# BROAD-SCALE PATTERNS OF THE DISTRIBUTION OF BODY SIZES OF INDIVIDUALS IN ECOLOGICAL COMMUNITIES

INTRODUCTION

Body size is one of the most fundamental properties of an organism, affecting metabolic rate, lifespan, home range size, reproductive rate, and many other ecological attributes (e.g., Peters 1983, Calder 1984, Schmidt-Nielsen 1984). Ecologists have long been interested in using these regular patterns to gain insight into ecological systems (Peters 1983), and it is increasingly recognized that body size has important consequences for understanding many aspects of community structure and function (e.g., Siemann et al. 1996, Cohen et al. 2003, Brown et al. 2004). However, information on continental and global scale variation in body size at the level of entire communities is extremely limited and, given size's key role in functional biology, quantifying and understanding broad scale patterns in body size represents a major challenge for ecology (Kaspari 2005).

The few broad scale analyses of community size patterns that do exist typically focus on very simple characterizations of the distribution of body sizes within a community, most typically average body mass (Li 2002, Kaspari 2005). However, much richer information is available by looking at the entire distribution of body sizes of individuals within the community, otherwise known as the size-spectrum (Figure 1, Sheldon et al. 1972, Griffiths 1986, Kerr and Dickie 2001). Both the general and specific form of this size-spectrum can provide substantial information about the processes structuring communities by reflecting the influences of competition, predation and resource use (Griffiths 1986, Enquist and Niklas 2001, Kerr and Dickie 2001). In addition, understanding broad-scale patterns in the spectrum is essential for linking patterns relating individual properties and body size, to patterns in assemblages, communities, and ecosystems. This information is necessary because in order to link community-level behavior with that of the component individuals, relationships between properties of the individual and body size must be summed across all of the individuals in a community, which requires information on the distribution of individual body sizes (e.g., Peters 1983, Enquist et al. 2003, Savage 2004).

In this project, I will take advantage of three continental-scale databases, to evaluate broad scale patterns in the distribution of body sizes of individuals. I will focus on terrestrial size-spectra, which have received very little attention, even at local scales (Kerr and Dickie 2001). I will evaluate patterns in the general and specific shape of these distributions as well as simpler, more traditional measures of the spectrum such as average mass. If we are to successfully link individuals and ecosystems, and understand the processes structuring local communities, it will be necessary to understand the patterns of, and forces structuring, the distribution of body sizes of individuals within and among assemblages, communities and ecosystems. Evaluation and quantification of broad scale patterns represents a first important step in this direction.
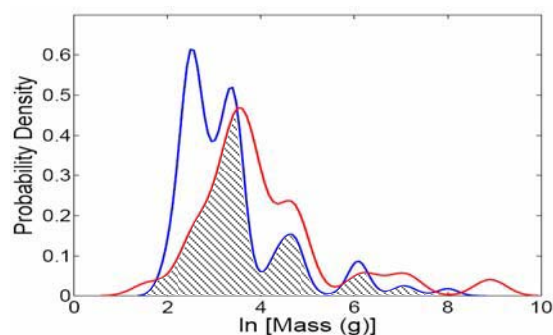
Figure 1. Example of size-spectra from two BBS routes in NM (one route in red – mean mass 387 g, the other in blue – mean mass 100 g). Spectra were generated using kernel density estimation. Overlap between the two spectra (discussed in text) is indicated by grey hatching. The overlap value for these spectra is ~ 50%.

Objectives: The primary objective of this study will be to detect and quantify broad-scale patterns in the distribution of body sizes of individuals, and to determine if these patterns can be explained by a combination of environmental and ecological factors.

Question (1) Is there a general form of terrestrial size-spectra (e.g., unimodal, multimodal, monotonic)? Does this form vary geographically or taxonomically, or is it approximately invariant? What are the best ways to quantify and compare these potentially complex distributions?

