et al. (2002) and Wang and Rennolls (2005). Moreover, we found a better performance for the Weibull when compared to the MW, an unexpected result based on Lai et al. (2003) findings. Besides providing worst fits then the best performance models, the Gamma and MW had great overlap with them, so the decision of selecting these models should be based on criteria other than fit (see discussion below). The EXP, LN and BS models clearly performed worst, suggesting that these three models will only provide good descriptions under specific forest structure settings (see Bliss and Reinker, 1964; Nanang, 1998; Podlaski, 2008 for examples).

Overall, the LL and OW distributions were the models that often provided plausible fits regardless of the ΔAIC threshold adopted, meaning that even when they were not the best fit, their fit was generally plausible. Although both models performed well in cases of high skewness and kurtosis (the two probably cover wide skewness-kurtosis ranges), they performed better under different cases. The OW described better distributions that were more symmetrical or with trends toward bimodal or rotated sigmoid shapes. This may be the case of

early-successional species (Knight, 1975; Wright et al., 2003), species or stands presenting peaks of growth and/or mortality (Goff and West, 1975; Leak, 2002), even-aged stands (Schmelz and Lindsey, 1965), forests under thinning (Maltamo et al., 2000), or harvesting focused on specific sizes. This last suggestion was evident for *Euterpe edulis* (Figure 3c), a palm species that suffers from illegal harvesting in the rainforest plot. Moreover, this generalized Weibull distribution was able to deal with trends towards bimodality using less parameters than finite Weibull mixtures (Liu et al., 2002; Westphal et al., 2006). Conversely, the LL can be indicated for typical right-skewed patterns, which was the probable reason why the LL performed even better at stand level. In particular, the LL described well species with large contributions of small and extreme values at the same time (i.e., heavy-tailed distributions; Figure 2b). This may be the case of large-sized species with low growth and/or high survival of juveniles (i.e. shade-tolerants – cf. Bailey and Dell, 1973; Knight, 1975) or of species with massive recruitment followed by high juvenile mortality. Thus, when distributions depart from the typical right-skewed pattern the OW may be more appropriate than the LL, and vice-versa.

A good fit to data, however, should not be the sole criterion to select one distribution over another. One may choose models with worst fits if there are theoretical or practical reasons to do so (Burnham and Anderson, 2002). Another important criterion is theory. Each model has assumptions on the processes generating the observed patterns, and if these processes make sense to a given species or forest, models can serve as alternative hypothesis of these underlying mechanisms. For instance, the Exponential and Weibull generate specific patterns of population growth and mortality (Muller-Landau et al., 2006), the BS assumes tree mortality as a fatigue (fluctuating stress) process (Podlaski, 2008), while the LN and Gamma assumes mortality as a cumulative damage or wear-out process. However, because different processes may result in similar patterns, there is always a risk of a model describing well a pattern, but that this pattern has been generated by a mechanism different from the one assumed in the original model formulation. In addition to theory, foresters may give preference to models with meaningful parameters or easy to interpret. Models with two shape parameters such as the LL, OW, MW, Beta and JSB are definitely more flexible but they are more difficult to interpret biologically. Moreover, one should ask if bounded distributions (e.g. LL, Beta and JSB) make sense to your data. Their fit is quite sensitive to the definition of upper and lower bounds, which are generally defined arbitrarily, vary with sample size or change over time in planted forests (Palahí et al., 2007; Taubert et al., 2013; Zhou and McTague, 1996). Although the range parameter could provide extra information on tree

maximum sizes, it often works purely as a phenomenological parameter with unrealistic estimates, such as ranges >500 cm.

Other criteria may also help foresters' choice. Besides flexible density functions, some distributions have very flexible hazard or mortality functions (see Table 1 and Nadarajah, 2009 for examples). This flexibility is important because it covers several types of tree mortality patterns (Coomes and Allen, 2007; Harcombe, 1987; Lorimer and Frelich, 1984; Lorimer et al., 2001). Thus, whenever tree dbh and age are proportional (e.g. even-aged forests – Bailey and Dell, 1973), the flexibility of distributions to describe mortality curves is also a useful criterion. Additionally, our results also showed that the performance of some distributions was less influenced by sample properties than others. For instance, the lack of relationship with forest type and the negative relationship with sample size were two positive features of the OW over other models. If optimization is part of the parameter estimation procedure, the frequency of convergence problems and sensibility to bad start parameters can also be taken into account. Unfortunately, the definition of candidate models based on criteria other than fit is much more subjective and great relies on the researcher previous knowledge on the models and on the system being modeled (Burnham and Anderson, 2002)

Although the evaluation ended in straightforward results, the overall performance of the ten candidate distributions should be interpreted with care. We could have had different outcomes if dbh data had come from managed forests (Li et al., 2002; Nanang, 1998), even-aged stands (Nord-Larsen and Cao, 2006), different cutoff criteria (Goff and West, 1975) or different development stages (Podlaski, 2006). Besides these limitations, other aspects not covered here can be addressed in future research. What would be the effect of different criteria for defining the location or range parameters? For the BS and LN distributions, not fixing the location led to better fits, but unrealistic negative estimates often occurred. In addition, the use of truncated versions of the BS and LN PDFs could have resulted in better performances for these and perhaps other models. Another promising approach is the assessment of the LL and OW, the two distributions with best performances, to model bivariate distributions such as tree dbh and height (Li et al., 2002) or distributions in time (Knoebel and Burkhart, 1991). Because of theoretical and empirical evidence in favor of the MLE (see methods section), it seems unlikely that the use of different parameter estimators would have resulted in different outcomes. However, it would be interesting to explore the use of maximum likelihood methods that could account for measurement errors (Taubert et al., 2013) or weight observations with larger dbh (Cousineau, 2008), which despite the smaller number, greatly contribute to forest structure and biomass (Chave et al., 2003). Thus,

future research has many opportunities to deepen and broaden the use of statistical models for modeling diameter distributions.

Acknowledgements

We are grateful to the researchers responsible for data collection in the four sites: A.A. Oliveira, A.R. Duarte, A.P. Savassi-Coutinho, M.T.Z. Toniato, D. Sampaio, G.A.D.C. Franco, G. Durigan, N.M. Ivanauskas, R.R. Rodrigues, T.E. Barreto, and V.C. Souza. Financial support for data collection was provided by FAPESP (process no. 99/09635-0 and 09/53413-5). First author was supported by CAPES (Demanda Social program), and P.I.P. was supported by CNPq (Bolsa de Produtividade 303878/2008-8).

Appendices

Appendix A. Probability density functions (PDF) of the ten candidate distributions and an example of the maximum likelihood procedures used to fit them to dbh data.

Appendix B. Direct comparison between the Odd Weibull and Logit-logistic performances.

References

- Bailey, R.L., 1980. Individual tree growth derived from diameter distribution models. *Forest Sci.* 26, 626-632.
- Bailey, R.L., Dell, T.R., 1973. Quantifying diameter distributions with the Weibull function. *Forest Sci.* 19, 97-104.
- Bates, D., Maechler, M., Bolker, B., 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. <http://CRAN.R-project.org/package=lme4>
- Bliss, C.I., Reinker, K.A., 1964. A lognormal approach to diameter distributions in even-aged stands. *Forest Sci.* 10, 350-360.
- Bolker, B.M., 2008. *Ecological models and data in R*. Princeton University Press, Princeton.
- Bolker, B.M., 2010. bbmle: Tools for general maximum likelihood estimation. R package version 0.9.5.1. URL: <http://CRAN.R-project.org/package=bbmle>.
- Borders, B.E., Patterson, W.D., 1990. Projecting stand tables: a comparison of the Weibull diameter distribution method, a percentile-based projection method and a basal area growth projection method. *Forest Sci.* 36, 413-424.
- Burnham, K.P., Anderson, D.R. 2002. *Model selection and multi-model inference: a practical information-theoretic approach*. 2nd ed. Springer, New York.

- Carrasco, J.M.F., Ortega, E.M.M., Cordeiro, G.M., 2008. A generalized modified Weibull distribution for lifetime modeling. *Comput. Stat. Data. Anal.* 53, 450-462.
- Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B., Hubbell, S.P., 2003. Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J. Ecol.* 91, 240-252.
- Coomes, D.A., Allen, R.B., 2007. Mortality and tree-size distributions in natural mixed-age forests. *J. Ecol.* 95, 27-40.
- Cooray, K., 2006. Generalization of the Weibull distribution: the odd Weibull family. *Stat. Modelling* 6, 265-277.
- Condit, R., 1998. Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots. Springer-Verlag, Georgetown.
- Condit, R., Sukumar, R., Hubbell, S.P., Foster, R.B., 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. *Am. Nat.* 152, 495-509.
- Cousineau, D., 2008. Fitting the three-parameter Weibull distribution: Review and evaluation of existing and new methods. *IEEE T. Dielectr. Electr. Insul.* 16, 281-288.
- Goff, F.G., West, D., 1975. Canopy-understory interaction effects on forest population structure. *Forest Sci.* 21, 98-108.
- Hafley, W.L., Schreuder, H.T., 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* 7, 481-487.
- Harcombe, P.A., 1987. Tree life tables. *Bioscience* 37, 557-568.
- Hyink, D.M., Moser, J.W., Jr. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. *Forest Sci.* 29, 85-95.
- Knight, D.H., 1975. A phytosociological analysis of species-rich tropical forest on Barro Colorado Island, Panama. *Ecol. Monogr.* 45, 259-284.
- Knoebel, B.R., Burkhardt, H.E. 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. *Biometrics* 47, 241-253.
- Lai, C.D., Xie, M., Murthy, D.N.P., 2003. A Modified Weibull distribution. *IEEE T. Reliab.* 52, 33-37.
- Leak, W.B. 1965. The J-shaped probability distribution. *Forest Sci.* 11, 405-409.
- Leak, W.B. 2002. Origin of sigmoid diameter distributions. Research Paper NE-178. USDA Forest Service, Newtown Square.
- Leiva, V., Barros, M., Paula, G.A., 2009. Generalized Birnbaum-Saunders models using R. Brazilian Statistical Association, São Paulo.

- Li, F., Zhang, L., Davis, C.J., 2002. Modeling the joint distribution of tree diameters and heights by bivariate Generalized Beta distribution. *Forest Sci.* 48, 47-58.
- Liu, C., Zhang, L., Davis, C.J., Solomon, D.S., Gove, J.H., 2002. A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *Forest Sci.* 48, 653-661.
- Lorimer, C.G., Dahir, S.E., Nordheim, E.V., 2001. Tree mortality rates and longevity in mature and old-growth hemlock-hardwood forests. *J. Ecol.* 89, 960-971.
- Lorimer, C.G., Frelich, L.E., 1984. A simulation of equilibrium diameter distributions of Sugar Maple (*Acer saccharum*). *B. Torrey Bot. Club* 111, 193-199.
- Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T., Saramaki, J., 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. *Forest Ecol. Manage.* 133, 263-274.
- Meyer, H.A., 1952. Structure, growth, and drain in balanced uneven-aged forests. *J. Forest.* 50, 85-92.
- Muller-Landau, H.C., Condit, R., Harms, K.E., et al., 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol. Lett.* 9, 589-602.
- Nadarajah, S., 2009. Bathtub-shaped failure rate functions. *Qual. Quant.* 43, 855-863.
- Nanang, D.M., 1998. Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of Neem plantations in Northern Ghana. *Forest Ecol. Manage.* 103, 1-7.
- Newton, P.F., Lei, Y., Zhang, S.Y., 2005. Stand-level diameter distribution yield model for black spruce plantations. *Forest Ecol. Manage.* 209, 181-192.
- Nord-Larsen, T., Cao, Q.V., 2006. A diameter distribution model for even-aged beech in Denmark. *Forest Ecol. Manage.* 231, 218-225.
- Palahí, M., Pukkala, T., Blasco, E., Trasobares, A., 2007. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). *Eur. J. For. Res.* 126, 563-571.
- Pinheiro, J., Bates, D., 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York.
- Podlaski, R., 2006. Suitability of the selected statistical distributions for fitting diameter data in distinguished development stages and phases of near-natural mixed forests in the Swietokrzyski National Park (Poland). *Forest Ecol. Manage.* 236, 393-402.
- Podlaski, R., 2008. Characterization of diameter distribution data in near-natural forests using the Birnbaum-Saunders distribution. *Can. J. For. Res.* 38, 518-527.

- Richards, S.A., 2005. Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* 86, 2805-2814.
- Royle, J.A., Dorazio, R.M., 2008. Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities. Academic Press, San Diego.
- Schmelz, D.V., Lindsey, A.A., 1965. Size-class structure of old-growth forests in Indiana. *Forest Sci.* 11, 258-264.
- Smith, B.D., Botsford, L.W., Wing, S.R. 1998. Estimation of growth and mortality parameters from size frequency distributions lacking age patterns: the red sea urchin (*Strongylocentrotus franciscanus*) as an example. *Can. J. Fish. Aquat. Sci.* 55, 1236–1247.
- Taubert, F., Hartig, F., Dobner, H.-J., Huth, A., 2013. On the challenge of fitting tree size distributions in ecology. *Plos One* 8, e58036.
- Van Sickle, J., 1977. Mortality rates from size distributions: The application of a conservation law. *Oecologia* 27, 311–318.
- Wang, M., Rennolls, K., 2005. Tree diameter distribution modelling: introducing the logit–logistic distribution. *Can. J. For. Res.* 35, 1313.
- Westphal, C., Trener, N., von Oheimb, G., Hansen, J., von Gadow, K., Hardtle, W., 2006. Is the reverse J-shaped diameter distribution universally applicable in European virgin beech forests? *Forest Ecol. Manage.* 223, 75-83.
- Wright, S.J., Muller-Landau, H.C., Condit, R., Hubbell, S.P., 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84, 3174–3185.
- Zhang, L., Packard, K.C., Liu, C., 2003. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce–fir stands in northeastern North America. *Can. J. For. Res.* 33, 1340-1347.
- Zhou, B., McTague, J.P., 1996. Comparison and evaluation of five methods of estimation of the Johnson system parameters. *Can. J. For. Res.* 26, 928-935.

Tables

Table 1. Main properties of the ten candidate distributions used to model tree diameter distributions. All distributions have an additional location parameter, which can be set to zero or to any desired dbh value depending on the diameter cutoff criteria. The density function ($f(x)$) defines the probability of having dbh x , while the hazard function ($h(x)$) defines the probability of death in dbh x .

Model	Parameters	Bounded	$f(x)$	$h(x)$
Log-normal	shape (σ), scale (μ)	no	dec, mod	dec, mod
Exponential	scale (= 1/rate)	no	dec	cte
Weibull	shape, scale	no	dec, mod	cte, inc, dec
Odd Weibull	shape ₁ , shape ₂ , scale	no	dec, mod, sin	cte, inc, dec, mod, bt
Mod. Weibull	shape ₁ , shape ₂ , scale	no	dec, mod, sin	cte, inc, dec, bt
Gamma	shape, scale	no	dec, mod	cte, inc, dec
B.-Saunders	shape, scale	no	dec, mod	dec, mod, sin(?)
Beta	shape ₁ , shape ₂ , range	yes	cte, inc, dec, mod, bt, sin	cte, inc, dec, mod, bt, sin(?)
Johnson's SB	shape ₁ , shape ₂ , range	yes	dec, mod, bt	dec, inc, mod, bt, sin*
LL	shape ₁ , shape ₂ , range	yes	dec, mod, bt	dec, inc, mod, bt, sin*

Types of shapes: cte= constant; dec= decreasing; inc= increasing; mod= unimodal; bt= bath-tube (U-shaped); sin= sinuous (S-shaped).

* For these models, we were not able to obtain the hazard functions, so the assignment of shapes for this function may not be entirely correct.

Table 2. Performance of the ten candidate distributions used to fit species and stand level diameter distributions. Values represent the percentage of times when the distribution provided a plausible fit among the set of candidate models. Model ranking was made using three different thresholds of ΔAIC (i.e. 2, 4, and 7). $M[\Delta AIC]$ is the sample median of the ΔAIC distribution for each model; lower $M[\Delta AIC]$ values mean better performances of the models.

Model	Species ($N= 225$)				Stands ($N= 256$)			
	<2	<4	<7	$M[\Delta AIC]$	<2	<4	<7	$M[\Delta AIC]$
Log-normal	9.0	15.2	23.2	17.7	6.3	11.4	18.9	18.6
Exponential	19.6	34.2	47.6	8.5	11.3	19.1	26.2	14.1
Weibull	43.5	56.0	72.0	3.0	27.7	38.3	52.3	7.0
Odd Weibull	47.1	76.4	89.3	2.2	45.3	70.7	87.9	2.7
Mod. Weibull	28.0	61.3	76.4	3.6	7.03	27.3	42.9	8.9
Gamma	39.1	49.8	67.1	4.0	27.3	34.4	44.7	8.2
B.-Saunders	11.7	18.3	28.2	18.9	3.6	7.2	10.6	32.6
Beta	28.9	65.3	78.7	3.2	7.8	26.2	37.9	10.5
Johnson's SB	24.4	43.6	60.0	3.7	16.4	25.0	37.9	7.6
Logit-Logistic	52.0	65.8	76.4	1.8	60.2	73.1	88.3	0.5

Table 3. Results of the mixed-effects logistic regressions used to assess the effect of different sample properties and forest type on the performance of each distribution. Values correspond to the difference in AIC (ΔAIC) between the model without the covariate (constant model) and the model with the covariate. If ΔAIC values are positive it means that the model with the covariate was a better fit even after being penalized by its number of parameters. Absolute ΔAIC values smaller than two are not strong evidence of difference between two model's fit. Symbols indicate whether there was evidence of a positive/increasing ($\ddagger$) or negative/decreasing ($\dagger$) relationship between the performance of the distribution and the covariate. The results for forest type are given in decreasing order of performance from left to right.

Model	<i>n</i>	<i>BA</i>	med	var	skew	kurt	forest type
Log-normal	14.7 $\ddagger$	3.0 $\ddagger$	9.6 $\ddagger$	3.7 $\dagger$	52.4 $\dagger$	39.1 $\dagger$	33.6: Rf< Sv= Sf= Ws
Exponential	29.9 $\ddagger$	31.2 $\ddagger$	-1.9	-0.2	-1.3	-0.6	-4.0
Weibull	1.6	30.9 $\ddagger$	0.0	35.9 $\ddagger$	12.5 $\ddagger$	9.8 $\ddagger$	14.5: Sv= Sf< Ws= Rf
Odd Weibull	1.9	-1.1	1.4	4.7 $\ddagger$	9.9 $\dagger$	32.8 $\dagger$	-2.7
Mod. Weibull	-1.1	-1.2	4.2 $\dagger$	15.3 $\ddagger$	63.9 $\ddagger$	31.2 $\ddagger$	31.3: Sf= Sv= Ws< Rf
Gamma	3.1 $\ddagger$	54.9 $\ddagger$	8.4 $\ddagger$	63.4 $\ddagger$	10.2 $\ddagger$	12.1 $\ddagger$	21.4: Sv< Sf= Rf= Ws
B.-Saunders	21.9 $\ddagger$	6.5 $\ddagger$	2.1 $\ddagger$	1.2 $\dagger$	17.3 $\dagger$	13.7 $\dagger$	40.4: Rf= Sv< Sf= Ws
Beta	4.3 $\ddagger$	16.8 $\ddagger$	-0.3	18.5 $\ddagger$	92.3 $\ddagger$	81.2 $\ddagger$	6.3: Sv= Sf< Ws= Rf
Johnson's SB	26.1 $\ddagger$	16.8 $\ddagger$	-1.6	-2.2	10.2 $\ddagger$	20.0 $\ddagger$	38.7: Rf= Sv< Ws< Sf
Logit-Logistic	36.1 $\ddagger$	4.9 $\ddagger$	1.7	13.3 $\dagger$	2.8 $\dagger$	-2.0	20.4: Rf< Ws= Sv= Sf

Legend: Sample size (*n*), basal area (*BA*), median (med), variance (var), skewness (skew) and kurtosis (kurt). Forest type: Sv=Savanna; Ws= White-sand; Sf= Seasonal forest; Rf= Rainforest.

Figures

Figure 1. Venn diagrams of the four distributions with best fit to species (top) and stand (bottom) diameter distributions. OW= Odd Weibull, JSB= Johnson's Special Bounded, LL= Logit-Logistic.

Figure 2. Examples of the observed dbh distributions and their respective fits by the four models with best fit to data. (a) *Aspidosperma polyneuron* (Seasonal forest); (b) *Trichilia catigua* (Seasonal forest); (c) *Maprounea guianensis* (Savanna forest); (d) *Ocotea corymbosa* (Savanna). Note that the x -axis is presented in log-scale.

Figure 3. Examples of the observed dbh distributions and their respective fits by the four distributions with best fit to data. (a) *Ocotea pulchella* (White-sand forest); (b) *Myrcia ilheosensis* (White-sand forest); (c) *Euterpe edulis* (Rain forest); (d) *Calophyllum brasiliensis* (White-sand forest).

Appendices

Appendix A. Probability density functions (PDF) of the ten candidate distributions and an example of the maximum likelihood procedures used to fit them to dbh data. Instead of the regular mathematical presentation of the formulas, functions are given in R language, the language used to build codes used in this study. For most candidate distributions we provide basic references for the functions and/or examples of their application in tree diameter distribution modelling. It should be note that since dbh data is generally truncated at the dbh cutoff criteria, we used versions of the PDFs including a location parameter that could be fixed at minimum dbh as we did in this study. For bounded distributions (i.e. Beta, Johnson's SB and Logit-Logistic) the minimum parameter 'min' is equivalent to the location parameter 'loc' of unbounded distributions. Moreover, since the method of finding maximum likelihood estimates (MLE) of the distribution parameters was numerical optimization, PDFs below generally had a conditional term defining what to return in cases when the optimizer provide values lying outside of the domain of the parameters.

```
#####
### R codes for the PDFs of the distributions ###
#####

### Exponential ###
# Meyer (1952)
# Obs.: rate = 1/scale

dEXP = function (x, rate, loc) {
  p = rate*exp(-rate*(x-loc))
  ifelse (rate<=0, return(-Inf), return(p))
}

### Log-normal ###
# Bliss and Reinker (1964) and Nanang (1998)
# Obs.: In the following code, meanlog = scale and sdlog = shape

dLN = function (x, scale, shape, loc) {
  a = 1/(sqrt(2*pi)*shape*(x-loc))
  b = exp(-1/(2*(shape^2))*(log(x-loc)-scale)^2)
  p = a*b
  return(p)
}

### Weibull ###
# Bailey and Dell (1973)

dWeibull = function (x, shape, scale, loc) {
  a = shape/scale*(((x-loc)/scale)^(shape-1))
```

```

        b = exp(-((x-loc)/scale)^shape)
        p = a*b
        ifelse (scale<=0|shape<=0,return(-Inf),return(p))
    }

### Odd Weibull ###
# Cooray (2006)

dOW = function (x, shapel, scale, shape2, loc) {
    a = (shapel*shape2/(x-loc))*(((x-loc)/scale)^shapel)*exp(((x-
loc)/scale)^shapel)
    b = ((exp(((x-loc)/scale)^shapel)-1)^(shape2-1))
    c = ((1+(exp(((x-loc)/scale)^shapel)-1)^shape2)^(2))
    p = (a*b)/c
    return(p)
}

### New Modified Weibull ###
# Lai et al. (2003)

dMW = function (x, shapel, scale, shape2, loc) {
    a = scale*(shapel+shape2*(x-loc))
    b = ((x-loc)^(shapel-1))*exp(shape2*(x-loc))
    c = exp(-scale*((x-loc)^shapel)*exp(shape2*(x-loc)))
    p = a*b*c
    ifelse (scale<=0|shapel<0|shape2<=0,return(-Inf),return(p))
}

### Gamma ###
# Podlaski (2008)

dGamma = function (x, shape, scale, loc) {
    a = ((x-loc)^(shape-1)/((scale^shape)*(gamma(shape))))
    b = (exp(-((x-loc)/scale)))
    p = a*b
    ifelse (shape<0,return(-Inf),return(p))
}

### Birnbaum-Saunders ###
# Podlaski (2008) and Leiva et al. (2009)

dBS = function(x, shape, scale, loc) {
    a = 1/(sqrt(2*pi))
    b = exp(-0.5*((1/shape)*(sqrt((x-loc)/scale)-sqrt(scale/(x-loc))))^2)
    c = ((x-loc)^(-3/2))*((x-loc)+ scale)/(2*shape*sqrt( scale))
    p = a*b*c
    ifelse (shape <= 0, return(-Inf),return(p))
}

### Generalized Beta ###
# Karian et al. (1996) and Li et al. (2002)

dBeta = function(x, shapel, shape2, min, range) {
    a = gamma(shapel+shape2+2)/(gamma(shapel+1)*gamma(shape2+1))
    b = ((x-min)^shapel)*((min+range-x)^shape2)
    c = (range^(shapel+shape2+1))
    p = a*b/c
    ifelse (shapel<=(-1)|shape2<=(-1),return(-Inf),return(p))
}

```

```

### Johnson's Special Bounded ###
# Johnson (1949) and Wang and Rennolls (2005)

dSB = function (x, shape1, shape2, min, range, ...) {
  a = shape2/sqrt(2*pi)
  b = range/((x-min)*(min+range-x))
  c = shape1 + shape2*log((x-min)/(min+range-x))
  d = exp(-0.5*(c^2))
  p = a*b*d
  ifelse (shape2<=0, return(-Inf), return(p))
}

### Logit-Logistic ###
# Wang and Rennolls (2005)

# PDF
dLL = function (x, shape1, shape2, min, range) {
  a = (range/shape2)*(1/((x-min)*(min+range-x)))
  b = (x-min)/(min+range-x)
  c = 1/(exp(-shape1/shape2)*b^(1/shape2)+exp(shape1/shape2)*b^(-
1/shape2)+2)
  p = a*c
  ifelse (shape2<=0, return(-Inf), return(p))
}

#####
### Example: fitting the Exponential function to dbh data ###
#####

# See Bolker (2008)

# Loading package 'bbmle' (Bolker 2010) for maximum likelihood #estimation
(function 'mle2')
require(bbmle)

# Setting an arbitrary dbh data
dbh = c(17.1,8.5,16.7,8.3,16.0,6.3,6.1,5.5,10.5,5.7,16.4,15.5,6.1,8.3,
7.9,6.0,6.3,5.3,5.9,5.6,5.4,7.2,24.1,8.4,9.7,10.0,5.0,8.4,12.7,6.6,9.8,
12.3,12.0,7.5,10.1,9.8,13.2,25.2,11.1,18.1)

# Setting minimum dbh and measurement precision
min.dbh = 4.8 #cm
prec = 0.1/2 #cm

# Defining the likelihood function using the Exponential PDF (function
#'dEXP', see above)
ll = function (rate) {
  -sum(log(dEXP(x= dbh, rate= rate, loc= min.dbh-prec)))
}

```

```

}

# Finding the MLEs using numerical optimization
MLE = mle2 (ll, start = list(rate= 1/mean(dbh-min.dbh)))

# Inspecting model fit to data
hist (dbh, prob=T)
curve(dEXP(x, rate= coef(MLE), loc= min.dbh-prec), add= TRUE, lwd= 2)

```

References

- Bailey, R.L., Dell, T.R., 1973. Quantifying diameter distributions with the Weibull function. *Forest Sci.* 19, 97-104.
- Bliss, C.I., Reinker, K.A., 1964. A lognormal approach to diameter distributions in even-aged stands. *Forest Sci.* 10, 350-360.
- Bolker, B.M., 2008. *Ecological models and data in R*. Princeton University Press, Princeton, US.
- Bolker, B.M., 2010. *bbmle: Tools for general maximum likelihood estimation*. R package version 0.9.5.1. URL: <http://CRAN.R-project.org/package=bbmle>.
- Cooray, K., 2006. Generalization of the Weibull distribution: the odd Weibull family. *Stat. Modelling* 6, 265–277.
- Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149-176.
- Karian, Z.A., Dudewicz, E.J., McDonald, P., 1996. The extended generalized lambda distribution system for fitting distributions to data. *Commun. Stat. Simul. Comput.* 25, 611–642.
- Lai, C. D., Xie, M., Murthy, D.N.P., 2003. A Modified Weibull distribution. *IEEE T. Relia.* 52, 33-37.
- Leiva, V., Barros, M., Paula, G.A., 2009. *Generalized Birnbaum-Saunders models using R*. Brazilian Statistical Association, São Paulo.
- Li, F., Zhang, L., Davis, C.J., 2002. Modeling the joint distribution of tree diameters and heights by bivariate Generalized Beta distribution. *Forest Sci.* 48, 47-58.
- Meyer, H.A., 1952. Structure, growth, and drain in balanced uneven-aged forests. *J. Forest.* 50, 85-92.
- Nanang, D.M., 1998. Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of Neem plantations in Northern Ghana. *Forest Ecol. Manage.* 103, 1-7.
- Podlaski, R., 2008. Characterization of diameter distribution data in near-natural forests using the Birnbaum-Saunders distribution. *Can. J. For. Res.* 38, 518-527.
- Wang, M., Rennolls, K., 2005. Tree diameter distribution modelling: introducing the logit–logistic distribution. *Can. J. For. Res.* 35, 1313.

