Universidade Federal de Goiás
Instituto de Ciências Biológicas
Pós-Graduação em Ecologia e Evolução

PADRÕES ESPACIAIS DA RIQUEZA DE ESPÉCIES NO CERRADO ATRAVÉS DA SELEÇÃO DE MODELOS UTILIZANDO O CRITÉRIO DE AKAIKE (AIC)

Marcel Müller Fernandes Pereira da Silva

Dissertação apresentada ao Programa de Pós-graduação em Ecologia e Evolução, do Instituto de Ciências Biológicas da Universidade Federal de Goiás, como parte dos requisitos para a obtenção do título de Mestre em Ecologia.

Goiânia,
Março de 2008

Universidade Federal de Goiás
Instituto de Ciências Biológicas
Pós-Graduação em Ecologia e Evolução

PADRÕES ESPACIAIS DA RIQUEZA DE ESPÉCIES NO CERRADO ATRAVÉS DA SELEÇÃO DE MODELOS UTILIZANDO O CRITÉRIO DE AKAIKE (AIC)

Marcel Müller Fernandes Pereira da Silva

Dissertação apresentada ao Programa de Pós-graduação em Ecologia e Evolução, do Instituto de Ciências Biológicas da Universidade Federal de Goiás, como parte dos requisitos para a obtenção do título de Mestre em Ecologia.

Orientador:
Prof Dr. José Alexandre Felizola Diniz-Filho

Goiânia,
Março de 2008

Dados Internacionais de Catalogação-na-Publicação (CIP)

(GPT/BC/UFG)

Silva, Marcel Müller Fernandes Pereira da.
S586p Padrões espaciais da riqueza de espécies no cerrado através da
seleção de modelos utilizando o critério de Akaike (AIC)
[manuscrito] / Marcel Müller Fernandes Pereira da Silva. – 2008.
110 f. : il., tabs.

Orientador: Prof. Dr José Alexandre Felizola Diniz-Filho.

Dissertação (Mestrado) – Universidade Federal de Goiás. Instituto de Ciências Biológicas, 2008.

Bibliografia.

I. Ecologia animal – Cerrados. 2. Espécies endêmicas – Cerrados 3. Riqueza de espécies – Cerrados 4. Espécies – Padrões espaciais I. Diniz-Filho, José Alexandre Felizola II. Universidade Federal de Goiás. **Instituto de Ciências Biológicas**. III. Titulo.

CDU:591.5(213.54)

Agradecimentos

Agradeço primeiramente ao Prof. Dr. José Alexandre, por ter aberto mais uma vaga para mim durante a seleção do mestrado quando suas vagas já estavam lotadas. Se não fosse esta vaga adicional talvez eu nem teria entrado de vez neste mestrado ou teria feito uma dissertação com um assunto completamente diferente. E este laboratório que fiz parte me fez aumentar ainda mais meu interesse e conhecimento não apenas pela Ecologia em grandes escalas, mas também pela Bioestatística, através de diversas discussões e durante as "horas do cafezinho".

O mestrado também só foi possível porque sempre tinha minha família me financiando para sustento aqui em Goiânia, a paciência pela minha impaciência e mau-humor e através de apoios pessoais. Além do pessoal de casa também agradeço aos meus primos Ramirez, Romero e ao amigo Diego pelos bons momentos e das risadas dos finais de semana. Não posso esquecer dos divertidos parentes que freqüentam a casa da tia Ana e pelo apoio da tia Onaide aqui em Goiânia.

Ao pessoal que ajudou bastante através de críticas, ou sugestões, ou artigos, ou mesmo por ter acompanhado durante a composição da dissertação: Alexandre Siqueira, Bruno Barreto, Bruno Spacek, Fernanda Cassemiro, Guilherme de Oliveira, Joaquín Hortal, José Alexandre, Luís Maurício, Natália Mundim, Paulo de Marco, Renata Alves, Rogério Pereira e Taís Borges.

Aos participantes da banca examinadora: Alexandre Siqueira, Joaquín Hortal e José Alexandre.

Á boa convivência, amizade e descontração com o pessoal do Laboratório de Ecologia Teórica e Síntese, as turmas de mestrado e doutorado do Programa de Pós-Graduação em Ecologia e Evolução e também à galera do futebol e do vôlei de toda semana.

Sumário

Resumo.....6
Abstract.....7
Apresentação.....8
Objetivos e justificativa geral.....13
Bibliografia.....14

Capítulo I:.....17
Conceitos Teóricos Estatísticos e Ecológicos

Conceitos Estatísticos.....18
Conceitos Ecológicos.....34
Bibliografia.....39

Capítulo II:.....46
Seleção de modelos espaciais para a diversidade de vertebrados do Cerrado através do critério de Akaike.

Resumo.....47
Introdução.....48
Metodologia.....51
Dados Ecológicos.....51
Modelos de regressão.....54
Seleção de modelos e o critério de Akaike.....56
Avaliação de pressupostos.....58
Resultados.....59
Discussão.....73
Conclusão.....77
Bibliografia.....78

Capítulo III:.....83
O efeito do domínio médio e padrões espaciais de espécies endêmicas no bioma Cerrado: gradientes geográficos ou mero acaso?

Resumo.....84
Introdução.....85
Metodologia.....88
Dados ecológicos.....88
Análise de dados.....91
Resultados.....93
Discussão.....98
Conclusão.....104
Bibliografia.....104

Conclusões gerais.....110

Resumo

A predição da riqueza de espécies em uma região é baseada em modelos estatísticos nos quais o número de espécies está em função de diversas variáveis ambientais, sócio-econômicas e climáticas. Se a análise é realizada em uma escala regional ou continental, deve se considerar o efeito de padrões espaciais nos dados, já que se estes forem significativos, os resultados da estatística convencional serão incorretos devido à autocorrelação espacial, aumentando o erro tipo I de cada coeficiente em relação a H_0. Neste caso, utilizam-se os modelos de regressão espacial que corrigem a autocorrelação dos resíduos e permitem uma melhor avaliação da significância estatística dos coeficientes. Neste trabalho, foram construídos modelos estatísticos para explicar a variação geográfica da riqueza de espécies de vertebrados no Cerrado, com base em variáveis ambientais, sócio-econômicas e climáticas. Esses modelos foram comparados entre si através do Critério de Akaike, para avaliar quais modelos melhor explicam a riqueza de mamíferos, aves, répteis e anfíbios, além daqueles do conjunto de espécies endêmicas desses grupos no Cerrado.

Palavras-chave: regressão múltipla, autocorrelação espacial, modelos espaciais autoregressivos, Critério de Akaike (AIC), efeito do domínio médio, Cerrado.

Abstract

The prediction of the species richness in a region is based on statistical models in which the number of species is a function of several environmental, socioeconomic and climatic variables. If the analysis is accomplished over a regional or continental scale, it should consider the effect of spatial patterns in the data, since if these are significant, the conventional statistics results will be incorrect due to the spatial autocorrelation increasing the error type I of each coefficient in relation to H_0. In this case, the models of spatial regression are used because they correct the autocorrelation of the residuals and allow a better evaluation of the statistical significance of the coefficients. In this work, statistical models were built to explain the geographical variation of the species richness of vertebrates in the Cerrado, based on environmental, socioeconomic and climatic variables. The models were compared amongst themselves through the Akaike Information Criterion (AIC), to evaluate which models better explain the mammals, birds, reptiles and amphibians richness, besides that of the group of endemic species of those groups in the Savannah Cerrado.

Key words: Ordinary Least Square (OLS), spatial autocorrelation, spatial autoregressive models, Akaike Information Criterion (AIC), mid-domain effect, Savannah Cerrado.

Apresentação

A busca por padrões de riqueza de espécies tem procurado entender os aspectos ecológico-evolutivos da biodiversidade. Para isto, muitas hipóteses foram elaboradas nos últimos 200 anos (Rohde, 1992), embora Rahbek & Graves (2001) e Willig *et al*. (2003) tenham diminuído esta lista a fim de considerar as hipóteses com o maior poder de explicação. Com os dados de diversidade que já são conhecidos atualmente, os resultados das hipóteses podem vir a ser falhos devido à falta de conhecimento adequado da biodiversidade em algumas regiões tanto em relação à sua ecologia quanto a falta de descrição das espécies (Whittaker *et al.*, 2005). Por isto é necessário conhecer mais sobre a diversidade em todos seus aspectos e deste modo associar os dados reais às teorias biológicas.

Um modo de compreender melhor a distribuição de espécies, em grandes extensões territoriais, é através de modelos que realizam descrições. Neste caso, utilizam-se os parâmetros que mais se associam à distribuição ou mesmo a utilização de modelos nulos nos quais descreve um padrão de riqueza de espécies na ausência dos principais fatores que estariam explicando uma distribuição de espécies (Gotelli & Graves, 1996; Colwell & Lees, 2000).

O bioma Cerrado (Figura 1), que foi a região de estudo deste trabalho, é um ecossistema de múltiplas fisionomias com características gerais de uma savana tropical, com extremos desde um campo limpo até florestas fechadas (cerradão) (Oliveira & Marquis, 2002). Ocupa aproximadamente 22% do território nacional, nas regiões Centro-Oeste, Norte, Nordeste e Sudeste, embora sua área de cobertura original venha sendo reduzida constantemente para uso agrícola. As estimativas variam, e as mais pessimistas sugerem que cerca de 55% de sua área

já tenham sido modificada por ação humana (Oliveira & Marquis, 2002; Klink & Machado, 2005).

Esta intensa ocupação humana é responsável pela diminuição da viabilidade do Cerrado para as espécies ocupantes, porque causa fragmentação de hábitats, perda de biodiversidade, degradação de paisagens, erosão do solo e provável modificação climática regional (Klink & Machado, 2005). Por causa desta degradação e também devido ao alto endemismo (principalmente de plantas), este bioma é considerado um dos "hotspots" mundiais para a conservação da biodiversidade (Myers *et al*., 2000). São necessários esforços urgentes de conservação para diminuir o impacto sobre este bioma, não apenas em termos acadêmicos (i.e., aumentando o conhecimento sobre a biodiversidade na região) mas também em termos de atitudes políticas e sócio-econômicas.

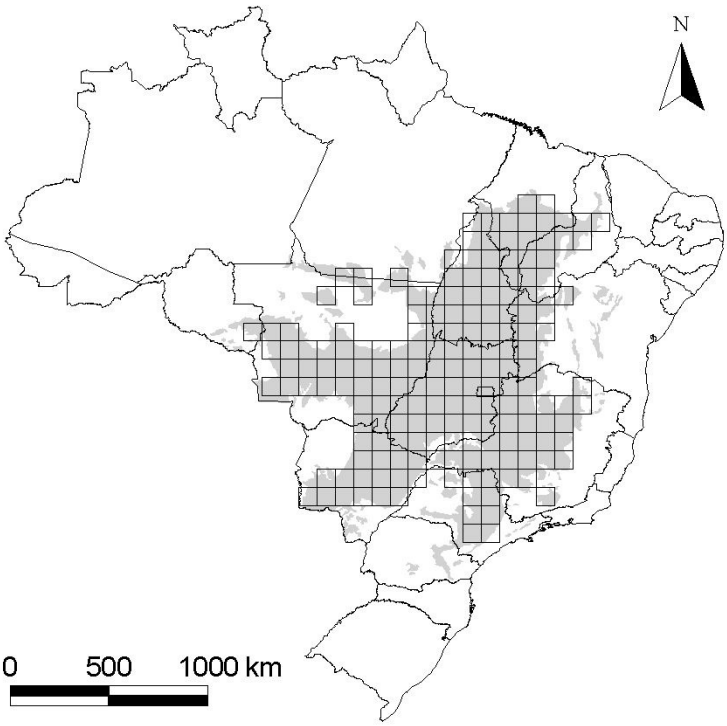


Figura 1. Distribuição do bioma Cerrado, dividido em 181 quadrículas de 1º de latitude x 1º de longitude, sobre o território brasileiro.

A distribuição da riqueza de vertebrados e das espécies endêmicas no Cerrado, que foi o tema deste estudo, está sob a influência de diversas variáveis. As variáveis utilizadas para predizer as espécies são em geral climáticas, ambientais e sociais, e possuem variação espacial se forem consideradas amplas escalas regionais. Para as espécies endêmicas do Cerrado, foi considerada também a hipótese "*mid-domain effect*" (Colwell & Lees, 2000), baseada em modelos nulos. De acordo com esta hipótese, a região central do "domínio" (bioma

Cerrado, por exemplo) terá maior diversidade de espécies simplesmente pela distribuição aleatória das espécies no domínio que é restrito por limites de borda.

As distribuições geográficas das espécies estão correlacionadas mais fortemente com algumas variáveis que outras, de modo que essas variações espaciais modificam a riqueza (que é na verdade a sobreposição das distribuições dessas espécies). Quando se consideram escalas espaciais amplas, deve-se levar em consideração a autocorrelação espacial porque quando esta ocorre há dependência entre pontos de amostragens, o que pode levar a falsas conclusões sobre os fatores que explicam a riqueza e a distribuição de espécies, especialmente em termos da significância estatística desses efeitos (Legendre, 1993; Diniz-Filho *et al.*, 2003). Recentemente, ecólogos e biogeógrafos vêm utilizando cada vez mais o efeito da autocorrelação espacial em seus estudos (e.g., Legendre, 1993; Lennon, 2000; Diniz-Filho *et al.*, 2003; Rangel *et al.,* 2006).

Para avaliar quais variáveis mais influenciam a distribuição de vertebrados no bioma Cerrado, foram utilizados neste trabalho modelos para descrever a riqueza de espécies que seria esperada pelo conjunto de variáveis disponíveis. Serão considerados os seguintes tipos de regressão: a regressão múltipla (OLS - *Ordinary Least Squares*) (Haining, 1990; Sokal & Rohlf, 1995), que não avalia a autocorrelação espacial, e a autoregressão espacial simultânea (SAR - *Simultaneous Autoregression)* (Haining, 1990; Fotheringham *et al.,* 2002), que incluem em seus cálculos os efeitos da variação espacial nos resíduos dos modelos. Se realmente existe uma estrutura espacial significativa, o modelo espacial possuirá uma melhor performance e seus resultados serão mais confiáveis em relação à regressão múltipla OLS, como estudos recentes têm

demonstrado (Tognelli & Kelt, 2004; Svenning *et al.*, 2006; Hawkins *et al.*, 2007; Kühn, 2007).

Como o potencial computacional tem aumentado constantemente, a ecologia teve um acréscimo na performance de modelagens, banco de dados e análises estatísticas de um grande volume de informações. No estudo de padrões de riqueza de espécies, como podem ser gerados diversos modelos devido às várias combinações possíveis das variáveis disponíveis, a seleção do melhor modelo, com o maior poder descritivo e que fornece as variáveis mais importantes, será feita utilizando o critério de Akaike, uma ferramenta avançada que não inclui testes de hipóteses em um sentido clássico (estatístico). Contudo, ele oferece uma idéia da evidência de cada modelo em descrever a variável resposta (Burnham & Anderson, 2004). Deste modo, cada modelo gerado pela regressão múltipla OLS e pelo modelo regressivo espacial (SAR) foi analisado por este critério de seleção, sendo comparados uns aos outros.

No primeiro capítulo desta dissertação foi feita uma revisão e conceituação das ferramentas estatísticas que serão utilizadas nos demais trabalhos, bem como uma revisão das principais hipóteses que poderiam explicar a distribuição de espécies no bioma Cerrado. No segundo capítulo, há uma análise das principais variáveis que influenciam a distribuição de vertebrados (anuros, répteis, aves e mamíferos) neste bioma, considerando a autocorrelação espacial nos cálculos e a seleção de modelos através do critério de Akaike. No terceiro capítulo utiliza-se o mesmo procedimento a fim de se avaliar a hipótese do efeito do domínio médio ("*mid-domain effect*") ao se considerar o quanto um modelo nulo de distribuição

das espécies poderia explicar a distribuição de espécies endêmicas, em adição às outras hipóteses ambientais e sociais.

Objetivos e Justificativa Geral

O critério de Akaike, como uma nova ferramenta de seleção de modelos, foi utilizado preferivelmente em relação à seleção tradicional (com base em testes de significância do modelo ou dos diferentes descritores) para determinar os melhores modelos baseados não apenas em variáveis "significativas", mas também na perda de informação do melhor modelo gerado pelo próprio processo de parametrização. Neste estudo também foi considerado o efeito espacial dos dados ecológicos, ao utilizar o modelo SAR em relação ao OLS, aumentando assim a performance e a precisão dos resultados. Essas análises permitem avaliar:

1) qual, ou quais, modelos regressivos (contendo diferentes combinações de variáveis preditoras), melhor explicam a distribuição da riqueza dos quatro grupos taxonômicos (répteis, anfíbios, aves e mamíferos), no bioma Cerrado, aqui considerados;

2) Para as espécies endêmicas, será adicionada a variável *GCM* no conjunto de dados, que indica a riqueza esperada de espécies de acordo com o modelo nulo de domínio médio (MDE), permitindo avaliar as variáveis que mais influenciam a distribuição das espécies características do Cerrado.

Assim, em termos gerais, os aspectos inovativos deste trabalho em relação à maior parte dos estudos que procuram avaliar padrões de riqueza são: 1) a

utilização da abordagem da hipótese "*mid-domain effect*", através da variável *GCM*, como uma variável descritora das espécies endêmicas; 2) utilização do modelo SAR por considerar o efeito espacial; 3) utilização do critério de Akaike como ferramenta de seleção de modelos; 4) avaliação da multicolinearidade e sua possível correção.

Bibliografia

Burnham, K.P. & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33 (2): 261-304.

Diniz-Filho, J.A.F.; Bini, L.M.; Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography,* 12: 53-64.

Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution,* 15 (2): 70-76.

Fotheringham, A.S.; Brunsdon, C.; Charlton, M. (2002) *Quantitative Geography : Perspectives on Spatial Analysis.* SAGE Publications.

Gotelli, N.J. & Graves, G.R. (1996) *Nulls models in ecology*. Smithsonian Institute Press.

Haining, R. (1990) *Spatial data analisys in the social and environmental sciences*. Cambridge University Press, Cambridge.

Hawkins, B.A.; Diniz-Filho, J.A.F.; Bini, L.M.; De Marco, P.; Blackburn, T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography,* 30 (3): 375-384.

Klink, C.A. & Machado, R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology,* 19 (3): 707-713.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology,* 74: 1659-1673.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. Elsevier Science.

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography,* 23 (1): 101-113.

Myers, N.; Mittermeier R.A.; Mittermeier C.G.; Fonseca G.A.B.; Kent J. (2000) Biodiversity hotspots for conservation priorites. *Nature,* 43: 853-858.

Oliveira, P.S. & Marquis, R.J. (2002) *The cerrados of Brazil: ecology and natural history of a neotropical savanna.* Columbia University Press.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences,* USA, 98: 4534-4539.

Rangel, T.F.L.V.B.; Diniz-Filho, J.A.F.; Bini, L.M. (2006) Towards an integrated computacional tool for spatial analysis in macroecology and biogeography. *Global Ecology & Biogeography,* 15: 321-327.

Rhode, K. (1992) Latitudinal gradients in species diversity: the search for the primary cause. *Oikos,* 65 (5): 14-27.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research.* W.H. Freeman and Company.

Svenning, J.; Engelbrecht, B.M.J.; Kinner, D.A.; Kursar, T.A.; Stallard, R.F.; Wright, S.J. (2006) The relative roles of environment, history and local dispersal in controlling the distributions of common tree and shrub species in a tropical forest landscape, Panama. *Journal of Tropical Ecology,* 22: 575-586.

Togneli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, 27: 427-436.

Whittaker , R.J.; Araújo, M.B.; Paul, J.; Ladle, R.J.; Watson, J.E.M.; Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, 11: 3-23.

Willig, M.R.; Kaufman, D.M.; Stevens, R.D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Syst.,* 34: 273-309.

CAPÍTULO I

CONCEITOS TEÓRICOS ESTATÍSTICOS E ECOLÓGICOS

Conceitos Teóricos Estatísticos

Os modelos em ecologia são aproximações simplificadas da realidade, ou da ”verdade”, ou seja, não há modelos que refletem perfeitamente toda a realidade (esta contém toda a informação sobre o processo ecológico de interesse) (Burnham & Anderson, 2004). Há diferentes tipos de modelos a serem considerados: (1) o modelo conceitual refere-se à expressão de um conjunto de idéias sobre como opera um sistema em particular; (2) ao transformar estas idéias em palavras temos um modelo verbal; (3) transformando idéias ou palavras em simbologias e equações matemáticas, temos modelos matemáticos (Williams, 2001; Hobbs & Hilborn, 2006). O uso de modelos ecológicos tem como objetivo inicial prever o comportamento futuro (ou inferir padrões no passado) das variáveis de interesse. Em princípio, é necessário utilizar modelos matemáticos em ecologia devido à complexidade da natureza, que dificulta a utilização de modelos simples e verbais. As utilizações destes modelos conferem alguma idéia e direção do que realmente deveríamos tentar medir na natureza (Gotelli, 1998). A análise de dados ecológicos também tem o objetivo de identificar qual modelo, ou quais, melhor descrevem o conjunto de dados com o interesse de identificar e explicar os fatores que mais influenciam a variável resposta, como por exemplo a riqueza de espécies. Outra informação que pode ser extraída é tentar extrair informações de como os processos ecológicos geraram os dados (Richards, 2005).