Question (2) What are the geographic patterns of the size distribution and can these patterns be related to ecological and environmental factors? Specifically, can changes in shape (e.g. number of modes) and magnitude (e.g. average mass) of the distribution be described using multivariate models that include environmental (e.g. temperature, precipitation, land use) and ecological (e.g. species richness, abundance) factors?

Question (3) Are there clues in the local and geographic patterns that point towards a general understanding of the processes underlying terrestrial size-spectra? How strong are the relative influences of local and regional processes on structuring the size-spectra? Are higher abundances in some size classes generated by greater numbers of species in those classes or greater numbers of individuals in the species that are present? How are differences in size-spectra related to differences in species composition?

Data: Broad-scale data that include information on the abundance and size of organisms are extremely rare. My research will focus on the three terrestrial datasets that can provide this type of information: 1) Breeding Bird Survey of North America (BBS) – The BBS contains estimates of breeding season abundance for over 500 species of birds at over 2500 locations throughout the United States and Canada. This data is obtained using consistent sampling protocols and sampling effort and as such is readily comparable at geographic scales. 2) Christmas Bird Count (CBC) – The CBC contains estimates of winter abundance for over 1500 locations in the United States. BBS and CBC data will be combined with data on the body size of bird species from the CRC Handbook of Avian Body Masses (Dunning 1992) using a novel technique for approximating the distribution of body sizes of individuals when only species level data on body mass is available (see *Approach*). 3) Synthesis and Analysis of Local Vegetation Inventories Across Scales (SALVIAS) – SALVIAS contains a global compendium of plant plot data with size measurements of every individual. It includes newly cleaned and standardized versions of the famous A.H. Gentry plots and an additional 2500 plant community inventories, many of which include individual size measurements. Finally, I will assemble a database from the literature for taxonomic groups not represented in these broad-scale surveys. Collectively these community databases contain well over 5 million records, which I will manage using relational databases.

This assemblage data will be combined with environmental data to establish correlates of the size-spectrum. This environmental data will include temperature and precipitation (e.g., USHCN; lwf.ncdc.noaa.gov/), production (GLO-PEM; www.geog.umd.edu/glopem), NDVI (AVHRR; xtreme.gsfc.nasa.gov/), land cover (e.g., MLCC; www.epa.gov/mrlc/), biome type (CEC/EPA; www.epa.gov/wed/pages/ecoregions/), elevation (e.g., NALCCDB; edcdaac.usgs.gov/glcc/), and measures of elevational and land cover heterogeneity (calculated using spatial windows and the basic CEC/EPA and NALCCDB data). This will be done using ARCs Geographic Information Systems combined with Matlab's Spatail Statistics Toolbox (www.spatial-statistics.com) allowing for the efficient analysis of these large datasets.

Approach: One reason why broad-scale research on terrestrial size-spectra has not been conducted is that the number of terrestrial communities for which we have information on the body mass of every individual at the site is very small. I will deal with this problem by taking advantage of the fact that body size within a species should be approximately normally distributed as a result of stabilizing selection and that the relationship between the mean and variance of these distributions appears to behave like that of a random variable (preliminary data on 62 bird species: $Var \sim Mean^{1.99}$; $r^2 = 0.93$). This relationship allows for the approximation of the community-level distribution of body sizes of individuals using data on species' abundances and the average mass of a species. I will validate this approach using communities for which both species abundances and individual masses are available (preliminary results suggest >99% correspondence between real and approximated distributions). This approach facilitates a substantial increase in the amount of ecological data that can be brought to bear on these questions.

The next step in quantifying a size-spectrum is to choose a method for approximating the underlying probability density function (pdf). Ecologists typically do this by simply binning individuals into different size classes (i.e., by creating a histogram). However, this approach can suffer from several weaknesses (Malamud et al. 1998), and a number of more advanced techniques are available in the statistics literature including kernel density estimation, nearest neighbor density estimation, and taking numerical derivatives of the cumulative density function (e.g., Silverman 1986, Sugihara and May 1990, Malamud et al. 1998). In collaboration with Dr. Green I will compare the performance of these density estimation techniques (and the more common linear and logarithmic binning approaches) for characterizing size-spectra using Monte Carlo simulations from known distributions and by comparing the results for observed size-spectra from natural communities.

