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

Simulação Orientada por Padrões e Análise Espacial dos Gradientes de Diversidade sob um Processo de Conservação de Nicho

Thiago Fernando Lopes Valle de Britto Rangel

Goiânia,
Fevereiro de 2006

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

Simulação Orientada por Padrões e Análise Espacial dos Gradientes de Diversidade sob um Processo de Conservação de Nicho

Thiago Fernando Lopes Valle de Britto Rangel

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,
Fevereiro de 2006

Agradecimentos

Às mulheres da minha vida, Marta e Fernanda, pelo amor e carinho que me proporcionaram durante o tempo. À Marta por sempre ter sido, simultaneamente, o farol que me guia e o motor que me propele. À Fernanda por me dar estabilidade e alegria ao longo do caminho.

Ao Alex, pela amizade e companhia, pela eterna paciência e bom humor (até com enxaqueca), e pela maestria com que ensina, incentiva e desafia a todo tempo. Sem dúvida, muito do que aprendi e fiz ao longo dos últimos seis anos foi inspirado e motivado por ele.

Ao Maurício, por sempre estar disponível para ensinar, ajudar e aconselhar. Ao Maurício e ao Cleiber pelas discussões e sugestões, que contribuíram com o desenvolvimento do modelo do capítulo 1. Ao Maurício novamente por várias sugestões e idéias para o desenvolvimento do SAM. A todos os colegas de laboratório, pela companhia e compartilhamento de momentos alegres e descontraídos.

Aos membros da banca de avaliação desta dissertação Prof. Dr. José Alexandre Felizola Diniz Filho, Prof. Dr. Wellington Santos Martins e Prof. Dr. Robert Knigh Colwell, por todas as sugestões e comentários que hei de escutar.

À família como um todo, Édison, Thomaz, Gustavo, Janes, bem como todos os demais, pelo apoio e companhia. Sou grato especialmente à motivação do Luiz Antônio para me ensinar e incentivar a dar os primeiros passos no aprendizado de programação computacional, ainda quando eu era adolescente, pois sem isto esta dissertação teria outras introduções, metodologias, resultados e discussões.

Por fim, à Tora, Vega, Luma e o Titan, pela constante disposição de brincar, bagunçar e divertirem alegremente comigo. Sem dúvida estes momentos de descontração fora importantes ao longo dos últimos dois anos.

Sumário

Apresentação

Gradientes Latitudinais na Riqueza de Espécies: Hipóteses

Muito se tem debatido sobre quais são os mecanismos ecológicos que determinam os padrões espaciais de riqueza de espécies. Na verdade, esses padrões, que já são conhecidos há dois séculos (Hawkins 2001), de uma forma geral seguem a tendência do aumento no número de espécies em forma de gradiente, desde os pólos (altas latitudes) até a linha do Equador (baixas latitudes). Entretanto, foi apenas nas últimas décadas que este se tornou um problema central em ecologia geográfica e biogeografia.

O estudo de padrões espaciais globais de diversidade é complexo. Um dos pontos principais é que não há possibilidade de qualquer inferência através de técnicas experimentais ou por controle de variáveis, já que (*i*) a extensão espacial não comporta tais procedimentos, (*ii*) o número de variáveis ambientais a serem controladas é praticamente infinito, além de que (*iii*) quanto maior for a extensão (escala) espacial considerada, maior espera-se que seja a riqueza biológica. Assim, ouve um eclipse do estudo de padrões espaciais de riqueza de espécies por mais de uma década no final do último século, de tal maneira que as perguntas e hipóteses relacionadas ao estudo destes padrões ficaram esquecidas. Naquela época, além de ter que enfrentar a complexidade dos padrões e do tipo de estudo, os pesquisadores também careciam de ferramentas analíticas apropriadas.

Entretanto, um ressurgimento vigoroso destes estudos passou a acontecer nos anos 90, muito provavelmente impulsionado pelo novo programa de pesquisa em Macroecologia (Brown e Maurer 1989; Brown 1995). É preciso destacar também que avanços tecnológicos, tais como os sistemas de informação geográfica (SIGs) permitem a manipulação de dados biológicos distribuídos espacialmente. Neste caso, a extensão da área estudada, e a resolução espacial empregada no estudo, tornam-se detalhes de menor importância, ou mesmo o próprio foco do estudo (Rahbek e