Com os modelos em mãos, devem ser considerados três princípios gerais ao se realizar inferências:

(1) Simplicidade e parcimônia: deve-se utilizar a explicação mais simples como sendo a preferencial e, no caso de modelos lineares, essa idéia seria

expressa pelo balanço entre simplicidade (i.e., número de variáveis) e poder de explicação;

(2) As diversas hipóteses a serem consideradas: os vários modelos descritos são avaliados para que possa ser feita a análise de quais destes melhor oferecem suporte aos dados, podendo ser adicionadas à análise outras hipóteses e retiradas aquelas que oferecem baixo poder explicativo;

(3) Avaliação do poder de cada modelo: fornece uma medida quantitativa para avaliar o poder de descrição e explicação de cada modelo (Burnhan & Anderson, 2004; Mazerolle, 2006). Estes princípios abordam a seleção de modelos sob o critério de Akaike (mais adiante).

As análises ecológicas de diversidade têm utilizado modelos na forma de equações matemáticas para realizar predições sobre a riqueza de espécies, em termos de quando esta variável está sob a influência de vários fatores, sendo o modelo mais simples denominado de regressão múltipla OLS. Esta equação revela as relações entre uma variável resposta e duas ou mais variáveis ecológicas, as variáveis explanatórias, independentes ou preditoras. A equação linear convencional da regressão múltipla é escrita como:

$$\hat{Y} = a + b_{y1}.X_1 + b_{y2}.X_2 + ... + b_{yk}.X_k$$ **Equação 1**

Sendo sua notação matricial dada por

$$Y = X\beta + \varepsilon \qquad \hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$ **Equação 2**

em que a estimativa da variável dependente $\hat{Y}$ está sob a função de *k* variáveis independentes X_1, X_2, ..., X_k. O termo $b_{\hat{y}}$ refere-se ao coeficiente de regressão da variável X_j , considerando que todas as outras variáveis da equação de regressão sejam mantidas matematicamente constantes, sendo denominado de coeficiente parcial de regressão. O valor de *a* é o intercepto da função, ou seja, o valor de $\hat{Y}$ quando todos os valores das variáveis independentes são iguais a zero (Sokal & Rohlf, 1995). Na equação 2 o vetor ***Y*** refere-se aos elementos da variável resposta que estão em função dos elementos da matriz ***X***, que contém as variáveis independentes (variáveis explanatórias), multiplicado pelo vetor β que contém os coeficientes de regressão parciais, descrevendo o quanto a variável resposta possui influência sobre as variáveis independentes. O vetor ε são os resíduos da regressão. Na verdade, assume-se que a matriz de covariância dos resíduos assume a seguinte forma: $\sigma^2 I$, na qual $\mathbf{I}$ é a matriz identidade e σ^2 representa a variância entre os valores residuais provenientes de cada ponto amostral (Haining, 1990; Fotheringham *et al*, 2002).

O objetivo da regressão múltipla é tentar ao máximo aumentar a proporção da variação de Y explicada pelo modelo regressivo e diminuir a proporção da variação de Y que não pode ser explicada pela equação, o resíduo. Este é medido pela diferença entre o valor observado e o valor estimado pela regressão de Y por X (Sokal & Rohlf, 1995).

O uso de regressões múltiplas em ecologia para questões de modelagem possui dois pontos de vista: a procura pelo melhor modelo descritivo e a abordagem da precisão e magnitude em cada modelo possível de ser gerado,

utilizando intervalos de confiança e coeficientes de determinação (r^2) para o desenvolvimento de novas idéias e pesquisas em relação à análise de processos ecológicos (Mac Nally, 2000).

Para que possa ser feita uma análise correta pela regressão múltipla os pressupostos a serem atendidos são: (1) os valores do resíduo, ou erro, devem possuir uma distribuição normal; (2) homogeneidade da variância, ou seja, as diferentes combinações das variáveis independentes devem possuir distribuições de amplitude dos valores do resíduo iguais estatisticamente entre si; (3) os valores do resíduo são independentes entre si (não há dependência espacial ou temporal nos resíduos, o que será explorado em mais detalhas adiante) (Zar, 1996).

Dado o conjunto de observações (uma variável resposta e duas ou mais variáveis explanatórias em cada ponto de amostragem) e realizados os cálculos matemáticos, obtém-se a equação de regressão considerando-se todas as variáveis explanatórias. Este modelo é ajustado para que a soma dos quadrados dos resíduos tenha o menor valor possível. Deste modo, a significância do efeito das variáveis explanatórias é avaliada sobre o poder de descrição em relação à variável Y. Caso o poder descritivo seja baixo, esta variável X_j é usualmente retirada do modelo e um novo ajuste é então realizado, até alcançar um modelo na qual todas as variáveis X_j sejam significativas. Este processo é denominado de seleção de variáveis, ou seja, a identificação de quais variáveis descrevem melhor os dados ecológicos, indo de encontro ao melhor modelo (Zar, 1996; Mac Nally, 2000).

Há várias técnicas disponíveis para a seleção de variáveis em regressão múltipla. Entre as disponíveis são os procedimentos automáticos de "*stepwise*

selection", que por sua vez podem ser realizados de diferentes formas. Há, por exemplo, um procedimento de "*forward selection*", que inicia-se com a equação mais simples (i.e., uma única variável explanatória), e gradualmente adiciona as variáveis independentes que sejam significativas para a equação de regressão. A outra possibilidade oposta é adotar um procedimento de "*backward elimination*", na qual inicia-se com a equação mais completa e as variáveis independentes menos significativas irão sendo sucessivamente retiradas. Há procedimentos mais complexos nas quais as variáveis podem tanto ser adicionadas como retiradas do modelo após os testes de significância (Zar, 1996). Entretanto, há várias críticas sendo feitas em relação a estes métodos (Stephens *et al.*, 2005), visto que procedimentos automáticos como "*stepwise*" podem incluir no modelo regressivo as variáveis independentes erradas, podem criar modelos com um coeficiente de determinação aumentado artificialmente e podem, em geral, permitir facilmente a geração de modelos falsos, sem significado biológico ou ecológico. Isto pode ocorrer por vários motivos: (1) quando não há uma seleção criteriosa *a priori* das variáveis de interesse do assunto em questão, deixando apenas os testes estatísticos selecionarem as variáveis (o que demonstra a importância do conhecimento teórico do sistema de estudo que está sendo abordado); (2) quando o número de observações é menor que a quantidade de variáveis disponíveis (Flack & Chang, 1987); (3) viés no teste de parâmetros através dos testes de significância, causados, por exemplo, por efeitos de autocorrelação espacial ou temporal; (4) o critério de seleção não envolve apenas um teste de hipóteses, mas vários simultaneamente, o que aumenta o erro tipo I; (5) neste caso de seleção automática de modelos, não há como avaliar o quanto a melhor regressão ajusta-

se bem aos dados, ou seja, nem sempre o melhor modelo selecionado possui a melhor performance porque outros futuros modelos podem ajustar tão bem quanto o anterior (Whittingham *et al.*, 2006). Apesar de todas estas críticas, o uso de hipóteses nulas e testes de significância é bastante útil e aceito em estudos experimentais e em alguns estudos observacionais, não podendo ser descartada das análises estatísticas (Stephens, 2005).

Outra questão a ser tratada em regressões é quando há associações entre as variáveis utilizadas (multicolinearidade) e, como conseqüência, isso pode gerar estimativas enviesadas dos parâmetros (Philippi, 1993). A informação é enviesada porque o erro padrão dos coeficientes de regressão aumentam na presença de relações entre variáveis e, deste modo, há desvios dos valores dos coeficientes de regressão que seriam esperados na ausência de multicolinearidade (Graham, 2003). Um modo direto de detectar a presença da multicolinearidade é avaliar as matrizes de correlação entre as variáveis descritoras. As medidas destas relações também podem ser contabilizadas pelo fator de aumento da variância (*variance inflation factor – VIF*) em cada variável preditora, que é dado pela seguinte equação:

$$VIF = \frac{1}{\left(1 - R^2\right)}$$ **Equação 3**

na qual este coeficiente de determinação R^2 é obtido pela regressão de uma variável explanatória como resposta às demais variáveis preditoras no modelo. O VIF indica o aumento da variância do coeficiente de regressão comparado quando

não há colinearidade. Embora tenha sido comumente descrito que valores maiores que dez sejam indicativos de problemas de colinearidade, Graham (2003) diz que mesmo baixos valores podem trazer problemas.

Atualmente, os pesquisadores utilizam freqüentemente a abordagem estatística tradicional de seleção de modelos para realizar inferências sobre processos biológicos (Hobbs, 2006). Neste processo, uma hipótese nula é gerada e avalia-se se esta hipótese pode ser rejeitada dado o conjunto de observações. A rejeição ocorre quando a probabilidade de algum evento em questão ocorrer simplesmente ao acaso é baixa e então a hipótese alternativa é aceita (Sokal & Rohlf, 1995). Contudo, em muitos casos, os pesquisadores estão interessados apenas em encontrar modelos com maior poder de explicação e descrição que estejam mais próximos da realidade e que possuam uma boa base teórica, procurando identificar os modelos mais parsimoniosos e o que melhor explicam os dados (Mac Nally, 2000).

O aumento dos suportes computacionais tornou mais fácil o uso de técnicas mais avançadas e desde então o uso dessas técnicas tem aumentado por parte dos pesquisadores. Estes novos procedimentos oferecem direções para a pesquisa ecológica, com análises mais completas e conclusões que se aproximam mais da realidade ao estudar um sistema em particular (Liebhold & Gurevitch, 2002; Johnson & Omland, 2004; Hobbs & Hilborn, 2006). Deste modo, uma maneira de realizar a seleção de modelos sem considerar testes de hipóteses é fundamentada na teoria da informação, que oferece um modo de extrair um poder de explicação de um conjunto de hipóteses, representados por modelos matemáticos. Neste caso, testes de hipóteses não são utilizados. Os testes

estatísticos tradicionais, através da significância de variáveis, têm resultados limitados, especialmente em dados observacionais, pouco informativos porque apenas revelam a significância e são mais prováveis de obterem resultados errados (Burnham & Anderson, 2004; Mazerolle, 2006; Hobbs & Hilborn, 2006).

A abordagem da seleção de modelos pelo Critério de Akaike possui três vantagens: (1) os diferentes modelos gerados são comparados uns aos outros, avaliando o suporte que cada modelo oferece aos dados observados; (2) cada modelo oferece uma medida quantitativa de quão provável seja o melhor modelo estimado; (3) em casos nas quais os modelos oferecem o mesmo suporte aos dados, utilizam-se técnicas para realizar a estimação e predição de forma mais completa (Johnson & Omland, 2004).

Entretanto, antes de realizar as análises ecológicas, é preciso estabelecer inicialmente o conjunto de modelos a ser analisado. Primeiramente, obtém-se um conjunto de hipóteses biológicas que são transformadas em modelos, o que requer a identificação das variáveis e a seleção das funções matemáticas que descrevem os processos ecológicos (Johnson & Omland, 2004; Hobbs & Hilborn, 2006).

O critério de Akaike (AIC) utiliza a abordagem da perda de informação de Kullback-Leibler. Esta abordagem refere-se à perda de informação quando um modelo é usado para aproximar a realidade, porque é impossível um modelo obter toda a informação de um processo ecológico complexo. A distância de Kullback-Leibler pode ser entendida como a distância entre a completa realidade e um modelo (Burnham & Anderson, 2004; Johnson & Omland, 2004; Richards, 2005; Hobbs & Hilborn, 2006; Mazerolle, 2006). Quando se faz seleção de modelos

utilizando o critério de Akaike, obtém-se no final os modelos que foram mais importantes ao oferecerem suporte aos dados e também evidenciam as variáveis de maior interesse. Como já foi dito, neste processo não utilizam-se testes de hipóteses como é feita na seleção de variáveis da regressão múltipla por um processo de "*stepwise*".

Ao realizar a seleção de modelos pelo Critério de Akaike utilizando os resultados de uma regressão múltipla, os valores de AIC são calculados utilizando as seguintes equações:

$$AIC = n\ln\left(\hat{\sigma}^2\right) + 2K \qquad \hat{\sigma}^2 = \sum\left(\hat{\varepsilon}_i\right)^2 \Big/ n$$

Equação 4

na qual $\hat{\sigma}^2$ refere-se à variância dos resíduos de cada modelo regressivo. A letra *K* indica quantos parâmetros estão incluídos em cada regressão, então neste caso será a quantidade de variáveis, o intercepto de cada modelo e também a variância residual (Burnham & Anderson, 2004; Richards, 2005). Os pressupostos da equação acima requerem os mesmo cuidados da regressão múltipla e, deste modo, é necessário que os valores dos resíduos sejam independentes entre si, distribuídos normalmente e que a variância seja homogênea.

Cada modelo construído terá um valor de AIC, sendo que os modelos com os menores valores de AIC serão considerados os melhores modelos para explicar um determinado padrão ecológico, entre todos aqueles que foram considerados. Os valores individuais do AIC para cada modelo são pouco interpretáveis, exceto quando são comparados uns aos outros, já que esses

valores são afetados por constantes de escalas desconhecidas, pelo tamanho da amostra e podem conter um amplo intervalo de valores. Para realizar uma melhor comparação entre os valores de cada modelo a seguinte transformação deve ser feita

$$\Delta_i = AIC_i - AIC_{MIN}$$ **Equação 5**

na qual AIC_{MIN} é o menor valor de AIC entre todos os modelos considerados (Richards, 2005; Hobbs & Hilborn, 2006). Neste caso, o melhor modelo terá um valor Δ igual a zero, e os outros modelos terão valores positivos. Esta equação fornece uma medida relativa ao melhor modelo e seus valores posicionados permitem uma rápida avaliação da comparação dos poderes de evidência, em relação aos dados, para cada modelo. Burnham & Anderson (2004) sugeriram a seguinte regra: valores de ΔAIC menores que 2 são considerados os modelos que oferecem os melhores suportes relativos aos dados e deveriam ser usados para fazer futuras inferências sobre o processo ecológico em questão.

Outra medida para que se possa realizar comparação entre modelos e avaliar o poder de evidência, é ponderar o valor de cada modelo (w_i) utilizando os valores ΔAIC, através da seguinte equação.

$$w_i = \frac{\exp\left(-\Delta_i / 2\right)}{\sum_{r=1}^{R} \exp\left(-\Delta_r / 2\right)}$$ **Equação 6**

Os valores Δ indicam a diferença de valores do Critério de Akaike relativos ao melhor modelo e *R* refere-se ao conjunto de todos os modelos considerados.

O resultado desta equação indica a chance de que o modelo *i* seja o melhor entre todo o conjunto de modelos candidatos, caso o estudo seja repetido diversas vezes (Johnson & Omland, 2004). Como exemplo, se o valor ponderado de um modelo for 0,50, isto significa que este modelo tem uma chance de 50% de ser o melhor modelo caso o estudo seja repetido diversas vezes.

Entretanto, é importante considerar que o critério de Akaike também pode levar a conclusões falsas se não forem consideradas as variações espaciais dos parâmetros. Quando há forte autocorrelação, a dependência espacial entre os pontos de amostragem, um pressuposto foge à regra da regressão múltipla, a de que os resíduos são independentes. Isto ocorre porque pontos adjacentes tendem a ter dados similares e neste caso os resíduos serão dependentes entre si, superestimando os graus de liberdade e os valores dos testes estatísticos convencionais (Legendre, 1993; Diniz-Filho *et al.*, 2003). Deste modo não será adequado usar regressão múltipla simples quando há autocorrelação espacial para realizar a seleção de modelos utilizando o critério de Akaike, mas sim um modelo autoregressivo (tópico adiante) que considera dados espacialmente estruturados (Keitt *et al.*, 2002)

A autocorrelação espacial pode ser positiva, quando pontos de amostragem localizados a uma dada distância apresentam valores, de um parâmetro qualquer, mais similares entre si do que esperados ao acaso, ou negativa, quando esses pontos de amostragem apresentam valores mais distintos do que esperado ao acaso (Legendre, 1993)

A magnitude da autocorrelação espacial é freqüentemente calculada em genética evolutiva e ecologia geográfica pelo Índice *I* de Moran através da seguinte equação (Legendre, 1998)

$$I(d) = \frac{\frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - \bar{y}\right)\left(y_j - \bar{y}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad i \neq j \qquad \textbf{Equação 7}$$

na qual *n* é o número de pontos de amostragem ou quadrículas; y_i e y_j são valores de uma dada variável Y qualquer nos pontos *i* e *j* ; $\bar{y}$ é o valor médio de *y* ; w_{ij} é valor que pertence à matriz *W* , esta sendo uma matriz de conexões entre os pontos de amostragem assumindo valor igual a 1 quando estes dois pontos de amostragem estão conectados dentro de um intervalo de distância e valor igual a zero quando este dois pontos de amostragem não estão na mesma classe de distância. O índice *I* de Moran assume valores entre +1, perfeita autocorrelação positiva, até -1 e neste caso há máxima autocorrelação negativa.

A partir de uma variável são calculados os valores de *I* para grandes distâncias, separando esta em intervalos iguais ou não, e os resultados são plotados em gráficos denominados correlogramas nas quais os valores de autocorrelação são colocados no eixo das ordenadas e as classes de distância no eixo das abscissas. Com este gráfico pode-se avaliar o grau de autocorrelação espacial de um parâmetro entre pares de pontos de amostragem e como varia entre distâncias próximas e distantes (Legendre, 1998; Tognelli & Keit, 2004).

Quando os valores do erro do modelo OLS são independentemente distribuídos, assume-se a seguinte notação:

$$\varepsilon \sim MVN\left(0,\sigma^2 I\right)$$

Isto significa que os valores do erro apresentam uma distribuição multivariada normal com média zero e a matriz de variância-covariância é representada pelo segundo termo (Fotheringham *et al.*, 2002; Lichstein *et al.*, 2002) . Esta matriz diz que a variância dos erros entre as diferentes localidades são iguais e a covariância é igual a zero, isto indica que há independência dos resíduos. Contudo, se um modelo de regressão múltipla OLS apresenta autocorrelação espacial, os valores do erro serão dependentes entre si em algum grau e, neste caso, os testes de significância e, eventualmente, as estimativas dos parâmetros, serão afetadas (ver Hawkins *et al.*, 2007). Neste caso, deve-se implementar outros modelos que lidam com este efeito espacial, e uma alternativa é incorporar a autocorrelação diretamente nos resíduos dos modelos (Rangel *et al*, 2006).

Neste estudo foi utilizado o modelo autoregressivo simultâneo (SAR – *Simultaneous Autoregressive Models*). Esse modelo é denominado "autoregressivo" porque, além das variáveis preditoras serem incluídas no modelo, uma variável, que considera o efeito de localizações vizinhas, é adicionada (neste caso o vetor **Y** é incluído nos cálculos) (Fotheringham *et al.*, 2002). Basicamente a nova variável é uma matriz de conexões (**W**) entre as localidades multiplicado pelos valores do vetor **Y** e, que também terá seu coeficiente de regressão (ρ), que é chamado de coeficiente autoregressivo. Tanto a matriz **W** quanto o

coeficiente autoregressivo são incorporadas à estrutura de covariância entre os resíduos do modelo.

Os valores da matriz **W** apresentam valores iguais a zero quando duas localidades não estão conectadas ou adjacentes entre si, ou quando consideram duas mesmas localidades (diagonal principal), e valores menores que 1 (ou mesmo assumir valor igual a 1 de acordo com o critério de conexão utilizado) quando estão ligadas entre si e assumem ter relações de dependência. Ou seja, o valor desta conexão irá diminuir com o aumento da distância entre os pares de amostragens e aumentará quanto mais próximos estiverem entre si, sendo que valores iguais a zero serão obtidos se os pares de amostragens estiverem fora da região de influência de um local sobre o outro (este último assume independência espacial) (Lichstein *et al.,* 2002; Fotheringham *et al.*, 2002; Tognelli & Kelt, 2004). Quanto maiores são os valores w_{ij} na matriz **W**, mais próximos são os valores obtidos de uma variável qualquer destas duas localidades.

O parâmetro ρ é o coeficiente espacial autoregressivo e como está ligado a uma variável que considera o efeito de localidades vizinhas para predizer valores, ele indica que algum grau de autocorrelação espacial ocorre. Esse valor deve ser estatisticamente significativo, porque em caso contrário o SAR volta a ser um modelo OLS. Se este possui um valor maior que zero, assume-se a existência de uma autocorrelação positiva nos resíduos do modelo OLS (Lichstein *et al.*, 2002; Fotheringham *et al.*, 2002).