After validating the techniques for estimating community patterns from species abundances and for characterizing body size distribution pdfs, I will then be able to address **Question (1)**. Three major classes of patterns have been observed for size-spectra: 1) monotonically decreasing, which has typically been described by power laws, where $frequency \sim mass^{-\lambda}$ (Enquist and Niklas 2001, Kerr and Dickie 2001, Li 2002); 2) unimodal, approximately described by log-normal distributions (Siemann et al. 1996, 1999); and 3) multimodal, where the distribution is characterized by multiple peaks (Griffiths 1986). In order to answer **Question (1)** I will start by evaluating the presence of multi-modality using established statistical techniques (Silverman 1981, 1986, Manly 1996) that have not yet been applied to size-spectra. Monotonic distributions will be distinguished from those exhibiting an interior maximum by competitive fitting of statistical models with those characteristics. Geographic and taxonomic patterns in the form of the distribution will then be evaluated. Attempts will also be made to quantify these distributions using simple statistical models and comparisons for model parameters and features such as the position of modes conducted.

Question (2) will then be addressed by constructing statistical models of environmental and ecological factors to explain: 1) the general form of the distribution; 2) similarities between different distributions; and 3) certain key measures of the distribution (e.g., average mass). Environmental and ecological correlates of the general form of the distribution will be determined using discriminant analysis and regression trees. Overall similarities between distributions will be determined by calculating the percentage overlap of those distributions (Figure 1) and this metric will be compared to similarity in environmental and ecological factors using partial-mantel tests (this eliminates the need for simple consistent quantification of potentially complex distributions). Finally, I will look for patterns in the average mass and average metabolic rate of an individual as well as the percentage of individuals occurring in particular body size quantiles using multiple regression and general linear models. Where

possible the performance of the models will be validated using bootstrap resampling, and the influence of spatial autocorrelation will be quantified and controlled for.

Question (3): Having quantified geographic patterns and underlying environmental relationships, it is desirable to gain additional insight into the processes that may be structuring size-spectra within communities. One approach is to evaluate the contributions of local and regional processes in generating the observed local distributions. The relationship between a local size-spectrum and the size-spectrum of the regional pool will be evaluated using Monte Carlo simulated size-distributions of the appropriate local abundance from the regional size-spectrum (defined as the sum of all spectra within a series of circular plots of different scales surrounding the focal community). Because of the large number of communities, their relatively high abundances, and need for thousands of randomizations to estimate confidence limits, this will require substantial computational power and an emphasis on efficient computational techniques. If the geographic patterns of the size-spectrum do not result entirely from random processes linking the regional pool to the local communities then more local scale processes will be necessary to understand the observed patterns. One way to look for clues to the local processes is to recognize that high abundances in a given area of the body size axis can result from two distinct but non-exclusive processes: 1) high numbers of species in that area of the axis; and 2) large numbers of individuals of the species present in that area of the axis (Griffiths 1986). I will evaluate the contributions of these two processes by: 1) looking for clumped distributions of species along the body size axis using available statistical techniques (Silverman 1986, Efron and Tibshirani 1993, Manly 1996) and by evaluating regular trends in species richness as a function of body size (Siemann et al. 1996, 1999); and 2) comparing rank abundance distributions in high and low abundance regions of the body size axis. Finally, I will evaluate the relationship between community composition and size-spectra using distance based methods to quantify differences in composition (e.g., Legendre and Legendre 1998) and the overlap measure mentioned above to quantify differences in the spectra (see also Ernest 2005). The observed relationship will be compared to that expected from composition alone by randomizing species identities.