Appendix B. Direct comparison between the Odd Weibull and Logit-logistic performances.

The Odd Weibull (OW) and the Logit-logistic (LL) were the two candidate distributions with best performance in terms of fit to data. Therefore, we performed a direct comparison to examine under which circumstances each model performed better than the other. We assess if the value of sample size, median, variance, skewness and kurtosis were different for species whose diameter distributions were well fitted by the LL, by the OW, by both distributions or by other distributions. In this case, we assigned a single best candidate distribution for the dbh distribution of each species. We compared the means between this groups using standard analysis of variance (ANOVA) on log transformed variables. For all sample properties except median there was a significant interaction term between the performance of the distribution and the level of analysis (i.e. species and stand levels). Differences between group means were determined using Tukey Honest significant difference test with 95% confidence level. However, because the same species could be well fitted by the two distributions at the same time, each species could be attributed to the two models at the same time. We performed two sample Student's *t*-tests excluding the species well fitted by both models at the same time and the differences between the two candidate distributions were qualitatively the same of the ANOVA (not shown).

The OW performed better than the LL when dbh distributions had larger sample median and smaller variance (Table B1) which characterizes more symmetric distributions. In addition, the OW performed better for larger sample sizes, what could suggest that the best fit of LL and other models over the OW could be a matter of sample size. Overall, there was no evidence for a differentiation in models performance related to sample skewness and kurtosis. But both models tended to performed better than other models when dbh distributions had high skewness and kurtosis. Since sample skewness and kurtosis were highly correlated (Pearson's correlation $r = 0.938$ [0.927–0.948], $t = 59.4$, $df = 479$, $p < 0.00001$) the results found for sample skewness was always the same as for kurtosis. The same is true for variance which was significantly correlated with sample maximum dbh (Pearson's correlation $r = 0.755$ [0.713–0.791], $t = 25.2$, $df = 479$, $p < 0.00001$).

Table B1. Comparison between sample properties of dbh distributions for which LL, OW, both or other distributions were plausible fits. Equal means are indicated by equal letter and all tests were statistically significant ($p < 0.01$). Data was log-transformed prior to analysis of variance, although the mean $\pm$ standard deviation presented refer to untransformed data.

	LL (N= 183)	OW (N= 115)	LL&OW (N= 94)	Others (89)	<i>F</i> *
<i>n</i>	185.5 $\pm$ 118.9 A	380.5 $\pm$ 570.1 B	165.6 $\pm$ 97.1 A	270.8 $\pm$ 491.4 A	29.9
median	9.9 $\pm$ 3.0 A	11.0 $\pm$ 4.3 B	9.8 $\pm$ 3.1 A	9.8 $\pm$ 3.5 A	8.50
variance	108.6 $\pm$ 122.2 A	68.4 $\pm$ 79.5 B	106.8 $\pm$ 104.2 A	82.4 $\pm$ 120.4 AB	15.4
skewness	1.83 $\pm$ 0.71 A	1.91 $\pm$ 1.16 A	2.63 $\pm$ 1.44 B	1.94 $\pm$ 0.90 A	54.1
kurtosis	4.09 $\pm$ 3.92 A	7.12 $\pm$ 9.26 AB	11.36 $\pm$ 12.09 b	5.97 $\pm$ 5.98 AB	40.5

* Value of the *F*-statistic (degrees of freedom= 473)

CAPÍTULO 4: DISTRIBUIÇÃO DIAMÉTRICA DOS FUSTES

4.2 How well do demographic rates predict tree size distributions? Evidence over 30 years of forest monitoring

RENATO A. FERREIRA DE LIMA^{1,2,4}, HELENE C. MULLER-LANDAU³; PAULO I. PRADO² AND
RICHARD CONDIT³

¹ *Programa de pós-graduação em Ecologia, Instituto de Biociências – Universidade de São Paulo (IB–USP), São Paulo, Brazil.*

² *Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, IB–USP, CEP 05508-090, São Paulo, Brazil.*

³ *Center for Tropical Forest Science (CTFS), Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Ancón, Republic of Panama.*

⁴ E-mail: raflima@usp.br

Abstract

Size distributions are frequently used in ecology to describe the structure populations or communities. On first principles, they should reflect the basic demography of species, including patterns in growth and mortality. However, there is certain debate on whether size distributions are influenced by other factors as well, such as population growth rates, or under which circumstances size distributions can be predicted from demographic rates. There are still very few studies that tested these assumptions at species-level and using long term monitoring of species demographic rates, especially for long-lived organisms such as trees. Using the 30 years of data on individual tree monitoring from the Barro Colorado Island (BCI) forest plot, we empirically examined how current tree size distributions of over 100 individual species are related to their demographic rates. More specifically, we evaluate how the shape of diameter distributions related to growth–diameter or mortality–diameter curves, mean recruitment and population growth rates. We also test whether observed diameter distributions match the expected under demographic equilibrium. We found that growth and recruitment, but not mortality, shape the dbh distribution of BCI shrub and tree species. On average, decreasing growth-diameter curves (i.e. higher juvenile growth) were associated to less steep diameter distributions, whereas high recruitment rates were related to steeper distributions. The shape of the mortality-diameter curve did not affect the shape of the diameter distribution. Thus, although the predictions of demographic equilibrium theory based on species growth and mortality were positively related to observed size distributions, there was lots of variation, making predictions quite unreliable. The main source of variation was probably related to fluctuations on recruitment patterns over time, which suggests that the majority of the BCI species are not at equilibrium. One of the main causes of departures from equilibrium is probably related to strong El Niño events, suggesting that such cyclic climate events are able to alter forest structure and dynamics and to maintain species distant from the expected equilibrium of theoretical formulations.

Keywords: Barro Colorado Island, El Niño effects, demographic equilibrium theory, diameter distribution, forest structure, population growth, Weibull distribution

Introduction

Size distributions, represented by the number of individuals in different size classes, are frequently used to describe populations or communities. The reason behind this usage is the consensus that the size distribution is a response to changes in fundamental processes

affecting the dynamics of populations or communities. In forest ecosystems, where size is traditionally obtained by measuring tree diameter as a proxy to size, the assessment of diameter distributions is generally associated with an attempt to infer forest stability, species-specific strategies of regeneration, or population trends (Lorimer & Frelich 1984, Condit et al. 1998, Kohira & Ninomiya 2003, Wright et al. 2003, Feeley et al. 2007). In addition, Goff & West (1975) have once suggested the dominant type of competition between trees in a forest (i.e. competition for space and/or light) could be inferred from diameter distributions.

On first principles, tree diameter distributions are determined by the interplay between species demographic rates, such as size-dependent growth and mortality and population growth (Condit et al. 1998, Kohyama et al. 2003, Muller-Landau et al. 2006). These principles are supported by the growing body of empirical evidence (Lorimer et al. 2001, Coomes & Allen 2007) and simulation studies (Lorimer & Frelich 1984, Leak 2002). The relationship between tree diameter distributions and demographic rates is easy to understand considering that diameter distributions change depending on the rate at which individuals enter a given size class by recruitment or growth as well as by the rate that they exit by mortality or growth (Van Sickle 1977). For instance, one can predict that species with slower growth would have steeper diameter distributions, all other things being equal (Condit et al. 1998). Therefore, any factor capable of altering baseline patterns in recruitment, growth and mortality will eventually affect diameter distributions. Often cited examples of these factors are disturbance events and competition (Goff & West 1975, Coomes & Allen 2007, Toledo et al. 2013). And since demographic rates are often correlated with species traits (Poorter et al. 2008), an indirect relationship between species traits and the shape of diameter distributions is also expected (Wright et al. 2003).

Although there is great variation in growth and mortality related to site, neighborhood, and genetic effects, one of the most important effects certainly is tree size. The way in which growth and mortality are related to tree size is important in shaping diameter distributions (Goff & West 1975, Kohyama et al. 2003, Bin et al. 2012). Simulation studies have shown that some patterns of growth and mortality can lead to different shapes of diameter distribution (Lorimer & Frelich 1984, Leak 2002). Empirical studies also found similar results, but they were not very reliable at stand-level (Muller-Landau et al. 2006) or non-existent for some species (Condit et al. 1998). More specifically, Muller-Landau et al. (2006) provided explicit theoretical predictions of size distributions based on realized growth-dbh and mortality-dbh curves. But besides the possibility of growth-mortality relationships (Wyckoff & Clark 2000, Rüger et al. 2011), there is a debate on under which circumstances

the inference of size distributions from demographic rates (or vice versa) would be appropriate. Harcombe (1987) once stated that such inference would only be possible if populations are at demographic equilibrium, a view shared by other researchers (Kohyama et al. 2003, Muller-Landau et al. 2006). Anyhow, it would be very useful to infer general demographic patterns from static size distributions (Van Sickle 1977), especially because estimating tree demography directly takes a lot of time, labor, and money.

In this study, we empirically test how current tree size distributions of individual species are related to their demographic rates. More specifically, we assess how species diameter distributions are related to the shape of their mortality and growth curves and to their trends in recruitment over time. We use 30 years of data on individual tree recruitment, growth, and mortality from the 50 ha Barro Colorado Island (BCI) forest plot, allowing the evaluation of long-term trends for over 100 tree species. We evaluate three basic predictions: (i) decreases in growth or mortality with diameter are associated with less steep diameter distributions; (ii) higher recruitment inputs are associated with steeper size distributions; (iii) since population growth occurs by the increase of recruits, increasing populations will also have more steep distributions; and (iv) trees with larger maximum heights will have steeper size distributions than shorter trees, because taller trees tend to have increasing growth-dbh curves (Kohyama et al. 2003, King et al. 2006). In addition, we test whether species diameter distributions are related to growth and mortality parameters as would be expected under demographic equilibrium (Muller-Landau et al. 2006).

Methods

Study site and plot census

Barro Colorado Island (BCI) is located in Gatun Lake, Panama, and is covered in semi-deciduous moist tropical forest. Mean temperature is 31°C and mean rainfall is 2551 mm y⁻¹ with a four-month dry season from mid-December until mid-April (Leigh et al. 2004). The forest dynamics plot has an area of 50 ha (1000×500 m) and is located in a fairly flat area (120 to 160 m above sea level) in the center of the island (9°9' N, 79°51' W). The plot is mostly old-growth forest undisturbed for at least 400 years, except for approximately 2 ha covered by secondary forest of around 100 years old. These and other details on the physical environment, forest structure and composition can be found in Leigh et al. (2004). During 1981-1983, all free-standing, woody stems ≥10 mm dbh (diameter at breast height – 1.3 m, or above buttresses) inside the 50-ha plot were tagged, identified to species, and measured for dbh with precision of 1 mm (detailed methods are given in Condit 1998). New censuses were

carried out in 1985 and then every five years until 2010, resulting in a total of seven censuses. At each recensus, the number of new recruits and new deaths were recorded and all recruits and surviving individual had their dbh (re)measured.

Fitting size distributions

Year 2010 diameter distributions were fitted for all species with secondary stem growth (no palms), 30 or more individuals and whose adults are generally taller than 4 m and thicker than 80 mm. The last criterion excludes small shrubs whose life cycles are often completed below the dbh cutoff, thus concealing a large proportion of their size distribution. A total of 180 species met these criteria for analysis. Species were divided into four growth forms based on height of the adult plant (Hubbell & Foster 1986): large shrubs (hereafter referred to simply as shrubs), understorey, midstorey, and canopy trees.

Species dbh distributions were fitted using the Weibull distribution (Bailey & Dell 1973), a commonly used distribution which provides a good description of tree diameter distributions (Muller-Landau et al. 2006, Lima et al. 2013). Because the dbh data is left censored, we used the truncated Weibull that takes the form:

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-x_{min}}{\alpha}\right)^{\beta}},$$

where β and α are the shape and scale parameters, and x_{min} is the minimum size for which dbh data are available. The Weibull reduces to an Exponential (or reverse-J) distribution when β equals one, to steeper distributions when $\beta < 1$ and unimodal distributions when $\beta > 1$. When β approaches 3.6, the Weibull approximates the Normal distribution (Bailey & Dell 1973). The parameter α is closely related to the median of the distribution. We set x_{min} at 9.5 mm, which simply is minimum dbh (10 mm) minus half of the dbh measurement precision (0.5 mm), to account for the fact that trees measured as 10 mm dbh in fact include trees 9.5 – 10.5 mm dbh. In the case of multiple stemmed individuals, we used only the dbh of the largest stem.

Demographic analyses

For each species at each census interval we obtained the annual diameter growth (g), mortality (m), recruitment (r) and population growth (λ) that were calculated as follows (Condit et al. 1999):

$$g = \frac{\text{dbh}_t - \text{dbh}_0}{t},$$

$$m = \frac{\ln N_0 - \ln S_t}{t},$$

$$r = \frac{\ln N_t - \ln S_t}{t},$$

$$\lambda = \frac{\ln N_t - \ln N_0}{t},$$

where dbh_t and dbh_0 are the individual dbh at times t and at time 0, N_t and N_0 are the population sizes at times t and 0, and S_t is the number of survivors at time t . To calculate g , the first census interval was excluded due to measurements taken around buttresses and due to dbh rounding for trees less than 55 mm dbh. To account for dbh rounding in the second census, the dbh of the third census was also rounded down before calculating g , a procedure which introduces an acceptable amount of bias in g estimates of small individuals (Condit et al. 1993). In addition, we excluded from growth analysis extreme cases in which individual stems shrunk $>25\%$ or grew $>75 \text{ mm y}^{-1}$ in dbh (Condit et al. 1999, Rüger et al. 2011). Values of population growth $\lambda < 0$ means population decline, while $\lambda > 0$ means the opposite.

Species growth-dbh curves were constructed by averaging individual growth rates per size class, curves that were obtained separately for the six census intervals. The use of averages was adopted to deal with negative growth observations related to measurement errors. The size of the bins was defined for each species at each census interval based on the quantiles of the dbh distribution. The number of bins was defined according to species abundances, namely four, five, six and ten bins for species <100 , 100–500, 500–1000, and >1000 individuals, respectively. Binning growth into size classes provided estimates that were reasonably correlated with the ones obtained by Rüger & Condit (2012) in a more complicated analysis that explicitly accounted for dbh measurement errors (Appendix 1). In contrast, size-class binning was not used to obtain mortality-dbh curves. We simply ascribed to each tree an initial dbh and a dead or alive status at the end of each census interval. This procedure may have resulted in biases towards small individuals that generally dominate size distributions. However, size-class binning often resulted in dbh classes with average mortality equals zero, especially for species with few deaths, making the power-law fitting (see below) unfeasible. For species with large numbers of deaths, estimates based on individuals and based on dbh class averages were generally similar.

The shape of growth and mortality rates as functions of tree dbh was described using power-law functions, a relationship between growth-dbh and mortality-dbh assumed in previous theoretical studies (Muller-Landau et al. 2006). Mean growth rates per size-class were fit as a power function of mean size-class dbh using non-linear regression with normal errors. Individual tree mortality was fit as a power function of initial tree dbh using non-linear

logistic regression with binomial errors. More specifically, we modeled mean growth and mortality as follows:

$$\text{Mean } g \sim \text{Normal} (\mu = a_g \text{dbh}^{b_g}, \sigma^2),$$

$$\text{Death} \sim \text{Binomial} (p = [1/(e^{a_m + \ln(\text{dbh})^{b_m}} + 1)]/t, n = 1),$$

where μ and σ^2 are the mean and variance parameters of the Normal distribution, p and n are the probability of success (deaths in this case) and number of trials of the Binomial distribution and t is the time interval between consecutive censuses. Note that dbh is the size-class mean of initial tree dbh for growth, and the individual initial tree dbh for deaths. Models were fitted separately for each census interval and the b_g and b_m exponents were averaged over census intervals. As for dbh analysis, we have only fitted growth and mortality curves for census intervals which had a least 30 individuals of a given species. However, for a few species we pooled census intervals together to have enough death observations for a confident estimate of b_m . We excluded species that had less than 30 individuals in all censuses prior to 2010 or less than 30 deaths in the 30 years of monitoring. In total, 170 species met these criteria for analysis.

Relating size distributions and demography

The relationship between dbh distributions and demographic rates was evaluated by regressing the Weibull shape parameter β on b_g , b_m and recruitment rate using generalized linear models, which were conducted for all species and for each growth form separately. Because β had a skewed distribution, we used generalized linear models with Gamma error distribution. The amount of variation explained by the regressions was evaluated by the pseudo- R^2 (i.e. $1 - \text{residual deviance}/\text{null deviance}$; Zuur et al. 2009) and F -tests were performed to test the effects of demographic variables on β . The same approach was used in all other regressions. Although the Weibull α was generally strongly correlated with the mean/median of the dbh distribution and, consequently, with growth form, α estimates often resulted in really small values ($0.0004 < \alpha < 5$) for heavily right-skewed distributions (observed $\beta < 0.35$; total of 34 species). The value of 0.35 was set as a rough threshold from which α tended towards such small values. We fixed α at zero to check whether these very low estimates were related to estimates being on the boundary of the parameter space. It turned out that fits using α fixed at zero generally provided worse fits, meaning that estimates were not on the boundary. But since these low- α estimates were related to a restriction of the

truncated Weibull in describing distributions with great contributions of small and extreme values at the same time, α was often omitted from the analysis.

The theoretical predictions regarding the relationship between growth and mortality and the Weibull parameters were assessed using the following equations (Muller-Landau et al. 2006):

$$\hat{\beta} = 1 - b_g,$$

$$\hat{\alpha} = \left(\frac{a_g \times (1 - b_g)}{\bar{m}} \right)^{1/(1 - b_g)},$$

where b_g and a_g are the parameters of the fitted power-law growth curve and $\bar{m}$ is the mean mortality rate obtained by averaging the mortality rate over size-classes of all census intervals. Since there is an uncertainty associated with the estimates of both predicted and observed Weibull parameters, the relationship between them was tested using standardized or reduced major axis regression (Smith et al. 2009), with both variables log-transformed first. Although the majority of species had monotonic growth-dbh curves that were well described by power-laws, some species had non-monotonic curves (see Appendix 2 for details). Given that the theoretical predictions above are based on power-law growth (Muller-Landau et al. 2006), these species were eliminated from this analysis. The shape of the recruitment curve over time was determined in the same way as employed for growth and mortality curves as a function of size (Appendix 2). Truncated Weibull and power-laws were fitted using maximum likelihood techniques and numerical optimization (Burnham & Anderson 2002, Bolker 2008). All fits were visually inspected to check for convergence problems. When such problems were detected, we varied the starting values of parameters and/or the optimization method (Bolker 2008) until convergence was reached. All analyses were performed using R version 2.15.1(www.R-project.org).

Results

Species varied widely in their diameter distributions. Diameter distributions were markedly right-skewed with 75% of species having Weibull parameter $\beta < 1$ (Figure S3, Appendix 3). Estimates of β ranged from values ≤ 0.15 (e.g. *Ocotea puberula* and *Platymiscium pinnatum*) to > 2 (e.g. *Gustavia superba*, *Herrania purpurea* and *Talisia nervosa*). No species had truly bell-shaped distributions ($\beta \approx 3.6$ - Appendix 4), a feature restricted to palms (not included in the analysis). Shrub and understorey species had more modal distributions (higher β) while midstorey and canopy species tended to be more right-skewed (Table 1). There was no

influence of species abundance on the shape of the diameter distribution ($F=0.70$, $df=178$, $p=0.40$; abundance was log-transformed).

For the majority of species (ca. 80%) growth rate with dbh increased whereas mortality decreased with dbh for about two thirds of the species (Figure S4, Appendix 3). Observed mortality curves were much more variable among census intervals than growth curves, as reflected in higher values of standard deviation between censuses for the power-law exponent b_m than for b_g (Welch t -test: $t=2.1$, $df=134.1$, $p=0.035$). The growth and mortality exponents were not significantly related to each other (Figure S5, Appendix 3) and they varied between growth forms, especially b_m (Table 1). The exponent b_g was also not related to species abundance ($F=1.8$, $p=0.18$), but the same was not true for b_m ($F=23.8$, $p<0.001$), which showed a positive relationship with abundance. Overall, b_g was positively related to species overall growth $F=9.5$, $p=0.003$, but this relationship did not hold for shrubs and understorey species; in contrast, b_m was negatively related to overall mortality for all growth forms ($F=6.9$, $p=0.01$).

Around 60% of the species had negative mean population growth rate λ over the 30 years of monitoring. However, the assessment of λ over time showed that this rate was not always negative and there was a community-wide tendency towards positive rates in the last census intervals (Figure 1). Population decline reflects a lack of recruits to replace deaths and thus can be caused by decreases in recruitment or increases in mortality. In the BCI plot, it seems that community-wide fluctuation in λ was more explained by recruitment than by mortality, which was more constant over the past 30 years. At species-level, 76% of the species had the lowest value of recruitment between 1990 and 2000. The same was true for λ for 60% of the species, but not for mortality (17% of the species). Consequently, λ was much more correlated with recruitment (partial $R^2=0.89$) than mortality (partial $R^2=0.01$ – multiple regression of mean λ on log-transformed means of r and m : $F=726.0$, $df=167$, $p<0.0001$, $R^2=0.90$). There was no evidence of an association between λ and overall growth rate ($F=1.384$, $p=0.241$). In addition, we found that 21.7%, 12.2% and 39.4% of species had decreasing, increasing and u-shaped net recruitment curves, respectively, while the remainder (26.7%) showed no clear trend over time. There was only weak evidence of a relationship between mean recruitment and b_g or b_m (Figure S5, Appendix 3).

There was strong evidence that the shape of growth curves and mean recruitment rate influenced the shape parameter β of the fitted Weibull distribution, but there was no evidence of β being influenced by the shape mortality curves (Figure 2). Overall, species with decreasing growth curves had less steep diameter distributions (higher β), while species with

higher recruitment rate had steeper distributions (lower β). Indeed, the species with highest β , *T. nervosa*, was the one with lowest b_g (Figure 3). In addition, there were many examples of highly skewed distributions associated with high recruitment (e.g. *Miconia argentea*) and vice-versa (e.g. *Celtis schippii* – Appendix 3). These overall results were practically the same when we consider growth forms separately (Appendix 5). The three demographic rates combined explained a large proportion of the variation in β (multiple regression: $F=98.4$, $df=132$, $p<0.001$, $\text{pseudo-}R^2=0.69$). The growth exponent explained a larger proportion of the variation in β than recruitment. Overall results using types of shape instead of the exponents of the growth and mortality curves were very similar ($F=42.4$, $df=130$, $p<0.0001$, $\text{pseudo-}R^2=0.61$). The above results are for the 136 species that showed monotonic growth and mortality curves. Results were the same including non-monotonic species ($F=121.3$, $df=166$, $p<0.0001$, $\text{pseudo-}R^2=0.68$), except that the effect of b_m that became marginally significant. And since λ and recruitment were highly correlated, declining populations tended to present less steep diameter distributions, but this relationship was highly variable (Figure 1). If we consider only species with growth and mortality constant in respect to tree dbh ($N=15$), this relationship becomes much stronger ($R^2=36\%$, $p=0.009$).

We found that the observed Weibull parameters were proportional to those expected based on the analytical derivations of Muller-Landau et al. (2006). Although the slope of the major axis regression was significantly different than one for β , it seems that the relationship between observed and predicted estimates was more predictable for this parameter than for α (Figure 4). It is noteworthy that besides species with non-monotonic growth curves (see Appendix 2 for how we defined this), species with highly skewed distributions (observed $\beta < 0.35$) were not included in this analysis. The inclusion of these species in the analysis does not change the results for β , but the slope of the relationship between observed and predicted α deviates greatly from one (results not shown).

Discussion

As it could be expected from first principles, current tree size distributions were indeed related to species demographic rates. Almost 70% of the variation in the shape of species size distributions was explained by the shape of growth-dbh curves and by recruitment over time. These findings support the suggestion of previous studies that size-dependent growth rate is the primary determinant of the shape of tree size distributions (Condit et al. 1998, Leak 2002, Kohyama et al. 2003, Muller-Landau et al. 2006). This result partially supports our first prediction of less steep diameter distributions being associated to decreasing growth and

mortality curves, because size-dependent mortality was not related to the shape of size distributions. The high variability in the mortality patterns between censuses may be one of the causes of low relationship with the shape of size distributions. Although there was a 30 years census interval, many species had mortality curves so variable among census intervals or so few death events that they did not present any evident pattern, making it difficult to have a good estimate of its shape (i.e. b_m) and thus a reliable correlation with size distributions. For long-lived organisms such as trees, it seems that even longer time intervals will be necessary to properly describe mortality patterns for several BCI species.

We have also shown here that besides size-dependent growth, the mean recruitment rate of species also plays an important role in shaping tree size distributions. Higher recruitment inputs were associated with steeper size distributions, thus supporting our second prediction and the suggestions of Bin et al. (2012), which found that species with modal diameter distributions had lower recruitment rates. It is straightforward to understand that if recruitment exceeds the number of individuals that die or grow out from the small size classes, there will be a packing of individuals in these size classes, making the size distribution to become steeper. The relationship between recruitment and size distribution reinforce the result found here that there was no relationship between recruitment rates and the shape of growth curves (i.e. b_g), otherwise higher juvenile growth (decreasing growth curves) would balance out this packing in small size classes. Therefore, fluctuations in recruitment rates may be important to shape of tree size distributions. Since El Niño events enhance fruit production at BCI (Wright et al. 1999), the relatively high rate of recruitment in between 1983-85 and the increase from 1995 on may be related to the strong 1982-83, 1991-92 and 1997-98 El Niño events, whereas the low recruitment rate between 1990 and 1995 may be related to the strong La Niña of 1988-1989 (NOAA 2013). One could expect higher recruitment and therefore steeper size distributions in the years following strong El Niño events. However, since these climate events may also affect patterns of tree growth and mortality in respect to dbh (Condit et al. 1995, 2004, Nakagawa et al. 2000, but see Williamson et al. 2001), we must consider the possible interplay among recruitment, growth and mortality to assess El Niño effects on size distributions. One must also consider that El Niño effects on mortality and growth are prompt, while such effects are delayed for recruitment.