Graves 2001). Além disto, o próprio dado biológico, como por exemplo a posição geográfica da área de distribuição da espécie, registros de ocorrência com detalhes taxonômicos, os acervos de museus, estão hoje sendo cada vez mais disponíveis através da rede mundial de computadores para inúmeros grupos de espécies. Isto, já hoje, e muito mais nos próximos anos, favorece o estudo dos padrões espaciais de diversidade biológica em escala espaciais amplas. Com relação ao conhecimento do próprio ambiente em si, a popularização do uso de sensoriamento remoto e a multiplicação dos algoritmos de estimativa de variáveis ecológicas por imagens de satélite, tornaram a obtenção de variáveis ambientais muito mais fácil, ao passo que antes eram obtidas apenas através de árduas medições em campo e extrapolações questionáveis. Hoje é possível encontrar os dados ambientais e climáticos necessários para avaliar as diferentes hipóteses com uma resolução temporal e espacial antes inimaginável (Kerr e Ostrovsky 2003). Além disto, os computadores permitem atualmente a utilização de ferramentas estatísticas complexas para processamento de grandes conjuntos de dados espaciais, de tal forma que os padrões de autocorrelação espacial entre as unidades podem servir para compreender os processos e mecanismos subjacentes à variação da riqueza, bem como para corrigir efeitos de correlação espúria entre riqueza e variáveis ambientais (Diniz-Filho *et al.* 2003). Modelos e simulações computacionais para avaliar hipóteses hoje podem ser criados e processados em qualquer computador doméstico, em questão de horas ou minutos (Rangel & Diniz-Filho 2003). Considerando esses avanços, é preciso também reconhecer que a análise dos padrões complexos de riqueza em grandes escalas dificilmente ainda precisa ser realizada simplesmente ao longo de uma única dimensão "latitudinal" (Hawkins & Diniz-Filho 2004). Essas novas ferramentas e o arcabouço teórico da macroecologia permitem um teste mais efetivo dos diferentes mecanismos ecológicos e históricos que devem explicar os gradientes de diversidade ou riqueza de espécies (Hawkins *et al.* 2003a, Diniz-Filho *et al.* 2004).

Estima-se que mais de uma centena hipóteses já tenham sido propostas desde o século XVIII para explicar os gradientes de diversidade em escala continental, mas muitas delas falham por basearem-se em pressupostos lógicos questionáveis, por circularidade de raciocínio ou por ruptura com as evidências empíricas (Rohde 1992; Hawkins *et al.* 2003a). Certamente, a complexidade destes padrões está ligada a fatores de causalidade múltipla, colinearidade entre as variáveis e eventos históricos difíceis de serem medidos e testados. No passado foi bastante forte o apelo para que um padrão tão comum de gradiente biológico ao longo da latitude tivesse também uma explicação ligada a uma medida ou variável comum no planeta. Ricklefs (1973) assim escreveu: "*O padrão latitudinal geral no número de espécies deve estar relacionado a algum fator climático, ou a combinação de fatores, que mudam de forma consistente com a latitude. Vários fatores poderiam servir como candidatos adequados: temperatura média, precipitação anual, sazonalidade, para dizer alguns, mas ecólogos têm falhado em encontrar um elo convincente entre diversidade orgânica e padrões no ambiente físico*". Em uma redução significativa realizada recentemente, hoje costuma-se pensar em cinco grupos de hipóteses tidas como as "mais aceitáveis" para explicar os gradientes de diversidade: clima, tempo evolutivo, heterogeneidade de habitats, disponibilidade de área e as restrições geométricas (Jetz e Rahbek 2001, Whittaker *et al.* 2001; Willig *et al.* 2003). Para algumas hipóteses em específico não há um claro grupo onde se encaixe, mas sim em dois ou mais deles.

i - Clima

A hipótese energética é tida como uma boa explicação para os padrões de diversidade e geralmente é vista como a principal hipótese climática (Hawkins *et al.* 2003b). Estudos recentes apontam para um alto poder de explicação da riqueza de espécies por variáveis medindo energia e água no ambiente, tais como temperatura (disponibilidade energia), precipitação (disponibilidade hídrica), evapotranspiração potencial (balanço hídrico/energético), produtividade primária e evapotranspiração real (energia disponível na cadeia trófica).