O modelo SAR incui em seu modelo a autocorrelação espacial nos resíduos ao assumir no cálculo da matriz de covariância o parâmetro ρ como indicando a seguir.

$$C = \sigma^2 \left[\left(I - \rho W \right)^T \left(I - \rho W \right) \right]^{-1}$$ **Equação 8**

O efeito espacial tende a mudar a influência das variáveis explanatórias sobre a variável resposta e, deste modo, a importância relativa de uma variável em predizer valores pode ser alterada. Variáveis que são bastantes autocorrelacionadas tendem a ter coeficientes mais fracos durante a predição (Lichstein *et al*, 2002). Neste caso novos coeficientes de regressão devem ser estimados através de uma equação proveniente de um modelo GLS (*Generalized Least Square*), que é uma classe de modelos que incorporam a autocorrelação nos resíduos e que diferenciam no cálculo dos coeficientes de regressão através de diferentes especificações da matriz de covariância (Haining, 1990). O cálculo para estimar o vetor dos coeficientes é feito pela equação:

$$\hat{\beta} = \left(X^T C^{-1} X \right)^{-1} X^T C^{-1} Y$$ **Equação 9**

que é mais geral do que a equação do OLS (equação 2). Estudos recentes têm incluído estes modelos para ajustar uma relação entre riqueza de espécies, ou abundância, e as variáveis climático-ambientais (Keitt *et al.*, 2002; Lichstein *et al.*, 2002; Tognelli & Kelt, 2004). Quando os resíduos provenientes da regressão múltipla OLS apresentam dados significativamente autocorrelacionados, utilizou-se a autoregressão espacial para considerar este efeito em seus cálculos. Nestes estudos, os resultados mostram que os modelos autoregressivos apresentaram

melhores performances em relação ao modelo OLS, menor autocorrelação dos resíduos e maior capacidade de explicação dos dados biológicos. Isto evidencia a importância da utilização de modelos autoregressivos, quando necessário, para estabelecer uma relação de causa e efeito mais próxima da realidade (entretanto, ver Hawkins *et al*., 2007.)

Conceitos Teóricos Ecológicos

Os diferentes grupos de vertebrados analisados neste trabalho possuem distintas distribuições espaciais sobre o bioma Cerrado, em parte porque respondem às variáveis ambientais de acordo com suas características biológicas e ecológicas. Mesmo dentro de um grupo taxonômico, as diferentes espécies irão se distribuir a partir de seus requerimentos em termos de seu nicho ecológico, dos padrões de dispersão e de outros componentes históricos em diferentes escalas de tempo.

A hipótese mais trabalhada em estudos de riqueza de espécies é baseada em energia ambiental, já que diversos autores confirmam esta medida na influência de padrões de riqueza em grande escala (Wright, 1983; Currie, 1987; Currie, 1991; Hawkins *et al.,* 2003a,b). Entretanto, ainda não há uma idéia clara sobre as relações de causa e efeito da hipótese espécie-energia (Evans & Gaston, 2005) em determinar o padrão de riqueza mais comumente conhecido, a variação latitudinal em grande escala.

Esta hipótese leva em consideração a entrada de radiação solar que entra na superfície da Terra, a temperatura (medidas diretas), a produtividade primária e as dinâmicas da água através da evapotranspiração (medidas indiretas), que de modo geral estão ligadas à hipótese climática. A relação da hipótese energética em plantas é que uma maior entrada de radiação solar possibilitaria maior captação desta energia solar por parte das plantas e a temperatura aumentaria a velocidade das reações químicas permitindo maior produção primária e conversão de produtos inorgânicos em substâncias orgânicas (Wright, 1983; Pianka, 1994),

sendo limitada pela disponibilidade de água. Esta dinâmica água-energia pode ter favorecido uma maior especiação em plantas.

Já em animais, há duas versões sobre como os efeitos da energia poderiam atuar na diversidade de espécies (Hawkins *et al.*, 2003a), considerando também a hipótese espécie-energia. Inicialmente, a energia causaria um aumento na produtividade em plantas e assim possibilitaria maiores quantidades de recursos disponíveis, permitindo que mais herbívoros ocupem nichos ecológicos e deste modo permitiriam a persistência de mais espécies nos próximos níveis tróficos. As espécies estariam limitadas pela produtividade dos níveis inferiores e as plantas pela disponibilidade de água e energia. Medidas de produtividade, tais como evapotranspiração atual (*AET*) e o índice de vegetação (*EVI*), geralmente oferecem alta proporção da explicação da riqueza de espécies em grande parte do globo terrestre (Currie & Paquin, 1987; Hawkins, *et al.,* 2003a,b; Hulbert & Haskell, 2003). O efeito dessas hipóteses seria sobre o tamanho da população, visto que o aumento deste diminuiriam as taxas de extinção e podem atuar na especiação (Hawkins *et al.,* 2003b).

Por outro lado, pode-se imaginar que a temperatura (*TEMP*) ou a evapotranspiração potencial (*PET*), que são medidas diretas de energia ambiental, podem atuar nos requerimentos fisiológicos dos organismos e também influenciar taxas de mutação, metabólicas e no tempo de geração (Evans & Gaston, 2005). Os organismos apresentam requerimentos de temperaturas, em determinados intervalos, para sua ótima performance em crescimento, reprodução e sobrevivência. Animais ectotérmicos apresentam seu espaço de atividades sob determinadas condições de temperatura e animais homotérmicos necessitam

gastar mais de sua energia em ambientes mais frios para manter a termorregulação, alocando recursos que seriam utilizados de outra forma para reprodução (Begon *et al.*, 1996). Em altas latitudes, a entrada de energia no ambiente, através da temperatura ao invés da produtividade, explica uma boa parte da riqueza para diversos grupos taxonômicos de animais (Currie, 1991; Hawkins *et al.*, 2003b).

Hawkins *et al.* (2003b) realizou uma revisão através de diversos estudos para avaliar a distribuição em grandes escalas de diversos grupos taxonômicos e verificou que nas regiões tropicais e quentes as variáveis relacionadas diretamente à água foram os melhores preditores da diversidade enquanto que nas regiões de alta latitude, mais frias, a entrada de energia atuou de maneira mais forte. A entrada de água por sua vez pode ser contabilizada através da hipótese climática. Esta hipótese estaria atuando indiretamente para maiores índices de produtividade (Hawkins, 2003a,b) e considera-se ser a melhor hipótese preditora de plantas em regiões tropicais e sub-tropicais.

Com relação à geografia, Kerr & Packer (1997) avaliaram que uma boa parte da variação na riqueza de mamíferos na América do Norte é bem predita pela variedade topográfica em associação com a energia ambiental. Rahbek & Graves (2001) avaliaram a diversidade de aves na América do Sul através de variáveis ambientais e espaciais e constataram que os locais mais diversos eram as regiões montanhosas dos Andes, onde as amplitudes dos relevos eram maiores, ao invés dos locais com maior entrada de energia. A diversidade de hábitats pode ser evidente quando há variação na topografia visto que a temperatura do ar diminui com a altitude cerca de 6º C a cada 1.000 metros de

elevação (Begon *et al.*, 1996). Outra atuação da topografia em áreas montanhosas é que a inclinação e a exposição ao sol podem tornar o ambiente bastante estressante e as formas de vida remanescentes são resistentes à seca e deste modo limitam também a diversidade, enquanto que as regiões mais baixas podem estar saturadas de água. Os locais que possuem maior diversidade ambiental (maior amplitude topográfica) podem sustentar mais espécies por causa dos distintos requerimentos ecológicos, de acordo com a teoria do nicho ecológico. Entretanto, esta amplitude da topografia e outras medidas de diversidade ambiental não são sempre fortemente correlacionadas (Rahbek & Graves, 2001).

Atualmente, o uso de indicadores sócio-econômicos para questões de conservação e a coincidência de riqueza de espécies com a população humana são fontes para diversos estudos (Balmford *et al.*, 2001; Araújo, 2003; Gaston & Evans, 2004; Diniz-Filho *et al.,* 2006). A idéia neste caso é avaliar o quanto a riqueza de espécies está associada a indicadores de ocupação humana. Deste modo, pode-se analisar conflitos de conservação, impactos sobre a diversidade, bem como seu maior conhecimento em locais mais densamente povoados. A ocupação humana no bioma Cerrado caracteriza-se por ser recente e altamente tecnológica, bastante associada com modernas técnicas agrícolas que tornaram o solo mais produtivo e devido também às oportunidades de emprego com a construção da capital federal, Brasília (Klink & Moreira, 2002). De acordo com trabalhos anteriores (Diniz-Filho *et al.*, 2005; Bini *et al.*, 2006), há uma correlação positiva entre espécies de anuros conhecidas e tamanho populacional humano, isto porque há maiores chances de serem descritas e maiores esforços de

amostragem em locais mais densamente povoados e deste modo estes locais possuem maiores quantidades de inventários faunísticos.

Entre as novas metodologias para predizer a riqueza de espécies estão o efeito do domínio médio ("*mid-domain effect*"), cuja predição é dada de forma computacional pelo modelo geométrico de Jetz & Rahbek (2001) ("*geometric constraints model*", *GCM*). No *GCM* a distribuição geográfica das espécies é geometricamente restrita pelos limites dentro de um domínio, o que leva a aumento da sobreposição das espécies em direção ao centro da região e deste modo a diversidade será maior na posição central da região (Colwell & Lees, 2000). Esta hipótese é considerada como um modelo nulo (Gotelli & McGill, 2006), ou seja, a distribuição de espécies preditas pelo modelo irá ocorrer principalmente no centro do domínio simplesmente ao acaso dentro dos limites geométricos do domínio, sem o efeito de outras variáveis que influenciam esta distribuição. Este modelo pode vir a explicar alguns padrões espaciais de riqueza de espécies endêmicas, sendo que os valores preditos por este modelo podem ser usados como uma variável para explicar a distribuição espacial de espécies (Colwell *et al.*, 2004). Como espécies endêmicas estão restritas a um bioma ou região, sua distribuição poderia ser explicada por este modelo. Este modelo, entretanto, possui posições contrárias e a favor em diversos artigos (Colwell & Lees, 2000; Hawkins & Diniz-Filho, 2002; Colwell *et al.*, 2004; Hawkins *et al.*, 2005), visto que há argumentos de que não possui significado ecológico para os gradientes de diversidade ou possui pouco poder de explicação, enquanto que por outro lado defende-se que há processos estocásticos influenciando a distribuição de espécies. Os resultados da hipótese do efeito do domínio médio também podem

ser alcançados utilizando simulações de um processo espacialmente explícito de dinâmica neutra de Hubbel (2001).

Bibliografia

Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography,* 12: 5-12.

Balmford, A.; Moore, J.L.; Brooks, T.; Burguess, N.; Hansen, L.A.; Williams, P.; Rahbek, C. (2001). Conservation conflicts across Africa. *Science,* 291: 2616-2619.

Begon, M.; Harper, J.L.; Townsend, C.R. (1996) *Ecology.* Blackwell Science.

Bini, L.M.; Diniz-Filho, J.A.F.; Rangel. T.F.L.V.B.; Bastos, R.P.; Pinto, M.P. (2006) Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distribution,* 12: 475-482.

Burnham, K.P. & Anderson, D.R. (2002). Model selection and multimodel inference. Springer, NY.

Burnham, K.P. & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33 (2): 261-304.

Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution,* 15 (2): 70-76.

Colwell, R.K.; Rahbek, C.; Gotelli, N.J. (2004) The mid-domain effect and species richness patterns: what have we learned so far? *American Naturalist,* 163 e-article.

Currie, D.J. & Paquin. V. (1987) Large-scale biogegraphical patterns of species richness of trees. *Nature,* 329: 326-327.

Currie, D.J. (1991) Energy and large-scale patterns of animal- and plant- species richness. *The American Naturalist,* 137 (1): 27-49.

Diniz-Filho, J.A.F.; Bini, L.M.; Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography,* 12: 53-64.

Diniz-Filho, J.A.F.; Bastos, R.P.; Rangel, T.F.L.V.B.; Bini. L.M.; Carvalho, P.; Silva, R.J. (2005) Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology & Biogeography,* 14: 469-477.

Diniz-Filho, J. A. F., L. M. Bini, M. P. Pinto, T. F. L. V. B. Rangel, P. Carvalho & R. P. Bastos (2006) Anuran Species Richness, Complementarity And Conservation Conflicts in Brazilian Cerrado. *Acta Oecologica,* 29: 9-15

Evans, K.L. & Gaston, K.J. (2005a). Can the evolutionary-rates hypothesis explain species-energy relationships? *Functional Ecology,* 19: 899-915.

Flack, V.F. & Chang, P.C. (1987) Frequency of selecting noise variables in subset regression analysis: a simulation study. *The American Statistician,* 41 (1): 84-86.

Fotheringham, A.S.; Brunsdon, C.; Charlton, M. (2002) *Quantitative Geography : Perspectives on Spatial Analysis.* SAGE Publications.

Gaston, K.J. & Evans, K.L. (2004) Birds and people in Europe. *Proceedings of the royal of London series B-Biological Sciences,* 271: 1649-1655.

Gotelli, N.J. & McGill, B.J. (2006) Null versus neutral models: what's the difference? *Ecography,* 29: 793-800.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology,* 84 (11): 2809-2815.

Haining, R. (1990) *Spatial data analisys in the social and environmental sciences*. Cambridge University Press, Cambridge.

Hawkins, B.A.;Porter, E.E.; Diniz-Filho, J.A.F. (2003a) Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology,* 84: 1608-1623.

Hawkins, B.A.; Field, R.; Cornell, H.V.; Currie, D.J.; Guégan, J.F.; Kaufman, D.M.; Kerr, J.T.; Mittelbach, G.G.; Oberdorff, T.; O'Brien, E.M.; Porter, E.E.; Turner, J.R.G. (2003b) Energy, water, and broad-scale geographic patterns of species richness. *Ecology,* 84 (12): 3105-3117.

Hawkins, B.A.; Diniz-Filho, J.A.F.; Bini, L.M.; De Marco, P.; Blackburn, T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography,* 30 (3): 375-384.

Hobbs, N.T. & Hilborn, R. (2006) Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications*, 16 (1): 5-19.

Hulbert, A.H. & Haskell, J.P. (2003) The effect of energy and seasonality on avian species richness and community composition. *American Naturalist,* 161: 83-97.

Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (10): 5661-5666.

Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, 19 (2): 101-108.

Keitt, T.H.; Bjφrnstad, O.N.; Dixon, P.M.; Pousty, S.C. (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography,* 25: 616-625.

Kerr, J.T. & Packer, L. (1997) Habitat heterogeneity as a determinant of mammals species richness in high-energy regions. *Nature,* 385: 252-254.

Klink, C.A. & Machado, R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology,* 19 (3): 707-713

Lichstein, J.W.; Simons, T.R.; Shriner, S.A.; Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs,* 72 (3): 445-463.

Liebhold, A.M. & Gurevitch J. (2002) Integrating the statistical analysisof spatial data in ecology. *Ecography,* 25: 553-557.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659-1673.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. Elsevier Science.

Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – "predictive" and "explanatory" models. *Biodiversity and Conservation,* 9: 655-671.

Mazerolle, M.J. (2006) Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses. *Amphibia-Reptilia*, 27: 169-180.

Myers, N.; Mittermeier R.A.; Mittermeier C.G.; Fonseca G.A.B.; Kent J. (2000) Biodiversity hotspots for conservation priorites. *Nature,* 43: 853-858.

Oliveira, P.S. & Marquis, R.J.; *The cerrados of Brazil: ecology and natural history of a neotropical savanna. Columbia University Press*.

Philippi, T.E. (1993) Multiple regression: herbivory. *Design and analysis of ecological experiments*: 183-210. Chapman & Hall, New York.

Pianka. E.R. (1994). *Evolutionary Ecology.* HarperCollins College Publishers.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences,* USA, 98: 4534-4539.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2005) Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. *Ecology Letters,* 8: 783-790.

Rangel, T.F.L.V.B.; Diniz-Filho, J.A.F.; Bini, L.M. (2006) Towards an integrated computacional tool for spatial analysis in macroecology and biogeography. *Global Ecology & Biogeography,* 15: 321-327.

Richards, S.A. (2005) Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology*, 86 (10): 2805-2814.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research.* W.H. Freeman and Company.

Stephens, P.A.; Buskirk, S.W.; Hayward, G.D.; Del Rio, C.M. (2005) Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology,* 42: 4-12.

Svenning, J.; Engelbrecht, B.M.J.; Kinner, D.A.; Kursar, T.A.; Stallard, R.F.; Wright, S.J. (2006) The relative roles of environment, history and local dispersal in controlling the distributions of common tree and shrub species in a tropical forest landscape, Panama. *Journal of Tropical Ecology,* 22: 575-586.

Togneli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, 27: 427-436.

Whittingham, M.J.; Stephens, P.A.; Bradbury, R.B.; Freckleton, R.P. (2006) Why do we still use stepwise modeling in ecology and behaviour? *Journal of Animal Ecology,* 75: 1182-1189.

Williams, B.K.; Nichols, J.D.; Conroy, M.J. (2002) *Analysis and Management of Animal Populations – Modeling, Estimation and Decision Making.* Academic Press.

Willig, M.R.; Kaufman, D.M.; Stevens, R.D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Syst.,* 34: 273-309.

Willis, K.J. & Whittaker, R.J. (2002). Species Diversity-Scale Matters. *Science,* 295: 1245-1248.

Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. *Oikos,* 41: 496-506.

Zar, J. H. 1999. *Biostatistical Analysis*. Fourth Edition. Prentice Hall, Upper Saddle River.

CAPÍTULO II

SELEÇÃO DE MODELOS ESPACIAIS PARA A DIVERSIDADE DE VERTEBRADOS NO CERRADO ATRAVÉS DO CRITÉRIO DE AKAIKE (AIC)

Resumo

No estudo da riqueza de espécies em grandes extensões territoriais devem-se levar em consideração os possíveis efeitos da autocorrelação espacial, que traz resultados incorretos sobre a estimativa do efeito de variáveis descritoras sobre a riqueza de espécies. Nesses casos, utilizam-se estatísticas que considerem este efeito espacial de modo que as estimativas dos parâmetros dos modelos que explicam a riqueza de espécies serão mais corretas. Neste trabalho, foram avaliados os padrões de riqueza para quatro grupos taxonômicos de vertebrados presentes no bioma Cerrado, em termos de estabelecer quais variáveis melhor associam-se com a distribuição destas espécies. Para isto, utilizou-se um modelo espacial autoregressivo simultâneo (SAR) e um modelo não-espacial (OLS), e em cada um deles as variáveis foram escolhidas utilizando-se a abordagem da seleção de modelos através do Critério de Akaike. As variáveis mais importantes foram relacionadas à produtividade. O Critério de Akaike selecionou os modelos que possuem maiores poderes de explicação. Os resultados observados mostraram bom desempenho do SAR em comparação ao OLS e permitiu estabelecer quais variáveis poderiam descrever os padrões espaciais de riqueza de espécies nos quatro grupos de vertebrados no Cerrado.

Palavras-chave: regressão múltipla (OLS), modelos espaciais autoregressivos, Critério de Akaike (AIC), Cerrado.

Introdução

A predição da riqueza de espécies, a fim de buscar os fatores ecológicos e evolutivos envolvidos na biodiversidade, pode ser realizada através de modelos regressivos nos quais o número de espécies está em função das variáveis que acreditam-se sejam as mais importantes em influenciar a distribuição de espécies sobre grandes extensões territoriais. Como podem ser gerados muitos modelos diferentes através de múltiplas combinações de variáveis, é comum avaliar, através de testes estatísticos de hipótese nula sobre as significâncias das variáveis, se estes descritores possuem influência sobre a distribuição de espécies de uma região. Isto é comumente realizado através da seleção de variáveis "*stepwise selection*", na qual ao longo do processo seletivo as variáveis estão sendo adicionadas ou retiradas de acordo com sua significância (Zar, 1996). Entretanto, isso não significa necessariamente um aumento do poder preditivo dos modelos considerados. Para a seleção de modelos, portanto, buscam-se as soluções com os maiores poderes de explicação sobre a distribuição de espécies e bem estruturados com a teoria subjacente aos padrões estudados (Mac Nally, 2000).

Neste sentido, alternativas que não lidam com testes de hipóteses, mas que realizam comparações entre os diferentes modelos à procura de resultados mais próximos da realidade, passaram a ser utilizadas. Com esses métodos, acredita-se que seja possível fazer melhores inferências e oferecer direções para a pesquisa ecológica (Johnson & Omland, 2004). Um procedimento que vem sendo bastante utilizado é o Critério de Akaike (AIC), que no caso de regressões apenas utiliza a variância residual e o número de parâmetros de cada modelo gerado

(Burnham & Anderson, 2004; Richards, 2005; Hoeting *et al.*, 2006; Svenning, *et al.*, 2006; Mazerolle, 2006; Vieira *et al.*, 2006; algumas aplicações em Kühn, 2007). Neste caso, no melhor modelo também pode haver a inclusão de variáveis não significativas pelo teste de hipóteses (ver Stephens *et al.,* 2005 e Hobbs & Hilborn, 2006 para uma revisão acerca de testes de nulidade contra a seleção de modelos baseado na teoria da informação).