Predicted and fitted Weibull parameters presented a positive linear relationship, reinforcing that the Weibull is not only a good phenomenological description it is also a valid model of the demographic process underlying species tree size distributions (Muller-Landau

et al. 2006). Despite of this positive relationship and the slope not different than one for the scale parameter, there was a lack of strong relationship for both Weibull parameters, which has some possible explanations. First, species are not at equilibrium and/or their mortality-dbh curves are not constant in respect to tree size, which represent violations of the underlying assumptions of demographic equilibrium theory used to predict the parameters from demographic data (Muller-Landau et al. 2006). For instance, the predicted parameters do not account for the recruitment variation effect on the shape of the size distribution, which may be seen as an indication of non-equilibrium. Besides recruitment, mortality also tended to be high or to increase after strong El-Niño events (Figure 1; see also Nakagawa et al. 2000), or at least for certain dbh classes (Condit et al. 1995), suggesting that such events may be a primary cause of departures from equilibrium at BCI. On the other hand, around 65% of species did not have dbh independent mortality and since the shape of mortality curves was estimated for census intervals separately, it is not likely that dbh dependent mortality is an indirect effect of El-Niño events. Therefore, although there was no community-wide relationship between shape of size distributions and mortality curves, this effect may have been important for individual species. On the other hand, predicted and fitted parameters were obtained under certain assumptions. First, we assumed that the Weibull is a good description for dbh distributions, which is likely true for many but not all species (Appendix 3). For certain species, more flexible functions may be needed for a proper characterization of diameter distributions (Muller-Landau et al. 2006, Lima et al. 2013). We also assumed that power-laws were good description of both growth and mortality curves, which is a quite strong assumption. In addition, we assumed here that growth or mortality were functions of mean size-class dbh and initial dbh at the beginning of each census interval, whereas Muller-Landau et al. (2006) assumed instantaneous growth and mortality were functions of dbh at continuous time, which are the basis for the demographic equilibrium predictions of size distributions. Although this may have produce some noise in the relationship between predicted versus observed Weibull parameters, we think that this effect is relatively small. All the above-mentioned sources of variation, particularly fluctuations in the recruitment rate, make it difficult to predict growth and mortality from size distributions. However, the overall relationship between predicted and fitted Weibull parameters was fairly strong and can be used to predict size distributions at least for species at demographic equilibrium.

Although we have found a trend towards declining populations having more modal distributions, this result was only true among canopy species. For other species, the combination of higher juvenile growth (decreasing growth-dbh curve) with higher adult

mortality (increasing mortality-dbh curves), was able to generate modal diameter distributions as well. One good example was *Herrania purpurea*, a shrub with modal distribution and increasing population associated to decreasing growth-dbh curves and increasing mortality curves. An opposite example is *Poulsenia armata*, a canopy species with modal distribution and decreasing growth-dbh curve, but with negative population growth over the last 30 years. Interestingly, the strong relationship between recruitment and population growth was not translated into a strong relationship between population growth and size distributions. Weak overall relationships between size distributions and population growth were also found in other forests and have been attributed to density-dependent mortality within recruits, sporadic or poor recruitment and fast juvenile growth (Kohira & Ninomiya 2003, Feeley et al. 2007, Bin et al. 2012). Bin et al. (2012) also suggested that u-shaped growth and/or mortality curves are also able to generate modal size distributions. In this study, we found no evidence of species with u-shaped mortality curves having more modal distributions than species with constant or increasing mortality curves (Appendix 2). Anyway, the results found here and elsewhere reinforce that modal size distributions do not necessarily mean that populations are in decline, which could be valid for some growth forms or for cases of constant growth and mortality curves (Condit et al. 1998).

Consistent with previous studies (King et al. 2006), different growth forms had different patterns of growth and mortality, resulting in a clear distinction in the shape of size distribution among them. The inclination of the growth-dbh curves (b_g) was lowest for shrubs and highest for canopy species, revealing opposite strategies of investments in growth with tree diameter. The shape of mortality-dbh curve (b_m) varied less among growth forms and only canopy species presented more steeply declining curves than others. Species with increasing growth curves and decreasing mortality-curves may reflect decreasing competition for light (Coomes & Allen 2007), a strategy that fitted well for canopy species. We also found that species with n-shaped and u-shaped growth and mortality curves, respectively, were mainly midstorey and canopy species (Appendix 2). This is in accordance with the suggestion that after a certain size senescence and exogenous disturbance dominates over competition as the main source of constraint for tree growth and survival (Goff and West 1975, Coomes & Allen 2007). However, since these differences only affect the right-end of the distribution, we found no strong evidence that non-monotonic growth or mortality curves had an effect on the fitted parameters of the Weibull distributions. These results support the idea that trees with different life-history strategies may have different size distributions (Wright et al. 2003). But it also reinforces that the differences between such strategies are in fact related to their

demographic patterns. Therefore, one could expect differences in size distributions between forests with similar proportion of life-histories but different conditions for growth (e.g. canopy openness) or, to a lesser extent, different mortality drivers.

Acknowledgements

We would like to thank the Center for Tropical Forest Science (CTFS), especially Stuart Davies, for support. We are grateful to R. Foster, S. Lao, R. Pérez, S. Aguilar, A. Hernandez, and many field assistants for species identifications, tree measurements, and database management. We also thank Nadja Rüger for her useful comments and for providing the digital version of the growth rate estimates of BCI species. The 30 years of monitoring of the BCI plot was funded by the US National Science Foundation, the John D. and Catherine T. MacArthur Foundation, the World Wildlife Fund, the Earthwatch Center for Field Studies and the Smithsonian Tropical Research Institute. RAFL was supported by CAPES and PIP was supported by CNPq (no. 303878/2008-8). This manuscript was substantially advanced during the 2012 CTFS/SIGEO analytical workshop funded by the US National Science Foundation (DEB-1046113).

References

- Bailey, R. L., and T. R. Dell. 1973. Quantifying diameter distributions with the Weibull function. *Forest Science* 19:97-104.
- Bin, Y., W. Ye, H. C. Muller-Landau, L. Wu, J. Lian, and H. Cao. 2012. Unimodal tree size distributions possibly result from relatively strong conservatism in intermediate size classes. *Plos One* 7:e52596.
- Bolker, B. M. 2008. *Ecological models and data in R*. Princeton University Press, Princeton.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multi-model inference: a practical information-theoretic approach*. 2nd ed. Springer, New York.
- Condit, R. 1998. *Tropical Forest Census Plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots*. Springer-Verlag, Georgetown.
- Condit, R., S. P. Hubbell, and R. B. Foster. 1993. Identifying fast-growing native trees from the Neotropics using data from a large, permanent census plot. *Forest Ecology and Management* 62:123-143
- Condit, R., S. P. Hubbell, and R. B. Foster. 1995. Mortality rates of 205 neotropical tree and shrub species and the impact of a severe drought. *Ecological Monographs* 65:419-439.

- Condit, R., R. Sukumar, S. P. Hubbell, and R. B. Foster. 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. *American Naturalist* 152:495-509.
- Condit, R., P. S. Ashton, N. Manokaran, J. V. LaFrankie, S. P. Hubbell, and R. B. Foster. 1999. Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50 -ha plots. *Philosophical Transactions of the Royal Society of London (Series B)* 354:1739-1748.
- Condit, R., S. Aguilar, A. Hernandez, R. Perez, S. Lao, G. Angehr, S. P. Hubbell, and R. B. Foster. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Nino dry season. *Journal of Tropical Ecology* 20:51-72.
- Coomes, D. A., and R. B. Allen. 2007. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology* 95:27-40.
- Feeley, K. J., S. J. Davies, M. Noor, A. R. Kassim, and S. Tan. 2007. Do current stem size distributions predict future population changes? An empirical test of intraspecific patterns in tropical trees at two spatial scales. *Journal of Tropical Ecology* 23:191-198.
- Goff, F. G., and D. West. 1975. Canopy-understory interaction effects on forest population structure. *Forest Science* 21:98-108.
- Harcombe, P. A. 1987. Tree life tables. *Bioscience* 37:557-568.
- Hubbell, S. P., and R. B. Foster. 1986. Commonness and rarity in a Neotropical forest: implications for tropical tree conservation. Pages 205-231 *in* M. E. Soulé editor. *Conservation biology: The science of scarcity and diversity*. Sinauer Associates, Sunderland.
- Hyink, D. M., and J. W. Moser, Jr. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. *Forest Science* 29:85-95.
- Kohira, M., and I. Ninomiya. 2003. Detecting tree populations at risk for forest conservation management: using single-year vs. long-term inventory data. *Forest Ecology and Management* 174:423-435.
- Kohyama, T., E. Suzuki, T. Partomihardjo, T. Yamada, and T. Kubo. 2003. Tree species differentiation in growth, recruitment and allometry in relation to maximum height in a Bornean mixed dipterocarp forest. *Journal of Ecology* 91:797-806.
- Leak W. B. 2002. Origin of sigmoid diameter distributions. Research Paper NE-178, 1-10. USDA Forest Service, Newtown Square, PA, USA.
- Leigh Jr., E. G., S. L. Lao, R. Condit, S. P. Hubbell, R. B. Foster, and R. Perez, 2004. Barro Colorado Island forest dynamics plot, Panama. Pages in E. C. Losos, and E. G. Leigh Jr.

- E.G. editors. Tropical forest diversity and dynamism: Findings from a large-scale plot network. University of Chicago Press, Chicago.
- Lima, R. A. F., J. L. F. Batista, and P. I. Prado. 2013. Modeling tree diameter distributions in tropical forests. *Forest Ecology and Management* (submitted).
- Lorimer, C. G., and L. E. Frelich. 1984. A simulation of equilibrium diameter distributions of Sugar Maple (*Acer saccharum*). *Bulletin of the Torrey Botanical Club* 111:193-199.
- Lorimer, C. G., S. E. Dahir, and E. V. Nordheim. 2001. Tree mortality rates and longevity in mature and old-growth hemlock-hardwood forests. *Journal of Ecology* 89:960-971.
- Muller-Landau, H. C., R. Condit, K. E. Harms, C. O. Marks, S. C. Thomas, S. Bunyavejchewin, G. Chuyong, L. Co, S. Davies, R. B. Foster, S. Gunatilleke, N. Gunatilleke, T. B. Hart, S. P. Hubbell, A. Itoh, A. R. Kassim, D. Kenfack, J. V. LaFrankie, D. Lagunzard, H. S. Lee, E. Losos, J. Makana, T. Ohkubo, C. Samper, R. Sakumar, I-F. Sun, N. Supardi, S. Tan, D. Thomas, J. Thompson, R. Valencia, M. I. Vallejo, G. V. Muñoz, T. Yamakura, J. K. Zimmerman, H. S. Dattaraja, S. Esufali, P. Hall, F. He, C. Hernández, S. Kiratiprayoon, H. S. Suresh, C. Wills, and P. Ashton. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology Letters* 9:589-602.
- Nakagawa, M., K. Tanaka, T. Nakashizuka, T. Ohkubo, T. Kato, T. Maeda, K. Sato, H. Miguchi, H. Nagamasu, K. Ogino, S. Teo, A. A. Hamid, and L. H. Seng. 2000. Impact of severe drought associated with the 1997-1998 El Nino in a tropical forest in Sarawak. *Journal of Tropical Ecology* 16:355-367.
- Newton, P. F., Y. Lei, and S. Y. Zhang. 2005. Stand-level diameter distribution yield model for black spruce plantations. *Forest Ecology and Management* 209:181-192.
- NOAA - National Weather Service, National Centers for Environmental Prediction, Climate Prediction Center 2013. Historical El Nino/La Nina episodes (1950-present). At: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml
- Poorter, L., S. J. Wright, H. Paz, D. D. Ackerly, R. Condit, G. Ibarra-Manríquez, K. E. Harms, J. C. Licona, M. Martinez-Ramos, S. J. Mazer, H. C. Muller-Landau, M. Peña-Claros, C. O. Webb, and I. J. Wright. 2008. Are functional traits good predictors of demographic rates? Evidence from five neotropical forests. *Ecology* 89:1908-1920.
- Rüger, N., A. Huth, S. P. Hubbell, and R. Condit. 2011. Determinants of mortality across a tropical lowland rainforest community. *Oikos* 120:1047-1056.
- Rüger, N., and R. Condit. 2012. Testing metabolic theory with models of tree growth that include light competition. *Functional Ecology* 26:759-765.

- Smith, R. J. 2009. Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology* 140:476–486.
- Toledo, J. J., W.E. Magnusson, and C.V. Castilho. 2013. Competition, exogenous disturbances and senescence shape tree size distribution in tropical forest: evidence from tree mode of death in Central Amazonia. *Journal of Vegetation Science* 24:651–663.
- Van Sickle, J. 1977. Mortality rates from size distributions. The application of a conservation law. *Oecologia* 27:311–318.
- Williamson, G. B., W. F. Laurance, A. A. Oliveira, P. Delamonica, C. Gascon, T. E. Lovejoy, and L. Pohl. 2000. Amazonian tree mortality during the 1997 El Nino drought. *Conservation Biology* 14:1538–1542.
- Wright, S. J., C. Carrasco, O. Calderon, and S. Paton. 1999. The El Nino Southern Oscillation variable fruit production, and famine in a tropical forest. *Ecology* 80:1632–1647.
- Wright, S. J., H. C. Muller-Landau, R. Condit, and S. P. Hubbell. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84:3174–3185.
- Wyckoff, P. H., and J. S. Clark. 2000. Predicting tree mortality from diameter growth: a comparison of maximum likelihood and Bayesian approaches. *Canadian Journal of Forest Research* 30:156–167.
- Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York.

Supplemental Material

Appendix 1. Comparison of power-law fits of growth used in this study with other approaches.

Appendix 2. Classification of growth and mortality curves into different types of shape and the comparison with their power-law fits.

Appendix 3. Descriptive results of Weibull fitted to species diameter distributions, of the power-law function fitted to growth and mortality curves and of the estimates of average recruitment.

Appendix 4. Diameter distributions of the 180 species evaluated in this study and their truncated Weibull fits.

Appendix 5. Results of the analyses separated by species growth forms.

Tables

Table 1. Summary of the Weibull parameters (β and α), of the power-law exponents fitted to growth-dbh (b_g) and mortality-dbh curves (b_m) and of the overall recruitment (r) and population growth rates (λ) per growth forms. Values are the mean $\pm$ standard deviation and F -test is the result of the analysis of deviance. Lowercase letters indicate whether group averages differ significantly. Values of r and λ are multiplied by one hundred. Since the power-law could not be fitted for some of the growth and/or mortality curves, the number of degrees of freedom (df) varied between b_g and b_m .

Parameter	Growth form				F -test (df)
	Shrub	Understorey	Midstorey	Canopy	
n	13	39	51	77	
Shape (β)	$1.20 \pm 0.37a$	$1.02 \pm 0.57ab$	$0.83 \pm 0.44b$	$0.55 \pm 0.34c$	15.3 (176)
Scale (α)	$20.1 \pm 9.5a$	$25.0 \pm 21.3a$	$49.2 \pm 50.1ab$	$73.3 \pm 131.9b$	6.4 (176)
b_g	$-0.10 \pm 0.18a$	$0.12 \pm 0.31b$	$0.33 \pm 0.31c$	$0.45 \pm 0.18d$	67.9 (144)
b_m	$-0.04 \pm 0.36a$	$-0.12 \pm 0.40a$	$-0.12 \pm 0.41a$	$-0.31 \pm 0.39b$	3.3 (146)
r	3.70 ± 3.24	3.52 ± 2.95	4.04 ± 4.50	3.63 ± 2.47	0.2 (166)
λ	-0.70 ± 0.70	-0.22 ± 1.88	-0.42 ± 1.71	-0.31 ± 1.64	0.3 (166)

Figures

Figure 1. Results of the (a) community-wide demographic trends of 180 tree species over almost 30 years and of the (b) regression of the shape of size distribution (β) on mean population growth rate (λ). In the left panel, lines connect the average annual rates of mortality (dashed line), growth (dotted line) and population growth (solid line). The vertical lines indicate the 95% confidence intervals around these averages. Rates were multiplied by one hundred for clarity. The line at zero separates declining populations ($\lambda < 0$) from increasing populations ($\lambda > 0$). In the right panel, the solid line is the trend line of the generalized linear model whose main results are given in the legend. Note that in this panel, the y-axis is in log-scale for clarity.

Figure 2. Partial regression plots of the multiple generalized linear regression of the Weibull shape parameter β on (a) the growth-dbh curve exponent (b_g), on (b) the mortality-dbh curve exponent (b_m) and on (c) the values of mean recruitment rate. The growth-dbh and mortality-dbh curves are increasing when the exponent is greater than zero and decreasing when it is smaller than zero. The y-axis represents the residuals of β , while the x-axes represent the residuals of each demographic variable given other independent variables in the model. The solid line is the result of the generalized linear models whose partial R^2 and p -value are given.

Figure 3. Examples of species diameter at breast height (dbh) distributions together with their power-law fits of growth (solid line) and mortality (dashed line) as functions of dbh. (a) *Prioria copaifera*, (b) *Xylopia macrantha*, (c) *Maquira guianensis* and (d) *Talisia nervosa*. Maximum likelihood estimates of the Weibull shape (β) and scale (α) parameters are presented as well. Note that axes have different ranges among species.

Figure 4. Predicted vs. observed shape (left) and scale (right) parameters of the Weibull distribution together with major axis regression (solid line) and 1:1 lines (dashed line). Observed parameters were obtained from the fitted diameter distributions while predicted parameters were obtained from fitted growth- and mortality-dbh curves (see methods for details). The legend indicates the slope b of the major axis regression, its 95% confidence intervals in parenthesis and the squared correlation coefficient.

Appendix 1. Comparison of power-law fits of growth used in this study with other approaches.

In this study, we used a simple approach to fit power-law functions to growth-dbh curves of BCI species, which includes the binning of growth rates into size classes to deal with negative growth observations. Other studies have used the more appropriated approaches to estimate the power-law parameters to growth as functions of dbh, in particular the work of Rüger & Condit (2012). They took into account measurement errors to assess true growth rates and used hierarchical Bayesian models capable of fitting power-laws also for rare species. We decided not to use this approach because it is very compute-intensive. In the original paper the authors present the estimates of the effect of dbh on growth in models that take into account other predictive variables that could affect growth, such as light. So, we had to obtain estimates of dbh effect on models having only dbh as predictive variable, which were kindly provided by the authors.

Here we compare the estimates of the power-law exponent b of growth curves of Rüger & Condit (2012) to the ones obtained in this study to assess how well both estimates are correlated. Only the 1985-1990 and 1990-1995 census intervals were available for comparison and for a total of 84 BCI species. Although the slope estimates for both census intervals did not include one (i.e. perfect linear relationship), there was as strong positive relationship between the power-law estimates obtained taking into account error measurements (Rüger & Condit 2012) and the ones obtained in this study (Figure S1). In addition, since we have averaged the estimates between the six census intervals available in this study we expect that our estimates are less subject to inter-census variation in growth. And since we are interested in community-wide tendencies and not on species-specific estimates, we suppose that the conclusions of this study would remain unaltered if we have accounted for measurement errors to estimate growth rates.

Figure S1. The power law exponent b obtained in this study (x -axis) compared to the power-law exponent from hierarchical models accounting for measurement errors (y -axis), for two census intervals: 1985-1990 and 1990-1995. The solid line is the linear regression model which has slope b (95% confidence intervals in parenthesis). The dashed line is the 1:1 line.

Appendix 2. Classification of growth and mortality curves into different types of shape and the comparison with their power-law fits.

Curves of growth and mortality as a function of tree size can have different shapes, among which the most common are monotonic (i) constant, (ii) decreasing or (iii) increasing shapes and the non-monotonic (iv) n-shaped or (v) u-shaped curves (Harcombe 1987, Coomes & Allen 2007, Rüger et al. 2011). In this study, although the majority of the species could be well described by monotonic functions such as power-laws, other species had curves with clearly non-monotonic shapes. Therefore, a model selection procedure was performed to classify growth and mortality curves into monotonic and non-monotonic types of shape. This selection was carried by competing the fit of four candidate functions, namely constant ($y = \text{constant}$), quadratic ($y \sim a + bx + cx^2$), saturating ($y \sim bx/(a + x)$) and power-law ($y \sim ax^b$). First, functions were fitted to growth and mortality data using maximum likelihood procedures (Bolker 2008). To do so we used mean growth and mortality rates per dbh size classes (quantiles) and data from all census intervals were considered as a single dataset. As for the main analyses, the number of bins was defined according to species abundances, namely four, five, six and ten bins for species <100, 100–500, 500–1000, and >1000 individuals, respectively. Then, growth or mortality curves were classified into the five types of curve depending on the simultaneous comparison of the candidate models' performance, which was based on their Akaike Information Criteria (AIC) value (Burnham & Anderson 2002). Furthermore, we were extra parsimonious in cases of draws, i.e., differences between AIC values of the candidate models (ΔAIC) smaller than $\ln(8)$. In such cases, curves were assigned the model with fewer parameters.

Therefore, growth and mortality curves were classified as follows. Whenever the constant model was selected as a plausible model, growth and mortality curves were classified as constant (independent of dbh). Species were classified as increasing or decreasing when the constant model was not plausible, and when the saturating and/or power-law models were selected as plausible fits, independent of the performance of the quadratic model. The classification into increasing or decreasing depended on the sign of the b parameter of the saturating or power-law models. When the quadratic model was the sole plausible model, the curve was classified as n-shaped (concave) or u-shaped (convex) depending on the sign of the parameter c . Note that the inclusion of the saturating model was only to have an unequivocal separation between non-monotonic curves from saturating curves that were classified as decreasing or increasing here.

As a result we found that 13% and 11% of growth and mortality were non-monotonic in their shape (Table S1). Only one species had u-shaped growth curve or no species had n-shaped mortality curve. The vast majority of species with u-shaped mortality curves were canopy (55%) or midstorey (41%). For species with n-shaped with growth curves, we found a very similar results with canopy and midstorey representing 40% of the cases each. There was no evidence of shrubs with non-monotonic growth or mortality curves.

There was a good agreement between the classification in types of shape and the description of growth and mortality curves using power-laws, except by the non-monotonic curves for which the power-law tended to fit growth and mortality curves that were increasing and that were more or less constant in shape, respectively (Figure S2). Thus, species with non-monotonic growth and mortality curves were not included in most of the analysis presented in this study.

Table S1. Summary of the types of shape of the growth-dbh and mortality-dbh curves and the mean $\pm$ standard deviation of shape and scale parameters for each type of shape. *N*: number of species in each type of shape. NA: growth or mortality curve estimates not possible.

	Type of shape	<i>N</i> (%)	Shape (β)	Scale (α)
Growth	Decreasing	6.7	1.52 ± 0.45	49.7 ± 30.1
	Constant	18.9	1.08 ± 0.66	30.7 ± 37.1
	Increasing	59.4	0.60 ± 0.35	52.0 ± 107.3
	n-shaped	12.8	0.72 ± 0.34	82.8 ± 96.2
	u-shaped	0.6	–	–
	NA	1.6	–	–
Mortality	Decreasing	39.4	0.68 ± 0.46	71.6 ± 129.3
	Constant	31.7	0.81 ± 0.54	36.2 ± 47.8
	Increasing	13.9	0.89 ± 0.46	19.8 ± 14.1
	n-shaped	–	–	–
	u-shaped	11.1	0.93 ± 0.39	72.2 ± 86.4
	NA	3.9	–	–

Figure S2. Notched box plots showing the distribution the power-law exponent b for species classified into the five types of growth (left panel) and mortality curves (right panel). The width of the box plots is proportional to the number of species within each group. Types of shape are decreasing (DECR), constant (CNST), increasing (INCR), n-shaped (N-SHP) or u-shaped (U-SHP). The dotted line separates the values of b that describes increasing ($b > 0$) and decreasing curves ($b < 0$). The single species with u-shaped growth curve was omitted.

Appendix 3. Descriptive results of Weibull fitted to species diameter distributions, of the power-law function fitted to growth and mortality curves and of the estimates of average recruitment.

The fitted parameters of Weibull and power-law functions, as well as mean rates of recruitment to population growth, are given in Table S2 for each of the 180 species. The two Weibull parameters had asymmetrical distributions for the 180 shrub and tree species evaluated in this study, particularly α , both of them being well approximated by a Log-normal distribution (Figure S3). The estimates of the power-law exponents of the growth (b_g) and mortality curves (b_m) were more symmetrical, with a clear tendency of growth curves having more positive estimates (increasing growth curves). The mean $\pm$ standard deviation of the two exponents was 0.29 ± 0.29 and -0.18 ± 0.40 , respectively. The estimates for recruitment rate over the past 30 years were asymmetrical as well, and the majority of the species presented rates below 0.05.

As presented in the main text, there was no indication of a significant relationship between b_g and b_m or between these variables with mean recruitment rate (Figure S5). This means that although there may be a trade-off between overall growth and mortality rates or between overall growth and recruitment rates (Rüger et al. 2011), there was no evident relationship in the way that growth and mortality rates are related to tree size or that these relationships are related to species recruitment rates.

Figure S3. Frequency distribution of the Weibull (a) shape β and (b) scale α parameters fitted to 180 species diameter distributions. Both distributions were well approximated by Log-normal distributions (solid lines), whose maximum likelihood estimates are presented in the legend of each panel. For the shape parameter, the vertical dashed line separates the values of β that describes distributions with ($\beta > 1$) and without a mode ($\beta < 1$).

Figure S4. Frequency distribution of the fitted power-law exponents of the growth (a: b_g) and mortality curves (b: b_m), as well as the distribution of mean recruitment rates (c). For panels a and b, the vertical dashed lines separates the values of b that describes increasing ($b > 0$) and decreasing curves ($b < 0$).

Figure S5. Relationship between (a) the fitted power-law exponents of the growth (b_g) and mortality curves (b_m) to each other, and between (b) b_g and (c) b_m with mean recruitment rates (r). Each species is represented by a circle. None of the generalized regression model suggested a significant relationship between variables. Degrees of freedom were 134 for all analyses.

Table S2. Fitted parameters of the size distributions (shape β and scale α), growth and mortality curves (a_g , b_g and a_m , b_m), mean recruitment (r) and population growth rates (λ) for the 180 species included in this study. Note that although the growth and mortality fits are presented for all species, not all of them could be described by power-laws (see Appendix 2). Values refers to species with more than 30 individuals at the 2010 census. Species for which the fit has not been possible are indicated by NA.