Neste momento é preciso fazer uma distinção entre dois mecanismos que a hipótese climática invoca para explicar padrões espaciais de diversidade. Em um primeiro momento, pode-se esperar que a “quantidade” de energia chegando ao sistema via temperatura e insolação serve aos organismos fotossintetizantes, produtores de energia, que alimentam as bases da cadeia trófica e por conseqüência todos os demais consumidores. Desta forma, seria esperado que quanto mais energia chegasse ao sistema, maior seria a disponibilidade para os produtores, alargando as bases da pirâmide energética, que suportaria mais tanto mais níveis tróficos quanto mais indivíduos e espécies. Entretanto, é também possível pensar que, para algumas espécies, a energia é capaz de ativar o metabolismo diretamente, acelerando as taxas metabólicas, o ciclos de vida, e possivelmente também as taxas evolutivas (especiação e extinção) (Currie *et al*. 2004). Certamente os organismos que mais devem responder a isto são aqueles ditos ectotérmicos. Estes seres dependem diretamente da temperatura ambiental para o funcionamento de seu metabolismo, e não seria um contra-senso pensar que, dentro de determinados limites, quanto mais energia disponível, mais “rápido” eles “funcionariam”, acelerando reprodução e taxas evolutivas. Isto tem sido bastante discutido recentemente por James H. Brown e seu grupo de trabalho, que se propuseram a criar as bases teóricas da “Ecologia Metabólica” (Brown *et al.* 2004). É interessante notar que existem evidências teóricas baseadas em cinética celular e enzimática, princípios de difusão energética, leis físicas sobre perda e acumulação de energia em forma de temperatura. Dentro deste esforço eles também mostraram que há alguma evidência empírica que leva à conclusões fortes e quantitativas (no sentido de modelos matemáticos), sendo até possível predizer a taxa de aumento no número de espécies de ectotérmicos de acordo com o aumento na temperatura (Allen *et al.* 2002; Kaspari *et al.* 2004).

Ainda no tocante à disponibilidade energética para as espécies, é intuitivo pensar organismos na verdade não sobrevivem de “disponibilidade energética” tão somente, sendo que esta na verdade se manifesta na forma de temperatura ou luz. Neste sentido Hawkins *et al*. (2003b) mostra através de uma meta-análise da bibliografia que, na verdade, é o balanço entre disponibilidade hídrica e energética que apresenta um maior poder de explicação dos padrões espaciais de diversidade biológica. Existem diversas variáveis capazes de medir este balanço, mas as mais comuns são a evapo-transpiração real e a evapo-transpiração potencial.

A primeira é uma medida direta de energia ambiental. Representa a quantidade de água que evaporou de uma superfície em uma determinada área, sabendo que para a evaporação é necessária uma condição que envolva disponibilidade de água e energia suficiente para que ocorra a evaporação. Um lugar úmido e frio é tão improvável de ocorrer evaporação quanto um lugar seco e quente. Desta forma esta variável é uma forma direta de se medir o balanço e a relação disponibilidade hídrica/energética. Em um enfoque mais metabólico, sabe-se também que este medida engloba toda aquela água que é expelida (evaporada ou gotejada) pela planta através de seus estômatos no ato de capturar moléculas de gás carbônico. Assim, a própria variável de evapo-transpiração real relaciona-se diretamente também com a transpiração do metabolismo vegetal (que é através da fotossíntese o principal responsável pela entrada de energia orgânica na cadeia trófica), pois quanto maior for a evapo-transpiração local, maior a possibilidade de que as plantas que ali estão realizarem fotossíntese. Esta por sua vez se liga aos mesmos fatores que a evapo-transpiração mede, que são a disponibilidade de energia na forma de calor (para dar sustentação ao acontecimento metabólico), a disponibilidade de energia na forma de luz (para ativar a fotossíntese através de seus intricados caminhos enzimáticos), e a disponibilidade de água no sistema (tanto através pela umidade do solo, de onde a planta retira nutrientes e água, quanto pela umidade relativa do ar, que permite a planta abrir seus estômatos para respiração e captação de gás carbônico).

A evapo-transpiração potencial é na verdade uma variável estimada, sendo impossível de ser determinada na prática. Ela está mais diretamente relacionada com a disponibilidade energética por si só, já que representa o quanto de água teria sido evaporado e transpirado caso houvesse uma superfície sempre saturada. Neste caso, assumindo a existência de uma superfície sempre saturada, a quantidade de água transpirada e/ou evaporada dali é determinada unicamente pela temperatura daquela água, que por sua vez é determinada pela quantidade de energia térmica ambiental.