Entretanto, a seleção de modelos, utilizando-se o critério de Akaike pode levar a falsas conclusões se na predição da riqueza de espécies não for considerada a autocorrelação espacial, caso esta exista e seja importante na estruturação dos resíduos dos modelos de regressão. Quando se consideram dados observacionais em grandes extensões territoriais, os pontos de coleta próximos (ou quadrículas) podem estar espacialmente autocorrelacionados, ou seja, os valores obtidos de pontos de coletas próximos entre si possuem valores mais similares do que esperado por acaso (Legendre, 1993). Quando isso ocorre, há uma violação dos pressupostos da estimativa dos erros dos coeficientes de regressão, porque cada amostra deve ser independente uma da outra e na autocorrelação há dependência entre elas. Deste modo, o tamanho amostral é sobrestimado e as inferências estatísticas serão incorretas, levando a falsas inferências (Diniz-Filho *et al.*, 2003).

A autocorrelação espacial é comumente avaliada nos valores residuais dos modelos regressivos gerados, ou seja, pontos de coleta próximos com estes valores similares entre si indicam a presença deste efeito espacial. Se a autocorrelação for significativa, utilizam-se normalmente modelos espaciais que considerem este efeito em seus cálculos. Recentemente, diversos autores

avaliaram como a autocorrelação espacial poderia alterar as conclusões de uma análise e sua interpretação sobre os dados ecológicos reais, e corrigiram os resultados através de estatísticas espaciais (Keitt *et al.*, 2002; Lichstein *et al.*, 2002; Fotheringham *et al.*, 2003; Diniz-Filho *et al.*, 2003; Tognelli & Kelt, 2004; Svenning *et al.*, 2006; Hawkins *et al.*, 2007; Kühn, 2007). Em princípio, este reconhecimento da autocorrelação espacial nos cálculos da predição pode corrigir os resultados da seleção de modelos trazendo resultados mais compatíveis com os dados reais.

Há diversas maneiras de incorporar o efeito do espaço geográfico nos modelos de regressão. Um dos mais utilizados em macroecologia e ecologia geográfica é a autoregressão simultânea (*SAR – Simultaneous Autoregressive Models*), que inclui a autocorrelação espacial no cálculo da matriz de covariância do resíduo (Haining, 1990). Isto pode trazer melhores resultados do que o OLS e também o Critério de Akaike irá selecionar estes modelos espaciais como a melhor abordagem para a distribuição de espécies.

Através destas metodologias estatísticas, o objetivo deste trabalho é utilizar a abordagem da seleção de modelos através do critério de Akaike para avaliar quais modelos possuem os melhores descritores da distribuição de vertebrados (anuros, répteis, aves e mamíferos) conhecidos do bioma Cerrado. Para isto, serão utilizadas as variáveis ambientais que melhor predizem os padrões de riqueza atuais (Hawkins *et al.* 2003a,b, 2004) e também parâmetros sociais relativos ao tamanho populacional humano e o número de inventários faunísticos deste bioma (Diniz-Filho *et al.*, 2005)

Metodologia

Dados ecológicos

A área a ser considerada é o bioma Cerrado, em toda sua extensão sobre o território brasileiro, que foi dividido em 181 quadrículas de 1º de latitude por 1º de longitude (110 Km x 110 Km) (Figura 1). Cada quadrícula é considerada como um ponto de amostragem e suas coordenadas geográficas são definidas pelo seu centro geográfico. Os dados referentes à distribuição das espécies no Cerrado e para a construção de mapas são de 138 mamíferos, 751 aves, 193 répteis e 131 anfíbios (Diniz-Filho *et al.*, 2006 *in press*). Verifica-se a extensão de uma espécie calculando em quantas quadrículas ela ocorre e quanto maior a sobreposição de espécies sobre uma quadrícula, maior será sua diversidade.

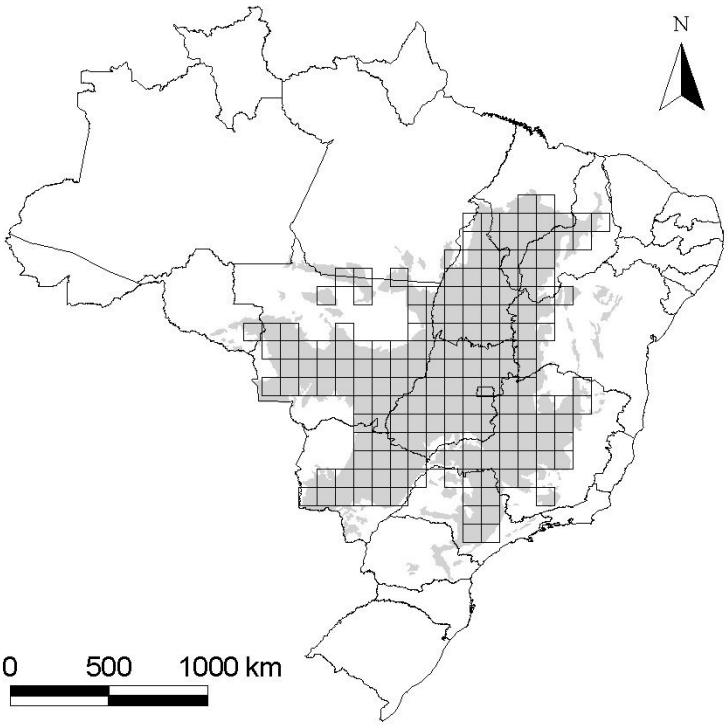


Figura 1: Mapa do bioma Cerrado e sua divisão em quadrículas sobre o território brasileiro.

As variáveis mais utilizadas para predições em grandes escalas podem ser classificadas dentro de hipóteses das quais as mais comuns são: energia ambiental (evapotranspiração potencial – *PET*, temperatura média anual - *TEMP*), produtividade (evapotranspiração atual – *AET*, índice de vegetação - *EVI*), heterogeneidade ambiental (topografia - *TOPOG*) condições climáticas (umidade relativa do ar – *HUMREL*) (Currie, 1991; Hawkins *et al.*, 2003a,b). Estes autores e diversos outros (e.g., Rahbek & Graves, 2001; Diniz-Filho *et al.*, 2003; Tognelli & Kelt, 2004; Hawkins *et al.*, 2005) já relacionaram essas hipóteses com a

distribuição da riqueza de espécies, em diferentes escalas espaciais, tanto em ambientes temperados quanto em tropicais, e verificaram que para as diferentes regiões consideradas há maior influência de uma hipótese sobre outra. Por exemplo, Hawkins *et al.,* (2003b) fez uma revisão de diversos estudos, em grandes escalas, para verificar quais variáveis são as mais importantes em descrever a distribuição de plantas, invertebrados e vertebrados. Para todos estes grupos, as variáveis relacionadas à energia e à entrada de água, bem como sua interação, foram consideradas as que melhor explicaram os padrões de riqueza.

Além das hipóteses já conhecidas, serão utilizadas neste estudo variáveis relacionadas à ocupação humana, incluindo o logaritmo do número de habitantes em cada célula (*HUMAN*) (IBGE, 2000) e ao número de inventários faunísticos (*KNOW*), este último relativo ao conhecimento da biodiversidade (Cavalcanti & Joly, 2002). Há atualmente uma grande tendência em utilizar estas variáveis em modelos de riqueza de espécies porque há um grande interesse em predições, questões de conflito entre conservação e desenvolvimento humano e também devido ao impacto que a sua presença causa em ambientes, tanto em pequena como em grande escala (Balmford *et al.*, 2001; Evans & Gaston, 2005). Diversos artigos revelam haver uma correlação positiva entre riqueza de espécies e a densidade populacional humana devida à hipótese de que os mesmos processos que atuam em aumentar a riqueza de espécies, principalmente a idéia da energia ambiental, também influenciam em maior abundância humana devido à procura de locais com mais disponibilidade de recursos e com maior produtividade, por exemplo, a agrícola (Balmford *et al.*, 2001; Araújo, 2003; Gaston & Evans, 2004; Diniz-Filho *et al.,* 2006).

Modelos de Regressão

Como o Cerrado possui grandes extensões territoriais, ocupando aproximadamente 22% do território brasileiro (Oliveira & Marquis, 2002), é necessário considerar os possíveis efeitos espaciais, a autocorrelação espacial (Legendre, 1993).

A magnitude da autocorrelação, para cada variável, é freqüentemente calculada pelo índice I de Moran e facilmente visualizada em correlogramas, nos quais em um sistema de coordenadas cartesianas o eixo x assume as distâncias entre as quadrículas e no eixo y os valores do índice I de Moran variam com o aumento das distâncias consideradas (Legendre & Legendre, 1998). Caso a variável não possua autocorrelação, os valores deste índice não irão variar no espaço geográfico. Verifica-se assim se as predições de riqueza de espécies dos modelos podem estar erradas através dos correlogramas dos resíduos da regressão destes modelos, avaliando-se particularmente se há autocorrelação positiva para as primeiras classes de distância são significativas. A significância de cada valor deste índice foi calculada através de testes de aleatorização, utilizando 1000 aleatorizações para o vetor dos resíduos e calculando-se o índice I de Moran, e o coeficiente testado ao nível de 5%.

Para corrigir estes efeitos espaciais, utilizam-se técnicas de modelagem que consideram o efeito de localidades vizinhas em descrever valores. A técnica OLS (*Ordinary Least Square*) descreve a riqueza de espécies apenas deixando que as variáveis explanatórias calculem a distribuição esperada das espécies como variável resposta (Fotheringham *et al.*, 2003). Por isso é preferível utilizar regressões que incluem o efeito da autocorrelação espacial em seus cálculos, já

que esta estatística espacial traz maiores poderes de explicação e resultados mais realísticos. Será utilizado neste estudo a ferramenta SAR (*Simultaneous Autoregressive Models*), que além de incluir as variáveis descritoras, também considera o efeito da estrutura espacial entre as quadrículas, considerando as distâncias geográficas ou uma rede de conexões entre os pontos de amostragem, ou seja, o quão próxima ou conectada está uma localidade da outra (Lichstein *et al.*, 2002).

Para este estudo, a estrutura espacial foi definida por uma rede de conexões entre as quadrículas, estabelecida utilizando-se o critério de Gabriel (Legendre, 1998). De acordo com esse critério, os locais que estão ligados recebem valor 1 e caso contrário, valor 0, sendo que a ligação se dá se o círculo cujo diâmetro é a distância entre os centros de duas quadrículas não inclui o centro de nenhuma outra quadrícula. Na regressão SAR, a autocorrelação espacial está incluída na matriz de covariância dos resíduos e por causa disto os coeficientes das variáveis explanatórias serão melhores estimados (Haining, 1990). Em um estudo recente, Kühn (2007) utilizou a regressão SAR para modelar a riqueza de espécies de plantas na Alemanha e constatou um melhor desempenho do modelo SAR em relação ao modelo OLS.

Outro problema que pode afetar descrições em regressões é o problema da multicolinearidade, no qual duas ou mais variáveis se correlacionam entre si. Isto pode levar a resultados errados porque, na presença de correlação entre variáveis, ocorrem desvios dos valores dos coeficientes de regressão que seriam esperados na ausência de multicolinearidade. Conseqüentemente, o erro padrão

dos coeficientes de regressão aumenta linearmente com o aumento da relação de uma variável com outra (Graham, 2003). Um modo detectar este problema é através do fator de aumento da variância (*Variance Inflation Factor* – VIF), que indica o quanto a variância dos coeficientes de regressão foi aumentada comparado quando não há multicolinearidade (Philip, 1993). Caso este problema seja bastante severo, deve-se utilizar técnicas que reduzam as correlações entre variáveis (Graham, 2003).

Seleção de Modelos e o Critério de Akaike

O AIC é uma ferramenta de seleção de modelos baseado na teoria da informação, sendo entendida como a distância entre a realidade e um modelo (Burnahm & Anderson, 2004). Diversos modelos são testados contra o conjunto de dados reais e deste modo estima-se a chance de um modelo ser novamente escolhido como o melhor caso o estudo seja repetido diversas vezes. Para que isto ocorra, cada hipótese, ou modelo, deve possuir inicialmente um valor de AIC, que para os modelos regressivos calcula-se da seguinte forma:

$$AIC = n\ln\left(\hat{\sigma}^2\right) + 2K \qquad \hat{\sigma}^2 = \sum\left(\hat{\varepsilon}_i\right)^2 \Big/ n$$

nas quais os valores $\hat{\varepsilon}_i$ referem-se aos resíduos de cada modelo, *n* é o tamanho amostral e K é o número de parâmetros considerados, ou seja, a quantidade de variáveis mais o intercepto da função e a variância residual. Para cada hipótese, é possível calcular o valor deste critério, embora individualmente este valor não

signifique nada, exceto quando comparado com os valores dos outros modelos provenientes de um conjunto de modelos que poderiam estar explicando um processo ecológico de interesse. Por isso utiliza-se a seguinte equação:

$$\Delta_i = AIC_i - AIC_{MIN}$$

que é a diferença entre o valor de Akaike de um modelo *i* qualquer com o menor valor de AIC obtido. Quanto menor for esta diferença, indica que outros modelos aproximam-se da condição de melhores modelos. Para obter a chance de um modelo ser escolhido novamente como o melhor, calcula-se a seguinte equação para cada modelo:

$$w_i = \frac{\exp\left(-\Delta_i / 2\right)}{\sum_{r=1}^{R} \exp\left(-\Delta_r / 2\right)}$$

na qual Δ refere à diferença entre os valores do Critério de Akaike e *R* seria o conjunto dos possíveis modelos que poderiam explicar a distribuição de aves, por exemplo.

Todos estes cálculos foram utilizados para avaliar o poder de evidência de cada modelo e permitem fazer comparações entre eles. Neste caso, testes de hipóteses não são utilizados, mas mesmo assim os modelos serão comparados uns com os outros para avaliar a performance de cada modelo.

Após a análise inicial dos pressupostos (ver abaixo), foram construídos os modelos OLS e SAR pela combinação de variáveis, para cada grupo taxonômico, que melhor poderiam estar contribuindo para a distribuição das espécies. Este

conjunto de modelos disponíveis para anuros, répteis, aves e mamíferos, foi classificado pelo critério de Akaike.

Avaliação de Pressupostos

Antes de serem obtidos os modelos descritos acima e sua avaliação pelo AIC, é necessário verificar os pressupostos da regressão quanto à normalidade, pela estatística Kolmogorov-Smirnov, a homogeneidade através do teste Scheffé-Box (Sokal & Rohlf, 1995) e independência dos resíduos estimados. A variável resposta foi transformada (transformação Freeman-Tukey) para verificar se houve uma melhora do coeficiente de assimetria e curtosis.

O pressuposto de independência é usualmente avaliado pela autocorrelação espacial dos resíduos a partir de correlogramas nas quais os valores do índice I de Moran são plotados contra as distâncias geográficas. Estes serão testados nas melhores abordagens OLS e SAR adiante, na qual obtém-se os valores residuais destas regressões.

A multicolinearidade foi avaliada pelo fator de aumento da variância e caso indique severa correlação entre variáveis, devem ser tomadas precauções para evitar problemas nos resultados da regressão.

Com a obtenção dos modelos OLS e SAR, foram realizadas avaliações das diferenças entre os resultados provenientes destas duas formas de regressão

quanto à autocorrelação espacial, poder de explicação pelo critério de Akaike, seleção de variáveis e grau de multicolinearidade.

Todas as análises foram realizadas no software SAM – *Spatial Analysis in Macroecology* (Rangel *et al.*, 2006).

Resultados

O pressuposto de normalidade da distribuição dos resíduos não foi atendido para todos os modelos, a transformação da variável dependente pelo método de Freeman-Tukey não causou melhora na normalidade. Para os melhores modelos OLS, apenas os resíduos dos modelos para répteis não apresentaram uma distribuição normal, ao passo que na modelagem SAR os répteis e mamíferos apresentaram valores dos resíduos normalmente distribuídos. Já para a homogeneidade das variâncias, dentro de cada grupo as variâncias foram homogêneas.

A partir dos valores da Tabela 2 verifica-se que há diferenças entre os melhores modelos selecionados utilizando regressões OLS ou SAR. As variáveis selecionadas são distintas e os coeficientes de determinação dos modelos SAR são maiores. Os modelos SAR apresentam menores valores de AIC, não implícito na Tabela 2, e isto significa que o ajuste foi melhor para este tipo de modelagem do que a OLS. Esta diferença de valores do critério de Akaike do melhor modelo SAR para o melhor OLS pode ser observado na Tabela 1.

Tabela 1: Diferença do valor AIC entre o melhor modelo SAR para o OLS em cada grupo taxonômico, de acordo com a Tabela 2.

Grupo	Δ AIC
Anuros	175,554
Répteis	156,393
Aves	557,320
Mamíferos	193,399

Esta diferença de desempenho entre as duas técnicas de regressão é evidente se também forem considerados os correlogramas, que permitem analisar a autocorrelação espacial dos resíduos (ou a dependência destes), considerando apenas do melhor modelo OLS e SAR de cada grupo (Figura 2).

Tabela 2: Resultados da seleção de modelos OLS e SAR para os diferentes grupos taxonômicos, destacando os coeficientes de determinação (R^2), o delta-AIC, o poder de evidência de cada modelo (W) e as variáveis mais importantes.

	OLS			
Grupo	R^2	ΔAIC	W	Variáveis
Anuros	0,776	0	0,364	*TOPOG, TEMP, PET, AET, HUMREL*
	0,778	0,44	0,292	*TOPOG, TEMP, PET, AET, EVI, HUMREL*
	0,779	1,865	0,143	*TOPOG, TEMP, PET, EVI, HUMREL, AET, HUMAN*
	0,779	3,792	0,054	*MODELO GLOBAL*
Répteis	0,43	0	0,303	*AET, PET, HUMREL, TEMP, EVI, TOPOG*
	0,435	0,627	0,221	*AET, PET, HUMREL, TEMP, EVI, KNOW, TOPOG*
	0,44	1,273	0,16	*MODELO GLOBAL*
	0,425	1,793	0,123	*AET, PET, TEMP, EVI, TOPOG, KNOW*
Aves	0,828	0	0,45	*MODELO GLOBAL*
	0,822	1,236	0,242	*PET, TEMP, TOPOG, AET, HUMREL, EVI*
	0,823	2,823	0,109	*PET, TEMP, TOPOG, AET, HUMAN, HUMREL, EVI*
	0,818	3,186	0,091	*PET, TEMP, AET, HUMREL, EVI*
Mamíferos	0,601	0	0,289	*PET, TOPOG, TEMP, AET, HUMREL, EVI*
	0,596	0,233	0,257	*PET, TOPOG, TEMP, AET, HUMREL*
	0,602	1,962	0,108	*PET, TOPOG, TEMP, AET, HUMREL, HUMAN, EVI*
	0,596	2,288	0,092	*PET, TOPOG, TEMP, AET, HUMREL, HUMAN*

Continuação da Tabela 2.