Species	code	β	α	a_g	b_g	a_m	b_m	r	λ
<i>Acacia melanoceras</i>	acacme	0.98	24.45	NA	NA	NA	NA	9.30	7.80
<i>Acalypha diversifolia</i>	acaldi	1.08	14.67	2.38	-0.13	-2.43	-0.03	8.24	-1.80
<i>Acalypha macrostachya</i>	acalma	1.54	40.46	NA	NA	-3.11	0.28	10.42	-1.81
<i>Adelia triloba</i>	ade1tr	0.94	40.59	1.51	0.32	-1.12	-0.48	2.10	-3.09
<i>Aegiphila panamensis</i>	aegipa	1.54	93.03	0.38	0.34	0.21	-0.77	2.66	-4.21
<i>Alchornea costaricensis</i>	alchco	0.25	0.67	2.15	0.14	0.88	-0.89	4.06	-0.98
<i>Alibertia edulis</i>	alibed	0.75	7.97	0.35	0.26	-3.98	-0.05	3.73	1.24
<i>Allophylus psilospermus</i>	allops	0.51	10.53	0.29	0.45	-1.16	-0.59	2.94	-1.57
<i>Alseis blackiana</i>	alsebl	0.18	0.01	0.12	0.53	-3.22	-0.34	2.71	0.23
<i>Andira inermis</i>	andiin	1.30	62.81	0.25	0.22	-5.01	0.12	1.25	-0.77
<i>Annona acuminata</i>	annoac	0.78	6.50	0.57	-0.11	-4.03	0.10	3.57	0.00
<i>Annona spraguei</i>	annosp	0.41	3.15	4.10	-0.06	-0.50	-0.51	16.08	3.50
<i>Apeiba membranacea</i>	apeime	0.84	217.1	1.45	0.19	-1.31	-0.48	1.75	-1.02
<i>Apeiba tibourbou</i>	apeiti	0.91	93.22	29.01	-0.30	-0.78	-0.47	5.25	0.10
<i>Aspidosperma spruceanum</i>	aspicr	0.22	0.16	0.08	0.67	-4.37	-0.09	1.88	0.20
<i>Astronium graveolens</i>	ast2gr	0.26	2.07	0.29	0.46	-3.56	-0.19	3.26	1.55
<i>Beilschmiedia pendula</i>	beilpe	0.34	4.15	0.09	0.64	-2.86	-0.33	2.98	-0.37
<i>Brosimum alicastrum</i>	brosal	0.66	56.51	0.19	0.42	-5.68	0.19	1.09	0.09
<i>Calophyllum longifolium</i>	calolo	0.18	0.00	0.23	0.39	-3.90	0.10	6.63	3.61
<i>Capparis frondosa</i>	cappfr	1.83	29.48	0.47	-0.09	-3.53	-0.44	1.24	-1.24
<i>Casearia aculeata</i>	caseac	0.72	8.78	0.25	0.20	-3.34	-0.21	3.37	-0.12
<i>Casearia arborea</i>	casear	0.58	74.24	1.18	0.19	-1.73	-0.27	2.65	-2.77
<i>Casearia sylvestris</i>	casesy	0.60	27.93	0.56	0.23	-2.81	-0.12	2.67	-2.31
<i>Cassipourea elliptica</i>	cassel	0.56	11.94	0.19	0.53	-5.19	0.13	2.70	1.44
<i>Cavanillesia platanifolia</i>	cavapl	0.31	76.23	3.58	0.21	NA	NA	2.62	1.29
<i>Cecropia insignis</i>	cecrin	0.72	73.78	34.04	-0.34	0.80	-0.77	11.04	1.55
<i>Cecropia obtusifolia</i>	cecrob	1.14	97.35	121.9	-0.57	-0.95	-0.26	16.99	3.89
<i>Ceiba pentandra</i>	ceibpe	0.56	480.8	2.55	0.42	0.51	-0.80	3.91	-0.70
<i>Celtis schippii</i>	celtsc	1.69	121.1	0.51	0.37	-3.67	-0.04	0.52	-1.89
<i>Chamguava schippii</i>	cha2sc	0.77	11.34	0.42	0.23	-4.08	-0.09	5.16	3.91
<i>Chrysochlamys eclipses</i>	chr1ec	0.88	15.42	0.61	0.04	-2.04	-0.71	1.55	-0.47
<i>Chrysophyllum argenteum</i>	chr2ar	0.40	3.54	0.27	0.33	-4.23	0.04	3.91	1.84
<i>Chrysophyllum cainito</i>	chr2ca	0.18	0.00	0.07	0.74	-1.81	-0.82	5.65	3.28
<i>Coccoloba coronata</i>	cocco	0.26	0.02	0.20	0.74	-2.23	-0.63	3.71	-0.84
<i>Coccoloba manzinellensis</i>	coccma	1.59	26.16	0.68	-0.10	-3.55	-0.48	1.76	-0.80

Species	code	β	α	a_g	b_g	a_m	b_m	r	λ
<i>Cordia alliodora</i>	cordal	0.41	19.00	3.89	0.08	-1.17	-0.49	6.31	1.69
<i>Cordia bicolor</i>	cordbi	0.82	111.1	1.82	0.13	-0.69	-0.63	5.12	-0.09
<i>Cordia lasiocalyx</i>	cordla	1.02	51.06	2.20	-0.19	-3.82	0.10	2.71	-1.38
<i>Coussarea curvigemma</i>	cou2cu	0.90	23.34	0.58	0.16	-4.75	0.09	3.04	1.15
<i>Croton billbergianus</i>	crotbi	0.53	7.62	2.61	0.16	-0.45	-0.34	17.83	-0.12
<i>Cupania latifolia</i>	cupala	0.24	0.26	0.30	0.55	-3.97	0.09	3.58	-0.04
<i>Cupania rufescens</i>	cuparu	0.70	12.80	0.25	0.54	-5.74	0.56	6.45	2.48
<i>Cupania seemannii</i>	cupasy	0.73	7.93	0.17	0.51	-4.90	-0.27	3.05	1.07
<i>Dendropanax arboreus</i>	dendar	1.53	246.5	0.44	0.30	-2.75	-0.16	0.50	-2.86
<i>Desmopsis panamensis</i>	des2pa	1.24	20.03	0.71	-0.01	-5.94	0.71	2.82	-0.15
<i>Diospyros artanthifolia</i>	dio2ar	0.53	14.98	0.79	0.22	NA	NA	4.22	2.86
<i>Dipteryx oleifera</i>	diptpa	0.64	389.5	0.57	0.41	-0.85	-0.92	1.39	-0.40
<i>Drypetes standleyi</i>	drypst	0.63	26.31	0.13	0.56	-3.21	-0.39	1.56	-0.06
<i>Erythrina costaricensis</i>	ery1co	1.94	86.44	8.62	-0.34	-1.78	-0.29	0.78	-5.19
<i>Erythroxylum macrophyllum</i>	ery2ma	0.29	0.10	0.12	0.60	-3.41	0.06	4.54	-0.54
<i>Eugenia coloradoensis</i>	eugeco	0.55	15.62	0.12	0.64	-3.77	0.02	2.21	-0.88
<i>Eugenia galalonensis</i>	eugega	0.54	2.39	0.22	0.41	-6.49	0.68	4.54	2.69
<i>Eugenia nesiotica</i>	eugene	0.23	0.04	0.07	0.66	-5.65	0.24	2.24	-0.08
<i>Eugenia oerstediana</i>	euegoe	0.43	4.57	0.14	0.64	-3.38	-0.05	3.42	-0.34
<i>Famea occidentalis</i>	faraoc	1.26	46.47	0.46	0.17	-5.64	0.34	1.97	0.33
<i>Garcinia intermedia</i>	gar2in	0.74	13.88	0.21	0.42	-5.65	0.33	2.45	1.06
<i>Garcinia madruno</i>	gar2ma	0.90	20.21	0.34	0.26	-6.35	0.76	2.16	-1.55
<i>Genipa americana</i>	geniam	0.72	68.51	0.02	0.97	-3.66	-0.15	0.79	-1.06
<i>Guapira standleyana</i>	guapst	1.00	224.8	0.35	0.42	-1.58	-0.56	0.70	-1.49
<i>Guarea grandifolia</i>	guagr	0.32	4.81	0.15	0.60	-2.17	-0.50	2.80	0.71
<i>Guarea guidonia</i>	guargu	0.39	6.15	0.25	0.33	-3.14	-0.27	2.77	0.21
<i>Guarea bullata</i>	guarsp	0.70	18.23	0.17	0.25	-3.12	-0.06	2.05	-2.62
<i>Guatteria dumetorum</i>	guatdu	0.44	18.11	0.17	0.61	-3.28	-0.02	2.50	-2.06
<i>Guazuma ulmifolia</i>	guazul	0.70	113.4	4.02	0.16	0.56	-0.97	3.87	0.84
<i>Guettarda foliacea</i>	guetfo	1.38	54.45	0.90	-0.02	-3.06	-0.25	0.81	-1.41
<i>Gustavia superba</i>	gustsu	2.47	204.1	1.35	-0.05	-1.62	-0.69	0.73	-0.91
<i>Hampea appendiculata</i>	hampap	0.53	12.23	4.12	0.18	-2.08	0.06	17.76	2.18
<i>Hasseltia floribunda</i>	hassfl	1.11	87.11	0.39	0.34	-1.39	-0.44	1.27	-3.56
<i>Heisteria acuminata</i>	heisac	1.93	63.90	1.18	-0.02	-4.90	0.08	1.12	-0.49
<i>Heisteria concinna</i>	heisco	1.16	95.90	0.75	0.13	-4.80	-0.01	1.20	0.16
<i>Herrania purpurea</i>	herrpu	2.01	32.46	2.15	-0.39	-4.81	0.06	2.57	0.16
<i>Hirtella triandra</i>	hirttr	0.74	38.31	0.22	0.38	-3.51	-0.25	2.12	0.38
<i>Hura crepitans</i>	huracr	1.43	817.1	0.75	0.39	-0.60	-0.80	0.26	-1.05
<i>Hybanthus prunifolius</i>	hybapr	1.65	20.32	0.69	-0.14	-5.29	0.59	2.54	-1.16
<i>Hieronyma alchorneoides</i>	hyeral	0.26	4.23	0.43	0.43	-0.16	-0.81	4.17	0.49
<i>Inga cocleensis</i>	ingaco	0.25	0.09	0.36	0.56	-4.59	0.45	4.46	-3.34
<i>Inga laurina</i>	ingafa	0.95	48.38	0.51	0.43	-2.83	-0.37	2.47	-0.68
<i>Inga goldmanii</i>	ingago	0.85	43.05	0.27	0.46	-4.93	0.40	2.14	-1.83

Species	code	β	α	a_g	b_g	a_m	b_m	r	λ
<i>Inga marginata</i>	ingama	0.19	0.00	0.51	0.59	-1.87	-0.20	8.38	-0.73
<i>Inga peizifera</i>	ingape	0.67	23.63	0.46	0.51	-5.70	0.63	3.86	-1.16
<i>Inga punctata</i>	ingapu	0.36	3.19	0.32	0.67	-2.74	0.05	9.03	4.02
<i>Inga nobilis</i>	ingaqu	1.47	62.16	0.32	0.36	-4.06	0.00	1.39	-1.12
<i>Inga acuminata</i>	ingas1	0.39	1.85	0.63	0.42	-4.86	0.12	5.38	3.79
<i>Inga sapindoides</i>	ingasa	1.03	70.55	0.46	0.44	-3.98	0.08	1.65	-2.49
<i>Inga thibaudiana</i>	ingath	0.53	14.07	2.32	0.22	NA	NA	8.15	5.18
<i>Inga umbellifera</i>	ingaum	1.28	35.43	0.32	0.25	-5.49	0.49	2.33	-0.65
<i>Jacaranda copaia</i>	jac1co	0.85	279.6	5.72	0.05	-0.07	-0.74	2.19	-0.30
<i>Lacistema aggregatum</i>	laciag	0.76	11.96	0.15	0.40	-3.17	-0.10	3.45	-0.58
<i>Lacmellea panamensis</i>	lacmpa	1.38	151.7	0.86	0.31	-6.02	0.02	1.25	0.34
<i>Laetia thamnina</i>	laetth	0.83	23.82	0.38	0.25	-4.05	0.17	3.01	-0.73
<i>Licania hypoleuca</i>	licahy	0.35	1.18	0.19	0.29	-5.78	0.37	2.37	0.54
<i>Licania platypus</i>	licapl	0.54	7.94	0.20	0.30	-4.03	0.05	2.85	0.11
<i>Lonchocarpus heptaphyllus</i>	loncla	0.54	19.56	0.29	0.37	-3.81	0.03	2.98	-0.80
<i>Luehea seemannii</i>	luehse	0.37	22.67	0.37	0.46	1.47	-1.21	5.80	0.42
<i>Macrocnemum roseum</i>	macrgl	0.76	62.86	0.11	0.61	-2.43	-0.44	1.64	-0.47
<i>Mosannonna garwoodii</i>	malmsp	0.80	19.29	0.55	0.00	-8.67	0.86	3.23	2.25
<i>Maquira guianensis</i>	maquco	1.48	57.82	1.39	-0.03	-6.72	0.65	1.54	-0.23
<i>Maytenus schippii</i>	maytsc	0.74	59.65	0.24	0.63	-5.45	0.20	1.07	-0.46
<i>Miconia affinis</i>	micoaf	0.93	18.28	1.88	-0.02	-5.27	0.64	6.25	0.69
<i>Miconia argentea</i>	micoar	0.20	0.01	1.24	0.28	-0.45	-0.49	14.97	1.21
<i>Miconia hondurensis</i>	micoho	0.74	19.54	1.31	0.10	-4.11	0.09	6.23	2.74
<i>Mouriri myrtilloides</i>	mourmy	1.33	14.81	0.34	0.13	-3.86	0.14	3.73	0.14
<i>Myrcia gatunensis</i>	myrcga	0.76	14.72	0.62	0.25	-2.69	-0.35	4.71	1.37
<i>Nectandra cissiflora</i>	nectci	0.40	2.35	0.26	0.65	-1.47	-0.75	2.78	-2.32
<i>Nectandra lineata</i>	nectgl	0.28	0.02	0.16	0.82	-1.67	-0.58	9.25	2.06
<i>Nectandra purpurea</i>	nectpu	0.80	10.33	0.37	0.46	-1.71	-0.69	3.30	-0.37
<i>Ocotea cernua</i>	ocotce	0.31	0.45	0.32	0.62	-3.18	-0.32	4.72	-0.73
<i>Ocotea oblonga</i>	ocotob	0.21	0.06	1.21	0.39	-1.45	-0.36	8.54	0.06
<i>Ocotea puberula</i>	ocotpu	0.13	0.00	0.26	0.64	-3.05	-0.09	5.43	-1.94
<i>Ocotea whitei</i>	ocotwh	0.43	24.55	0.38	0.48	-0.89	-0.56	2.18	-3.73
<i>Trophis caucana</i>	olmeas	1.15	53.20	1.12	0.08	-0.53	-0.69	1.10	-4.10
<i>Ormosia coccinea</i>	ormocr	0.46	5.52	0.79	0.12	NA	NA	3.55	2.44
<i>Ormosia macrocalyx</i>	ormoma	0.33	0.79	0.54	0.10	NA	NA	3.74	1.48
<i>Ouratea lucens</i>	ouralu	1.20	14.77	0.48	-0.07	-5.32	0.25	2.51	0.38
<i>Palicourea guianensis</i>	paligu	1.23	17.06	7.20	-0.32	-1.96	0.16	27.40	4.66
<i>Pentagonia macrophylla</i>	pentma	1.41	25.53	0.65	0.08	-1.56	-0.67	2.22	-2.23
<i>Perebea xanthochyma</i>	perexa	1.37	57.74	0.54	0.15	-2.39	-0.52	1.74	-0.57
<i>Cinnamomum triplinerve</i>	phoeci	0.30	0.86	0.43	0.60	-0.83	-0.86	5.81	-0.70
<i>Picramnia latifolia</i>	picrla	0.49	3.60	0.26	0.28	-3.42	-0.04	3.22	-0.27
<i>Piper reticulatum</i>	pipere	0.89	17.75	2.02	-0.12	-1.15	-0.65	2.91	-1.06
<i>Platymiscium pinnatum</i>	pla1pi	0.15	0.00	0.17	0.57	-1.67	-0.38	5.19	-0.55

Species	code	β	α	a_g	b_g	a_m	b_m	r	λ
<i>Platypodium elegans</i>	pla2el	0.33	14.19	0.21	0.49	-2.01	-0.29	2.89	-1.73
<i>Posoqueria latifolia</i>	posola	0.63	29.64	0.29	0.26	-3.14	-0.40	1.24	-0.04
<i>Poulsenia armata</i>	poular	1.32	151.5	0.43	0.35	-2.32	-0.17	0.54	-4.47
<i>Pourouma bicolor</i>	pourbi	0.20	0.02	1.03	0.37	-3.22	-0.27	6.62	4.12
<i>Pouteria reticulata</i>	poutre	0.28	1.17	0.09	0.72	-2.67	-0.36	2.62	-1.35
<i>Pouteria stipitata</i>	poutst	0.77	97.85	0.35	0.40	-4.02	-0.09	1.43	-0.26
<i>Prioria copaifera</i>	pri2co	0.44	36.13	0.15	0.61	-4.22	-0.22	0.95	0.02
<i>Protium costaricense</i>	protco	0.88	39.89	0.59	0.24	-3.96	0.11	2.52	-1.03
<i>Protium panamense</i>	protpa	1.08	24.78	0.28	0.23	-5.28	0.40	2.79	0.41
<i>Protium tenuifolium</i>	protte	0.46	9.71	0.32	0.39	-4.11	-0.01	2.90	0.46
<i>Pseudobombax septenatum</i>	pse1se	0.44	64.85	0.92	0.35	0.01	-0.68	8.06	1.39
<i>Psidium friedrichsthalianum</i>	psidfr	0.37	1.67	0.28	0.19	-1.66	-0.96	2.90	1.33
<i>Psychotria grandis</i>	psycg3	0.57	4.63	0.72	0.53	-1.39	-0.40	6.64	-2.08
<i>Pterocarpus rohrii</i>	pterro	0.89	24.96	0.30	0.24	-4.40	0.18	3.13	-0.80
<i>Quararibea asterolepis</i>	quaras	0.53	60.42	0.23	0.49	-3.98	-0.08	1.39	-0.34
<i>Quassia amara</i>	quasam	1.54	54.23	1.33	-0.26	-3.71	-0.19	0.70	-1.09
<i>Randia armata</i>	randar	0.97	55.92	0.28	0.24	-3.22	-0.22	1.50	-0.58
<i>Rinorea sylvatica</i>	rinosy	1.32	16.49	0.37	-0.03	-3.94	-0.04	2.31	-0.41
<i>Sapium glandulosum</i>	sapiau	0.50	39.78	1.36	0.29	0.75	-0.79	7.56	-0.85
<i>Senna dariensis</i>	sennda	1.16	20.39	20.76	-0.49	-1.34	-0.17	11.98	-1.47
<i>Simarouba amara</i>	simaam	0.30	1.00	0.45	0.48	-1.96	-0.24	7.04	0.81
<i>Siparuna pauciflora</i>	sipapa	0.35	0.08	0.52	0.14	-2.02	-0.60	4.65	-0.91
<i>Sloanea terniflora</i>	sloate	0.59	31.72	0.04	0.74	-4.31	-0.07	0.72	-1.10
<i>Solanum hayesii</i>	solaha	0.76	31.10	6.71	0.01	-0.26	-0.35	15.74	-2.83
<i>Sorocea affinis</i>	soroaf	1.72	39.24	0.62	-0.05	-5.12	0.37	1.56	-1.18
<i>Spondias mombin</i>	sponmo	0.49	45.73	0.91	0.36	-0.74	-0.53	9.35	2.71
<i>Spondias radlkoferi</i>	sponra	0.29	2.36	0.26	0.51	-0.24	-0.73	7.54	0.74
<i>Sterculia apetala</i>	sterap	0.64	215.0	0.68	0.33	-0.88	-0.81	2.34	-1.24
<i>Stylogyne turbacensis</i>	stylst	0.88	7.71	0.59	-0.09	-4.16	0.07	4.12	-0.01
<i>Swartzia simplex</i>	swars1	0.74	19.57	0.11	0.38	-6.92	0.37	1.64	0.60
<i>Symphonia globulifera</i>	sympgl	0.40	5.00	0.29	0.45	-4.53	0.30	2.97	-0.74
<i>Tabebuia guayacan</i>	tab1gu	0.50	85.76	0.09	0.63	-3.64	-0.14	1.14	-0.63
<i>Tabebuia rosea</i>	tab1ro	0.40	15.28	0.15	0.51	-2.19	-0.36	2.86	-1.08
<i>Tabernaemontana arborea</i>	tab2ar	0.38	9.72	0.40	0.43	-4.10	-0.05	3.13	1.03
<i>Tachigali versicolor</i>	tachve	0.41	2.73	0.08	0.64	-3.09	-0.03	3.70	-1.10
<i>Talisia nervosa</i>	taline	3.19	38.42	5.37	-0.82	-5.19	0.08	0.85	-0.46
<i>Talisia princeps</i>	talipr	0.92	14.96	0.62	-0.25	-5.90	0.24	3.07	0.50
<i>Terminalia amazonia</i>	termam	0.92	217.8	0.35	0.52	-2.46	-0.42	1.36	-0.95
<i>Terminalia oblonga</i>	termob	0.67	151.9	0.35	0.42	-2.31	-0.43	1.25	-0.16
<i>Tetragastris panamensis</i>	tet2pa	0.18	0.00	0.16	0.51	-3.71	-0.18	3.61	1.34
<i>Thevetia ahouai</i>	thevah	1.20	30.20	4.09	-0.30	-3.23	-0.09	3.53	-0.96
<i>Trattinnickia aspera</i>	tratas	0.93	169.8	1.58	0.25	-3.64	0.05	2.35	-1.48
<i>Trema micrantha</i>	tremin	1.19	165.4	NA	NA	NA	NA	19.75	19.75

Species	code	β	α	a_g	b_g	a_m	b_m	r	λ
<i>Trichilia pallida</i>	tri2pa	0.77	34.64	0.48	0.29	-2.79	-0.31	2.35	-0.54
<i>Trichilia tuberculata</i>	tri2tu	0.40	6.02	0.07	0.61	-3.90	0.05	2.31	-0.57
<i>Triplaris cumingiana</i>	tripcu	1.45	152.1	1.13	0.18	-1.31	-0.64	1.37	-1.75
<i>Trophis racemosa</i>	tropca	0.64	19.24	0.34	0.24	-5.42	0.44	2.27	-1.04
<i>Turpinia occidentalis</i>	turpoc	0.55	42.92	2.31	0.24	-2.88	-0.05	3.24	-2.24
<i>Unonopsis pittieri</i>	unonpi	0.78	47.59	0.54	0.29	-2.23	-0.40	2.60	-0.60
<i>Virola sebifera</i>	virose	1.17	117.5	0.32	0.37	-2.94	-0.18	0.59	-2.21
<i>Virola multiflora</i>	virosp	1.16	205.9	0.52	0.42	-3.68	-0.11	0.72	-1.29
<i>Virola surinamensis</i>	virusu	1.34	363.9	0.73	0.38	-2.78	-0.20	0.29	-2.05
<i>Vismia baccifera</i>	vismba	0.91	17.23	14.99	-0.47	-2.64	-0.01	6.26	-0.93
<i>Vochysia ferruginea</i>	vochfe	0.17	0.02	2.44	0.54	-3.19	0.08	4.65	-0.55
<i>Xylopia macrantha</i>	xyl1ma	0.61	17.99	0.59	0.22	-4.88	-0.03	3.61	2.50
<i>Zanthoxylum ekmanii</i>	zantbe	0.80	177.6	17.21	-0.06	0.38	-0.65	8.70	0.83
<i>Zanthoxylum panamense</i>	zantp1	0.40	10.44	0.62	0.41	-1.22	-0.39	5.24	-1.15
<i>Zanthoxylum acuminatum</i>	zantpr	0.74	63.63	0.64	0.50	-3.01	-0.01	2.15	-2.72
<i>Zuelania guidonia</i>	zuelgu	0.30	1.58	0.13	0.58	-1.68	-0.41	4.86	-1.03

Appendix 4. Diameter distributions of the 180 species evaluated in this study and their truncated Weibull fits (bold line). Since the truncated Weibull was fitted directly to data and not to the histograms show below, therefore there may be some disparities between them due to histogram binning. Each histogram contains the density on the y-axis and the diameter in mm on the x-axis. Species full names are given in Table S2.

Appendix 5. Results of the analyses separated by species growth forms.

In the main text, we have carried the analysis by pooling species from all growth forms together. Here we present the same generalized regression models, but for growth forms individually. For the exponent b_g (Figures S6) and for mean recruitment rate (Figures S7) there was no difference from the results presented in the main text, with both of them being negatively correlated to the shape of the diameter distribution (β). The only exception was the result for shrubs that also presented trends similar to other growth forms but these trends were not significant, probably due to the small number of shrubs included in the analyses. For population growth (λ), results were somewhat different from the main text. The weak relationship between β and λ found in the main text, was no longer significant for midstorey species and was highly significant for canopy species (Figure S8). For canopy species, even if we take into account the shape of the growth-dbh curve (b_g) the relationship between β and λ remains significant. This is contrary to Condit et al. (1998), who found only a weak association for canopy species that disappeared when b_g is taken into account, a contrast that was probably caused by the different metric to describe size distributions of for the longer time span used here. For the regression of β on the exponent b_m , no growth forms presented evidence of significant relationships, the same result as well as for the results pooling growth forms together. Thus, these results are not presented. We have also carried the multiple regressions as presented in the main text but separately for each growth form. The results were qualitatively the same except for midstorey species in which the shape of mortality-dbh curve became marginally significant ($p=0.075$)

Literature cited in the appendices

- Bolker, B.M. 2008. Ecological models and data in R. Princeton University Press, Princeton.
- Burnham, K.P. and D.R. Anderson. 2002. Model selection and multi-model inference: a practical information-theoretic approach. 2nd ed. Springer, New York.
- Condit, R., R. Sukumar, S. P. Hubbell, and R. B. Foster. 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. *American Naturalist* 152:495-509.
- Coomes, D.A., and R.B. Allen. 2007. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology* 95:27-40.
- Harcombe, P.A. 1987. Tree life tables. *Bioscience* 37:557-568.
- Rüger, N. & Condit, R. 2012. Testing metabolic theory with models of tree growth that include light competition. *Functional Ecology* 26:759-765.
- Rüger, N., A. Huth, S.P. Hubbell, and R. Condit. 2011. Determinants of mortality across a tropical lowland rainforest community. *Oikos* 120:1047-1056.

Figure S6. Relationship between the shape (β) of the size distribution and the exponent of the growth-dbh curve (b_g) for different growth forms. The trend line is the result of the generalized linear model, which is presented with the pseudo- R^2 and p -value in the legend.

Figure S7. Relationship between the shape (β) of the size distribution and the mean recruitment rate for different growth forms. The trend line is the result of the generalized linear model, which is presented with the pseudo- R^2 and p -value in the legend.

Figure S8. Relationship between the shape (β) of the size distribution and the mean population growth rate for different growth forms. The trend line is the result of the generalized linear model, which is presented with the pseudo- R^2 and p -value in the legend.