Significance

Understanding the processes involved in generating observed distributions of body sizes of individuals is a major challenge for ecology. It is essential for scaling from individuals to ecosystems and has the potential to provide novel insights into the structure and assembly of communities. A major impediment to this understanding is that there is very little knowledge of geographic patterns for these distributions, and practically no understanding of how they are related to different features of the ecology and environment. Once these patterns are quantified, we can seek to discover the processes that generate and maintain them.

Specifically, these patterns can tell us whether individuals are distributed patchily along the body size axis as predicted by some community assembly models (Holling 1992), or evenly distributed as predicted by others (Bowers and Brown 1982). Other assembly models predict monotonic decreases in abundance based on geometry or metabolism (Morse et al. 1985, Enquist and Niklas 2001, Brown et al. 2004). While some of these models have been evaluated at the species-level they are clearly derived by considering processes at the level of the individual, and broad-scale analysis of size-spectra will tell us much about which of these models, and the processes they represent, are operating in ecological systems. In addition, by combining these newly quantified patterns with patterns of and models for abundance, this work could result in a validated multivariate model that predicts abundance, biomass, production, and turnover rates of communities using only environmental data. A project of this magnitude, requiring the combination of enormous ecological and environmental datasets, and complex statistical and computational analyses, would not be possible without bioinformatics.

TRAINING OBJECTIVES AND CAREER DEVELOPMENT

In the course of this research I will master a new set of skills, and develop a new set of interests, that will complement my current strengths and areas of expertise. To date, the majority of my research has focused on temporal patterns of species richness and composition. As such, I have very limited experience with studying spatial patterns and exploring patterns related to body size.

Given my interest in pursuing the fields of macroecology and bioinformatics, it is important that I develop as large a research "tool-kit" as possible. Because this project integrates spatial analyses, randomization/simulation and probability density estimation, it will allow me to develop substantial statistical and analytical tools that will serve me well in my career. In particular I will gain experience in sophisticated GIS and statistical modeling. As more and more broad-scale geographically referenced data becomes available competency in using GIS will become increasingly necessary for cutting edge macroecological research. I will also develop a detailed understanding of probability density estimation – a field of statistics that is traditionally underutilized by ecologists – and in computational simulation and randomization techniques that are broadly applicable to a variety of ecological questions.

I am committed to pursuing a career in ecological research, ideally as a tenure-track faculty member at a leading research university. By combining the quantitative skills that I will develop during this postdoctoral fellowship with my more traditional educational background (B.A. and Ph.D. in Biology), this research experience will broaden my training and my abilities to conduct a wide-range of ecological research in the future. In the academic setting I hope to represent a link between theoretical ecologists (traditionally trained as mathematicians) and empirically oriented field ecologists.

JUSTIFICATION OF SPONSORING SCIENTIST AND HOST INSTITUTION

My proposed sponsor, Dr. Brian Enquist, has made substantial, high profile contributions to macroecology, my general area of interest. He is also one of the few people who has actively studied the distribution of body sizes of individuals in terrestrial communities. He authored a pioneering paper looking at patterns in the size-spectra of trees (Enquist and Niklas 2001) and this paper inspired my current interest in taking a detailed look at these patterns in terrestrial systems. Dr. Enquist would be an excellent mentor because of his experience in addressing interesting ecological questions by quantifying patterns in broad-scale data. In addition, he is excited about this potential collaboration and the combination of my background working with animal communities and his experience with plant communities will undoubtedly lead to an inspired and synergistic collaboration. There are also several postdocs and graduate students in his lab with whom I would collaborate and interact. More broadly, the University of Arizona as a whole represents an excellent location for me to broaden my perspectives and my training. The university also offers a Quantitative Biology Initiative which is designed to bring together faculty and students from different areas of the university including mathematics, statistics, and biology, to work collaboratively on studying big picture questions in quantitative biology.