	SAR			
Grupo	R^2	ΔAIC	W	Variáveis
Anuros	0,945	0	0,662	*PET, AET*
	0,945	1,347	0,337	*TEMP, TOPOG, AET*
	0,938	22,196	$1{,}00 \times 10^{-5}$	*TEMP, AET, HUMREL*
	0,937	31,46	9.76×10^{-8}	*TEMP, TOPOG, PET, HUMREL, AET*
Répteis	0,807	0	0,211	*TOPOG, HUMAN, EVI, AET, PET*
	0,809	0,316	0,180	*TOPOG, HUMAN, KNOW, EVI, PET, AET*
	0,799	1,475	0,101	*TOPOG, KNOW*
	0,798	2,373	0,061	*TOPOG, HUMAN*
Aves	0,992	0	0,881	*AET, TEMP, EVI, PET*
	0,991	4,113	0,112	*AET, TEMP, PET*
	0,991	10,578	4×10^{-3}	*AET, TEMP*
	0,991	13,787	8×10^{-4}	*AET, TEMP, KNOW, HUMAN, EVI*
Mamíferos	0,851	0	0,455	*AET, PET, HUMREL, TEMP*
	0,861	0,675	0,324	*AET, PET, HUMREL, TEMP, EVI*
	0,856	1,898	0,176	*AET, PET, TEMP*
	0,859	5,502	0,029	*AET, HUMREL, TEMP, EVI, TOPOG, PET*

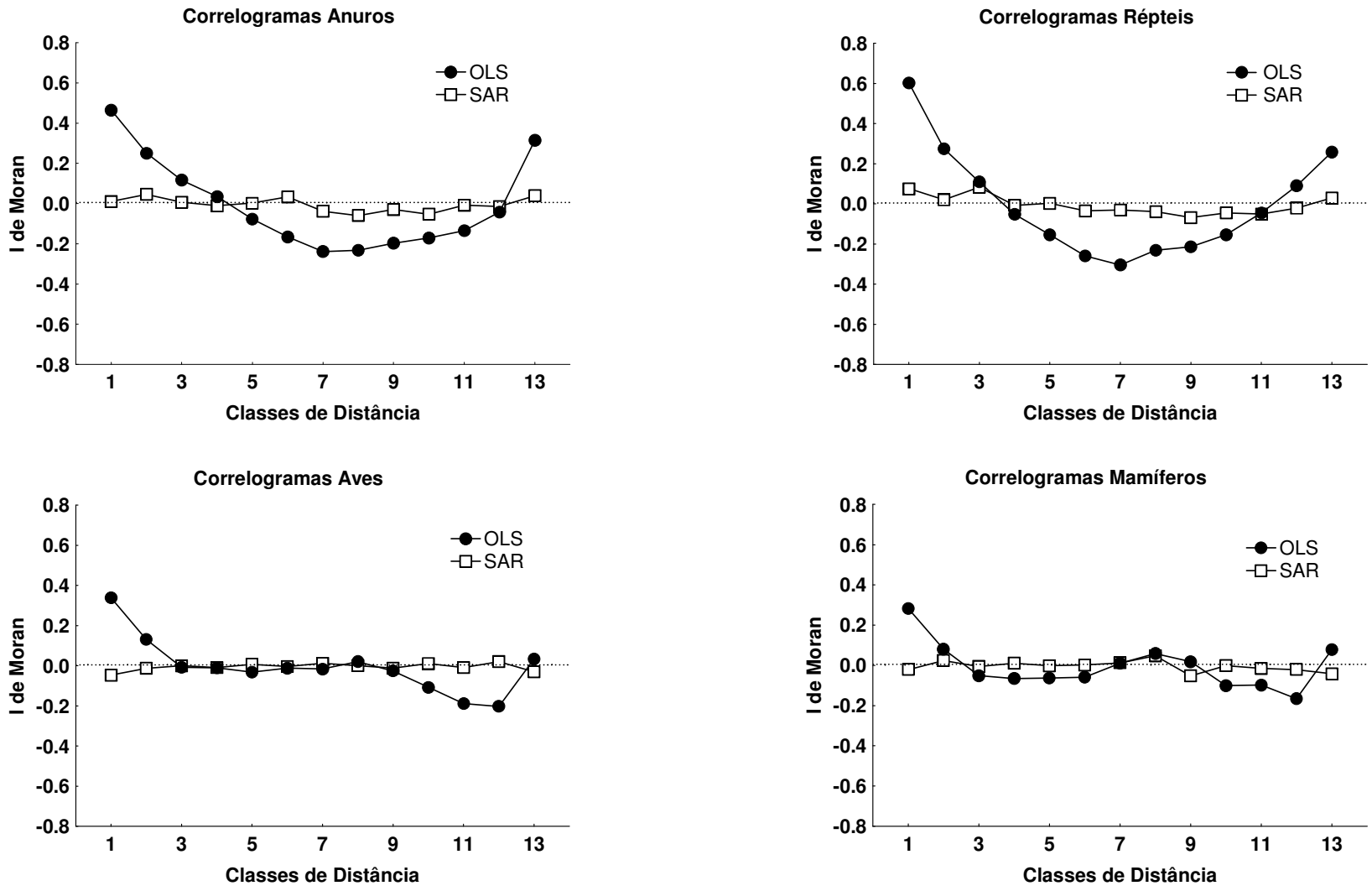


Figura 2: Correlogramas espaciais dos resíduos dos melhores modelos OLS e SAR, para cada grupo taxonômico.

As classes de distâncias referem-se às distâncias geográficas para o cálculo da autocorrelação espacial. As análises dos correlogramas dos resíduos dos modelos OLS mostraram que as primeiras classes de distância tinham valores de autocorrelações espaciais positivas e significativas (verificados através de testes de randomização, com $P < 0,05$), indicando que quadrículas mais próximas são mais similares entre si, para os valores dos resíduos da regressão, do que esperado simplesmente ao acaso, mesmo depois de levar em consideração os gradientes ambientais. Em longas distâncias a autocorrelação torna-se novamente positiva e significativa indicando valores residuais similares mais do que esperado ao acaso. Quando se utilizaram os resíduos provenientes da modelagem SAR, o correlograma apresentou um menor grau de autocorrelação espacial nas primeiras classes de distância e os I de Moran tendem a se manter próximos a zero em distâncias mais longas. Através de testes de randomização, para estes modelos SAR, verificou-se que estes valores para as primeiras classes de distância não são significativos, exceto para répteis. Isto revela o melhor desempenho da modelagem SAR para predições de riqueza de espécies.

Outra forma de avaliar a autocorrelação espacial dos resíduos é mapear estes valores sob as quadrículas do bioma Cerrado (Figura 3). Para o modelo OLS, há a formação de grupos de quadrículas próximos entre si com altos valores de resíduos, indicando que estas quadrículas agrupadas possuem valores similares devido à autocorrelação espacial e deste modo são mais dependentes entre si. Já no modelo SAR, não é evidente esta formação de grupos com altos valores de resíduos. Há grupos de quadrículas com valores residuais bastantes próximos a zero, e isto indica um melhor ajuste do modelo SAR porque os valores

de riqueza estimados, a partir deste modelo, são parecidos com os valores observados.

Quanto às variáveis selecionadas através do critério de Akaike, há diferenças dos modelos OLS e SAR para todos os grupos considerados (Tabela 2). Estes resultados, dispostos na Tabela 2, confirmam, em parte, os resultados esperados para as latitudes tropicais (Currie, 1991; Hawkins *et al.,* 2003b). Considerando apenas modelos SAR, para aves, as variáveis descritoras mais importantes foram relacionadas à dinâmica água-energia e produtividade (*AET* e *EVI*) e também à entrada de energia (*PET* e *TEMP*), embora estes tenham coeficientes negativos (Tabela 4). Os mamíferos foram similares às aves, exceto pelo fato de que a variável relacionada à entrada de água foi adicionada ao modelo, e não havendo efeito do índice de vegetação.

Em anuros e répteis, as variáveis relacionadas à entrada de energia possuem relações negativas com a riqueza, um contraste do que seria esperado pelo fato da energia ser uma importante variável para estes grupos devido a seus requerimentos de temperatura. Novamente, a dinâmica água-energia, expressa pela variável evapotranspiração atual, entraram como importantes descritores. A topografia, relacionada à heterogeneidade ambiental, e o tamanho populacional humano, entraram apenas no melhor modelo de répteis.

As variáveis selecionadas para explicar a riqueza em todos os grupos foram a evapotranspiração atual e potencial (*AET* e *PET*), que possuem relações distintas para a distribuição de espécies. Ambos são preditores reconhecidos para os biomas tropicais e fazem parte da hipótese geral de energia. A variável para inventários faunísticos (*KNOW*) não apareceu em nenhum melhor modelo SAR, ao

passo que nos modelos OLS foi selecionada apenas para explicar a riqueza de aves. Isto ocorreu devido a esta variável estar autocorrelacionada, porque os maiores índices desta variável encontra-se próximos a centros urbanos onde são feitos maiores esforços de amostragem. Esta variável não possui influência nos processos ecológicos, é apenas relacionada ao conhecimento da biodiversidade.

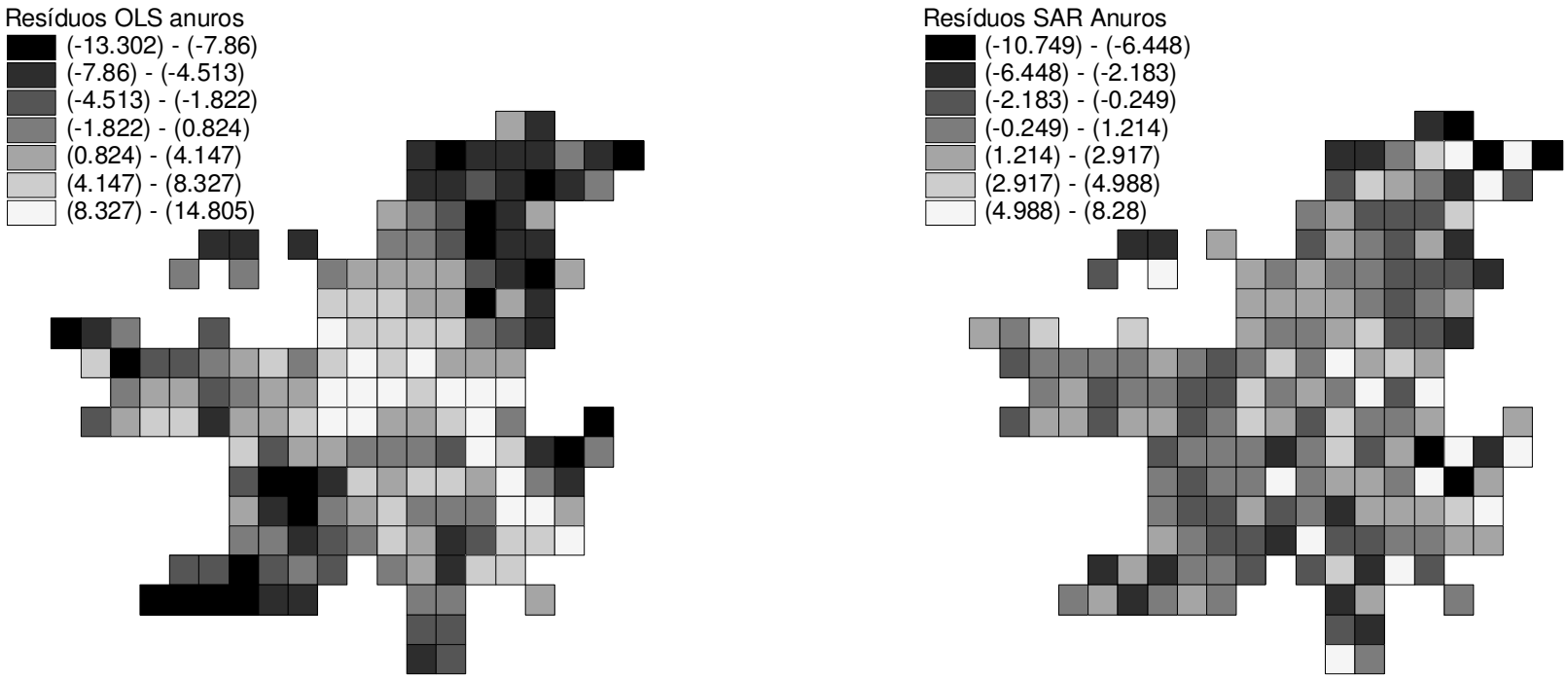
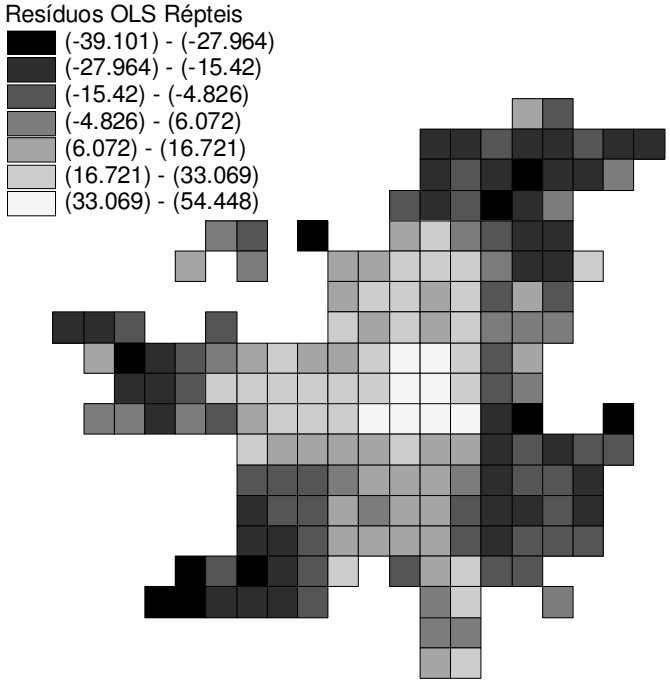
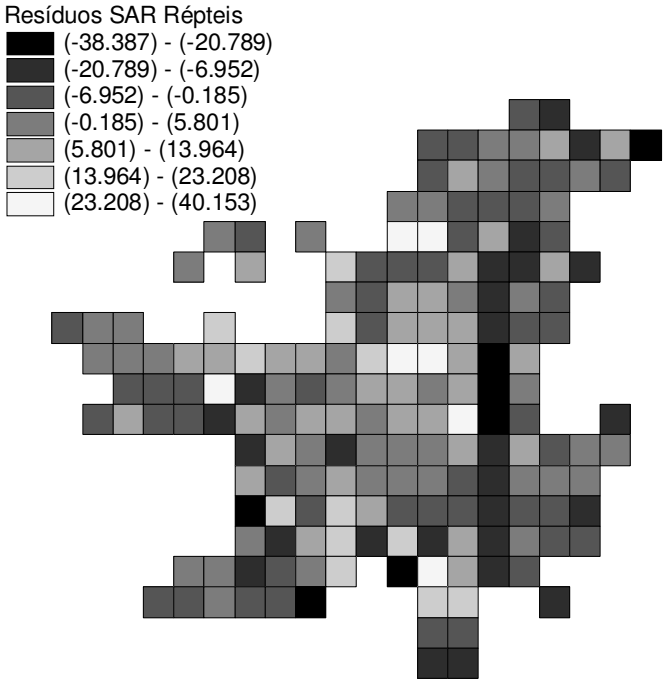
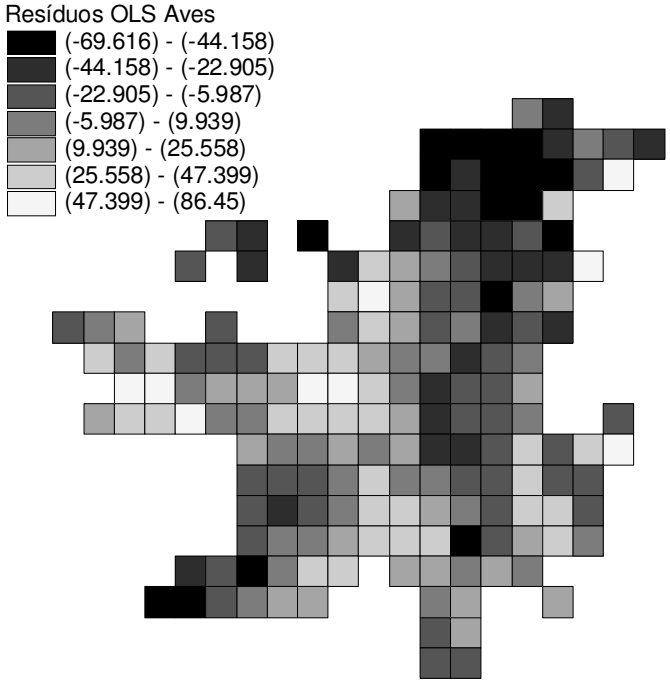


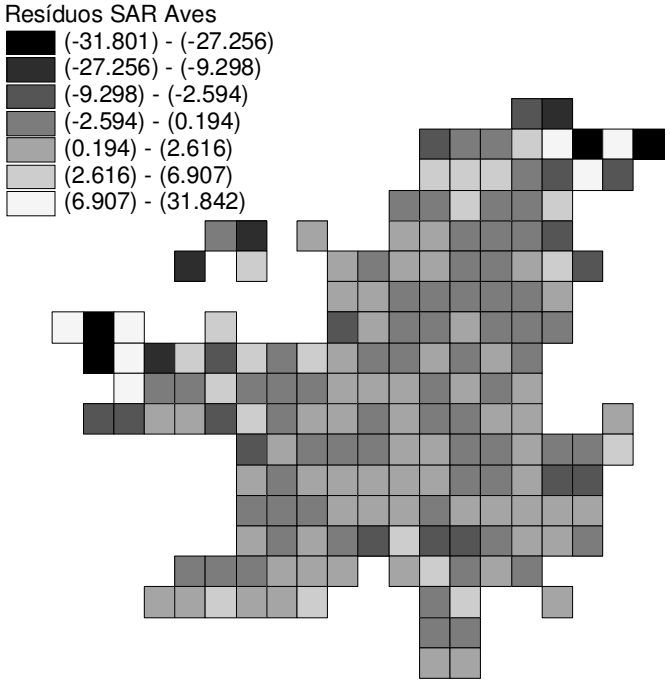
Figura 3: Distribuição espacial dos valores residuais sobre as quadrículas do bioma Cerrado, distinguindo entre os modelos OLS e SAR dos grupos taxonômicos considerados. Aqui os mapas para anuros.



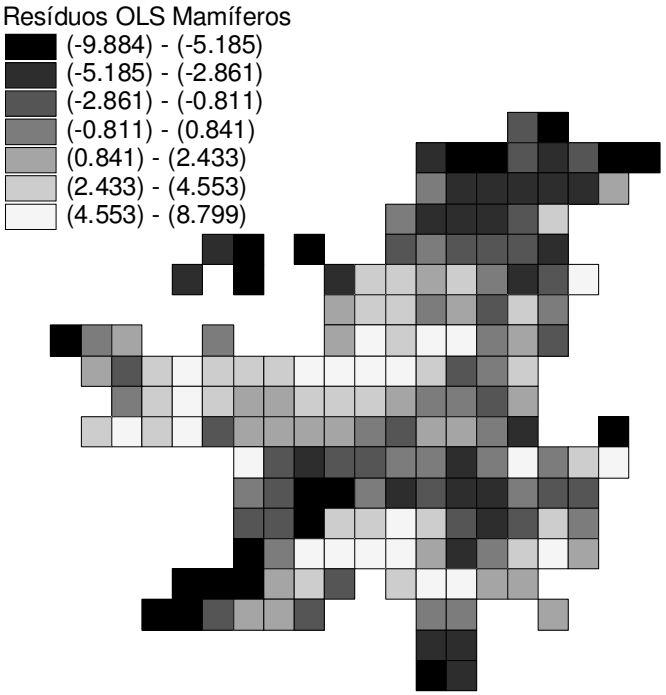


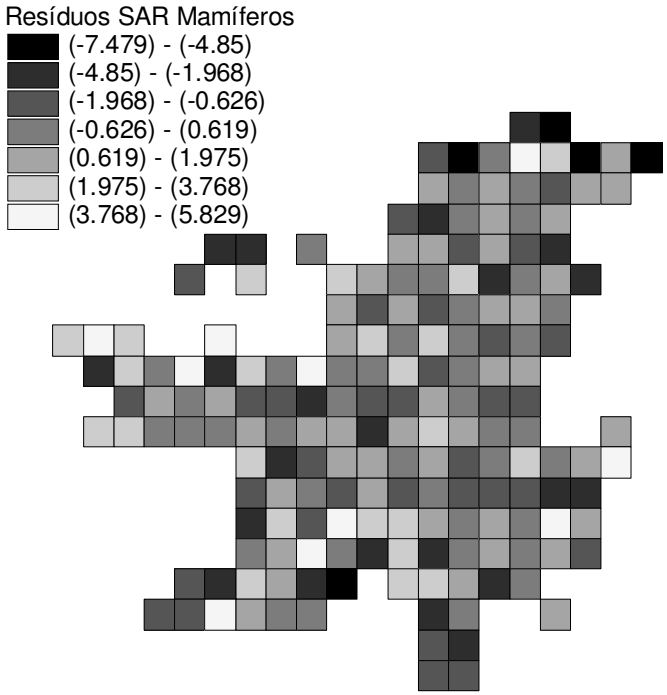
Continuação da Figura 3. Mapa para répteis.





Continuação da Figura 3. Mapa para aves.





Continuação da Figura 3. Mapas para mamíferos.

Se forem consideradas as correlações entre os descritores, verificou-se que as variáveis dentro de cada hipótese (energia ambiental, produtividade, condições climáticas e sociais) apresentam correlações positivas e significativas, embora também apareçam correlações das variáveis entre as hipóteses, e com isto surgem os problemas de multicolinearidade que podem afetar os resultados dos modelos. Através do fator de aumento da variância (*VIF*) detectou-se que os níveis de colinearidade entre os melhores modelos SAR foram menores do que entre os modelos OLS (Tabela 3). Isto ocorre porque, para o modelo OLS, o critério de Akaike selecionou mais variáveis, o que aumenta a probabilidade dos modelos conterem as variáveis que mais se correlacionam. Como os valores do fator de aumento da variância foram baixos, a descrição provavelmente não sofreu efeitos consideráveis.

Tabela 3: Valores do fator de aumento da variância para as variáveis que eram comuns aos melhores modelos OLS e SAR em cada grupo vertebrado.