CAPÍTULO 5: DISTRIBUIÇÃO ESPACIAL DAS ESPÉCIES

Comparando padrões espaciais de árvores em quatro florestas Neotropicais

Renato A. Ferreira de Lima^{1,2}, Paulo I. Prado², Ricardo R. Rodrigues³

¹ Programa de pós-graduação em Ecologia, Instituto de Biociências – Universidade de São Paulo (IB-USP)

² Laboratório de Ecologia Teórica (LET), Departamento de Ecologia, Instituto de Biociências – Universidade de São Paulo (IB-USP), CEP 05508-090, São Paulo, SP, Brazil

³ Departamento de Ciências Biológicas, Escola Superior de Agricultura ‘Luiz de Queiroz’ – Universidade de São Paulo (ESALQ-USP), CEP: 13018-900, Piracicaba, SP, Brazil

Autor para correspondência: Renato A. Ferreira de Lima (raflima@usp.br). Rua do Matão, trav. 14, 321, São Paulo, SP, CEP 05508-090, Brazil. Telefone: +55 11 30910969. Fax +55 11 30918096.

Contagem de palavras: 5835 palavras.

(Versão preliminar e em português do artigo a ser submetido à Plant Ecology)

Resumo A maneira com a qual os indivíduos se distribuem no espaço possui um papel central em ecologia, pois ela reflete processos que afetam, de maneira direta ou indireta, a dinâmica de populações e a manutenção da diversidade de espécies. Assim, descrever quais são os padrões espaciais das espécies de uma comunidade pode fornecer informações sobre quais são os processos que a organizam. Neste estudo, foram comparados os padrões de distribuição espacial de árvores em quatro parcelas permanentes instaladas no sudeste do Brasil sob condições ambientais distintas, visando avaliar como diferentes condições abióticas podem afetar estes padrões. Para cada espécie, a distribuição espacial foi descrita usando a função de distribuição radial de pares $g(r)$, que calcula a densidade de indivíduos coespecíficos em anéis concêntricos. Para as espécies com distribuição agregada, também foi estimada a força e distância de agregação espacial, assumindo processo homogêneo de Thomas. Os resultados mostraram que a maioria das espécies se distribuiu de maneira agregada nas quatro parcelas, com proporções relativamente menores na parcela de floresta estacional e de floresta ombrófila. Contudo, as parcelas apresentaram diferentes padrões de intensidade e tamanho de agregação, o que em parte pode ser explicado pelo maior grau de heterogeneidade ambiental encontrados no cerradão e restinga. Tendências semelhantes foram encontradas quando populações de uma mesma espécie foram comparadas entre parcelas, mas houve exceções, sugerindo que as condições abióticas não foram os únicos determinantes da distribuição espacial das espécies. Houve ainda forte evidência de que espécies mais raras tendem a apresentar distribuições aleatórias, o que sugere uma possível diferença de estratégia de regeneração entre espécies raras e comuns.

Palavras-chave: agregação espacial, Brasil, função de distribuição radial de pares, heterogeneidade ambiental, padrão de pontos, processo de Thomas

Abstract The way in which individuals are distributed in space has a central role in ecology, since it reflects processes that have direct or indirect influence on the dynamics of populations and on the maintenance of species diversity. Thus, the description of species spatial patterns in a given community can provide useful information on what are the processes that organize it. In this study, we compared the spatial distribution of tree species in four permanent forest plots placed in south-eastern Brazil under contrasting environmental conditions in order to assess how abiotic factors may affect spatial patterns. For each species, the spatial distribution was described using the pair correlation function $g(r)$, which estimates the density of conspecifics in concentric rings. For species with clumped distributions we also estimated the

intensity and size of clumps, assuming the homogeneous Thomas process. We found that most species had clumped distributions in the four plots, with relatively less clumped distributions at the seasonal and rain forest plots. However, the intensity and size of clumps were different among plots, which may in part be explained by the higher degree of heterogeneity found in savanna and white-sand plots. Similar trends were found when populations of the same species were compared between plots, but there were exceptions, suggesting that abiotic conditions were not an exclusive cause of species spatial patterns. There was also strong evidence for less abundant species being more aggregate, specially at short distances, which suggests different regeneration strategies between rare and common species.

Keywords: Brazil, environmental heterogeneity, pair-correlation function, point pattern, poisson cluster process, spatial clustering

Introdução

A maneira com a qual os indivíduos se distribuem no espaço possui um papel central em ecologia. Para organismos sésseis como as árvores, entender porque um indivíduo se estabelece em um determinado local ou a uma determinada distância de outros indivíduos pode revelar o papel de importantes processos que atuam sobre o indivíduo, mas que repercutem nas populações e, conseqüentemente, nas comunidades (Russo *et al.* 2005). Esta relação entre arranjos espaciais e processos ecológicos que atuam sobre o indivíduo levou à formulação de diferentes hipóteses e teorias que consideram, implicitamente ou explicitamente, que diferentes arranjos espaciais possuem um efeito indireto sobre a dinâmica de populações, manutenção da diversidade de espécies e sobre os padrões de distribuição de abundância de espécies, curvas de espécie-área e de decaimento de similaridade com a distância (Plotkin *et al.* 2000; Hubbell 2001; He & Legendre 2002; Plotkin & Muller-Landau 2002; Harte *et al.* 2005; Green & Plotkin 2007). Mais recentemente, McGill (2010) propôs uma teoria unificada da biodiversidade na qual os padrões espaciais das espécies têm um papel fundamental para explicar a diversidade de espécies. Por estes motivos, o interesse na distribuição espacial das espécies e, conseqüentemente, nas técnicas para descrever tais distribuições, possui um longo histórico e ecologia (Taylor *et al.* 1978; Hubbell 1979; Armesto *et al.* 1986; He *et al.* 1997; Condit *et al.* 2000; Seidler & Plotkin 2006; Li *et al.* 2009) e têm crescido consideravelmente nas últimas décadas (Wiegand *et al.* 2013).

Até os anos 80, havia uma crença de que espécies florestais tropicais se distribuíam aleatoriamente (Hubbell 1979). Contudo, Hubbell (1979) e estudos posteriores têm encontrado que a grande maioria das espécies arbóreas possui padrões espaciais agregados de distribuição (Armesto *et al.* 1986; He *et al.* 1997; Condit *et al.* 2000). Padrões agregados têm sido frequentemente citados como o resultado da associação de espécies com habitats, dispersão limitada de sementes, heterogeneidade na disponibilidade de recursos ou da facilitação/competição entre espécies. Contudo, esses fatores deixam marcas distintas nos padrões de distribuição das espécies. Há certo consenso de que a agregação em escalas pequenas a médias (<100 m) seja um produto da limitação de dispersão e/ou da heterogeneidade na distribuição de recursos limitantes (i.e. micro-sítios favoráveis), enquanto que a agregação em escalas médias a grandes seja fruto da associação das espécies com diferentes ambientes (e.g. tipos de solo, topografia, estágios sucessionais – Collins & Klahr 1991; Szwagrzyk & Czerwczak 1993; Condit *et al.* 2000; Valencia *et al.* 2004; Seidler & Plotkin 2006; Law *et al.* 2009; Li *et al.* 2009). Distribuições espaciais regulares têm sido raramente encontradas em florestas tropicais e subtropicais, mas são mais comuns em florestas temperadas (ver tabela 1 em Szwagrzyk & Czerwczak 1993). Eles resultariam de processos como competição intra ou inter-específica, alelopatia ou herbivoria (Collins & Klahr 1991; Barot *et al.* 1999). Por fim, padrões aleatórios de distribuição são esperados para espécies com boa dispersão e sem associação com habitat particulares (i.e. generalistas).

O solo e a topografia têm sido comumente citados como fatores abióticos capazes de gerar agregação em espécies arbóreas por associação a habitats (Phillips *et al.* 2003; Palmiotto *et al.* 2004; Lan *et al.* 2011). De fato, 30-80% das espécies tem suas distribuições espaciais de associadas a algum tipo de solo ou posição topográfica (Harms *et al.* 2001; Valencia *et al.* 2004; Russo *et al.* 2005; John *et al.* 2007). Assim, espera-se que quanto maior a heterogeneidade edáfica e/ou topográfica na floresta, maior será a proporção de espécies associadas a um determinado habitat (Condit *et al.* 2000). Além disso, como as manchas desses tipos de habitat geralmente são da ordem de 1000 m² a alguns hectares, espera-se que nessas florestas a agregação das espécies se estenda até distâncias maiores (i.e. limitação de dispersão + associação de habitat), enquanto que em áreas mais homogêneas a agregação se restrinja a distâncias pequenas (apenas limitação de dispersão) ou que ela não exista (sem limitação ou associação). Além disso, espera-se que, em ambientes sob solo menos propícios ao desenvolvimento vegetal, aumentos na disponibilidade de recursos sejam mais ‘percebidos’ pelas plantas e, portanto, mais ‘refletida’ em seus padrões de distribuição.

O objetivo deste estudo é descrever e comparar a distribuição espacial de árvores entre quatro parcelas permanentes de 10,24 ha instaladas em diferentes florestas do sudeste do Brasil. Posicionadas em áreas com diferentes tipos de solo e amplitudes topográficas, estas parcelas representam uma oportunidade para avaliar como diferentes condições abióticas podem afetar a distribuição das espécies. A distribuição espacial das espécies foi descrita através da função de distribuição radial de pares $g(r)$ (*pair-correlation function*), capaz de descrever de maneira eficiente a distribuição de árvores em diferentes escalas (Condit *et al.* 2000; Wiegand & Maloney 2004; Wiegand *et al.* 2013). Assim, nós avaliamos a distribuição espacial das populações de cada espécie nas quatro parcelas para responder as seguintes perguntas: (i) A proporção de espécies com distribuição agregada varia entre as quatro áreas? (ii) Para espécies agregadas, a intensidade e tamanho de agregação são diferentes entre áreas? (iii) As tendências encontradas para a comunidade como um todo se repetem para espécies ocorrentes em duas ou mais parcelas? Nós assumimos que nas parcelas com maior heterogeneidade edáfica e/ou topográfica, a agregação das espécies seja mais freqüente e se estenda a distâncias maiores do que em florestas mais homogêneas. Assumimos, também, que em áreas com solo menos férteis, a distribuição de muitas espécies deve estar mais associada a manchas de solo relativamente mais férteis. Assim, esperamos um maior número de espécies agregadas e uma maior intensidade de agregação na floresta ombrófila (solos + topografia), seguido das parcelas de cerradão e restinga (solos) e, por fim, da floresta estacional (baixa heterogeneidade ambiental). Esperamos, ainda, que espécies compartilhadas entre parcelas sigam a mesma tendência da comunidade como um todo, ou seja, que elas sejam mais agregadas em parcelas mais heterogêneas ou com solos mais restritivos.

Material e métodos

Áreas de estudo e mapeamento das árvores

As informações sobre a distribuição espacial das espécies foram coletadas em quatro parcelas permanentes instaladas nos principais tipos florestais do sudeste do Brasil: floresta de savana (cerradão), floresta estacional, floresta de restinga e floresta ombrófila. Essas áreas variam consideravelmente em termos de seu ambiente físico, principalmente solo e clima, e em termos da composição, estrutura e diversidade da floresta (Rodrigues 2005). As parcelas permanentes foram instaladas em diferentes unidades de conservação do Estado de São Paulo (coordenadas aproximadas: cerradão – 22°35'S e 50°22'W; floresta estacional – 22°24'S e 49°42'W; floresta de restinga – 25°04'S e 47°55'W; floresta ombrófila – 24°10'S e 47°55'W),

sempre buscando trechos de florestas bem conservados e típicos das quatro formações florestais consideradas. Cada parcela permanente possui 10,24 ha (320×320 m) e está subdividida em 256 sub-parcelas de 20×20 m. Todas as árvores, arbustos, palmeiras e samambaias com perímetro à altura do peito (PAP) ≥ 15 cm dentro destas parcelas foram marcadas, identificadas e tiveram seu PAP mensurado. A localização dos indivíduos foi realizada com uma precisão de 10 cm dentro de subdivisões menores de, no mínimo, 10×10 m. Esse procedimento foi realizado em 2002 e conferido em 2004/2005 e 2012/2011 para todas as parcelas, exceto para a floresta ombrófila. Assim, tanto a localização quanto a identificação das espécies foi conferida em diferentes oportunidades. Os dados usados neste estudo se referem ao censo 2004/2005, que resultou em uma densidade de 2142, 1176, 1492 e 1135 indivíduos por ha para o cerradão, floresta estacional, floresta de restinga e ombrófila. A riqueza total em cada parcela uma destas florestas foi de 118, 147, 117 e 224 espécies, respectivamente.

Análise dos dados

A análise da distribuição espacial das espécies foi realizada usando a função de distribuição radial de pares $g(r)$ (*pair-correlation function*). Esta função calcula o número de indivíduos da mesma espécie em anéis concêntricos centrados em cada indivíduo e divide esta contagem pela área de cada anel. O valor da função é padronizado dividindo a densidade nos anéis pela densidade esperada na área como um todo que é referida como λ . Assim, $g(r)$ descreve a densidade esperada de indivíduos da mesma espécie na vizinhança de cada árvore, sendo assim uma descrição com significado biológico claro e que possibilita a comparação da distribuição espacial entre espécies (Condit *et al.* 2000). Assumindo um processo estacionário (λ constante em toda a área de estudo – Wiegand *et al.* 2013), valores de $g(r) = 1$ representam distribuições aleatórias, $g(r) < 1$ distribuições regulares e $g(r) > 1$ distribuições agregadas. A função $g(r)$ tem relação matemática com a estatística *O-ring* – $O(r) = \lambda g(r)$ e com a função de Ripley – $K'(r) = 2\pi r g(r)$, onde $K'(r)$ é a primeira derivada da função $K(r)$ (Wiegand & Maloney 2004; Perry *et al.* 2006). A diferença básica entre estas funções é que a função $K(r)$ é cumulativa descrevendo a densidade esperada de vizinhos até x metros, enquanto que $O(r)$ e $g(r)$ descrevem a densidade esperada de vizinhos em uma classe de distância específica, definida entre x e $x + r$ metros. Esse é um dos principais benefícios dessas funções, pois permite analisar padrões de distribuição em classes de distâncias específicas. Adicionalmente, $g(r)$ é a função mais indicada para descrever a distribuição de árvores na floresta (Condit *et al.*

2000; Wiegand & Maloney 2004), mesmo quando há suspeitas de processos não estacionários (Wiegand *et al.* 2013).

Apesar de poucas árvores terem apresentado diâmetros relativamente grandes (>150 cm), considerou-se que os diâmetros das árvores eram suficientemente pequenos em relação às dimensões das parcelas (320×320 m) para serem considerados como pontos. Mesmo assim, a função $g(r)$ foi aplicada usando o raio mínimo de 2,5 m para o primeiro anel, visando minimizar vieses nas estimativas de espécies com diâmetros grandes nas parcelas. Esse valor de distância foi replicado para os demais anéis ($r = 2,5$ m) até uma distância máxima de 160 m, resultando num total de 64 anéis concêntricos por indivíduo de cada espécie. Essa distância máxima representa metade da dimensão das parcelas permanentes estudadas e visa evitar os vieses de estimação de $g(r)$, que se tornam consideráveis à medida que a distância aumenta. O efeito de borda causado pela ausência de observações fora da área delimitada de estudo também podem gerar estimativas enviesadas de $g(r)$. Assim, aplicou-se a correção isotrópica de Ripley, que basicamente corrige as estimativas de densidade baseado na proporção em área do anel que se encontra além dos limites das parcelas. Foram incluídos nas análises apenas espécies com 15 ou mais indivíduos dentro de cada parcela permanente, seguindo as recomendações de Baddeley & Turner (2005), o que resultou num total de 326 espécies analisadas. Para estas espécies, 999 simulações com reposição (i.e. *bootstraps*) das contribuições de cada indivíduo para a $g(r)$ foram usadas para gerar intervalos de confiança contendo 95% dos valores simulados para cada classe de distância. Para uma espécie de samambaia arborescente (*Alsophila sternbergii*), fustes com uma distância menor que 30 cm entre si foram considerados como parte de uma mesma touceira.

Em um segundo momento, as espécies com distribuições agregadas indicadas pelo testes de aderência descritos abaixo, foram avaliadas quanto à força e distância de agregação. Para isso, foram usados procedimentos distintos. O primeiro deles assume que as espécies com distribuição agregada seguem o processo homogêneo de Thomas, também chamado de *Poisson Cluster Process*. Este modelo assume que (i) os centros dos agregados estão distribuídos segundo um processo Poisson com intensidade κ , (ii) o número de indivíduos por agregado segue uma Poisson com média μ ($\mu = \lambda/\kappa$) e (iii) os indivíduos ao redor destes centros estão distribuídos segundo uma Normal com desvio padrão σ . O ajuste ao do processo ou modelo de Thomas à $g(r)$ é usando o método de mínimos contrastes e a seguinte função (Wiegand *et al.* 2007):

$$g(r) = 1 + \exp(-r^2/4\sigma^2)/4\pi\sigma^2\kappa.$$

Os valores de μ pode ser usado como uma métrica de força de agregação, enquanto que a distância 2σ expressa o tamanho da agregação, isto é, a área do agregado que contém 86% dos indivíduos (Wiegand *et al.* 2007, 2009). O procedimento alternativo para medir a força e distância de agregação foi definido pela soma dos valores médios ou dos intervalos de confiança de $g(r) - 1$, denominada aqui como A_c , e o valor de r onde $g(r) \leq 1$, isto é, onde $g(r)$ cruza a linha de completa aleatoriedade espacial, denominado como r_c . Este procedimento é mais parcimonioso, pois não assume nenhum tipo de modelo de agregação dos dados e produz métricas que equivalem aos valores dos parâmetros μ e σ do modelo de Thomas (Apêndice 1). Quanto maior a força de agregação maior será o valor de A_c e quanto maior a distância de agregação maior será o valor de r_c . Além disso, os parâmetros do modelo de Thomas não são independentes entre si; quanto menor o tamanho da agregação 2σ , maior é a força de agregação, isto é, mais próximos estarão os pontos em cada agrupamento (Wiegand *et al.* 2009). Para fins comparativos, foi utilizado também como métrica de agregação o índice Ω_{0-10} sugerido por Condit *et al.* (2000), que equivale ao valor de $g(r)$ quando $r = 10$.

Para determinar se a distribuição espacial das espécies era diferente do esperado sob a hipótese nula de completa aleatoriedade espacial, foram realizados dois tipos de testes de aderência, ambos baseados em 199 simulações de Monte Carlo e nos valores observados de $g(r)$. De maneira simplificada, o teste ‘desvio absoluto máximo’ (*maximum absolute deviation*, MAD) determina se há evidência de agregação em curtas distâncias, enquanto que o teste proposto por Loosmore & Ford (2006) que determina se há agregação em alguma classe de distância. Ambos os testes foram realizados para valores de r entre 2,5 e 80 m, visto que raramente os tamanhos de agregados se estenderam além desta distância. As relações entre os parâmetros e métricas das distribuições foram realizadas usando modelos de regressão. Funções de ligação não-lineares e modelagem de variância foram usadas sempre que necessário e a sua plausibilidade como um modelo alternativo também foi determinada baseada nos valores de AIC dos modelos lineares e/ou de variância constante. A comparação entre as distribuições espaciais das espécies ocorrentes em mais de uma parcela simultaneamente foi feita visualmente, baseado nos valores observados de $g(r)$ em cada parcela. Os valores observados de $g(r)$, intervalos de confiança, testes de aderência e ajuste do modelo de Thomas foram realizados usando o pacote ‘spatstat’ (Baddeley & Turner 2005) do ambiente R versão 2.15.1 (www.R-project.org). As estimativas dos parâmetros e comparação dos modelos de regressão foram obtidas com o auxílio do pacote ‘bbmle’ (Bolker 2010) também do ambiente R.

Resultados

Nas quatro florestas estudadas, a maioria das espécies apresentou distribuições agregadas em alguma classe de distância, um resultado que foi similar se considerarmos apenas agregações a curtas distâncias (Tabela 1). Isso quer dizer que houve uma tendência generalizada de declínio nos níveis de agregação com a distância entre as espécies estudadas (Figura 1, Apêndice 2). Na floresta estacional e ombrófila, contudo, a proporção de espécies com distribuição aleatória foi relativamente elevada. Não houve evidência de nenhuma espécie que sugerisse um padrão regular de distribuição ao longo de várias classes de distância. De fato foram raros os casos nos quais o valor de $g(r)$ esteve abaixo de um, especialmente a distâncias pequenas. O índice Ω_{0-10} proposto por Condit *et al.* (2000), que avalia agregação até 10 m, assinalou distribuições regulares em apenas cinco casos: *Clethra scabra* (restinga), *Enterolobium contortisiliquum* (F. estacional), *Ormosia arborea*, *Syagrus romanzoffiana* (restinga) e *Swartzia submarginata* (F. ombrófila). Houve forte evidência de que espécies mais raras tendem a apresentar distribuições aleatórias (regressão logística: $AIC_{nulo} - AIC_{logístico} = 39.9$). A abundância por parcela das 72 espécies com evidência de distribuição aleatória foi de $39,1 \pm 27,1$ (média $\pm$ desvio padrão), enquanto que para as 252 espécies agregadas a abundância foi $223,4 \pm 508,0$. Contudo, existiram exemplos contrários a esta tendência foram *Campomanesia xanthocarpa* ($n = 101$) e *Holocalyx balansae* ($n = 104$), ambas na floresta estacional, e *Handroanthus serratifolia* ($n = 103$), *Hieronyma alchorneoides* ($n = 148$) e *Marlierea suaveolens* ($n = 146$) na floresta ombrófila. Por outro lado, houve dezenas de exemplos de espécies com baixas abundâncias ($n < 30$), mas distribuições fortemente agregadas (Tabela S1).

Considerando apenas as espécies com distribuição agregada, as distribuições de frequência dos parâmetros do modelo de Thomas foram assimétricas. A maioria das espécies apresentou valores de densidade por agregado (μ) e o tamanho de agregação (2σ) menores que 10 indivíduos e 45 m, respectivamente (Figura 2). De maneira geral, os maiores valores de μ (> 100) e de 2σ (> 80 m) estiveram associados a espécies muito abundantes do cerradão, com exceção para *Savia dictyocarpa* na floresta estacional, que apresentou uma impressionante densidade em agregados relativamente pequenos ($\mu = 107,0$ indivíduos; $2\sigma = 40$ m). Tanto μ quanto 2σ estiveram relacionados com a abundância das espécies, de modo que quanto maior a abundância maior o valor dos parâmetros e sua previsibilidade (i.e. menor variância – Figura 3). Essa relação também existiu para as métricas alternativas de agregação, mas ao considerarmos A_c ou Ω_{0-10} esta relação se inverteu de modo que ambas métricas decresceram com o aumento da abundância das espécies (Figura S2). Houve uma relação forte e positiva

entre os parâmetros μ e 2σ do modelo de Thomas (Figura 3), um resultado que se manteve para a relação entre A_c e r_c ($F= 136.2$, $gl= 246$, $p< 0.001$, $R^2= 0,35$). Contudo, é importante lembrar que o ajuste deste modelo nem sempre foi bom para todas as classes de r ou para todas as espécies com distribuição agregada. O modelo frequentemente não foi capaz de descrever rápidos declínios de agregação em curtas distâncias seguidos de suaves declínios (e.g. Figura 1a) ou de aumentos na agregação em distâncias maiores.

Ainda considerando apenas as espécies com distribuição agregada, houve diferença na distribuição dos parâmetros μ e 2σ do modelo de Thomas entre as parcelas estudadas. Este resultado não se alterou ao incluir o efeito da abundância das espécies sobre a estimativa dos parâmetros do modelo de Thomas. No cerradão, as espécies tiveram maiores médias de μ e 2σ , seguidas pelas espécies na restinga e floresta estacional e, por fim, pela floresta ombrófila que apresentou as menores médias para ambos os parâmetros (Figura 2, Tabela 2). Apesar das tendências gerais terem sido mantidas, o resultado foi ligeiramente distinto usando as métricas alternativas A_c e r_c (Tabela 2). A comparação entre espécies ocorrentes em mais de uma parcela ao mesmo tempo foi possível para apenas 24 espécies. De maneira geral, as tendências para as espécies seguiram àquelas da comunidade como um todo, com agregados mais densos e extensos para as populações ocorrentes no cerradão (nove espécies), seguido da restinga (quatro), floresta estacional e ombrófila (três). Esse padrão ficou evidente para *Maytenus robusta*, que ocorreu em três parcelas (Apêndice 3). Os casos de espécies com distribuições similares entre parcelas diferentes foram menos frequentes (cinco casos), com destaque para as espécies de palmeiras e para espécies co-ocorrentes entre a restinga e a floresta ombrófila. Algumas espécies consideradas típicas de uma determinada formação florestal tiveram populações menos agregadas em suas formações de origem. Foi o caso das espécies *Casearia sylvestris*, *Croton floribundus* e *Mollinedia schottiana* espécies típicas da floresta estacional ou ombrófila, que tiveram distribuições bem mais agregadas ou na restinga (Fig. S3).

Discussão

Os resultados mostraram que a maioria das espécies arbóreas, avaliadas em parcelas representativas de quatro importantes tipos florestais do estado de São Paulo, se distribuiu de maneira agregada no espaço. Esse resultado tem sido repetidamente encontrado para florestas tropicais e subtropicais (Hubbell 1979; Armesto *et al.* 1986; Condit *et al.* 2000; Li *et al.* 2009), e indica que a distribuição das espécies nestas florestas geralmente não é ao acaso e deve responder a processos como associação de habitat e/ou limitação de dispersão (Tuomisto

et al. 2003). Adicionalmente, as baixas proporções de espécies com distribuições regulares em escalas bem pequenas (até 10m), sugerem que a competição intra-específica não deve ser um processo importante para a maioria das espécies estudadas. Aparentemente, estas sugestões não se estendem para florestas temperadas nas quais as distribuições aleatórias e regulares são mais frequentes (Szwagrzyk & Czerwczak 1993; Manabe *et al.* 2000).

A frequência de espécies com populações agregadas encontrada aqui para a parcela de restinga (13%) foi próxima àquela encontrados por Lima *et al.* (2011) na mesma área. Mas, como esses autores avaliaram a distribuição espacial em distâncias acima de 20 m, houve discrepâncias para algumas espécies, como *Myrcia brasiliensis* e *Ormosia arborea*, ambas agregadas apenas em escalas < 20 m. A análise da distribuição espacial para as comunidades como um todo (Capretz *et al.* 2011) revelou padrões similares aos encontrados aqui, exceto na floresta ombrófila, para a qual os autores encontraram um padrão mais agregado e distinto das demais parcelas. A maior agregação ao nível comunidade provavelmente é um efeito da topografia e de manchas de floresta com bambus, que influenciam a densidade de árvores nesta parcela (Lima *et al.* 2012). Se esta influencia for inespecífica (maior luminosidade nos topos de morro e maior taxa de danos nos bambuzais), o resultado será manchas com maior densidade de árvores ao nível da comunidade, mas não necessariamente ao nível populacional.