Um outro conjunto de variáveis também bastante utilizadas para estas medições de energia ambiental são aquelas ditas variáveis de produtividade. A primeira delas, a Produtividade Primária Bruta, representa o quanto de moléculas orgânicas foi sintetizado pelos organismos produtores (autótrofos), normalmente em escala de peso/área/tempo. Já que o anabolismo vegetal é em grande parte determinado pelo balanço entre a disponibilidade hídrica e energética, é possível inferir que a Produtividade Primária será determinada pelas duas variáveis discutidas anteriormente, bem como relacionada com a condição ambiental de temperatura, iluminação. Estas moléculas orgânicas são os pilares da pirâmide energética, e da entrada de energia na cadeia trófica pelos produtores, responsáveis por "sustentar" todo o consumo das demais espécies. Entretanto, nem tudo que os produtores sintetizam pode ser considerado energia disponível para os demais indivíduos e espécies, já que eles mesmos consomem parte disto. Esta porção de energia que fica de fato disponível para os demais, retirando-se o que os próprios produtores consomem para seu sustento, é conhecida como Produtividade Primária Líquida, geralmente representada na mesma escala da anterior. Desta forma, produtividade primária líquida passa a ser a variável de maior interesse e mais direta interpretação na influência e na abundância e diversidade das demais espécies consumidoras.

Em um estudo recente, Currie *et al.* (2004) estudam várias hipóteses climáticas (disponibilidade energética, tolerância fisiológica à variação climática, e taxas de especiação) e os mecanismos biológicos que cada uma delas invoca para explicar como a diversidade de espécies se organiza ou é determinada pelo ambiente. Sua base metodológica consiste em uma vasta revisão da literatura na busca de evidências empíricas que dêem suporte às várias facetas desta hipótese, bem como a dedução de predições a partir de conhecimentos prévios de outras áreas da biologia. Para a hipótese de disponibilidade energética determinando a diversidade de espécies, mostram que não há evidências empíricas de um mecanismo biológico que faça com que uma alta densidade de indivíduos implique em alta diversidade de espécies, já que a única relação causal que podemos traçar a partir de disponibilidade energética é com a capacidade de suporte de indivíduos em uma determinada área. Por outro lado, também discutem que a hipótese de tolerância fisiológica das espécies, postulada na forma de que a riqueza de espécies em uma determinada região é condicionada ao número de espécies capazes de suportar a condição ambiental ali presente. Segundo estes autores, há evidências de que a tolerância climática determine o número de espécies capazes de sobreviver à uma determinada região, porém eventualmente é possível encontrar espécies que, apesar de tolerar tais condições ambientais, encontram-se ausentes. Por último, há um apelo de que a disponibilidade energética, principalmente na forma de temperatura, possa influenciar as taxas de evolutivas, já que aceleraria as taxas de mutação e processos fisiológicos, bem como encurtaria o tempo de geração. Assim, o balanço entre as taxas de especiação e extinção deveria ser maior em direção aos trópicos do que aos pólos, para que o padrão esperado de riqueza de espécies fosse o que observamos hoje. Entretanto, não há evidências empíricas que dê suporte aos mecanismos invocados por esta hipótese.

ii - Tempo Evolutivo

A hipótese de tempo evolutivo parte de afirmações e pressupostos lógicos e muitas vezes bastante coerentes, porém difíceis ou até impossíveis de serem testados. Em um primeiro momento poderemos entender este “tempo” como sendo o tempo em que uma determinada área existe ou manteve-se sem nenhum distúrbio, ou seja, o tempo em que a comunidade que ali existe está sem nenhuma influência de catástrofe ambiental. Segundo o esperado, a diversidade de uma comunidade seria tão mais rica quanto há mais tempo ela existe (hipótese da disponibilidade temporal, ou tempo-desde-origem). Assim, alguns hábitats em específico poderiam até abrigar mais espécies do que abrigam atualmente (não saturados), apesar de que não tenha havido tempo suficiente para a especiação/evolução (tempo evolutivo), ou colonização (tempo ecológico) total destes habitats (Krebs 2001). Para dar suporte a esta hipótese podemos pensar em áreas que recentemente foram vítimas de distúrbios fortes, como incêndios, tempestades ou que foram invadidas por glaciações mostrariam um claro empobrecimento biológico. Entretanto, este teste tenderia muito mais à uma hipótese de tempo ecológico do que evolutivo em si, sendo que rapidamente após o distúrbio indivíduos presentes no *pool* regional poderiam logo migrar para a área, igualando novamente a diversidade na comunidade. Talvez o máximo do conhecimento histórico que tenhamos sobre uma determinada área não seja suficiente para definirmos se ela é ou não mais antiga que uma outra área, já que este conhecimento se dá em escala de tempo geológico.