	OLS				SAR			
Variáveis	Anuros	Répteis	Aves	Mamíferos	Anuros	Répteis	Aves	Mamíferos
AET	1,584	1,675	1,808	1,675	1,038	1,116	1,088	1,572
PET	__	3,745	3,921	3,745	__	2,283	3,322	3,412
HUMREL	__	__	__	2,415	__	__	__	1,582
TEMP	4,065	__	4,854	4,098	1,038	__	3,401	3,300
EVI	__	1,824	1,841	__	__	1,213	1,140	__
TOPOG	__	3,086	__	__	__	2,785	__	__

Tabela 4: Coeficientes de regressão dos melhores modelos OLS e SAR para os quatro grupos taxonômicos.

	OLS				SAR			
Variáveis	Anuros	Répteis	Aves	Mamíferos	Anuros	Répteis	Aves	Mamíferos
AET	0,342	1,301	5,092	0,343	0,098	0,166	2,01	0,148
PET	0,254	-1,41	-3,987	-0,33	-2,517	-0,095	-0,675	-0,051
HUMREL	-0,666	-1,15	-4,022	-0,267	__	__	__	0,057
TEMP	-2,001	13,894	-2,869	1,109	__	__	-11,074	-0,422
EVI	__	-61,773	122,956	8,21	__	-12,32	22,446	__
KNOW	__	__	-3,26	__	__	__	__	__
HUMAN	__	__	7,237	__	__	1,606	__	__
TOPOG	0,024	0,057	0,039	0,008	__	0,008	__	__

Discussão

Por meio dos resultados obtidos foi clara a melhor performance do modelo autoregressivo simultâneo (SAR) em relação à regressão múltipla não espacial (OLS). Em todos os quesitos avaliados, tais como a autocorrelação espacial, grau de multicolinearidade, seleção de variáveis e poder de explicação, os modelos SAR se apresentaram mais adequados para os dados de riqueza. Apenas no quesito da normalidade houve uma violação significativa para alguns modelos SAR, já que a maior parte dos resíduos possui distribuição leptocúrtica (ou seja, se concentraram ao redor de zero mais do que o esperado).

Quando se utiliza o modelo SAR, permite-se explicitamente que o efeito espacial seja controlado, já que nos resíduos do modelo a autocorrelação espacial está presente na matriz de covariância através do parâmetro autoregressivo ρ, que expressa a influência de localidades vizinhas em predizer valores (Lichstein *et al.*, 2002; Fotheringham *et al.*, 2003). Como o efeito espacial está presente nos dados reais, o modelo SAR está mais próximo das observações. As variáveis que são bastante autocorrelacionadas tendem a ter uma estimativa associada ao coeficiente mais fraca se o efeito espacial é controlado e deste modo os coeficientes são melhores calculados (Lichstein *et al.*, 2002). Os modelos OLS, por outro lado, muitas vezes violam o pressuposto da independência dos valores residuais e, deste modo, as variáveis que são altamente autocorrelacionadas, tendem a ter sua significância estatística estimada de maneira enviesada (Fotheringham *et al.*, 2003).

A autocorrelação espacial nos valores residuais foi detectada porque a distribuição geográfica de muitas espécies são maiores que o tamanho das quadrículas utilizadas e, deste modo, a riqueza entre quadrículas vizinhas é parecida. Entretanto, pode ser difícil descobrir as causas da autocorrelação se apenas a variável resposta for analisada. Isto ocorre porque autocorrelação positiva em pontos de coleta próximos podem ser causados por variáveis que poderiam explicar a variação da riqueza de espécies em locais vizinhos e que não foram medidas ou consideradas (Diniz-Filho *et al.*, 2003) A autocorrelação também pode ocorrer devido às variáveis utilizadas estarem autocorrelacionadas, ou seja, as quadrículas próximas são similares para uma outra variável.

Pela seleção dos melhores modelos utilizando-se o critério de Akaike, algumas variáveis perderam importância ao serem incorporadas nos modelos SAR, quando comparadas às analises tradicionais utilizando o OLS. Em anuros, foram retiradas as variáveis relacionadas à entrada de água, temperatura e heterogeneidade ambiental, destacando-se nos modelos SAR a dinâmica água-energia, com o efeito da entrada de energia. Este último, entretanto, apresentou coeficiente negativo. Os anfíbios possuem requerimentos de temperatura para o metabolismo (e.g., Allen *et al.*, 2002). Com os resultados do melhor modelo, a riqueza de anuros será maior em regiões mais frias. Isto pode ser explicado parcialmente por um problema da falta de dados para este grupo, porque a riqueza de anuros deveria ser maior na região noroeste do Cerrado, que possui maiores valores de produtividade e temperatura nestas regiões. Como o tamanho populacional humano é maior na região sudeste do Cerrado, o conhecimento

desta região através do número de inventários faunísticos é maior e deste modo mais espécies são conhecidas nesta região (Diniz-Filho *et al.*, 2005).

Para répteis ocorreu a mesma discordância que em anuros em relação ao coeficiente negativo para temperatura. Do modelo OLS para o SAR foram retiradas as variáveis relacionadas à entrada de água e a temperatura média anual, e acrescentado o efeito do tamanho populacional humano. O tamanho populacional humano reforça a idéia da medida de maior conhecimento faunístico e coincidente com a riqueza de espécies (Diniz-Filho *et al.*, 2005) embora não haja uma correlação elevada entre tamanho populacional humano e riqueza de répteis. Isso se dá principalmente porque, em termos de coincidência de padrões espaciais, a região sudeste do Brasil, que é mais densamente povoada, não possui maior riqueza de répteis. A topografia, outra variável selecionada para explicar a riqueza de espécies de répteis, é reconhecida como amplitude do relevo e quanto maior este valor dentro de uma quadrícula, maior seria a riqueza esperada (Kerr & Packer, 1997; Rahbek & Graves, 2001). No melhor modelo, entretanto, apresentou um coeficiente de regressão bastante baixo. Os resultados confirmam a falta de uma maior quantidade de dados para répteis visto que a maior riqueza observada encontra-se na região central do Cerrado, onde estão os maiores esforços para descrição deste grupo taxonômico (Diniz-Filho *et al.*, 2006 *in press*).

Para aves e mamíferos os resultados foram de acordo com estudos anteriores (Hawkins *et al.*, 2003a,b). Em ambos a variável mais importante, nos modelos SAR, foi a evapotranspiração atual (*AET*), o que sugere um importante efeito da dinâmica água-energia para a distribuição destes grupos. A única

diferença entre os melhores modelos de aves e mamíferos foi que, para aves, o índice de vegetação foi selecionado. Por outro lado, há um efeito da umidade relativa na riqueza de mamíferos, mas não na de aves. O índice de vegetação mede a cobertura vegetal, uma medida indireta da produtividade (Hulbert & Haskell, 2003), e seus maiores valores estão a nordeste no Cerrado e enquanto que a maior riqueza de aves encontra-se a sudeste. A umidade relaciona-se à entrada de água no sistema agindo como principal descritor da riqueza de plantas (Hawkins, 2003b) e isto pode aumentar a riqueza da cadeia trófica adiante.

Como nos grupos anteriores, em aves e mamíferos, também selecionaram-se mais variáveis utilizando o OLS do que o SAR. Isto torna-se favorável ao aparecimento da multicolinearidade, porque as variáveis que fazem parte das hipótese consideradas (produtividade, energia ambiental, condições climáticas e sociais) são correlacionadas entre si tanto positivamente ou negativamente. Tanto o coeficiente de regressão quanto o poder para detectar um efeito de uma variável diminuem de modo não-linear com o aumento da multicolinearidade (Graham, 2003). Contudo, se nos melhores modelos OLS e SAR forem retiradas as variáveis que se correlacionam, o valor do critério de Akaike aumenta e o poder de explicação do modelo diminui. Isto ocorre porque a contribuição da variável em explicar a distribuição de espécies está sendo omitida e mesmo as variáveis restantes no modelo podem ter coeficientes enviesados por motivos desconhecidos (Philip, 1993). No modelo SAR foram selecionados menos variáveis devido à autocorrelação espacial ser incluída na matriz de covariância dos resíduos e deste modo seu desempenho foi maior. Como os valores do fator de aumento de variância foram baixos, não há maiores preocupações relativos à

multicolinearidade, embora Graham (2003) cite que mesmo valores menores que dois da equação do fator de aumento da variância podem alterar os resultados. Um método que poderia ser utilizado para corrigir a multicolinearidade seria através de análises de componentes principais, na qual cada eixo obtido é independente um do outro, mas a principal limitação deste método é a dificuldade de interpretação biológica dos eixos multidimensionais gerados e também a perda ou generalização de informações dos dados observados (Graham, 2003).

Conclusão

O critério de Akaike como ferramenta de seleção de modelos selecionou os principais modelos de cada grupo taxonômico utilizando a variância residual, o tamanho amostral e o número de parâmetros da regressão considerados, que ao contrário da *stepwise regression*, selecionam-se apenas as variáveis significativas. As variáveis explanatórias mais importantes confirmaram, em parte, o que era esperado para explicar os padrões de riqueza do bioma Cerrado, visto a importância da dinâmica água-energia como variável relacionada à produtividade. Os resultados contrários ao esperados, pela literatura, para anfíbios e répteis devem ser melhor explicados devido à falta de dados para estes grupos, ao norte do bioma Cerrado onde os maiores índices de temperatura estariam de acordo com os requerimentos fisiológicos, do que o próprio critério de seleção, por isto deve haver maior busca de dados de campo para estes grupos.

Os resultados observados quando se considera a autocorrelação espacial, através do modelo autoregressivo, e a presença da multicolinearidade, demonstraram um aumento de desempenho e que o critério de Akaike conseguiu avaliar esta diferença em relação aos modelos não espaciais.

Bibliografia

Allen, A.P.; Brown, J.H.; Gyllooly, J.F. (2002). Global biodiversity, biochemical kinetics and the energetic equivalence rule. *Science* 297: 1545-1548.

Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography* 12: 5-12.

Balmford A, JL Moore, T Brooks, N Burgess, LA Hansen, P Williams, C Rahbek. 2001. Conservation conflicts across Africa. *Science* 291: 2616-2619.

Burnham, K.P. & Anderson, D.R. (2002). Model selection and multimodel inference. Springer, NY

Cavalcanti, R.B. & Joly, C.A. (2002) Biodiversity and conservation priorities in the Cerrado. *The Cerrado of Brasil: Ecology and Natural History of a Neotropical Savanna*: 351-367. Columbia University Press, New York.

Currie, D.J. (1991) Energy and large-scale patterns of animal- and plant- species richness. *The American Naturalist* 137 (1): 27-49.

Diniz-Filho, J.A.F.; Bini, L.M.; Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography* 12: 53-64.

Diniz-Filho JAF, RP Bastos, TFLVB Rangel, LM Bini, P Carvalho, RJ Silva. 2005. Macroecological correlates and spatial patterns of anurans description dates in Brazilian Cerrado. Global Ecol. Biogeogr. 14: 469-477.

Diniz-Filho, J.A.F.; Bini, L.M.; Vieira, C.M.; Blamires, D.; Terribile, L.C.; Bastos, R.P.; Oliveira. G.; Barreto, B. (2006) Spatial patterns of species richness in terrestrial vertebrates of brazilian Cerrado. *Zoological Studies* (*in press)*.

Evans, K.L. & Gaston, K.J. (2005) People, energy and avian species richness. *Global Ecology and Biogeography* 14: 187-196.

Fotheringham, A.S.; Brunsdon, C.; Charlton, M. (2002) *Quantitative Geography* : *Perspectives on Spatial Analysis.* SAGE Publications.

Gaston KJ & KL Evans. 2004. Birds and people in Europe. Proceedings of the Royal Society of London, Series B- Biological Sciences 271: 1649-1655.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, 84 (11): 2809-2815.

Haining, R. (1990) Spatial data analisys in the social and environmental sciences. Cambridge University Press, Cambridge.

Hawkins, B.A.; Porter, E.E.; Diniz-Filho, J.A.F. (2003a). Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology* 84 (6): 1608-1623.

Hawkins BA, R Field, HV Cornell, DJ Currie, JF Guegan, DM Kaufman, JT Kerr, GG Mittelbach, T Obedorff, EM O'Brien, EE Porter, JRG Turner. (2003b). Energy, water, and broad-scale geographic patterns of species richness. Ecology 84: 3105-3117.

Hawkins, B.A.; Diniz-Filho, J.A.F.; Soeller, S.A. (2005) Water links the historical and contemporary components of the Australian bird diversity gradient. *Journal of Biogeography* 32: 1-8.

Hawkins, B.A.; Diniz-Filho, J.A.F.; Bini, L.M.; De Marco, P.; Blackburn, T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography* 30 (3): 375-384.

Hobbs, N.T. & Hilborn, R. (2006) Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications*, 16 (1): 5-19.

Hoeting, J.A.; Davis, R.A.; Merton, A.A.; Thompson, S.A. (2006) Model selection for geostatistical models. *Ecological Applications*, 16 (1): 87-98.

Hulbert, A.H. & Haskell, J.P. (2003) The effect of energy and seasonality on avian species richness and community composition. *The American Naturalist* 161: 83-97.

IBGE (Instituto Brasileiro de Geografia e Estatística). 2000. Censo demográfico 2000: resultados preliminares. Rio de Janeiro: IBGE

Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, 19 (2): 101-108.

Keitt, T.H.; Bjφrnstad, O.N.; Dixon, P.M.; Pousty, S.C. (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography* 25: 616-625.

Kerr, J.T. & Packer, L. (1997) Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature* 385: 252-254.

Kühn, I. (2007) Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13: 66-69.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659-1673.

Legendre P & L Legendre. (1998). Numerical ecology. Amsterdam: Elsevier.

Lichstein, J.W.; Simons, T.R.; Shriner, S.A.; Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72 (3): 445-463.

Liebhold, A.M. & Gurevitch J. (2002) Integrating the statistical analysisof spatial data in ecology. *Ecography,* 25: 553-557.

Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology : The distinction between – and reconciliation of – "predictive" and "explanatory" models. *Biodiversity and Conservation,* 9: 655-671.

Mazerolle, M.J. (2006) Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses. *Amphibia-Reptilia*, 27: 169-180.

Oliveira PS & RJ Marques. 2002. The Cerrado of Brazil: Ecology and Natural History of Neotropical Savanna. Columbia: Columbia University Press.

Philippi, T.E. (1993) Multiple regression: herbivory. *Design and analysis of ecological experiments*: 183-210. Chapman & Hall, New York.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences,* USA, 98: 4534-4539.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research.* W.H. Freeman and Company.

Stephens, P.A.; Buskirk, S.W.; Hayward, G.D.; Del Rio, C.M. (2005) Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology* 42: 4-12.

Svenning, J.; Engelbrecht, B.M.J.; Kinner, D.A.; Kursar, T.A.; Stallard, R.F.; Wright, S.J. (2006) The relative roles of environment, history and local dispersal in controlling the distributions of common tree and shrub species in a tropical forest landscape, Panama. *Journal of Tropical Ecology* 22: 575-586.

Togneli, M.F. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, 27: 427-436.

Whittingham, M.A.; Stephens, P.A.; Bradbury, R.B.; Freckleton, R.P. (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology,* 75: 1182-1189.

CAPÍTULO III

O EFEITO DO DOMÍNIO MÉDIO E PADRÕES ESPACIAIS DE ESPÉCIES ENDÊMICAS NO BIOMA CERRADO: GRADIENTES GEOGRÁFICOS OU MERO ACASO?

Resumo

Os principais fatores atualmente usados para explicar os padrões globais de riqueza de espécies, tais como a entrada de energia no sistema, a disponibilidade de área, o tempo evolutivo e a heterogeneidade ambiental, têm sido avaliados diversas vezes na literatura ecológica. Entretanto, recentemente desenvolveu-se a hipótese do efeito do domínio médio, que vem sendo alvo de diversas discussões. Esta hipótese prediz que uma alta sobreposição de espécies ocorrerá no centro de uma região sob estudo, assumindo que não há variação espacial ou temporal nos efeitos dos gradientes ambientais. Neste trabalho foram avaliados diversos modelos para explicar a riqueza de espécies endêmicas de vertebrados terrestres (anfíbios, répteis, aves e mamíferos) no Cerrado, levando-se em consideração os possíveis efeitos da autocorrelação espacial. O objetivo foi avaliar as variáveis que estariam estruturando o padrão de distribuição de espécies endêmicas deste bioma. Através da seleção de modelos pelo Critério de Akaike, o melhor modelo inclui as variáveis relacionadas à hipótese do efeito do domínio médio, de produtividade e energia ambiental. Deste modo, a riqueza de espécies endêmicas poderia ser explicada por um balanço entre processos estocásticos atuais, históricos e ecológicos.

Palavras-chave: efeito do domínio médio (MDE), espécies endêmicas, Cerrado.

Introdução

Há grandes interesses em desvendar os processos ecológicos e evolutivos que podem determinar os gradientes de riqueza de espécies conhecidos atualmente (Willig *et al.*, 2003; Hawkins *et al.*, 2003a). Essa avaliação vem a ampliar o conhecimento acerca das relações em grandes escalas de tempo e espaço para grandes grupos taxonômicos e que podem servir a propostas de conservação (Brown, 1995). Quando esses processos são colocados em modelos que incorporam o efeito de múltiplas variáveis, obtêm-se as associações das variáveis com os padrões de riqueza de espécies e deste modo pode-se tentar explicar os dados observados.

Dentre as hipóteses mais utilizadas para se avaliar os padrões de riqueza, cinco oferecem os maiores poderes de explicação e associação: disponibilidade de energia, tempo evolutivo, heterogeneidade de hábitats, disponibilidade de área e as restrições geométricas (Rahbek & Graves, 2001). Uma revisão dessas hipóteses acerca dos padrões de riqueza distribuídos por todo o globo terrestre foi feita por Willig *et al.* (2003). A despeito do maior reconhecimento do efeito de outras hipóteses, especialmente os efeitos do clima (Hawkins *et al.*, 2003b), o efeito das restrições geométricas tem sido alvo de discussões sobre seu uso e pressupostos (Zapata *et al.*, 2003; Hawkins *et al.*, 2005; McClain *et al.*, 2007).

A hipótese das restrições geométricas, ou modelos do Efeito do Domínio Médio ("*Mid-Domain Effect*", ou MDE) (Colwell & Lees, 2000) diz que a distribuição de espécies sobre uma região é restrita pelos limites, ou bordas, deste domínio (por exemplo, continentes, biomas, extensões latitudinais ou longitudinais, elevação topográfica e profundidades marinhas) e, como conseqüência, a maior

diversidade de espécies será encontrada na localidade central e irá gradativamente diminuir em direção às bordas do domínio (Colwell & Lees, 2000; Colwell *et al.*, 2004). O pressuposto essencial do MDE, dado como um modelo nulo, é a ausência dos fatores espaciais e temporais atuando na origem e manutenção das distribuições das espécies, de modo que sua sobreposição (i.e., riqueza) ocorrerá simplesmente ao acaso em um ambiente homogêneo (Colwell & Lees, 2000). Assim, esta poderia ser colocada casualmente em qualquer lugar, mas sua expansão é limitada pelas restrições geométricas e, deste modo, a maior sobreposição das espécies estaria no centro do domínio.

O uso do MDE como modelo nulo serviria para comparar os padrões de riqueza de espécies esperados na ausência de fatores causais com os padrões observados. Constatada uma diferença entre estes padrões, isso indicaria que haveria algum processo que realmente estaria agindo sobre a distribuição de espécies. Entretanto, os proponentes desta hipótese também assumem seu uso simultâneo com as outras variáveis ambientais ou históricas, sendo que alguns estudos de fato admitem ser esta expectativa nula o principal preditor da riqueza, em detrimento das hipóteses ambientais largamente conhecidas (Jetz & Rahbek, 2001, 2002; Bellwood *et al.*, 2005; Brehm *et al.*, 2007).

Mesmo que alguns estudos considerem o MDE como um efeito importante a ser considerado em análises ecológicas, um conjunto de críticas limita seu uso porque citam que seus pressupostos não consideram descrições reais e adequadas do mundo, seus conceitos são inconsistentes, falhos e possuem fraquezas metodológicas (Hawkins & Diniz-Filho, 2002; Zapata *et al.*, 2003; Hawkins *et al.*, 2005). McClain *et al.* (2007) recentemente discutiram os desafios e

cuidados na aplicação das restrições geométricas em análises ecológicas. Em um artigo recente, Dunn *et al.* (2006) revisou diversos artigos e demonstrou em quais situações há altos poderes de explicação com a hipótese do efeito do domínio médio, ou seja, espécies com grandes distribuições relativas ao domínio e regiões com maiores extensões territoriais. Assim, acredita-se que a discussão sobre a validade do MDE e seu uso como descritor dos padrões de riqueza de espécies, como um processo puramente estocástico, continuará na literatura científica.