Houve diferença nos padrões de agregação entre as quatro florestas estudadas, sugerindo diferenças mesmo entre áreas com proporções similares de espécies agregadas. Em parte, esse resultado pode ser explicado pelas diferenças na heterogeneidade ambiental entre parcelas, uma vez que maior heterogeneidade ambiental leva a agregações mais fortes e extensas (Szwagrzyk & Czerwczak 1993; Condit *et al.* 2000). Esse pode ser o caso do cerrado, onde a existência de uma mancha de solo com maior capacidade retenção de água (Gleissolo) parece ter condicionado a ocorrência de várias espécies, em especial para as espécies típicas da floresta estacional (e.g. *Casearia sylvestris*, *Croton floribundus* e *Mollinedia schottiana*), e da restinga, onde manchas de solos periodicamente inundáveis parecem ter exercido um papel similar (Oliveira *et al.* 2013). Mas, contrariando as expectativas, a alta heterogeneidade ambiental da floresta ombrófila proporcionada pela topografia e pela ocorrência de bambuzais (Lima *et al.* 2012) não esteve associada a valores de intensidade e tamanho de agregação similares aos encontrados no cerrado e na restinga, ambientes mais restritivos quanto à sazonalidade da precipitação e/ou às propriedades do solo. Os menores tamanhos de agregação encontrados na floresta ombrófila sugerem duas possíveis explicações. Primeiro, as agregações nesta parcela estariam mais relacionadas à história de

vida das espécies (*e.g.* estratégias de dispersão de sementes), assumindo que não há restrições ambientais ao desenvolvimento das espécies na parcela como um todo (Condit *et al.* 2000). A segunda explicação seria o grão ambiental, ou seja, o tamanho das manchas dos diferentes tipos de ambiente, que seriam menores na floresta ombrófila que no cerradão. Se houverem espécies associadas a estas manchas, o resultado será espécies com menores tamanhos de agregação.

Contudo, a comparação da distribuição espacial entre populações de uma mesma espécie entre parcelas revelou uma tendência de encontrar populações mais agregadas no cerradão e restinga, seguindo os padrões gerais encontrado para as espécies como um todo. Como deve haver pouca variação intra-específica nos aspectos de história de vida, esse resultado sugere um maior controle ambiental para cerca de metade das 24 espécies que tiveram suas populações comparadas entre parcelas. Nesses casos, é provável que essas maiores agregações no cerradão e restinga estejam associadas à heterogeneidade edáfica, como exposto acima. Adicionalmente, a alta similaridade nas proporções de síndromes de dispersão entre a floresta estacional e cerradão e entre restinga e floresta ombrófila (R.A.F. Lima, dados não publicados) reforçam que as diferenças nos padrões entre parcelas não são apenas de cunho biológico, uma vez que diferentes síndromes geram diferentes padrões espaciais (Seidler & Plotkin 2006). Assim, encontramos evidências de que a heterogeneidade ambiental pode ser importante para explicar a distribuição de algumas espécies de árvores, mas não de todas. É possível, portanto, que esta heterogeneidade tenha maior peso sob a distribuição espacial em ambientes mais restritivos, onde boa parte das espécies tenderá a se associar, positiva ou negativamente, a determinados tipos de habitat (Valencia *et al.* 2004; Russo *et al.* 2005; John *et al.* 2007; Oliveira *et al.* 2013). No caso deste estudo, a tolerância das espécies quanto à falta ou ao excesso de água no solo seria um fator preponderante de suas distribuições, como sugerido por Engelbrecht *et al.* (2007).

Neste estudo, as métricas de intensidade (μ) e distância de agregação (2σ) estiveram positivamente relacionadas com a abundância das espécies (n). A relação entre μ e n foi mais forte e pode ser explicada mais facilmente, pois espera-se que espécies com maiores abundâncias (n maior) também terão maior densidade média de indivíduos por agregado (μ maior), desde que o número de agregado não cresça. A relação entre n e 2σ , contudo, foi mais variável e é possível que a relação entre n e 2σ para espécies mais abundantes seja um artefato da modelagem através do processo de Thomas. Acima de certas densidades, há grande sobreposição entre agregados e a distinção entre eles fica menos clara, o que pode ter feito com que a distância de agregação fosse superestimada. A relação encontrados entre r_c e n em

todas as classes de abundância, reforça que a relação positiva entre n e 2σ talvez seja mesmo válida. Vale lembrar que esse resultado não contradiz àqueles encontrados por Condit *et al.* (2000), que encontraram maiores intensidades de agregação a curtas distâncias (Ω_{0-10}) para espécies mais raras. Embora estes resultados pareçam indicar diferenças nos padrões de distribuição entre espécies raras e comuns, há estudos mostrando para outras medidas de agregação que essa relação é meramente uma questão amostral (Taylor *et al.* 1978; He & Gaston 2003). Estudos abordando a relação entre o ajuste do modelo de Thomas $g(r)$ e o tamanho da amostra são necessários para confirmar tais sugestões.

Agradecimentos Agradecemos aos pesquisadores responsáveis pelos censos realizados nas quatro parcelas: A.A. Oliveira, A.R. Duarte, A.P. Savassi-Coutinho, M.T.Z. Toniato, D. Sampaio, G.A.D.C. Franco, G. Durigan, N.M. Ivanauskas, R.R. Rodrigues, T.E. Barreto, e V.C. Souza. Agradecemos também a Camila Y. Mandai pelos comentários em versões prévias deste manuscrito. A coleta e checagem dos dados foram financiadas pela FAPESP (processo 99/09635-0 e 09/53413-5, respectivamente). RAFL recebeu bolsa CAPES (programa Demanda Social) e PIP recebeu bolsa CNPq (Bolsa de Produtividade 303878/2008-8).

Referências

- Armesto JJ, Mitchell JD, Villagran C (1986) A comparison of spatial patterns of trees in some tropical and temperate forests. *Biotropica* 18:1-11.
- Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. *J. Stat. Soft.* 12:1-42.
- Barot S, Gignoux J, Menaut JC (1999) Demography of a savanna palm tree: Predictions from comprehensive spatial pattern analyses. *Ecology* 80:1987-2005.
- Bolker, BM (2010) bbmle: Tools for general maximum likelihood estimation. R package version 0.9.5.1. <http://CRAN.R-project.org/package=bbmle>.
- Capretz RL, Batista JLF, Sotomayor JFM, Cunha CR, Nicoletti MF, Rodrigues RR (2012) Spatial pattern of four forest communities by Ripley's K function in Sao Paulo state. *Cienc. Flor.* 22:551-565.
- Collins SL, Klahr SC (1991) Tree dispersion in Oak-dominated forests along an environmental gradient. *Oecologia* 86:471-477.
- Condit R, Ashton PS, Baker P, Bunyavejchewin S, Gunatilleke CVS, Gunatilleke IAUN *et al.* (2000) Spatial patterns in the distribution of tropical tree species. *Science* 288:1414-1418.

- Engelbrecht BMJ, Comita LS, Condit R, Kursar TA, Tyree MT, Turner BL, Hubbell SP (2007) Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* 447:80-83.
- Green JL, Plotkin JB (2007) A statistical theory for sampling species abundances. *Ecol. Lett.* 10:1037-1045.
- Harms KE, Condit R, Hubbell SP, Foster RB (2001) Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *J. Ecol.* 89:947-959.
- Harte J, Conlisk E, Ostling A, Green JL, Smith AB (2005) A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol. Monog.* 75:179-197.
- He F, Legendre P, LaFrankie JV (1997) Distribution patterns of tree species in a Malaysian tropical rain forest. *J. Veg. Sci.* 8:105-114.
- He, F, Gaston KJ (2003) Occupancy, spatial variance, and the abundance of species. *Am. Nat.* 162:366-375.
- He, F, Legendre P (2002) Species diversity patterns derived from species-area models. *Ecology* 83:1185-1198.
- Hubbell, SP (1979) Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* 203:1299-1309.
- Hubbell, SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton.
- John R, Dalling JW, Harms KE, Yavitt JB, Stallard RF, Mirabello M. *et al.* (2007) Soil nutrients influence spatial distributions of tropical tree species. *P. Natl. Acad. Sci.* 104:864-869.
- Lan G, Hu Y, Cao M, Zhu H (2011) Topography related spatial distribution of dominant tree species in a tropical seasonal rain forest in China. *For. Ecol. Manag.* 262:1507-1513.
- Law R, Illian J, Burslem DF, Gratzner G, Gunatilleke CV, Gunatilleke I (2009) Ecological information from spatial patterns of plants: insights from point process theory. *J. Ecol.* 97:616-628.
- Li L, Huang Z, Ye W, Cao H, Wei S, Wang Z, Lian J, Sun I-F, Ma K, He F (2009) Spatial distributions of tree species in a subtropical forest of China. *Oikos* 118:495-502.
- Lima RAF, Oliveira AA, Martini AMZ, Sampaio D, Souza VC, Rodrigues RR (2011) Structure, diversity, and spatial patterns in a permanent plot of a high Restinga forest in Southeastern Brazil. *Acta Bot. Bras.* 25:633-645.

- Lima RAF, Rother DC, Muler AE, Lepsch IF, Rodrigues RR (2012) Bamboo overabundance alters forest structure and dynamics in the Atlantic Forest hotspot. *Biol. Conserv.* 147:32-39.
- Loosmore N, Ford ED (2006) Statistical inference using the G or K point pattern spatial statistics. *Ecology* 87:1925-1931.
- Manabe T, Nishimura N, Miura M, Yamamoto S (2000) Population structure and spatial patterns for trees in a temperate old-growth evergreen broad-leaved forest in Japan. *Plant Ecol.* 151:181-197.
- McGill BJ (2010) Towards a unification of unified theories of biodiversity. *Ecol. Lett.* 13:627-642.
- Oliveira AA, Vicentini A, Chave J, Lima RAF, Castanho CT, Martinit AMZ, Davies S, Rodrigues RR, Souza VC (2013) Habitat specialization and phylogenetic structure of tree species in a tropical rain forest on white sand. *J. Plant Ecol.* (submitted).
- Palmiotto PA, Davies SJ, Vogt KA, Ashton MS, Vogt DJ, Ashton PS (2004) Soil-related habitat specialization in dipterocarp rain forest tree species in Borneo. *J. Ecol.* 92: 609-623.
- Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecol.* 187:59-82.
- Phillips OL, Vargas PN, Monteagudo AL, Cruz AP, Zans MEC, Sanchez WG, Yli-Halla M, Rose S (2003) Habitat association among Amazonian tree species: a landscape-scale approach. *J. Ecol.* 91:757-775.
- Plotkin JB, Potts MD, Yu DW, Bunyavejchewin S, Condit R, Foster RB *et al.* (2000) Predicting species diversity in tropical forests. *P. Natl. Acad. Sci.* 97:10850-10854.
- Plotkin JB, Muller-Landau HC (2002) Sampling the species composition of a landscape. *Ecology* 83:3344-3356.
- Rodrigues RR (2005) Parcelas permanentes em 40ha de florestas do Estado de São Paulo: uma experiência interdisciplinar. <http://www.lerf.esalq.usp.br/parrel2005.php>.
- Russo SE, Davies SJ, King DA, Tan S (2005) Soil-related performance variation and distributions of tree species in a Bornean rain forest. *J. Ecol.* 93:879-889.
- Seidler TG, Plotkin JB (2006) Seed Dispersal and Spatial Pattern in Tropical Trees. *Plos Biol.* 4:e344.
- Szwagrzyk J, Czerwczak M (1993) Spatial patterns of trees in natural forests of East-Central Europe. *J. Veg. Sci.* 4:469-476.

- Taylor LR, Woiwod IP, Perry JN (1978) The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* 47:383-406.
- Tuomisto H, Ruokolainen K, Yli-Halla M (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*:241-244.
- Valencia R, Foster RB, Villa G, Condit R, Svenning JC, Hernández C, Romoleroux K, Losos E, Magård E, Balslev H (2004) Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *J. Ecol.* 92:214-229.
- Wiegand T, Moloney KA (2004) Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* 104:209-229.
- Wiegand T, Gunatilleke S, Gunatilleke N, Okuda T (2007) Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology* 88:3088-3102.
- Wiegand T, Martinez I, Huth A. (2009) Recruitment in tropical tree species: Revealing complex spatial patterns. *Am. Nat.* 174:e106-e140.
- Wiegand T, He F, Hubbell SP (2013) A systematic comparison of summary characteristics for quantifying point patterns in ecology. *Ecography* 36: 92-103.

Tabelas

Tabela 1 Proporção de espécies com distribuição aleatória e agregada por parcela. As indicações de agregação são indicadas pelos testes de aderência de Loosmore & Ford (2006 – LF), que avalia a agregação em qualquer escala, e *Maximum Absolute Deviation* (MAD), que avalia a agregação em curtas distâncias.

Parcela	n	teste LF		teste MAD	
		agregado	aleatório	agregado	aleatório
Cerradão	62	96,7	3,2	88,7	11,3
F. Estacional	61	65,6	34,4	54,1	45,9
F. de Restinga	80	82,5	17,5	70,0	30,0
F. Ombrófila	123	71,5	28,5	56,9	43,1

Tabela 2 Comparação das métricas de agregação espacial de populações arbóreas entre as quatro parcelas permanentes estudadas. Os valores de média $\pm$ desvio padrão se referem aos parâmetros μ e 2σ do modelo de Thomas e as métricas alternativas obtidas neste estudo (ver Apêndice 1). Apenas espécies com distribuições agregada foram consideradas nesta análise. Os resultados do teste F se referem ao modelo com melhor ajuste, que inclui tanto o efeito da parcela quanto o da abundância das populações sobre as respectivas métricas. Contudo, os modelos que desconsideram a abundância foram igualmente válidos. Letras diferentes indicam diferenças entre as médias de cada parcela permanente.

Parcela	n	μ	2σ (m)	A_c	r_c (m)
Cerradão	60	32,8 $\pm$ 48,5 A	47,3 $\pm$ 24,2 A	23,9 $\pm$ 25,3 A	48,5 $\pm$ 32,4 A
F. Estacional	40	8,5 $\pm$ 21,2 BC	28,8 $\pm$ 23,3 B	13,6 $\pm$ 30,6 B	15,5 $\pm$ 14,8 C
F. de Restinga	66	9,3 $\pm$ 12,9 B	31,8 $\pm$ 20,0 B	12,6 $\pm$ 15,3 B	28,8 $\pm$ 26,7 B
F. Ombrófila	88	3,5 $\pm$ 3,9 C	25,2 $\pm$ 16,7 B	10,1 $\pm$ 14,3 B	19,7 $\pm$ 23,5 BC
teste F^*	254	$F= 44,3$; $R^2= 0,41$	$F= 11,8$; $R^2= 0,15$	$F= 7,4$; $R^2= 0,09$	$F= 10,5$; $R^2= 0,14$

* Para todas as análises o número de grau de liberdade foi de 249 e o valor $p < 0.001$.

Fig. 1 Distribuição espacial de algumas espécies apresentadas junto com as suas funções de distribuição radial de pares $g(r)$ e o ajuste da função ao modelo de Thomas. A linha pontilhada indica o valor esperado sob completa aleatoriedade espacial ($g(r) = 1$), enquanto que a linha tracejada indica a distância de agregação ($r_c = 2\sigma$) que contém em média 86% dos pontos estão localizados. (A) *Protium heptaphyllum* (Burseraceae) no cerradão (distribuição agregada); (B) *Campomanesia xanthocarpa* (Myrtaceae) na floresta estacional (distribuição aleatória); (C) *Aparisthium cordatum* (Euphorbiaceae) na floresta de restinga (distribuição agregada).

Fig. 2 Distribuição de frequência dos parâmetros de intensidade (μ) e tamanho de agregação (2σ) do modelo de Thomas (A e B) para diferentes espécies arbóreas, e sua comparação entre as quatro florestas estudadas (C e D). Nas figuras acima a linha tracejada representa o terceiro quartil da distribuição, i.e., valor até o qual 75% dos dados se encontram. Nos *boxplots* abaixo, a linha em negrito e as dobras na caixa equivalem à mediana e ao intervalo de confiança 95% ao redor dela, enquanto que os limites da caixa e os extremos das linhas tracejadas equivalem aos quartis da distribuição. Os valores de μ e 2σ são dados em números de indivíduos e metros, respectivamente.

Fig. 3 Relação dos parâmetros μ e 2σ do modelo de Thomas com a abundância das espécies (A e B) e entre si (C). Cada círculo representa a população de uma espécie em uma das parcelas estudadas e a linha representa o previsto pelo modelo de regressão linear Gaussiana aplicado às variáveis log-transformadas. Apenas as espécies com distribuição agregada foram usadas nesta análise. Note que os eixos de todas as figuras estão em escala logarítmica. Os valores de μ e 2σ são dados em números de indivíduos e metros, respectivamente.

Apêndices

Apêndice 1 Descrição das métricas alternativas para descrever a agregação e os resultados das análises usando estas métricas.

Além dos parâmetros μ e σ do modelo de Thomas, foram obtidas métricas alternativas para descrever os padrões de agregação das espécies. Como a maioria das espécies com distribuição agregada mostrou valores decrescentes de $g(r)$ com a distância, definiu-se as primeiras métricas alternativas, denominadas aqui de A_c e r_c , que buscam medir a força e distância máxima de agregação, respectivamente. A primeira métrica é definida pela soma das áreas acima e abaixo da curva observada de $g(r)$ em relação a um, que representa o esperado para uma distribuição sob completa aleatoriedade espacial. Na prática, ao invés de calcular a integral sob as curvas, o valor de A_c foi calculado a partir da soma dos valores observados de $g(r) - 1$ até a distância máxima considerada. Valores positivos de A_c indicam agregação enquanto que valores A_c indicam regularidade. A segunda métrica representa a distância na qual a curva atinge um, ou seja, representa a distância até a qual há evidência de agregação espacial. Quanto maior o valor de r_c , maior a distância de agregação. Essas métricas podem ser obtidas de duas formas, uma a partir da média de $g(r)$ (Figura S1a) ou a partir dos intervalos de confiança ao redor da média de $g(r)$ (Figura S1b).

A terceira métrica de agregação, Ω_{0-10} , foi proposta por Condit *et al.* (2000) e representa a densidade relativa média de indivíduos de uma mesma espécie até 10 metros de distância. Assim, essa métrica pode ser vista como um índice de agregação a curtas distâncias. Na prática, o índice Ω_{0-10} foi definido como o valor de $g(r)$ quando $r=10$ e esteve positivamente relacionado à métrica A_c . Estas três métricas (A_c , r_c e Ω_{0-10}) são métrica empíricas e mais parcimoniosas, visto que elas não assumem nenhum tipo de modelo de agregação espacial. Além disso, elas não são interdependentes quanto à sua estimação, ao contrário dos parâmetros μ e σ do modelo de Thomas (Wiegand *et al.* 2009). Houve evidência também que os valores de A_c foram independentes da abundância da espécie ($F=0.580$, g.l= 246, $p=0.447$). Assim, essas métricas foram calculadas para validar os resultados obtidos usando os ajustes do modelo de Thomas. Os resultados para A_c e Ω_{0-10} foram inversos àqueles encontrados para μ , enquanto que as tendências encontradas entre σ e n foram basicamente às mesmas encontradas para r_c e n (Figura S2). Para esta última métrica, a retirada das espécies muito abundantes tornou a relação entre r_c e n não significativa ($F=2.01$, g.l.= 124, DF, $p=0.159$).

Fig. S1 Esquema ilustrativo de como foram obtidas as métricas A_c , r_c e Ω_{0-10} , calculadas a partir (a) dos valores médios e dos intervalos de confiança. O exemplo apresenta os resultados para a espécie *Terminalia brasiliensis* no Cerradão, cuja função $g(r)$ foi obtida para intervalos de 2.5 m e valor máximo de distância 100 m.

Fig. S2 Resultados para a relação entre as três métricas alternativas (a) A_c , (b) r_c e (c) Ω_{0-10} com a abundância das espécies. Cada ponto representa uma espécie diferente, enquanto que linha sólida representa o previsto pelo modelo de regressão linear aplicado sobre as variáveis log-transformadas, cujos resultados principais são apresentados na legenda. No gráfico mais à direita, os pontos cheios significam as espécies cujo intervalo de confiança ao redor de Ω_{0-10} não inclui um (linha tracejada), que representa o esperado sob completa aleatoriedade espacial; pontos vazios significam o oposto.

Apêndice 2 Valores ajustados dos parâmetros do modelo de Thomas e os resultados do teste de aderência que determinou se a distribuição $g(r)$ diferiu do esperado sob completa aleatoriedade espacial. O modelo de Thomas possui três parâmetros que são a densidade de agregados κ , a densidade de árvores por agregado μ e o raio de agregação σ . Já o teste de aderência proposto por Loosmore & Ford (2006) avalia se há agregação em alguma classe de distância, nesse caso avaliado entre 2.5 e 80 m.

Tabela S1 Valores de abundância (n), dos parâmetros do modelo de Thomas (κ , μ e σ) e do teste de aderência (estatística e valor de p) das populações avaliadas em cada parcela permanente. Os valores de κ e σ foram multiplicados por 100 e dois, respectivamente. Para espécies cujo teste de aderência não sugeriu evidências de agregação os parâmetros do modelo de Thomas não são apresentados.