Dr. Jessica Green is an early-career scientist who has already made significant contributions to the study of spatial patterns in ecological systems. Her training as a physicist and her interest in ecological questions have provided Dr. Green with extensive experience in applying tools from other fields to broad-scale ecological questions (Green et al. 2003, Green and Ostling 2003, Green et al. 2004). Her quantitative background will prove invaluable in navigating the statistical and computational complexities of this project. In addition to Dr. Green and her group, the University of California Merced houses the Center for Computational Biology. This research center will provide interaction with computer scientists and access to high performance computing.

BIBLIOGRAPHY

Bowers, M. A., and J. H. Brown. 1982. Body size and coesistence in desert rodents: chance or community structure. Ecology **63**:391-400.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology **85**:1771-1789.

Calder, W. A. 1984. Size, function, and life history. Dover, Mineola, N.Y.

Cohen, J. E., T. Jonsson, and S. R. Carpenter. 2003. Ecological community description using the food web, species abundance, and body size. Proceedings of the National Academy of Sciences of the United States of America **100**:1781-1786.

Dunning, J. B. J. 1992. CRC handbook of avian body masses. CRC Press, Boca Raton, FL.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Enquist, B., E. Economo, T. Huxman, A. Allen, D. Ignace, and J. Gillooly. 2003. Scaling metabolism from organisms to ecosystems. Nature **423**:639-642.

Enquist, B. J., and K. J. Niklas. 2001. Invariant Scaling Relations Across Tree-Dominated Communities. Nature **410**:655-660.

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

Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie. 2004. Spatial scaling of microbial eukaryote diversity. Nature **432**:747-750.

Green, J. L., and A. Ostling. 2003. Endemics-area relationships: The influence of species dominance and spatial aggregation. Ecology **84**:3090-3097.

Griffiths, D. 1986. Size-abundance relations in communities. American Naturalist **127**:140-166.

Holling, C. S. 1992. Cross-scale morphology, geometry, and dynamics of ecosystems. Ecological Monographs **62**:447-502.

Kaspari, M. 2005. Global energy gradients and size incolonial organisms: Worker mass and worker number in ant colonies. Proceedings of the National Academy of Sciences of the United States of America **102**:5079-5083.

Kerr, S. R., and L. M. Dickie. 2001. Biomass Spectrum. Columbia University Press, New York.

Legendre, P., and L. Legendre. 1998. Numerical Ecology, Second English edition. Elsevier Science B.V., Amsterdam.

Li, W. K. W. 2002. Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. Nature **419**:154-157.

Malamud, B., G. Morein, and D. Turcotte. 1998. Forest fires: An example of self-organized critical behavior. Science **281**:1840-1842.

Manly, B. F. J. 1996. Are there clumps in body-size distributions? Ecology **77**:81-86.

Morse, D. R., J. H. Lawton, M. M. Dodson, and M. H. Williamson. 1985. Fractal dimension of begetation and the distribution of arthropod body lengths. Nature **314**:731-733.

Peters, R. H. 1983. The ecological implications of body size. Cambridge University Press, New York, NY.

Savage, V. M. 2004. Improved approximations to scaling relationships for species, populations, and ecosystems across latitudinal and elevational gradients. Journal of Theoretical Biology **227**:525-534.

Schmidt-Nielsen, K. 1984. Scaling: Why is animal size so important. Cambridge University Press, Cambridge.

Sheldon, R. W., A. Prakash, and W. H. J. Sutcliffe. 1972. The size distribution of particles in the ocean. Limnology and Oceanography **17**:327-340.

Siemann, E., D. Tilman, and J. Haarstad. 1996. Insect species diversity, abundance and body size relationships. Nature **380**:704-706.

Siemann, E., D. Tilman, and J. Haarstad. 1999. Abundance, diversity and body size: patterns from a grassland arthropod community. Journal of Animal Ecology **68**:824-835.

Silverman, B. W. 1981. Using kernel density estimates to investigate multimodality. Journal of the Royal Statistical Society B **43**:97-99.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman & Hall/CRC, New York.

Sugihara, G., and R. M. May. 1990. Applications of Fractals in Ecology. Trends in Ecology & Evolution **5**:79-86.