A abordagem das restrições geométricas já foi utilizada em alguns estudos no continente sul-americano (Diniz-Filho *et al.*, 2002; Rangel & Diniz-Filho, 2004), com baixo porcentual de explicação da hipótese do efeito do domínio médio, e que consideravam todo o continente como domínio e suas bordas limitadas pelo oceano. Neste estudo, foram avaliados os padrões no bioma Cerrado, localizado na região central do território brasileiro, como domínio geral limitado pelo clima (no sentido de Colwell & Less, 2000). A hipótese do efeito do domínio médio foi incluída juntamente com as variáveis das hipóteses de energia ambiental, produtividade, heterogeneidade ambiental, climática, relacionadas à ocupação humana e ao número de inventários faunísticos a fim de explicar os padrões de riqueza de espécies endêmicas de vertebrados terrestres no Cerrado.

Dada a extensão territorial do Cerrado, foram utilizadas ferramentas que corrigem possíveis efeitos da autocorrelação espacial utilizando regressões espaciais (Diniz-Filho *et al.*, 2003; Fotheringham *et al.*, 2003). Através de combinações das variáveis nas análises de regressão, foram avaliados diferentes modelos para explicar a variação de riqueza nas espécies endêmicas, sendo estes avaliados no sentido de determinar aqueles que oferecem maiores suportes aos

dados observados. Esse processo de avaliação dos modelos foi realizado pelo Critério de Akaike (AIC), um método de seleção de modelos que não realiza testes de hipóteses, mas que avalia cada modelo de acordo com o tamanho amostral, a variância residual e o número de parâmetros de cada regressão (Burnham & Anderson, 2004).

Assim, este estudo teve como objetivo entender se a distribuição das espécies endêmicas de vertebrados terrestres no Cerrado é o resultado de processos estocásticos e ecológicos, com suas respectivas contribuições. Embora ainda faltem amostragens da fauna endêmica no bioma Cerrado (Cavalcanti & Joly, 2003), o que pode acarretar falsas expectativas de um modelo, a abordagem aqui utilizada pode ser considerada pioneira ao modelar a distribuição destas espécies e também devido ao uso da hipótese do efeito do domínio médio.

Metodologia

Dados Ecológicos

O domínio a ser considerado é o bioma Cerrado, dividido em 181 quadrículas de 1º de latitude por 1º de longitude (Figura 1). Cada quadrícula possui referência à localização geográfica, medidas das variáveis sócio-ambientais e valores da riqueza de espécies endêmicas observadas. Os dados sobre espécies endêmicas tais como sua extensão geográfica contínua e riqueza em cada quadrícula são de 127 espécies endêmicas, sendo 16 mamíferos, 29 aves, 35 répteis e 47 anfíbios (Diniz-Filho *et al.*, 2007, *in press*). Verifica-se a extensão de uma espécie calculando em quantas quadrículas ela ocorre e quanto

maior a sobreposição de espécies sobre uma quadrícula maior será sua diversidade. Os dados referentes à riqueza de espécies endêmicas foram logaritimizados para reduzir a assimetria (Sokal & Rohlf, 1995).

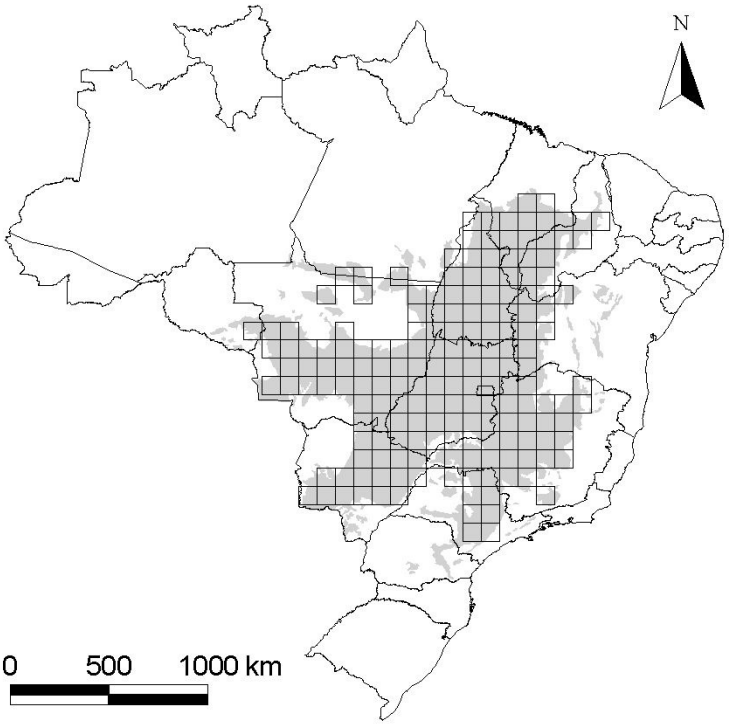


Figura 1: Mapa do bioma Cerrado, considerado aqui como um domínio com limites, e sua divisão em quadrículas sobre o território brasileiro.

Diversos autores destacam as principais hipóteses que, através de sua variação no espaço, poderiam estar influenciando a distribuição de espécies: produtividade (evapotranspiração atual – *AET*, índice de vegetação – *EVI*), heterogeneidade ambiental (neste caso a amplitude do relevo – *TOPOG*), energia ambiental (evapotranspiração potencial – *PET*, temperatura média anual – *TEMP*),

condições climáticas (umidade relativa do ar – *HUMREL*) e a hipótese relativa à ocupação humana (tamanho populacional presente no Cerrado no ano 2000 – *HUMAN*, número de inventários faunísticos ou conhecimento da biodiversidade – *KNOW*) (Currie, 1991; Hawkins *et al.*, 2003a,b; Diniz-Filho *et al.*, 2005; Rangel *et al.*, 2007).

Colwell *et al.* (2004) enfatizam que uma abordagem multivariada, ou seja, o uso de variáveis relacionadas ao MDE juntamente com outras variáveis sócio-ambientais, pode permitir uma melhor avaliação de quanto um processo estocástico, ocorrendo na ausência de um gradiente geográfico, poderia explicar a distribuição de espécies dentro de um domínio. Entretanto, poucos trabalhos consideraram simultaneamente efeitos ambientais e o MDE (por exemplo, Lees *et al.*, 1999; Jetz & Rahbek, 2002; Rangel & Diniz-Filho 2004; Kerr *et al.*, 2006; Brehm *et al.*, 2007).

Para avaliar o efeito do MDE foi gerada uma variável baseada no algoritmo "*spreading dye*" ("dispersão de tinta") de Jetz & Rahbek (2001), que é conhecido como modelo de restrição geométrica ("*geometric constraints model*", ou *GCM*). Este algoritmo inicia-se através da escolha aleatória de uma determinada quadrícula do domínio considerado e a introdução de uma espécie nesta quadrícula. Através da distribuição conhecida desta espécie (dados observados), sua ocupação se expande a partir dessa quadrícula inicial para os locais adjacentes, em direção aos limites do domínio, continuamente. O processo de expansão pára quando a espécie ocupa todas as quadrículas, contínuas, referentes ao tamanho da sua distribuição observada. Entretanto, neste algoritmo, a expansão não pode atravessar os limites do domínio e a distribuição da espécie

é colocada aleatoriamente no espaço, sem levar em consideração os possíveis gradientes geográficos relacionados aos nichos das espécies. Após simular esse processo para todas as espécies, obtém-se uma variável de riqueza, e toda a simulação é realizada mais uma vez. Uma vantagem desta variável é que sua unidade (número de espécies) é a mesma da variável resposta e deste modo uma predição direta da riqueza de espécies (Colwell *et al.*, 2004).

Análise de Dados

Antes dos modelos regressivos serem avaliados, foram feitos testes de homogeneidade (teste Scheffé-Box), normalidade (Kolmogorov-Smirnov) e independência dos valores residuais (autocorrelação espacial). Isto foi realizado através de modelos regressivos globais (que considera todas as variáveis em questão) via OLS (*Ordinary Least Square*), que não corrigem o efeito da autocorrelação espacial, e da abordagem SAR (*Simultaneous Autoregressive Models*), que incluem o efeito espacial na matriz de covariância dos resíduos (Haining, 1990). A autocorrelação espacial foi calculada através do índice I de Moran (Legendre & Legendre, 1998) e a variação destes em função das distâncias geográficas entre as células foi avaliada através de correlogramas, na qual o eixo das abscissas contém as classes de distâncias e o eixo das ordenadas os valores deste índice em cada classe de distância. Geralmente nas primeiras classes de distâncias os valores do I de Moran são positivos, indicando que quadrículas adjacentes são mais similares entre si do que esperado ao acaso, e neste caso viola-se ao pressuposto de amostras independentes. A significância de cada valor deste índice foi calculada através de testes de randomização, utilizando 1000

aleatorizações do vetor de resíduos da regressão, e o coeficiente testado ao nível de 5%. Caso em um modelo OLS seja detectada autocorrelação positiva significativa para as primeiras classes de distâncias nos resíduos, é preferível utilizar modelos autoregressivos, tal como o SAR. Ao utilizar esta estatística, a rede de conexões entre as quadrículas foi obtida a partir do algoritmo da rede de Gabriel (Legendre & Legendre, 1998), na qual a ligação entre os pontos de amostragens ocorre se o círculo cujo diâmetro é a distância entre os centros de duas quadrículas não inclui o centro de nenhuma outra quadrícula.

Através da regressão da variável resposta (riqueza de espécies endêmicas) sobre as variáveis sócio-ambientais e o *GCM*, foi possível construir um conjunto de modelos com os fatores que melhor descrevem a riqueza dessas espécies no bioma Cerrado. Devido à enorme quantidade de modelos possíveis para esse conjunto de variáveis, foi utilizado um critério de seleção de modelos que não se baseia em testes de hipóteses (i.e., como uma "*stepwise selection*"), mas pelo Critério de Akaike (AIC – *Akaike Information Criterion*), que neste caso o valor de AIC de cada modelo é dado por uma equação que relaciona a variância residual proveniente de cada regressão, o tamanho amostral e o número de parâmetros utilizados em um determinado modelo (Burnham & Anderson, 2004). A comparação entre modelos é feita primariamente pela diferença de um valor AIC com outro (ΔAIC). É considerado o melhor modelo aquele que possui o menor valor deste critério, sendo também possível calcular a chance (w) de um modelo ser escolhido como o melhor modelo caso o estudo repetido diversas vezes, uma medida do poder de evidência (para uma referência completa sobre o Critério de Akaike, ver Burnham & Anderson, 2004).

Todas as análises foram realizadas no software SAM – *Spatial Analysis in Macroecology* (Rangel *et al.*, 2006).

Resultados

Em relação aos testes dos pressupostos da regressão, o modelo OLS global apresentou normalidade dos seus resíduos, o que não foi significativo para o modelo SAR. Já em relação à homogeneidade das variâncias, estas foram consideradas homogêneas para ambos os tipos de modelos.

Após a construção do modelo global para a regressão OLS, foram obtidos os valores preditos e os resíduos, utilizando-se a riqueza de espécies endêmicas como variável resposta. O coeficiente de determinação (r^2) e o valor de AIC para este modelo foi respectivamente 0.906 e -589.443. A análise do correlograma do resíduo mostrou que a primeira classe de distâncias apresentou uma autocorrelação espacial significativa ($P < 0,05$) (Figura 2), indicando que células próximas são mais similares entre si do que esperado simplesmente ao acaso.

Quando utilizou-se o modelo SAR, com todos os parâmetros no modelo, os erros deste no correlograma demonstraram um menor grau de autocorrelação espacial nas primeiras classes de distância, com valores do índice de Moran oscilando próximo a zero nas distâncias intermediárias. A estrutura espacial não foi significativa para localidades próximas ($P > 0,05$), indicando que o modelo espacial possui uma melhor performance para descrever o padrão espacial na riqueza de espécies endêmicas no Cerrado. A preferência deste modelo também é reforçada porque o valor de r^2 aumentou (0.977) e diminuiu o AIC (-844.464).

Como já foi dito anteriormente, quanto menor o valor de AIC, tanto melhor é a aproximação da "realidade". A diferença de valores Akaike para estes modelos OLS e SAR foi de 255,021.

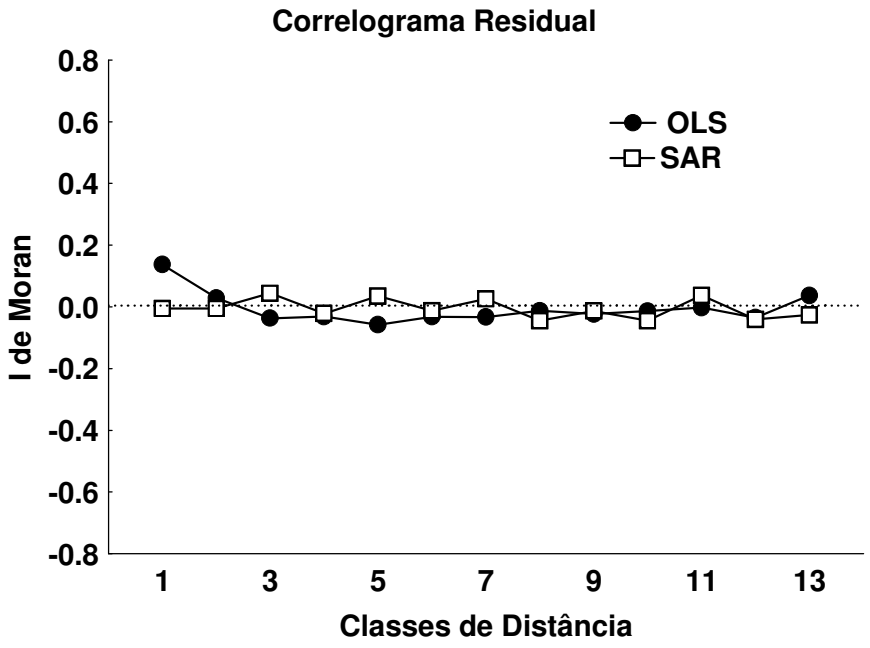


Figura 2: Correlogramas espaciais dos resíduos dos modelos globais OLS e SAR, considerando todas as variáveis.

Mapeando-se a distribuição dos resíduos dos modelos de riqueza no Cerrado (Figura 3), percebe-se que, para o modelo OLS, há mais grupos de valores de resíduos próximos entre si, indicando que estas quadrículas agrupadas possuem valores similares devido à autocorrelação espacial. Já no mapa do SAR, há menos grupos formados por valores próximos dos resíduos e também a amplitude do resíduo é menor, indicando que com este modelo a autocorrelação

espacial é menor e por isso melhor descreve a distribuição de espécies endêmicas no Cerrado, como observado no correlograma apresentado na Figura 2.

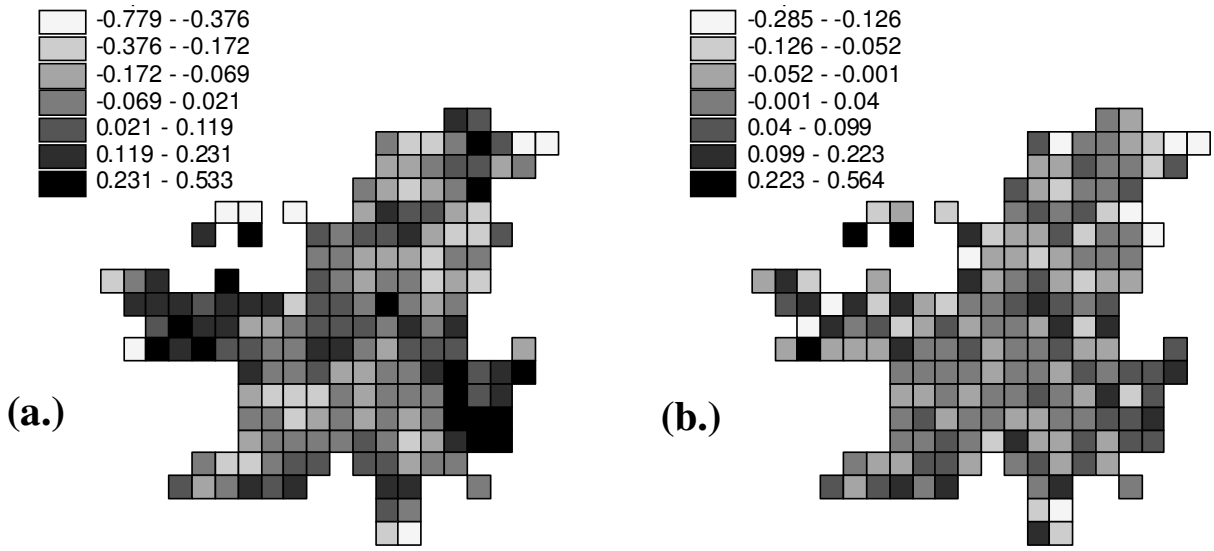


Figura 3: Distribuição dos resíduos dos modelos globais sobre as quadrículas do Cerrado. O mapa (a) representa os resíduos do modelo OLS e (b) são os resíduos do modelo SAR.

Deste modo, visto que os modelos espaciais são mais adequados, os resultados apresentados adiante referem-se exclusivamente aos obtidos no modelo espacial SAR. A partir do conjunto de modelos construídos combinando as 9 variáveis, demonstrou-se que a melhor equação possui valores de r^2 e de AIC respectivamente iguais a 0.982 e -900,867, nas quais as variáveis selecionadas foram: a predição de espécies pela variável *GCM*, índice de vegetação (*EVI*) e

temperatura média anual (*TEMP*). Este modelo possui uma probabilidade de 58% de ser escolhido novamente como o melhor modelo, entre todos considerados, se diversos estudos forem realizados novamente. Posicionando-se em ordem decrescente os valores do coeficiente de determinação e em ordem crescente os valores do Critério de Akaike (ambos posicionam os modelos do melhor para o pior), avaliou-se quais modelos melhor oferecem explicação aos dados (Tabela 1). Os três primeiros modelos possuem valores iguais de r^2, mas os valores do Critério de Akaike para estes dois primeiros foram diferentes, permitindo um melhor discernimento entre eles. O modelo global possui uma diferença do valor AIC de 56.43 em relação ao melhor modelo.

Tabela 1: Resultados da seleção de modelos SAR para as espécies endêmicas do Cerrado.

R^2	ΔAIC	W	modelo
0,982	0	0,58	*GCM, EVI, TEMP*
0,982	1,963	0,21	*GCM, EVI, TEMP, TOPOG*
0,982	2,187	0,19	*GCM, HUMAN, EVI, TEMP*
0,981	9,273	$5x10^{-3}$	*GCM, TEMP, TOPOG*
0,981	12,224	$1x10^{-3}$	*GCM, TEMP, TOPOG, KNOW, HUMAN*

A variável mais importante para prever as espécies endêmicas foi *GCM* (Tabela 2). Todos os três primeiros modelos, com valores de ΔAIC< 3, incorporaram esta variável e, considerando modelos com apenas uma variável, a

equação com a variável *GCM* (a = 1,652, p<0,05; b = 0,058, p<0,05; abordagem presente em Jetz & Rahbek, 2001) possui a melhor performance (r^2 = 0,962; AIC = -769,767). No mapa verifica-se que a distribuição de espécies endêmicas segue de fato uma tendência similar à distribuição de valores a partir da hipótese do efeito do domínio médio (Figura 3).

Entretanto, é importante considerar que, neste trabalho, o processo de seleção de modelos considerou o efeito das variáveis *GCM*, da temperatura média e do índice de vegetação. A equação com apenas a variável *GCM* está em uma posição muito além dos três primeiros colocados no posicionamento dos valores do Critério de Akaike para cada modelo.

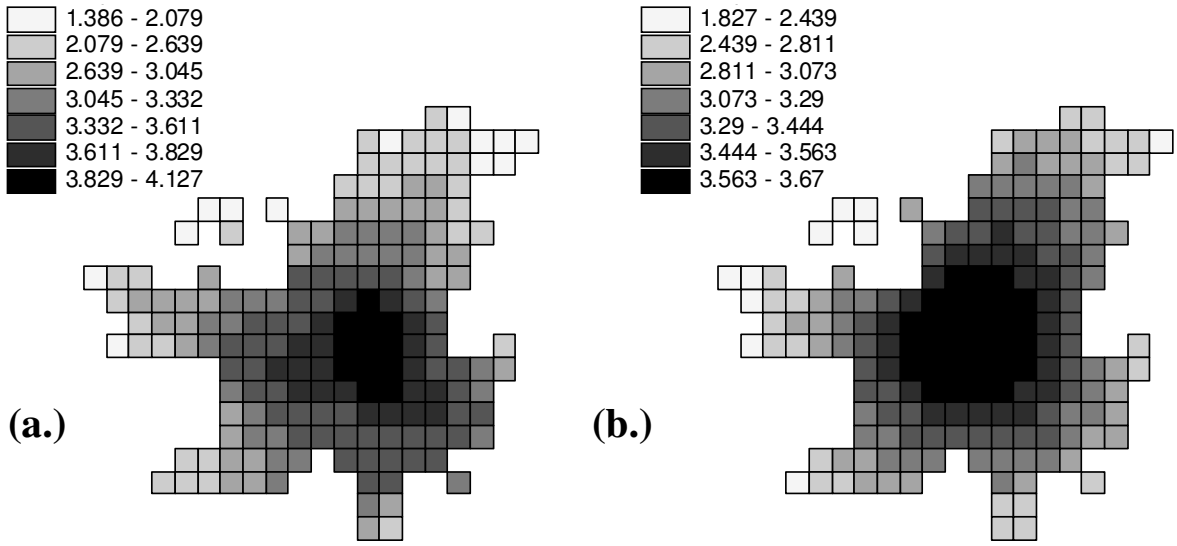


Figura 3: O mapa (a) é a distribuição das espécies endêmicas no Cerrado e (b) é a distribuição de acordo com a hipótese do efeito do domínio médio. Ambos os valores estão logaritimizados.