Espécie	Site	n	100 κ	2 σ	μ	LF	p
<i>Actinostemon conceptionis</i>	eec	24	—	—	—	21.9	0.39
<i>Actinostemon concolor</i>	eec	66	0.031	16.3	2.07	210.1	0.005
<i>Agonandra excelsa</i>	eec	19	0.041	12.9	0.46	196.1	0.02
<i>Aiouea saligna</i>	peic	45	0.029	33.2	1.52	24.3	0.025
<i>Albizia pedicellaris</i>	peic	168	0.473	9.6	0.35	3.6	0.015
<i>Albizia polycephala</i>	eec	17	0.074	12.7	0.22	90.7	0.095
<i>Alchornea glandulosa</i>	pecb	43	0.169	10.3	0.25	31.9	0.035
<i>Alchornea triplinervia</i>	peic	53	—	—	—	4.8	0.555
<i>Allophylus petiolulatus</i>	pecb	27	—	—	—	25.5	0.17
<i>Alseis floribunda</i>	pecb	87	0.240	10.0	0.35	11.2	0.01
<i>Alsophila sternbergii</i>	pecb	331	0.041	12.7	7.82	379.0	0.005
<i>Amaioua intermedia</i>	eea	380	0.012	45.0	29.96	45.3	0.005
<i>Amaioua intermedia</i>	peic	857	0.046	39.6	18.33	5.2	0.005
<i>Andira anthelmia</i>	peic	559	0.050	101.5	10.95	0.6	0.02
<i>Aniba viridis</i>	peic	66	0.007	48.9	9.60	128.8	0.005
<i>Annona neosericea</i>	pecb	28	0.025	15.4	1.08	327.8	0.005
<i>Annona sylvatica</i>	eec	15	—	—	—	9.7	0.955
<i>Aparisthium cordatum</i>	peic	182	0.026	22.4	6.82	90.3	0.005
<i>Aspidosperma polyneuron</i>	eec	653	0.136	28.9	4.69	1.4	0.005
<i>Astrocaryum aculeatissimum</i>	pecb	30	0.172	4.2	0.17	523.1	0.005
<i>Astrocaryum aculeatissimum</i>	peic	79	0.010	39.0	7.84	329.6	0.005
<i>Astronium graveolens</i>	eec	163	0.029	58.0	5.40	5.3	0.015
<i>Balfourodendron riedelianum</i>	eec	352	0.074	14.4	4.62	39.1	0.005
<i>Bathysa australis</i>	pecb	265	0.042	26.7	6.09	20.2	0.005
<i>Blepharocalyx salicifolius</i>	peic	51	0.012	54.9	4.05	42.6	0.005
<i>Bowdichia virgilioides</i>	eea	28	0.005	18.9	6.01	7575.5	0.005
<i>Brosimum glaziovii</i>	pecb	30	0.008	33.0	3.62	535.3	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Byrsonima laxiflora</i>	eea	309	0.009	83.2	32.78	13.3	0.005
<i>Byrsonima ligustrifolia</i>	peic	222	0.028	22.1	7.76	108.2	0.005
<i>Cabranea canjerana</i>	pecb	51	0.311	8.3	0.16	17.2	0.07
<i>Cabranea canjerana</i>	eec	26	–	–	–	18.3	0.43
<i>Calophyllum brasiliense</i>	peic	415	0.522	14.2	0.78	1.5	0.005
<i>Calyptanthus concinna</i>	peic	332	0.023	34.3	13.89	36.0	0.005
<i>Calyptanthus lanceolata</i>	pecb	48	0.059	12.4	0.79	101.9	0.005
<i>Campomanesia xanthocarpa</i>	eec	101	–	–	–	2.0	0.465
<i>Cariniana estrellensis</i>	eec	20	0.015	36.1	1.27	59.6	0.09
<i>Cariniana estrellensis</i>	pecb	58	–	–	–	4.4	0.57
<i>Casearia decandra</i>	pecb	49	0.846	5.0	0.06	18.9	0.07
<i>Casearia gossypiosperma</i>	eec	78	0.028	32.3	2.73	39.1	0.005
<i>Casearia obliqua</i>	pecb	21	–	–	–	43.5	0.16
<i>Casearia sylvestris</i>	eea	29	0.023	6.0	1.24	7886.6	0.005
<i>Casearia sylvestris</i>	pecb	40	0.041	18.0	0.95	72.9	0.015
<i>Casearia sylvestris</i>	eec	51	–	–	–	6.1	0.37
<i>Cecropia glaziovii</i>	eec	15	0.028	11.5	0.52	988.4	0.005
<i>Cecropia glaziovii</i>	pecb	93	0.085	9.4	1.07	155.4	0.005
<i>Cedrela fissilis</i>	pecb	18	–	–	–	32.4	0.315
<i>Cedrela fissilis</i>	eec	58	–	–	–	5.7	0.345
<i>Ceiba speciosa</i>	eec	40	0.028	39.3	1.37	34.6	0.04
<i>Centrolobium robustum</i>	pecb	18	0.004	45.9	4.93	285.4	0.03
<i>Centrolobium tomentosum</i>	eec	412	0.415	13.2	0.97	2.7	0.005
<i>Chionanthus filiformis</i>	peic	15	–	–	–	49.8	0.285
<i>Chrysophyllum gonocarpum</i>	eec	231	0.017	95.3	13.27	3.2	0.005
<i>Chrysophyllum inornatum</i>	pecb	93	0.013	20.3	7.21	720.3	0.005
<i>Chrysophyllum viride</i>	pecb	108	0.028	36.1	3.77	20.6	0.005
<i>Cinnamodendron dinisii</i>	pecb	39	0.099	15.1	0.38	40.1	0.02
<i>Citronella paniculata</i>	pecb	55	0.007	82.2	8.21	38.8	0.005
<i>Citronella paniculata</i>	eec	17	0.002	98.9	7.78	163.9	0.03
<i>Clethra scabra</i>	eea	29	0.002	21.5	12.77	31212.3	0.005
<i>Clethra scabra</i>	peic	26	–	–	–	18.5	0.335
<i>Clusia criuva</i>	peic	98	0.010	39.0	9.85	123.2	0.005
<i>Coccoloba fastigiata</i>	pecb	25	0.014	37.1	1.69	80.6	0.045
<i>Copaifera langsdorffii</i>	eea	4015	0.071	34.3	55.50	NA	NA
<i>Cordia americana</i>	eec	31	–	–	–	18.7	0.21
<i>Cordia ecalyculata</i>	eec	81	0.312	6.2	0.25	24.6	0.005
<i>Cordia sellowiana</i>	pecb	24	–	–	–	10.1	0.735
<i>Cordia silvestris</i>	pecb	41	–	–	–	5.4	0.72
<i>Cordia myrciifolia</i>	peic	92	0.045	23.8	2.00	31.2	0.005
<i>Coussapoa microcarpa</i>	pecb	37	–	–	–	17.9	0.155
<i>Coussapoa microcarpa</i>	peic	36	–	–	–	4.9	0.83
<i>Coussarea contracta</i>	pecb	144	0.009	32.9	15.30	278.2	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Croton floribundus</i>	eea	106	0.001	53.5	81.64	2894.6	0.005
<i>Croton floribundus</i>	eec	450	0.093	8.0	4.71	199.0	0.005
<i>Croton macrobothrys</i>	pecb	26	0.013	9.6	1.88	6226.4	0.005
<i>Cryptocarya mandioccana</i>	pecb	154	0.087	22.9	1.73	8.9	0.005
<i>Cupania oblongifolia</i>	pecb	30	0.030	21.2	0.99	94.9	0.015
<i>Cupania vernalis</i>	eec	43	–	–	–	6.9	0.425
<i>Cyathea corcovadensis</i>	pecb	27	0.003	25.8	9.13	6294.7	0.005
<i>Cyathea microdonta</i>	peic	27	0.025	17.1	1.07	249.2	0.015
<i>Cyathea phalerata</i>	pecb	137	0.057	12.2	2.36	118.0	0.005
<i>Cybianthus peruvianus</i>	peic	18	–	–	–	89.3	0.105
<i>Dahlstedtia pinnata</i>	pecb	75	0.007	24.4	9.99	993.8	0.005
<i>Daphnopsis utilis</i>	eea	215	0.011	47.7	19.87	54.9	0.005
<i>Diatenopteryx sorbifolia</i>	eec	48	0.009	29.2	5.40	504.3	0.005
<i>Diploon cuspidatum</i>	pecb	63	0.037	30.4	1.66	24.3	0.01
<i>Ecclinusa ramiflora</i>	pecb	93	0.015	36.4	6.00	76.2	0.005
<i>Endlicheria paniculata</i>	pecb	26	0.100	8.1	0.25	213.6	0.005
<i>Endlicheria paniculata</i>	peic	53	0.303	5.0	0.17	83.0	0.005
<i>Enterolobium contortisiliquum</i>	eec	16	–	–	–	31.8	0.52
<i>Erythroxylum amplifolium</i>	peic	61	0.004	25.6	16.31	3361.7	0.005
<i>Erythroxylum deciduum</i>	eea	22	0.002	67.5	10.77	1431.9	0.005
<i>Esenbeckia leiocarpa</i>	eec	189	0.002	28.4	83.40	6269.8	0.005
<i>Eugenia aurata</i>	eea	44	0.007	74.1	6.30	40.8	0.015
<i>Eugenia blastantha</i>	eec	33	0.006	53.8	5.23	130.9	0.005
<i>Eugenia brevistyla</i>	pecb	126	0.171	9.4	0.72	32.9	0.005
<i>Eugenia bunchosifolia</i>	pecb	18	0.011	45.4	1.63	141.8	0.04
<i>Eugenia capitulifera</i>	pecb	19	0.033	10.6	0.56	815.4	0.005
<i>Eugenia cerasiflora</i>	pecb	21	–	–	–	11.1	0.81
<i>Eugenia</i> cf. <i>verticillata</i>	pecb	42	0.008	33.6	5.10	388.3	0.005
<i>Eugenia cuprea</i>	pecb	189	0.024	48.3	7.56	14.4	0.005
<i>Eugenia melanogyna</i>	pecb	158	0.204	9.6	0.76	25.9	0.005
<i>Eugenia mosenii</i>	pecb	127	0.006	52.4	19.82	111.9	0.005
<i>Eugenia neoglomerata</i>	pecb	104	0.046	21.7	2.23	50.4	0.005
<i>Eugenia neoglomerata</i>	peic	17	0.017	22.1	0.99	631.7	0.02
<i>Eugenia neoverrucosa</i>	pecb	24	–	–	–	13.4	0.575
<i>Eugenia oblongata</i>	pecb	78	0.017	33.8	4.38	67.0	0.005
<i>Eugenia platysema</i>	pecb	42	0.019	45.3	2.20	36.1	0.03
<i>Eugenia pluriflora</i>	eea	19	0.083	10.5	0.22	98.4	0.065
<i>Eugenia prasina</i>	pecb	40	0.018	17.3	2.12	442.5	0.005
<i>Eugenia pruinosa</i>	pecb	32	–	–	–	19.5	0.16
<i>Eugenia ramboi</i>	eec	169	0.301	6.9	0.55	22.0	0.005
<i>Eugenia</i> sp. 1	pecb	20	–	–	–	25.3	0.485
<i>Eugenia stigmatica</i>	peic	39	0.059	26.0	0.65	18.9	0.08
<i>Eugenia subavenia</i>	pecb	54	0.018	23.0	2.95	356.4	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Eugenia subterminalis</i>	eec	19	0.015	28.5	1.22	136.8	0.045
<i>Eugenia sulcata</i>	peic	186	0.125	20.1	1.45	9.2	0.005
<i>Eugenia supraaxillaris</i>	pecb	93	0.121	20.7	0.75	7.7	0.05
<i>Eugenia umbelliflora</i>	peic	63	0.020	52.5	3.11	14.9	0.045
<i>Eugenia verticillata</i>	pecb	57	0.044	11.0	1.28	296.8	0.005
<i>Eugenia xiriricana</i>	pecb	96	0.021	42.9	4.48	23.2	0.005
<i>Euterpe edulis</i>	pecb	2608	0.235	22.5	10.82	1.2	0.005
<i>Euterpe edulis</i>	peic	3041	0.116	53.5	25.70	1.2	0.005
<i>Faramea montevidensis</i>	eea	309	0.014	30.5	21.16	125.9	0.005
<i>Gallesia integrifolia</i>	eec	35	0.009	28.7	3.88	483.8	0.005
<i>Garcinia gardneriana</i>	peic	253	0.016	40.3	15.07	40.4	0.005
<i>Garcinia gardneriana</i>	pecb	297	0.030	28.1	9.51	33.8	0.005
<i>Geonoma schottiana</i>	peic	17	0.010	13.8	1.64	3705.0	0.005
<i>Gochnatia polymorpha</i>	eea	197	0.034	36.1	5.59	13.7	0.005
<i>Guapira graciliflora</i>	eea	22	0.006	52.2	3.73	239.7	0.01
<i>Guapira noxia</i>	eea	17	0.002	81.9	6.98	300.2	0.015
<i>Guapira opposita</i>	eea	16	0.017	9.9	0.94	4261.2	0.005
<i>Guapira opposita</i>	pecb	463	0.053	22.0	8.60	21.5	0.005
<i>Guarea macrophylla</i>	peic	15	0.014	11.0	1.02	5078.0	0.005
<i>Guarea macrophylla</i>	pecb	77	0.053	34.8	1.41	9.0	0.025
<i>Guatteria australis</i>	peic	110	0.061	28.0	1.76	9.1	0.03
<i>Guatteria australis</i>	pecb	27	0.108	9.0	0.24	93.6	0.04
<i>Handroanthus albus</i>	peic	29	0.034	18.6	0.83	108.1	0.015
<i>Handroanthus</i> cf. <i>ochraceus</i>	eea	18	0.004	26.9	4.15	3429.6	0.005
<i>Handroanthus serratifolius</i>	pecb	103	0.028	59.7	3.58	5.2	0.055
<i>Hedyosmum brasiliense</i>	peic	16	0.019	16.3	0.84	734.4	0.015
<i>Heisteria silvianii</i>	peic	117	0.024	27.2	4.84	60.8	0.005
<i>Heisteria silvianii</i>	pecb	65	–	–	–	7.8	0.105
<i>Hieronyma alchorneoides</i>	pecb	148	2.044	4.2	0.07	2.9	0.075
<i>Hirtella angustifolia</i>	peic	21	0.003	23.3	6.38	5739.2	0.005
<i>Hirtella hebeclada</i>	pecb	45	–	–	–	14.9	0.12
<i>Holocalyx balansae</i>	eec	104	0.308	10.9	0.33	4.8	0.07
<i>Humirastrum dentatum</i>	peic	40	–	–	–	10.6	0.295
<i>Hymenaea courbaril</i>	pecb	19	0.009	53.7	2.09	82.1	0.09
<i>Hymenolobium janeirense</i>	peic	17	–	–	–	34.6	0.395
<i>Ilex dumosa</i>	peic	76	0.097	15.9	0.77	21.0	0.01
<i>Ilex paraguariensis</i>	eea	33	0.003	51.2	9.38	896.9	0.005
<i>Ilex theezans</i>	peic	177	0.003	45.2	58.69	831.8	0.005
<i>Inga lenticellata</i>	pecb	61	0.340	8.0	0.17	12.4	0.065
<i>Inga marginata</i>	eec	38	0.021	24.1	1.77	228.1	0.005
<i>Inga marginata</i>	pecb	45	0.076	15.9	0.58	31.6	0.03
<i>Inga sessilis</i>	pecb	31	0.004	50.6	8.22	460.1	0.005
<i>Inga striata</i>	eec	41	0.090	16.6	0.44	19.9	0.08

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Ixora brevifolia</i>	eea	58	0.002	31.9	24.31	4262.0	0.005
<i>Jacaranda micrantha</i>	eec	25	0.011	44.4	2.24	82.8	0.03
<i>Jacaranda puberula</i>	pecb	37	0.006	34.9	6.11	689.1	0.005
<i>Jacaranda puberula</i>	peic	59	–	–	–	3.8	0.615
<i>Jacaratia spinosa</i>	eec	26	–	–	–	10.1	0.7
<i>Lafoensia pacari</i>	eea	36	0.006	31.6	6.21	2035.8	0.005
<i>Laplacea fruticosa</i>	peic	15	–	–	–	39.1	0.59
<i>Leptolobium elegans</i>	eea	121	0.002	78.2	47.91	225.4	0.005
<i>Licaria armeniaca</i>	pecb	22	0.037	12.2	0.58	181.5	0.03
<i>Lonchocarpus cultratus</i>	eec	52	0.048	18.3	1.06	104.1	0.005
<i>Machaerium acutifolium</i>	eea	1032	0.007	107.7	152.8	9.4	0.005
<i>Machaerium brasiliense</i>	eea	156	0.008	53.0	19.78	81.1	0.005
<i>Machaerium nyctitans</i>	eec	16	–	–	–	33.7	0.335
<i>Machaerium stipitatum</i>	eec	104	0.145	7.7	0.70	76.9	0.005
<i>Magnolia ovata</i>	pecb	31	0.071	15.5	0.43	68.5	0.02
<i>Malouetia cestroides</i>	pecb	50	–	–	–	5.2	0.505
<i>Manilkara subsericea</i>	peic	231	0.018	46.5	12.55	21.7	0.005
<i>Maprounea guianensis</i>	eea	144	0.013	53.3	11.18	33.8	0.005
<i>Marlierea eugeniopsoides</i>	pecb	43	0.051	12.1	0.82	202.6	0.005
<i>Marlierea obscura</i>	pecb	87	0.025	34.5	3.46	34.9	0.005
<i>Marlierea racemosa</i>	peic	191	0.008	37.4	22.16	213.5	0.005
<i>Marlierea suaveolens</i>	pecb	146	0.014	144.2	10.50	3.3	0.055
<i>Marlierea tomentosa</i>	pecb	146	0.019	42.2	7.60	31.6	0.005
<i>Matayba elaeagnoides</i>	eea	83	0.048	6.4	1.69	983.2	0.005
<i>Matayba intermedia</i>	peic	90	0.033	36.7	2.66	12.2	0.005
<i>Matayba obovata</i>	pecb	71	0.024	41.7	2.84	24.0	0.015
<i>Maytenus robusta</i>	eea	48	0.014	30.9	3.23	252.4	0.005
<i>Maytenus robusta</i>	peic	162	0.081	11.4	1.96	96.7	0.005
<i>Maytenus robusta</i>	pecb	44	0.025	40.2	1.71	22.4	0.055
<i>Meliosma sellowii</i>	pecb	20	0.037	17.8	0.53	184.7	0.01
<i>Meriania sanchezii</i>	pecb	23	0.101	9.7	0.22	82.4	0.05
<i>Metrodorea nigra</i>	eec	4625	0.341	57.3	13.20	NA	NA
<i>Miconia cubatanensis</i>	peic	273	0.021	27.2	12.96	83.6	0.005
<i>Miconia ligustroides</i>	eea	138	0.029	42.8	4.66	11.0	0.005
<i>Miconia sellowiana</i>	eea	37	0.002	61.3	16.15	663.9	0.005
<i>Mollinedia schottiana</i>	peic	22	0.015	11.3	1.39	3107.7	0.005
<i>Mollinedia schottiana</i>	pecb	62	0.023	32.5	2.59	41.4	0.01
<i>Mollinedia uleana</i>	pecb	84	0.068	18.6	1.20	31.8	0.005
<i>Mollinedia widgrenii</i>	eec	37	–	–	–	17.3	0.16
<i>Myrceugenia kleinii</i>	pecb	24	0.018	37.6	1.32	50.4	0.075
<i>Myrceugenia myrcioides</i>	pecb	15	0.002	99.2	7.06	210.2	0.035
<i>Myrceugenia pilotantha</i>	pecb	21	0.024	22.2	0.85	121.8	0.03
<i>Myrcia anacardiifolia</i>	pecb	62	0.052	13.5	1.16	98.8	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Myrcia bella</i>	eea	26	0.010	43.1	2.66	159.0	0.015
<i>Myrcia bicarinata</i>	peic	77	0.007	32.2	10.43	403.3	0.005
<i>Myrcia brasiliensis</i>	peic	129	0.286	7.6	0.44	23.5	0.005
<i>Myrcia flagellaris</i>	pecb	171	0.065	24.5	2.56	12.5	0.005
<i>Myrcia glabra</i>	peic	87	0.044	22.5	1.95	38.8	0.005
<i>Myrcia grandiflora</i>	peic	62	0.018	29.0	3.43	92.2	0.005
<i>Myrcia guianensis</i>	eea	874	0.008	86.3	110.2	17.3	0.005
<i>Myrcia hebeptala</i>	peic	50	–	–	–	11.1	0.165
<i>Myrcia ilheosensis</i>	peic	133	0.004	48.2	36.12	447.4	0.005
<i>Myrcia linguaeformis</i>	eea	33	0.005	42.2	6.10	451.4	0.005
<i>Myrcia macrocarpa</i>	pecb	34	0.019	34.9	1.76	55.8	0.03
<i>Myrcia multiflora</i>	peic	113	0.016	26.0	6.78	173.2	0.005
<i>Myrcia multiflora</i>	eea	898	0.018	62.9	48.67	11.3	0.005
<i>Myrcia pubipetala</i>	pecb	42	–	–	–	5.6	0.645
<i>Myrcia racemosa</i>	peic	446	0.183	17.6	2.38	5.1	0.005
<i>Myrcia</i> sp.1	peic	145	0.130	19.8	1.09	8.0	0.01
<i>Myrcia spectabilis</i>	pecb	65	0.501	4.3	0.13	47.6	0.01
<i>Myrcia splendens</i>	eea	243	0.004	56.8	66.01	281.0	0.005
<i>Myrcia splendens</i>	peic	16	–	–	–	28.9	0.47
<i>Myrcia tenuivenosa</i>	pecb	15	0.114	7.6	0.13	239.8	0.02
<i>Myrcia tijucensis</i>	pecb	58	0.019	34.9	2.91	57.9	0.005
<i>Myrcia tomentosa</i>	eea	24	0.834	3.7	0.03	36.8	0.09
<i>Myrcia venulosa</i>	eea	339	0.065	29.9	5.07	7.8	0.005
<i>Myrciaria floribunda</i>	pecb	40	0.755	4.4	0.05	21.1	0.055
<i>Myrocarpus frondosus</i>	pecb	31	–	–	–	20.4	0.16
<i>Myrsine hermogenesii</i>	pecb	42	0.028	17.6	1.48	167.7	0.005
<i>Myrsine lancifolia</i>	eea	290	0.014	22.3	20.82	427.1	0.005
<i>Myrsine umbellata</i>	peic	46	0.013	24.4	3.55	346.4	0.005
<i>Myrsine umbellata</i>	eea	968	0.016	59.1	58.27	13.6	0.005
<i>Myrsine venosa</i>	peic	47	0.011	38.5	4.14	104.1	0.005
<i>Nectandra cuspidata</i>	eea	441	0.004	65.8	106.2	140.1	0.005
<i>Nectandra grandiflora</i>	peic	22	0.003	55.6	6.53	390.2	0.005
<i>Nectandra megapotamica</i>	eec	28	0.005	45.2	5.34	755.5	0.005
<i>Nectandra membranacea</i>	pecb	60	0.342	5.9	0.17	36.1	0.02
<i>Nectandra oppositifolia</i>	peic	29	0.032	11.7	0.88	520.1	0.005
<i>Neomitranthes glomerata</i>	eec	65	0.026	27.2	2.46	68.0	0.005
<i>Neomitranthes glomerata</i>	pecb	115	0.028	54.8	3.96	6.5	0.025
<i>Ocotea aciphylla</i>	peic	273	0.010	64.9	27.96	27.7	0.005
<i>Ocotea corymbosa</i>	eea	1526	0.057	68.5	25.98	0.7	0.005
<i>Ocotea dispersa</i>	pecb	20	0.015	21.2	1.30	455.4	0.005
<i>Ocotea dispersa</i>	peic	32	0.014	26.4	2.16	167.1	0.005
<i>Ocotea indecora</i>	eec	685	0.034	82.8	19.68	1.5	0.005
<i>Ocotea pulchella</i>	peic	542	0.010	65.6	52.77	22.0	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Ocotea pulchra</i>	peic	50	0.002	74.9	21.69	303.9	0.005
<i>Ocotea teleiandra</i>	pecb	62	0.043	24.5	1.42	49.2	0.01
<i>Ocotea venulosa</i>	peic	43	0.003	73.9	16.42	239.4	0.005
<i>Ormosia arborea</i>	eea	37	0.003	48.8	12.93	1347.2	0.005
<i>Ormosia arborea</i>	peic	95	1.381	3.5	0.07	10.7	0.01
<i>Ouratea spectabilis</i>	eea	32	0.003	52.5	9.85	487.9	0.005
<i>Parapiptadenia rigida</i>	eec	44	–	–	–	6.8	0.51
<i>Parinari brasiliensis</i>	pecb	33	0.017	28.1	1.91	126.6	0.005
<i>Peltophorum dubium</i>	eec	24	0.504	4.6	0.05	46.4	0.075
<i>Pera glabrata</i>	eea	510	0.025	36.1	19.55	21.5	0.005
<i>Pera glabrata</i>	peic	537	0.019	72.0	27.82	4.7	0.005
<i>Persea willdenovii</i>	eea	195	0.011	42.9	17.97	79.3	0.005
<i>Picramnia ramiflora</i>	eec	20	0.012	26.4	1.65	294.6	0.01
<i>Pilocarpus pauciflorus</i>	eec	43	0.007	12.3	5.89	10379.4	0.005
<i>Pimenta</i> cf. <i>pseudocaryophyllus</i>	peic	67	0.126	15.0	0.52	23.3	0.01
<i>Piptadenia gonoacantha</i>	eec	76	0.867	3.9	0.09	22.0	0.015
<i>Piptocarpha axillaris</i>	eea	82	0.007	93.6	11.04	19.8	0.005
<i>Pisonia ambigua</i>	pecb	20	–	–	–	56.2	0.11
<i>Platymiscium floribundum</i>	pecb	34	0.152	11.2	0.22	31.0	0.085
<i>Platypodium elegans</i>	eea	115	0.003	49.6	35.55	775.8	0.005
<i>Plinia complanata</i>	pecb	17	0.222	5.1	0.07	255.5	0.015
<i>Podocarpus sellowii</i>	peic	29	0.022	14.6	1.30	712.3	0.005
<i>Posoqueria latifolia</i>	peic	75	0.226	9.5	0.32	17.3	0.01
<i>Pourouma guianensis</i>	pecb	17	–	–	–	11.3	0.88
<i>Pouteria beaurepairei</i>	peic	249	0.022	74.4	10.98	3.5	0.005
<i>Pouteria caimito</i>	pecb	20	–	–	–	11.4	0.825
<i>Pouteria psammophila</i>	pecb	39	–	–	–	10.4	0.28
<i>Pouteria ramiflora</i>	eea	129	0.018	32.2	6.88	133.9	0.005
<i>Protium heptaphyllum</i>	eea	453	0.011	52.8	41.33	63.7	0.005
<i>Protium kleinii</i>	pecb	17	0.014	20.3	1.17	543.0	0.01
<i>Prunus myrtifolia</i>	eea	15	0.003	44.2	5.35	1910.3	0.005
<i>Pseudopiptadenia warmingii</i>	pecb	18	–	–	–	32.5	0.355
<i>Psidium cattleianum</i>	peic	241	0.039	49.2	6.05	7.4	0.005
<i>Psychotria nemorosa</i>	pecb	17	0.236	4.7	0.07	197.7	0.025
<i>Psychotria suterella</i>	pecb	51	0.248	6.5	0.20	67.2	0.015
<i>Pterocarpus rohrii</i>	pecb	118	0.864	7.7	0.13	5.1	0.035
<i>Qualea cordata</i>	eea	472	0.005	51.4	92.38	183.0	0.005
<i>Qualea grandiflora</i>	eea	55	0.021	15.9	2.55	640.3	0.005
<i>Quiina glazovii</i>	pecb	72	0.020	57.8	3.49	12.9	0.04
<i>Rhamnidium elaeocarpum</i>	eec	42	–	–	–	17.7	0.1
<i>Roupala montana</i>	eea	24	0.013	20.6	1.82	886.8	0.005
<i>Roupala montana</i>	pecb	32	0.031	23.2	1.01	64.0	0.025
<i>Rudgea jasminoides</i>	pecb	150	0.015	40.4	9.73	48.6	0.005

Espécie	Site	<i>n</i>	100κ	2σ	μ	LF	<i>p</i>
<i>Rudgea recurva</i>	pecb	69	0.007	23.8	9.24	1118.1	0.005
<i>Savia dictyocarpa</i>	eec	158	0.001	40.0	107.1	5013.4	0.005
<i>Schefflera angustissima</i>	peic	582	0.604	11.2	0.94	1.5	0.005
<i>Seguiera americana</i>	eec	35	–	–	–	9.6	0.53
<i>Senegalia polyphylla</i>	eec	64	0.055	23.5	1.14	18.7	0.015
<i>Simira corumbensis</i>	pecb	68	0.021	23.6	3.14	131.5	0.005
<i>Siparuna guianensis</i>	eea	287	0.008	40.8	35.79	153.8	0.005
<i>Siphoneugena guilfoyleiana</i>	peic	81	0.006	31.3	14.07	885.5	0.005
<i>Sloanea guianensis</i>	pecb	105	0.038	30.7	2.68	18.7	0.01
<i>Sloanea hirsuta</i>	pecb	44	0.124	11.1	0.35	48.3	0.015
<i>Solanum argenteum</i>	eec	44	0.013	21.1	3.43	568.9	0.005
<i>Sorocea bonplandii</i>	pecb	28	0.136	6.7	0.20	211.5	0.01
<i>Sorocea bonplandii</i>	eec	16	0.036	13.9	0.43	250.3	0.015
<i>Styrax camporum</i>	eea	25	0.004	52.4	6.83	777.2	0.005
<i>Styrax ferrugineus</i>	eea	19	0.017	13.2	1.08	2177.0	0.005
<i>Swartzia submarginata</i>	pecb	21	–	–	–	18.1	0.635
<i>Syagrus oleracea</i>	eec	154	0.076	16.5	1.98	33.1	0.005
<i>Syagrus romanzoffiana</i>	eec	304	1.840	5.2	0.16	2.1	0.01
<i>Syagrus romanzoffiana</i>	eea	269	1.587	7.0	0.17	1.6	0.015
<i>Syagrus romanzoffiana</i>	peic	19	–	–	–	35.0	0.25
<i>Symplocos mosenii</i>	eea	838	0.008	82.3	105.9	22.1	0.005
<i>Symplocos pubescens</i>	eea	16	0.009	19.2	1.82	1717.2	0.005
<i>Symplocos tenuifolia</i>	eea	71	0.003	38.8	24.71	1487.3	0.005
<i>Tapirira guianensis</i>	peic	520	0.099	25.3	5.13	5.1	0.005
<i>Tapirira guianensis</i>	eea	913	0.010	84.8	89.69	14.2	0.005
<i>Terminalia brasiliensis</i>	eea	191	0.038	16.8	4.87	95.2	0.005
<i>Ternstroemia brasiliensis</i>	peic	528	0.009	42.1	59.49	115.1	0.005
<i>Tetrastylidium grandifolium</i>	pecb	244	0.034	29.7	7.04	28.0	0.005
<i>Tetrorchidium rubrivenium</i>	pecb	39	–	–	–	7.8	0.51
<i>Trichilia catigua</i>	eec	592	1.160	5.2	0.50	2.9	0.005
<i>Trichilia claussenii</i>	eec	664	0.026	39.2	25.12	19.0	0.005
<i>Trichilia pallida</i>	eec	61	–	–	–	9.2	0.125
<i>Urera baccifera</i>	eec	30	0.010	42.2	2.84	115.3	0.005
<i>Virola bicuhyba</i>	pecb	173	0.042	32.1	4.06	13.7	0.005
<i>Vochysia tucanorum</i>	eea	1905	0.012	63.4	150.0	17.0	0.005
<i>Weinmannia paulliniifolia</i>	peic	24	0.004	28.2	6.11	2473.6	0.005
<i>Xylopia aromatica</i>	eea	1755	0.006	92.8	267.7	17.0	0.005
<i>Xylopia brasiliensis</i>	peic	25	–	–	–	30.7	0.15
<i>Xylopia langsdorffiana</i>	peic	955	0.123	29.6	7.59	1.7	0.005
<i>Zanthoxylum fagara</i>	eec	25	0.020	11.7	1.24	1522.2	0.005
<i>Zollernia ilicifolia</i>	pecb	63	–	–	–	1.9	0.905

Apêndice 3. Comparação da distribuição espacial de espécies ocorrentes em duas ou mais parcelas permanentes.

Algumas espécies apresentaram populações maiores que 20 indivíduos em mais de uma parcela ao mesmo tempo (Tabela A2), possibilitando a comparação dos valores observados de $g(r)$ entre parcelas. Como são populações que compartilham os mesmo aspectos de história de vida, é possível avaliar qual é o efeito do local no padrão de distribuição das espécies. Como o ajuste do modelo de Thomas foi realizado usando o método de mínimos contrastes do pacote *spatstat*, não foi possível obter intervalos de confiança ao redor das estimativas dos parâmetros. Assim, a comparação entre as distribuições espaciais das espécies foi feita visualmente, baseado nos valores observados de $g(r)$ em cada parcela.

Tabela S2. Relação das espécies cujas populações apresentaram mais de 20 indivíduos em duas ou mais parcelas permanentes.