Discussão

Os resultados dos testes dos pressupostos dos modelos de regressão indicaram apenas que os valores residuais dos modelos SAR não apresentaram normalidade, enquanto que as variâncias foram homogêneas e não foi identificado autocorrelação espacial em seus resíduos, ou seja, o valor do resíduo de uma quadrícula não é influenciado significativamente por sua vizinhança. A maior parte dos resíduos dos melhores modelos SAR permaneceu próximo de zero e não apresentam uma distribuição normal significativa. Isto foi diferente nos modelos OLS porque, apesar deste apresentar normalidade para os valores residuais, estes apresentam um certo grau de autocorrelação espacial significativa para a primeira classe de distância. Deste modo, preferiu-se seguir os modelos regressivos utilizando a autoregressão SAR, que possui um parâmetro que contabiliza o grau de autocorrelação espacial presente nos dados observados e que por isso tem estimativas dos coeficientes mais corretas e predições mais acuradas (Lichstein *et al.*, 2002).

O melhor modelo SAR explicando a riqueza de espécies endêmicas de vertebrados terrestres no Cerrado, utilizando-se o critério de AIC, incorporou as variáveis relacionadas às hipóteses de produtividade, energia ambiental e do efeito do domínio médio. Estes dados estariam de acordo ao esperado (Hawkins *et al.*, 2003b) para esta região tropical se não fosse a associação negativa das variáveis temperatura média anual e índice de vegetação em relação à riqueza de espécies endêmicas. Isto significa que os locais com maior produtividade primária e mais quentes não possuem maior diversidade do que seria esperada. Já a variável *GCM* possui associação positiva aos dados observados. Esta variável

sozinha, entretanto, não apresenta um ajuste tão bom quanto a obtida utilizando-se o melhor modelo revelado pelo Critério de Akaike.

As associações com os gradientes ambientais podem estar negativas (contrárias ao que seria esperado) devido à falta de amostragens e mais registros de espécies endêmicas sobre este bioma (Marinho-Filho *et al.*, 2002; Cavalcanti & Joly, 2002), ou por apresentarem-se como efeitos parciais após ter sido considerado o efeito dos gradientes em grande escala expressos na variável *GCM*. Outra hipótese é que as espécies adaptadas a este bioma estariam concentradas no centro do domínio, sem levar em consideração a hipótese do efeito do domínio médio (explicação adiante). Deve-se levar também em consideração os componentes históricos e biogeográficos na determinação do endemismo no Cerrado (esta última hipótese não será analisada neste estudo) (Rodrigues *et al.*, 2005).

Pelo fato do bioma Cerrado dividir fronteiras com outros biomas a maior diversidade de espécies endêmicas poderia ser de fato no centro do domínio. Se a maior riqueza de espécies fosse encontrada em locais com maiores índices de produtividade, temperatura e entrada de água, então a diversidade seria maior na região noroeste do Cerrado, que está próximo às transições de um bioma a outro (neste caso o bioma Amazônico). Neste caso, as distribuições das espécies poderiam ultrapassar os limites do domínio Cerrado e não seriam mais considerados como espécies endêmicas deste bioma (sendo, assim, excluídas da presente análise). Assim, as espécies adaptadas ao Cerrado se encontrariam principalmente mais próximos ao centro do domínio, onde há maiores áreas para a ocupação, embora encontrem-se algumas espécies endêmicas também próximas

às bordas, com distribuições mais restritas (na predição de espécies a partir da hipótese do efeito do domínio a alta sobreposição de espécies encontra-se próxima à borda oeste).

De qualquer modo, as taxas de endemismo para vertebrados são constatadas como baixas neste bioma (Myers *et al.*, 2000) sendo maiores para anuros e répteis do que para aves e mamíferos (Klink & Machado, 2005). A fauna presente no Cerrado compartilha traços biológicos comuns a outras formações abertas na América do Sul e é bastante influenciada pelos biomas de floresta vizinhos tais como a Mata Atlântica e a Amazônica (Marinho-Filho *et al.*, 2002).

No estudo de Kerr *et al.* (2006) em Madagascar, a maior riqueza de espécies de aves e mamíferos endêmicos está na borda leste deste domínio (portanto neste estudo a hipótese do efeito do domínio médio não foi um forte preditor de espécies) por causa da maior cobertura de floresta nesta região, onde há maior produtividade primária (entretanto ver o comentário de Lees & Colwell, 2007, que afirmam que os dados de Kerr *et al.* (2006) estavam incorretos e os resultados foram mal-interpretados). Entretanto, este domínio é considerado como de "difíceis limites para a dispersão" (Colwell & Lees, 2000; mas ver Zapata *et al.*, 2003, sobre os possíveis problemas acerca dos limites) por causa de seu isolamento geográfico. Já o bioma Cerrado é uma região que não está isolada, dividindo fronteiras com a Amazônia, Mata Atlântica, Pantanal e Caatinga. Por causa disto algumas espécies podem dispersar além dos limites e ocupar dois ou mais biomas.

Dunn *et al.* (2006) recentemente fizeram uma revisão de diversos estudos que consideram esta hipótese das restrições geométricas e verificaram que a

riqueza predita pela hipótese de MDE encontra maiores suportes aos dados observados quando as distribuições observadas das espécies são grandes, comparadas ao tamanho do domínio e também estas possuem maiores extensões, enquanto que em menores tamanhos de ocorrência (relativos) e extensões territoriais os fatores ecológicos e históricos parecem possuir uma marca mais acentuada na distribuição de espécies. No caso das espécies endêmicas consideradas aqui neste estudo, não são tantas espécies que ocupam vastas extensões geográficas comparadas ao domínio (Figura 4). Entretanto, a variável *GCM* pode ser o principal descritor, em detrimento às variáveis sócio-ambientais, talvez porque a maior parte das espécies endêmicas está localizada no centro do domínio onde estão as maiores áreas para ocupação.

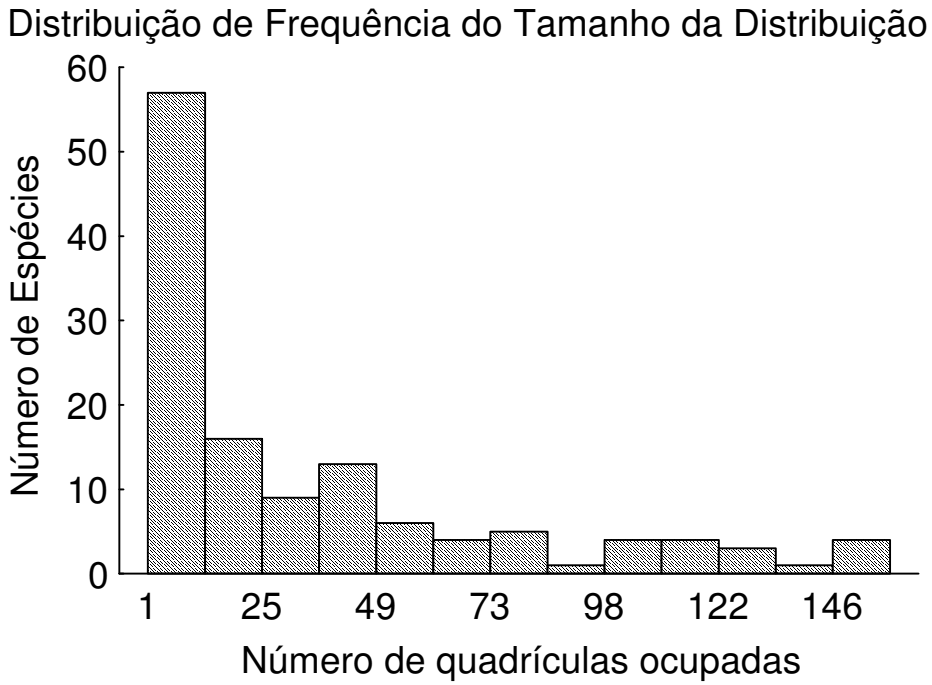


Figura 4: Histograma do número de quadrículas ocupadas pelas espécies endêmicas do bioma Cerrado.

Assim, por causa dos limites, a maior sobreposição de espécies endêmicas encontra-se no meio do domínio. Silva (1997), entretanto, identificou três sub-áreas de endemismo para aves no Cerrado, sendo que duas estão localizadas nas bordas do bioma, mas o autor utilizou apenas distribuições restritas de aves endêmicas e neste caso as áreas de endemismo podem estar no centro ou nas bordas do domínio.

Na hipótese do efeito do domínio médio, já que não há o efeito ambiental sobre as espécies, já que o ambiente seria visto como homogêneo em toda sua extensão, a distribuição das espécies, de acordo com o algoritmo "*spreading dye*", poderia ser iniciada em qualquer quadrícula do domínio, embora por definição esta não possa ultrapassar o domínio. Uma das críticas sendo feitas a esta hipótese refere-se à impossibilidade de gerar, no mundo real, as distribuições das espécies em um ambiente sem gradientes ambientais, já que estes seriam os principais fatores que as estruturam (Hawkins *et al.*, 2005). Em um ambiente homogêneo, este algoritmo faz a distribuição aleatória das espécies em um bioma sem levar em consideração a teoria do nicho, na qual as espécies permanecem em locais mais adequados para a persistência de suas populações. Pelo fato de serem utilizadas as distribuições observadas para o algoritmo, o tamanho da distribuição é afetado por poucas características biológicas (Dunn *et al.*, 2006), mas a distribuição é alocada no espaço geográfico por processos estocásticos. Ou seja, esses processos fariam com que o tamanho da distribuição geográfica de uma espécie fosse limitada por mecanismos biológicos inerentes ao grupo taxonômico e que sejam independentes de variação ambiental e temporal. Isso, por um lado, leva ao ponto de que o MDE assumiria algumas características biológicas

intrínsecas ao grupo estudado (Colwell *et al.*, 2004), contudo assumindo características estruturadas na ausência de efeitos espaciais e temporais.

Por outro lado, Hawkins & Diniz-Filho (2002) e Hawkins *et al.* (2005) argumentaram que isso seria inadequado, já que a distribuição estatística dos tamanhos das distribuições geográficas seria de fato determinada pelo ambiente, de modo que a utilização de padrões reais no *GCM* (ou em qualquer outro modelo para o MDE) seria tautológica. Na realidade, sob a teoria do nicho, a distribuição geográfica de uma espécie na ausência de fatores ambientais seria infinita, de modo que a riqueza de espécies seria constante sobre o espaço. Mesmo sob um modelo de dinâmica meta-populacional estocástico, a riqueza seria uma variável aleatória sobre o espaço, com padrões espaciais estacionários (ou seja, apenas em pequena escala espacial).

Por outro lado, Rangel & Diniz-Filho (2005) demonstraram recentemente, através de um modelo de simulação espacialmente explícito da dinâmica neutra de Hubbell (2001) em meta-comunidades, que o MDE poderia explicar a riqueza de espécies em dadas combinações de parâmetros (dispersão limitada e elevadas taxas de recolonização). Isso seria uma base teórica para a sugestão de Colwell *et al.* (2004) de que o MDE (por meio do *GCM*) poderia ser considerado não um modelo nulo a ser aceito ou rejeitado, mas sim como uma expressão de processos estocásticos associados aos padrões de riqueza.

Conclusão

A análise ecológica através de autoregressões, como o modelo SAR, se mostrou satisfatória porque corrigiu o efeito da autocorrelação espacial presente nos resíduos da regressão OLS e deste modo foi possível selecionar pelo Critério de Akaike os resultados mais corretos.

O melhor modelo selecionado parece demonstrar que a ocorrência de espécies endêmicas no Cerrado é estruturada em parte por processos estocásticos, juntamente com a influência ambiental. Na verdade, de acordo com as críticas, os pressupostos da hipótese do efeito do domínio médio são bastante simplistas, de modo que sua interpretação deve ser cuidadosa. Os desafios futuros para a melhoria desta hipótese são oferecer uma estrutura que esteja mais de acordo com as condições reais e também do próprio algoritmo que gera as distribuições das espécies no domínio.

Bibliografia

Bellwood, D.R.; Hughes, T.P.; Connolly, S.R.; Tanner, J. (2005) Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecology Letters*, 8: 643-651.

Brehm, G.; Colwell, R.K.; Kluge, J. (2007) The role of environment and mid-domain effect on moth species richness along a tropical elevational gradient. *Global Ecology and Biogeography*, 16: 205-219.

Brown, J.H. (1995) *Macroecology*. The University of Chicago Press.

Burnham, K.P. & Anderson, D.R. (2002). Model selection and multimodel inference. Springer, NY.

Cavalcanti, R.B. & Joly, C.A. (2002) Biodiversity and conservation priorities in the Cerrado. *The Cerrado of Brasil: Ecology and Natural History of a Neotropical Savanna*: 351-367. Columbia University Press, New York.

Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution,* 15 (2): 70-76.

Colwell, R.K.; Rahbek, C.; Gotelli, N.J. (2004) The mid-domain effect and species richness patterns: what have we learned so far? *American Naturalist,* 163: E1-E23.

Currie, D.J. (1991) Energy and large-scale patterns of animal- and plant- species richness. *The American Naturalist,* 137 (1): 27-49.

Diniz-Filho, J.A.F.; Sant'ana, C.E.R.; Souza, M.C.; Rangel, T.F.L.V.B. (2002) Null models and spatial patterns of species richness in South American birds of prey. *Ecology Letters*, 5: 47-55.

Diniz-Filho, J.A.F.; Bini, L.M.; Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography,* 12: 53-64.

Diniz-Filho JAF, RP Bastos, TFLVB Rangel, LM Bini, P Carvalho, RJ Silva. 2005. Macroecological correlates and spatial patterns of anurans description dates in Brazilian Cerrado. Global Ecol. Biogeogr. 14: 469-477.

Diniz-Filho, J.A.F.; Bini, L.M.; Pinto, M.P.; Terribile, L.C.; Oliveira, G.; Vieira, C.M.; Blamires, D.; Barreto, B.S.; Carvalho, P.; Rangel, T.F.L.V.B.; Tôrres,

N.M.; Bastos, R.P. (2007) Conservation biogeography and endemism of terrestrial vertebrates in Brazilian Cerrado. *ORYX* (*in press*).

Dunn, R.R.; McCain, C.M.; Sanders, N.J. When does diversity fit null model predictions? Scale and range size mediate the mid-domain effect. *Global Ecology and Biogeography* (in press).

Fotheringham, A.S.; Brunsdon, C.; Charlton, M. (2002) *Quantitative Geography : Perspectives on Spatial Analysis.* SAGE Publications.

Gotelli, N.J. & McGill, B.J. (2006) Null versus neutral models: what's the difference? *Ecography,* 29: 793-800.

Haining, R. (1990) Spatial data analisys in the social and environmental sciences. Cambridge University Press, Cambridge.

Hawkins, B.A. & Diniz-Filho, J.A.F. (2002) The mid-domain effect cannot explain the diversity gradient of Neartic birds. *Global Ecology and Biogeography*, 11: 419-426.

Hawkins, B.A.;Porter, E.E.; Diniz-Filho, J.A.F. (2003a) Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology,* 84: 1608-1623.

Hawkins BA, R Field, HV Cornell, DJ Currie, JF Guegan, DM Kaufman, JT Kerr, GG Mittelbach, T Obedorff, EM O'Brien, EE Porter, JRG Turner. (2003b). Energy, water, and broad-scale geographic patterns of species richness. Ecology, 84: 3105-3117.

Hawkins, B.A.; Diniz-Filho, J.A.F.; Weis, A.E. (2005) The mid-domain effect and diversity gradients: is there anything to learn? *The American Naturalist*, 166: E140-E143.

Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences of the USA,* 98: 5661-5666.

Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, 297: 1548-1551.

Kerr, J.T.; Perring, M.; Currie, D.J. (2006) The missing Madagascan mid-domain effect. *Ecology Letters*, 9: 149-159.

Klink, C.A. & Machado, R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology,* 19 (3): 707-713

Lees, D.C.; Kremen, C.; Andriamampianina, L. (1999) A null model for species diversity gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. *Biological Journal of the Linnean Society*, 67: 529-584.

Lees, D.C. & Colwell, R.K. (2007) A strong Madagascan rainforest MDE and no equatoward increase in species richness: re-analysis of the "The missing Madagascan mid-domain effect", by Kerr, J.T., Perring, M. & Currie D.J. (Ecology Letters 9:149-159, 2006). *Ecology Letters,* 10: E4-E8

Legendre, P. & Legendre, L. (1998). Numerical ecology. Amsterdam: Elsevier.

Lichstein, J.W.; Simons, T.R.; Shriner, S.A.; Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs,* 72 (3): 445-463.

Marinho-Filho, J.; Rodrigues, F.H.G.; Juarez, M.J. (2002) The Cerrado mammals: diversity, ecology and natural history. *The Cerrado of Brasil: Ecology and*

Natural History of a Neotropical Savanna: 266-284. Columbia University Press, New York.

McClain, C.R.; White, E.P.; Hurlbert, A.H. (2007) Challenges in the application of geometric constraint models. *Global Ecology and Biogeography*, 16: 257-264.

Myers, N.; Mittermeier R.A.; Mittermeier C.G.; Fonseca G.A.B.; Kent J. (2000) Biodiversity hotspots for conservation priorites. *Nature,* 43: 853-858.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences,* USA, 98: 4534-4539.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2004) An evolutionary tolerance model explaining spatial patterns in species under environmental gradients and geometric constraints. *Ecography*, 27: 1-11.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2005) Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. *Ecology Letters,* 8: 783-790.

Rangel, T.F.L.V.B.; Diniz-Filho, J.A.F.; Bini, L.M. (2006) Towards an integrated computacional tool for spatial analysis in macroecology and biogeography. *Global Ecology & Biogeography,* 15: 321-327.

Rangel, T.F.L.V.B.; Bini, L.M.; Diniz-Filho, J.A.F.; Pinto, M.P.; Carvalho, P.; Bastos, R.P. (2007) Human development and biodiversity conservation in Brazilian Cerrado. *Applied Geography*, 27: 14-27.

Rodrigues, M.T. (2005) A biodiversidade dos Cerrados: conhecimento atual e perspectivas, com uma hipótese sobre o papel das matas galerias na troca

faunística durante ciclos climáticos. *Cerrado: Ecologia, Biodiversidade e Conservação*, 235-246

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research.* W.H. Freeman and Company.

Willig, M.R.; Kaufman, D.M.; Stevens, R.D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Syst.,* 34: 273-309.

Zapata, F.A.; Gaston, K.J.; Chown, S.L. (2003) Mid-domain models of species richness gradientes: assumptions, methods and evidence. *Journal of Animal Ecology,* 72: 677-690.

Conclusões Gerais

-A autocorrelação espacial dos resíduos foi significativa para a regressão OLS, geralmente nas primeiras classes de distância. Isto indica que as quadrículas (pontos de coleta) mais próximas são dependentes entre si para os valores residuais, violando o pressuposto de independência dos resíduos dos modelos lineares;

-Os modelos SAR foram eficientes ao corrigirem os efeitos da autocorrelação espacial presente nos resíduos dos modelos. Deste modo, os coeficientes foram melhores calculados e assim as predições foram mais corretas;

-O critério de Akaike selecionou os melhores modelos, similares ao esperado pelas expectativas teóricas. Este critério não utiliza testes de hipóteses, apenas utilizando o tamanho amostral, a variância residual e o número de parâemtros de cada modelo regressivo para estimar um valor AIC, na qual o modelo que possui o menor valor deste critério possui a melhor performance;

-A multicolinearidade foi evidente primariamente nos modelos OLS porque estes apresentam mais variáveis nos seus melhores modelos do que em relação ao SAR;

- No Cerrado, os padrões de riqueza de aves e mamíferos são explicados principalmente pelas variáveis relacionadas à dinâmica água-energia e

produtividade enquanto que os padrões de anfíbios e répteis são melhores relacionados à entrada de energia e novamente à dinâmica água-energia;

-Com relação às espécies endêmicas, a variável descritora mais importante foi relacionada à hipótese das restrições geométricas, ou seja, a atuação da estocasticidade parece ser tão ou mais atuante do que a influência espacial e temporal do ambiente.