Espécie	código	EEA	EEC	PECB	PEIC
<i>Amaioua intermedia</i>	amaiinte	380	0	6	876
<i>Astrocaryum aculeatissimum</i>	astracul	0	0	32	81
<i>Cabralea canjerana</i>	cabrcanj	5	26	52	14
<i>Casearia sylvestris</i>	casesylv	29	51	40	0
<i>Clethra scabra</i>	cletscab	31	0	0	28
<i>Coussapoa microcarpa</i>	cousmicr	0	0	46	37
<i>Croton floribundus</i>	crotflor	107	481	0	0
<i>Endlicheria paniculata</i>	endlpani	4	12	29	59
<i>Euterpe edulis</i>	euteedul	0	14	2747	3110
<i>Garcinia gardneriana</i>	garcgard	0	0	305	263
<i>Guatteria australis</i>	guataust	0	0	28	136
<i>Heisteria silvianii</i>	heissilv	0	0	66	118
<i>Inga marginata</i>	ingamarg	0	43	48	0
<i>Jacaranda puberula</i>	jacapube	0	0	40	63
<i>Maytenus robusta</i>	maytrobu	52	5	44	171
<i>Mollinedia schottiana</i>	mollscho	0	0	63	27
<i>Myrcia multiflora</i>	myrcmult	917	0	0	116
<i>Myrsine umbellata</i>	myrsumbe	1019	6	0	48
<i>Neomitranthes glomerata</i>	neomglom	0	66	118	8
<i>Ormosia arborea</i>	ormoarbo	37	1	6	96
<i>Pera glabrata</i>	peraglab	518	0	10	548
<i>Roupala montana</i>	roupmont	25	4	32	0
<i>Syagrus romanzoffiana</i>	syagroma	272	312	0	19
<i>Tapirira guianensis</i>	tapiguia	993	0	8	524

Fig. S3 Valores médios observados da função $g(r)$ em relação à distância para populações de espécies ocorrentes em duas ou mais parcelas permanentes simultaneamente. As cores das linhas indicam as diferentes parcelas (EEA= vermelho, EEC= laranja, PEIC= azul, PECB= verde), enquanto que a linha tracejada indica o valor de $g(r)$ esperado sob completa aleatoriedade espacial. Nomes completos das espécies podem ser obtidos na Tabela S2.

Fig. S3 (cont.) Valores médios observados da função $g(r)$ em relação à distância para populações de espécies ocorrentes em duas ou mais parcelas permanentes simultaneamente. As cores das linhas indicam as diferentes parcelas (EEA= vermelho, EEC= laranja, PEIC= azul, PECB= verde), enquanto que a linha tracejada indica o valor de $g(r)$ esperado sob completa aleatoriedade espacial. Nomes completos das espécies são fornecidos na Tabela S2.

CAPÍTULO 6: CONSIDERAÇÕES FINAIS

Considerações finais

6.1 Descrevendo padrões de comunidades arbóreas em parcelas permanentes

Nesta tese, foram estudadas as comunidades arbóreas de quatro parcelas permanentes localizadas no Brasil e de uma parcela permanente localizada no Panamá. As comunidades brasileiras representavam diferentes formações florestais e foram avaliadas quanto aos seus descritores básicos de estrutura (abundância, área basal e riqueza por sub-parcela), quanto a suas distribuições de abundância de espécies (SAD), distribuições de diâmetro (SDD) e distribuições espaciais (SSD). Já a parcela panamenha foi descrita quanto à relação entre a SDD e as taxas demográficas das espécies, avaliadas ao longo de um período de 30 anos de monitoramento.

As parcelas permanentes brasileiras apresentaram padrões bem distintos umas das outras. Ou seja, além da composição de espécies estas comunidades possuem estruturas bem distintas. A parcela de Cerradão apresentou padrões de estrutura mais contrastantes das demais, representados por maiores valores de densidade e menores valores de área basal, o que se refletiu em uma SDD caracterizada por uma proporção maior de indivíduos <20 cm de DAP. Esse contraste parece ter sido mais marcante para a densidade de indivíduos do que para a área basal. Quanto à diversidade de espécies, os padrões foram um pouco distintos, com as parcelas em climas sazonais (Cerradão e Floresta Estacional) apresentando menores valores de diversidade, o que se refletiu em SADs mais inclinadas – menor uniformidade de abundância entre as espécies (i.e. equitatividade). Provavelmente, estes contrastes entre parcelas foram determinados pelas condições abióticas as quais estas comunidades estão sujeitas. Parece que a sazonalidade da precipitação e as restrições edáficas atuam em conjunto na determinação da estrutura e diversidade destas florestas. Contudo, houve evidências de que a sazonalidade foi preponderante em relação ao solo, em especial para os valores de área basal e diversidade. Não houve, entretanto, um número suficiente de parcelas sob as diferentes combinações de condições ambientais para testar estas sugestões ou para fazer generalizações para cada formação florestal, sendo necessário o confronto destas sugestões com outros estudos realizados em condições similares.

Em relação à distribuição espacial, a frequência de espécies agregadas, assim como os padrões de intensidade e tamanho de agregação destas espécies, foi distinta entre as quatro

parcelas brasileiras. As parcelas de Cerradão e Restinga foram aquelas que apresentaram as maiores intensidades e distâncias de agregação enquanto que a Floresta Estacional e Ombrófila apresentaram os menores. Estes resultados sugerem que os padrões de distribuição espacial das espécies respondem à heterogeneidade ambiental interna das parcelas, mas que o grau de restrição climática ou edáfica aos quais estas comunidades estão sujeitas possui um papel importante no peso desta resposta. A grande variabilidade encontrada para os valores de abundância por sub-parcela (Capítulo 2) parece ter sido um reflexo desta heterogeneidade ambiental para o Cerradão, mas não para a Floresta Ombrófila. Por outro lado, a maior proporção de espécies com distribuição aleatória e os menores valores de intensidade e tamanho de agregação na Floresta Estacional parecem estar ligados aos seus menores valores de densidade e diversidade de espécies por sub-parcela. Isso reforça a idéia de que os padrões de distribuição espacial das espécies se refletem nos padrões de diversidade das comunidades.

Na parcela panamenha, foram confirmadas as previsões teóricas de que as distribuições diamétricas atuais possuem forte relação com as taxas demográficas das espécies, em especial os padrões de crescimento e recrutamento. De maneira geral, espécies com maior crescimento juvenil e menor recrutamento tenderam a ter SDD menos inclinadas e mais simétricas. Apesar disso, houve grande variação entre as espécies, sugerindo que muitas espécies não têm populações em equilíbrio nesta parcela. Houve evidências de que fortes eventos climáticos como *El Niño*, que influenciam a reprodução, crescimento e mortalidade das espécies, podem ter sido uma das causas para os desvios das populações de seus estados de equilíbrio demográfico. Na prática, esta grande variação dificulta fazer inferências precisas sobre quais os padrões demográficos passados das espécies baseado nas SDD atuais. Contudo, essas inferências podem servir como uma análise rápida para detectar padrões demográficos gerais em comunidades arbóreas florestais.

6.2 Ajuste e seleção de modelos por máxima verossimilhança

Como dito anteriormente, um dos objetivos desta tese foi descrever comunidades arbóreas usando o ajuste e seleção de modelos, aliado à máxima verossimilhança e teoria da informação. Essa abordagem foi usada de maneira integral nos Capítulos 2 e 4.1, onde diferentes modelos estatísticos foram ajustados aos dados por máxima verossimilhança e comparados com o AIC. No Capítulo 4.2, foi realizado apenas o ajuste de funções pré-selecionadas aos dados de diâmetro, crescimento e mortalidade (Weibull e função de potência). A definição da forma das curvas de crescimento e morte das espécies, contudo, foi

um exemplo claro do uso do ajuste e seleção de modelos por máxima verossimilhança neste capítulo. No Capítulo 3, a proposta inicial era ajustar a distribuição Poisson–Lognormal às SADs e verificar se o parâmetro de intensidade amostral da distribuição era comparável com a redução no tamanho amostral médio entre escalas. Contudo, como as árvores possuem distribuições agregadas, assumir que as árvores são incluídas na amostra de maneira independente seria incorreto. Como o uso de uma ‘Binomial Negativa–Lognormal’ não possui implementação simples, optou-se por uma abordagem distinta para esse capítulo, baseada em simulações. Por fim, no Capítulo 5 foi ajustado o modelo de Thomas aos dados, porém esse ajuste foi realizado usando outro método de estimação de parâmetro, por questões de praticidade analítica.

De maneira geral, o uso da abordagem de ajuste e seleção de modelos por máxima verossimilhança se mostrou bastante flexível e apropriado para responder as perguntas de cada capítulo. Visto que há grande possibilidade de que os padrões encontrados nas parcelas estudadas tenham mais de uma explicação, a possibilidade de comparar simultaneamente múltiplos modelos foi uma das vantagens principais do uso desta abordagem nesta tese. Mesmo que de maneira fenomenológica, a vantagem desta comparação simultânea de vários modelos candidatos ficou clara no Capítulo 4.1. A comparação entre diferentes funções para descrever as curvas de crescimento e mortalidade em relação ao diâmetro das árvores, usada no Capítulo 4.2, foi outro exemplo. Por outro lado, a seleção de modelos que assumem mecanismos coerentes aos sistemas estudados foi um exercício conceitual importante na busca por processos capazes de explicar os padrões encontrados. Por exemplo, a escolha da função Weibull para o ajuste das SDD na parcela panamenha (Capítulo 4.2), mesmo que esta função não seja aquela com melhor ajuste aos dados de distribuição diamétrica (ver Capítulo 4.1) foi uma decisão mais teórica, visto que a Weibull deriva de populações em equilíbrio demográfico com determinados padrões de crescimento e mortalidade. O mesmo pode ser dito para a escolha do modelo de Thomas para modelar a distribuição espacial das espécies (Capítulo 5). Por fim, como os dados das comunidades arbóreas estudadas são de natureza observacional, a seleção candidatos baseado no suporte que os próprios dados observados oferecem aos modelos foi bastante apropriada e mais intuitiva.

Houve, contudo, algumas limitações, em sua maioria, operacionais. No Capítulo 1, a construção de funções de verossimilhança capazes de lidar com dependência espacial entre as observações demandaria desenvolvimentos analíticos e computacionais por especialistas, principalmente porque boa parte das variáveis não apresentou distribuição normal. No Capítulo 4.2, a função Weibull apresentou um comportamento peculiar quanto à estimação de

seu parâmetro de escala, comportamento que não esteve relacionado a problemas numéricos, mas que inviabilizou a análise deste parâmetro para algumas espécies. Além disso, como em alguns capítulos foram ajustados vários modelos aos dados de várias espécies em diferentes parcelas permanentes (*e.g.* Capítulos 4.1 e 4.2), a execução de algumas análises requereu a construção de funções capazes de automatizar o ajuste dos modelos aos dados de cada espécie. Em muitos casos, essa construção levou bastante tempo e sua execução foi computacionalmente intensiva, principalmente para espécies ou parcelas com grande volume de dados. Como aconteceu em alguns capítulos, o uso consorciado da seleção de modelos, máxima verossimilhança, outros métodos de estimação ou testes de hipótese se mostraram viáveis e, por vezes, tornaram as análises mais simples de serem conduzidas. Assim, o uso da seleção de modelos por máxima verossimilhança, consorciado ou não, foi uma alternativa viável e apropriada para a descrição das comunidades arbóreas estudadas.

ANEXOS

Anexo 1. Propriedades gerais das principais funções de densidade probabilísticas usadas neste trabalho.

Existem inúmeras funções de densidade probabilísticas (PDF). Elas são modelos para descrever dados de diferentes tipos (*e.g.* contagens, medidas ou percentagens) e com diferentes graus de assimetria. Cada distribuição que gerou as PDF possui seus princípios e pressupostos sobre os processos geradores dos dados. Idealmente, esses princípios devem ser usados para selecionar e interpretar os resultados do processo de seleção de modelos na descrição de um conjunto de dados qualquer. Por exemplo, ao assumir que uma variável segue uma distribuição Normal, assume-se também que os dados são gerados pela soma de vários efeitos e que a variação em torno da média dos dados são erros aleatórios. Contudo, essas distribuições também são usadas de maneira fenomenológica, por se ajustarem bem às distribuições empíricas. Mesmo assim, a escolha de uma PDF apropriada depende de alguns critérios ou recomendações. O tratamento das distribuições apresentado a seguir é bastante preliminar. Mais detalhes podem ser encontrados na literatura especializada (*e.g.* Johnson *et al.* 1994, Lindsey 2004, Otto & Day 2011). Aqui, procurou-se fazer um apanhado das informações e aplicações consideradas mais relevantes para os estudos em ecologia no contexto de seleção de modelos. Além disso, as descrições estiveram restritas às distribuições mais usadas durante esta tese. Detalhes sobre a Johnson's Special Bounded (JSB), Generalized Beta (Beta), Birnbaum-Saunders (BS), Logit-Logistic (LL), New Modified Weibull (MW) e Odd Weibull (OW), distribuições usadas em apenas um dos capítulos desta tese, podem ser obtidos nos trabalhos originais, cujas referências são fornecidas no Capítulo 4.1 desta tese.

Existe uma taxonomia das distribuições de probabilidade, que agrupa em famílias as distribuições que têm origens ou propriedades comuns, como é o caso das distribuições Normal, Exponencial e Gama, que pertencem à família Exponencial de distribuições. Existem ainda distribuições que são casos especiais de outras. Por exemplo, quando os parâmetros de forma das distribuições Weibull e Gama são iguais a um, elas se reduzem a uma Exponencial, assim como a Binomial Negativa que se reduz a uma Geométrica quando o parâmetro de forma é igual a um. Outras distribuições são casos limite de outras, ou seja, uma distribuição tende a outra conforme o valor do parâmetro se aproxima de um certo valor. É o caso da

Binomial negativa e da Gama/Weibull que tendem a uma Poisson e uma Normal, respectivamente, para grandes valores de parâmetro de forma. Esses e outros exemplos estão ilustrados de maneira bastante didática em Bolker (2008, p.137). Aqui a apresentação das PDFs é feita usando uma separação simples entre distribuições discretas e contínuas.

Poisson

Distribuição discreta que descreve contagens de eventos independentes que ocorrem aleatoriamente no tempo ou espaço, a uma taxa constante. Quanto à forma, a Poisson torna-se mais regular e simétrica à medida que a esperança da média aumenta. Por isso, quando a média é inferior a um, a moda da distribuição é zero, e quando a média atinge valores maiores que 10, a Poisson aproxima uma distribuição Normal (Bolker 2008). A Poisson possui apenas um parâmetro λ , que descreve a taxa ou intensidade do processo, que corresponde à média e também à variância da distribuição. Portanto, a distribuição Poisson é apropriada para dados de contagem cuja média é próxima de sua variância (Lindsey 2004). Esta distribuição é bastante usada como um padrão de aleatoriedade de eventos no tempo ou no espaço. Deve ser usada apenas para dados de contagem, por exemplo, número de indivíduos, sementes ou plântulas por unidade amostral (Bolker 2008).

Intervalo: $0 - \infty$

Assimetria: à direita

Parâmetros: valor esperado da contagem (λ ; parâmetro de locação e forma).

Domínio: $\lambda \geq 0$

PDF: $f(x; \lambda) = e^{-\lambda} \lambda^x / x!$

$E[x] = \lambda$

$\text{var}[x] = \lambda$

Binomial negativa

Esta distribuição foi inicialmente formulada para descrever o número de tentativas (tempo decorrido) até que um dado número de sucessos ocorra em linhas de montagem de automóveis. Ao contrário da Poisson, esta distribuição pode ter variância maior que a média. Por tal propriedade, esta distribuição têm sido frequentemente utilizada fenomenologicamente para substituir a Poisson no caso dados de contagem com agregação (Bolker 2008). Em outras palavras, este modelo se aplica quando os eventos não são independentes nas unidades de

tempo ou espaço da amostra (Lindsey 2004). Para tanto, usa-se uma parametrização diferente da inicialmente descrita (ver Otto & Day 2011 e Bolker 2008 para parametrizações alternativas), com parâmetros μ (média esperada por unidade amostral) e κ (parâmetro de dispersão). Nesse caso, κ é um parâmetro de forma que avalia o quão agregado estão os valores observados dentro das unidades amostrais. Quanto menor o valor de κ , mais o número total de indivíduos está concentrado em poucas amostras. A distribuição geométrica, criada para o número de tentativas até um evento ocorra (Bolker 2008), é um caso particular da binomial negativa no qual o parâmetro $\kappa = 1$. A binomial negativa também é usada para contagens por unidade amostral, mas quando se espera variância maior que a média.

Intervalo: $0 - \infty$

Assimetria: à direita

Parâmetros: valor esperado da contagem (μ ; parâmetro de locação)
dispersão (κ ; parâmetro de forma)

Domínio: $\mu, \kappa > 0$

PDF: $f(x; \mu, \kappa) = \frac{\Gamma(\kappa + x)}{\Gamma(\kappa)x!} \left(\frac{\kappa}{\kappa + \mu}\right)^\kappa \left(\frac{\mu}{\kappa + \mu}\right)^x$, onde Γ é a função gamma.

$E[x] = \mu$

$\text{var}[x] = \mu + \mu^2/\kappa$

Normal (ou Gaussiana)

Variáveis com distribuição normal são comuns em medições e resultam da soma de muitos efeitos (Lindsey 2004). Na prática, diz-se que quando somamos muitas observações independentes e igualmente distribuídas, a distribuição desta soma será aproximadamente normal, que é uma aplicação do Teorema do Limite Central (Johnson *et al.* 1994). Por isso, é comum que outras distribuições probabilísticas (*e.g.* Gama, Poisson) tendam a uma Normal para amostras muito grandes. Sem dúvida, esta é a distribuição probabilística mais utilizada na estatística tradicional, como nos testes de comparação de média e na análise de variância. Em termos matemáticos, a normal descreve valores cuja probabilidade decresce exponencialmente a partir da média, conferindo um formato completamente simétrico. Por tal motivo, a média, moda e mediana são coincidentes na Normal. Outra propriedade matemática bastante conveniente da Normal é que seus parâmetros μ e σ^2 equivalem diretamente a média e variância da distribuição. Assim, μ é o parâmetro de locação que desloca a distribuição ao

longo do eixo das abscissas, enquanto σ^2 molda sua forma; quanto maior σ^2 mais ampla é a distribuição. A variável Normal pode assumir qualquer valor real, sendo útil para lidar com valores negativos. Várias distribuições, inclusive as discretas, tendem à Normal sob certas circunstâncias. É comumente utilizada para descrever dados contínuos onde se assume várias fontes aditivas de erros de mensuração ou variação simétrica em torno de um valor central.

Intervalo: $-\infty - \infty$

Assimetria: nenhuma

Parâmetros: média (μ ; parâmetro de localização)
desvio padrão (σ ; parâmetro de forma)

Domínio: μ : real;
 σ : real e positivo

$$\text{PDF: } f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

$$E[x] = \mu$$

$$\text{var}[x] = \sigma^2$$

Exponencial

Descreve o tempo decorrido até que um dado evento ocorra sob a suposição de que os eventos ocorram a uma taxa constante (independente do tempo). Portanto, também se enquadra no grupo de distribuições de tempo de vida, sendo a equivalente contínua da distribuição Geométrica. Equivalente contínua da distribuição geométrica, a Exponencial possui apenas um parâmetro, λ , que controla a inclinação da distribuição que é decrescente. Este parâmetro é a taxa de ocorrência. É apresentada separadamente aqui por ser frequentemente usada em trabalhos de ecologia na descrição de sobrevivência de indivíduos, distribuições diamétricas de árvores e distâncias de dispersão de sementes a partir da árvore matriz.

Intervalo: $0 - \infty$

Assimetria: à direita

Parâmetros: $\lambda > 0$

Domínio: real e positivo

$$\text{PDF: } f(x; \lambda) = \lambda \exp(-\lambda x)$$

$$E[x] = 1/\lambda \text{ (tempo médio entre eventos)}$$

$$\text{var}[x] = 1/\lambda^2$$

Lognormal

Outra distribuição bastante comum na natureza, a Lognormal é usada para descrever valores que resultam de efeitos multiplicativos, e não aditivos como na Normal. Na prática, obtém-se uma distribuição Lognormal quando temos variáveis que são o produto de várias observações independentes (Bolker 2008). Em termos matemáticos, esta distribuição descreve variáveis cujo logaritmo é normalmente distribuído. Mas, ao contrário da distribuição Normal, sua média e variância não são independentes entre si. Trata-se de uma distribuição cuja assimetria tende a diminuir com o aumento da variância. É frequentemente usada para descrever valores contínuos de tamanhos ou massa. Existe uma polêmica sobre o uso da Log-normal para a descrição de distribuições de abundância de espécies (Williamson & Gaston 2005), contudo seu uso em ecologia tem sido comum para este fim.

Intervalo: $0 - \infty$

Assimetria: à direita

Parâmetros: log-média (μ ; parâmetro de locação)
log-desvio padrão (σ ; parâmetro de forma)

Domínio: μ : real; $\sigma > 0$

$$\text{PDF: } f(x; \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \frac{-(\log x - \mu)^2}{2\sigma^2}$$

$$E[x] = \exp (\mu + \sigma^2/2)$$

$$\text{var}[x] = \exp (2\mu + \sigma^2) (\exp(\sigma^2) - 1)$$

Weibull

Parte de um grupo de distribuições chamada de distribuições de tempo de vida (Lindsey 2004), a Weibull é comumente usada em avaliações de confiabilidade de processos industriais. Mais especificamente, ela descreve o intervalo de tempo contínuo até que um evento ocorra, no caso, uma falha. Se fizermos um paralelo com o crescimento de organismos, essa falha pode ser equivalente à morte de um organismo, permitindo usar a Weibull para análises de sobrevivência. A Weibull é uma distribuição bem flexível e possui dois parâmetros: um parâmetro de forma β , e um parâmetro de escala α . Quando $\beta = 1$, a

Weibull se reduz a outra distribuição freqüentemente usada em ecologia, a Exponencial, descrita acima. Nesse caso, a taxa de decaimento da Exponencial (λ) é inversamente proporcional ao parâmetro α (*i.e.* $\lambda = 1/\alpha$), evidenciando a Exponencial como um caso particular da Weibull.

O parâmetro de forma β é proporcional a taxa de falha ou mortalidade, que ao contrário da Poisson ou da Binomial, pode mudar ao longo do tempo. Essa propriedade do parâmetro β é bastante interessante, pois permite a uma interpretação com significado biológico bastante claro e direto: se $\beta < 1$, a mortalidade decresce com o tempo (maior mortalidade entre indivíduos jovens); se $\beta = 1$, a mortalidade é constante no tempo (causas de morte externas e aleatórias); e se $\beta > 1$, a mortalidade aumenta com o tempo (maior mortalidade entre indivíduos mais velhos). Partindo do pressuposto que há uma relação linear entre idade e tamanho, a Weibull passou a ser utilizada também para modelar distribuições de tamanho de indivíduos arbóreos. Contudo, essa relação geralmente não é linear, fazendo com que o uso da Weibull para esse fim seja meramente fenomenológico. Apesar disso, essa distribuição foi derivada a partir relações entre crescimento e mortalidade com o tamanho de indivíduos arbóreos, fazendo com que a Weibull tenha uma justificativa teórica baseada na dinâmica de populações sob equilíbrio demográfico (Muller-Landau *et al.* 2006).

Intervalo: $0 - \infty$

Assimetria: à direita ou esquerda

Parâmetros: β (parâmetro de forma)

α (parâmetro de escala)

γ (parâmetro de locação ou de posição)

Domínio $\beta, \alpha > 0$

$$\text{PDF: } f(x; \beta, \alpha) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp - \left(\frac{x}{\alpha}\right)^{\beta}$$

$$E[x] = \alpha \Gamma(1 + 1/\beta)$$

$$\text{var}[x] = \alpha^2 (\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta))$$

Gama

A Gama também faz parte das distribuições de tempo de vida e descreve o tempo decorrido até que um número a de eventos ocorram, dada uma taxa instantânea de ocorrência r . Assim, esta distribuição pode ser vista como a equivalente em tempo contínuo da Binomial Negativa,

tendo sido comumente usada para descrever valores positivos com assimetria à direita. Trata-se de uma distribuição bastante flexível e, por isso, bastante utilizada de forma fenomenológica. Dentro da formulação de sua PDF, o parâmetro a define sua forma, enquanto r se relaciona à escala. Para valores de $a \leq 1$, a moda da Gama é zero, e ela assume a forma de “J-invertido”. A medida que esse parâmetro cresce a distribuição apresenta moda em valores diferentes de zero e ela torna-se mais simétrica (Bolker 2008). Como em outras distribuições, os parâmetros a e r estão inter-relacionados na definição da média e variância da distribuição. A Gama possui duas parametrizações, sendo apresentada aqui a parametrização na qual a taxa de ocorrência $r = 1/s$. Assim como a Weibull, a Gama se reduz a uma Exponencial quando $a=1$. Mas, diferentemente da Weibull, cuja taxa de ocorrência do evento é decrescente ($\beta < 1$), constante ou crescente ($\beta > 1$), a taxa de ocorrência da Gama cresce ou decresce até um limite de a/s , uma diferença que pode ser uma justificativa teórica para usar uma ou outra distribuição (Lindsey 2004). É comumente usada para descrever valores contínuos com assimetria à direita, como tempos, distâncias ou taxas de crescimento.

Intervalo: $0 - \infty$

Assimetria: à direita

Parâmetros: a (parâmetro de forma): ocorrências por evento

s (parâmetro de escala): inverso da taxa de ocorrência ($s = 1/r$)

Domínio $a, s > 0$

$$\text{PDF: } f(x; a, s) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-x/s}$$

$$E[x] = as$$

$$\text{var}[x] = as^2$$

Referências bibliográficas

- Bolker, B.M. 2008. Ecological models and data in R. Princeton University Press, Princeton.
- Johnson, N.L., Kotz, S. & Balakrishnan, N. 1994. Continuous univariate distributions. 2ª edição. John Wiley and Sons, Nova York.
- Lindsey, J.K. 2004. Introduction to applied statistics: A modelling approach. Oxford University Press, Oxford.

- Muller-Landau, H.C., Condit, R., Harms, K.E., Marks, C.O., Thomas, S.C., Bunyavejchewin, S., Chuyong, G., Co, L., Davies, S., Foster, R.B., Gunatilleke, S., Gunatilleke, N., Hart, T.B., Hubbell, S.P., Itoh, A., Kassim, A.R., Kenfack, D., LaFrankie, J.V., Lagunzard, D., Lee, H.S., Losos, E., Makana, J., Ohkubo, T., Samper, C., Sakumar, R., Sun, I.-F., Supardi, N., Tan, S., Thomas, D., Thompson, J., Valencia, R., Vallejo, M.I., Muñoz, G.V., Yamakura, T., Zimmerman, J.K., Dattaraja, H.S., Esufali, S., Hall, P., He, F., Hernández, C., Kiratiprayoon, S., Suresh, H.S., Wills, C. & Ashton, P. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology Letters* 9, 589-602.
- Otto, S.P. & Day, T. 2011. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press, Princeton.
- Williamson, M. & Gaston, K.J. 2005. The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *Journal of Animal Ecology* 74, 409-422.