Por outro lado, poderíamos assumir que as taxas de extinção e especiação são constantes ao longo do espaço, e pensar que para alguns clados este "tempo evolutivo" corre mais rápido (velocidade evolutiva) do que para outros (Diniz-Filho *et al.* 2004). Assim, é possível fazer duas predições. Em primeiro lugar, se as taxas evolutivas forem iguais para todos os clados, aqueles que forem mais antigos deverão ser mais diversos, já que terão tido maior oportunidade de especiação. Entretanto este é um tema extremamente nebuloso, já que o resgate da história evolutiva de um clado é complicado pela falta de evidências sobre o passado, e não há evidências de que as taxas evolutivas sejam constantes entre diferentes clados. Além disto, teríamos que assumir que as taxas de extinção e especiação são constantes ao longo do espaço e que variações ambientais não influenciem nas taxas de especiação. Também seria preciso saber quais são os grupos mais recentes e os mais antigos, de tal forma que pudéssemos correlacionar idade evolutiva com diversidade taxonômica.

Por outro lado, poderíamos supor que características biológicas intrínsecas aos clados determinem suas próprias taxas evolutivas. Neste caso, por exemplo, é possível assumir que espécies de menor tamanho corporal possuam um tempo de geração mais curto, de tal forma que favoreça o acontecimento de mutações, o que por conseqüência levaria a especiação como um evento mais comum do que para espécies de grande porte. Neste caso, espera-se que, por exemplo, a velocidade evolutiva de um clado seja determinada pelo tamanho do corpo das espécies que o compõe.

Uma outra possibilidade seria assumir que as taxas de evolução são determinadas por padrões espaciais, como, por exemplo, temperatura ambiental. Neste caso, observamos um elo entre duas hipóteses, sendo elas a hipótese climática e hipótese evolutiva. Conforme o que foi discutido anteriormente, não há evidências suficientes de que fatores climáticos determinem taxas evolutivas.

É sempre um grande desafio medir tempo evolutivo, que dirá resgatar isto do passado. Então, como esta hipótese envolve tempo, seria de suma importância para o seu teste que tenhamos uma forma de medir o tempo passado, ou história, seja da região, seja do próprio clado. Esta talvez seja uma das hipóteses que mais careça de evidências empíricas, mas não por haver muitas evidências contra ela, mas por ser algo extremamente difícil (ou até impossível) de ser medido com precisão.

iii - Heterogeneidade de Habitats

Um padrão geral de aumento da complexidade ambiental pode ocorrer na medida em que se move a partir dos pólos e em direção aos trópicos. Os ambientes mais heterogêneos e complexos seriam capazes de oferecer um nicho potencial mais amplo, e neste caso abrigar uma maior variedade de espécies de plantas e animais, constituindo uma comunidade mais rica tanto em escala local como regional. Entretanto esta é provavelmente uma hipótese circular, já que supõe que ambientes heterogêneos e complexos abrigam uma maior diversidade, que por sua vez causa um ambiente mais heterogêneo e complexo. Onde neste ciclo inicia-se a causa está por ser elucidado. Ademais, a falta de um mecanismo ecológico que aponte a causa da heterogeneidade ambiental nos trópicos, aliado à falta evidências empíricas, tem ofuscado esta hipótese nos últimos anos.

Isolamento geográfico de determinadas áreas pode de fato gerar uma diversificação (heterogeneidade) ambiental ao longo do tempo. Entretanto, não há muitos modelos preditivos para o mecanismo de sucessão ecológica após o isolamento de áreas, e provavelmente o mecanismo mais atuante será de fato a pressão de seleção e adaptações ambientais das espécies que se isolaram. Com o tempo o isolamento deve levar à especiação, que aumenta a heterogeneidade. Entretanto, mais uma vez, isto parece ser muito mais a conseqüência do que a causa de uma diversidade mais alta quando olhada em escala regional.

Algumas porções tropicais abrigam a maior parte da variação altitudinal do planeta, provendo uma variedade de habitats de quente para temperado, e de temperado para frio, ao longo da variação de altitude em grandes cadeias de montanhas. De fato, na América do Sul, a região com maior riqueza de espécies está ao longo da cadeia montanhosa dos antes. A diminuição do número de espécies com o decréscimo da temperatura em altas altitudes é tão conspícuo quanto com o decréscimo da latitude, e por isto se diz que os gradientes altitudinais espelham os latitudianis. Entretanto, para a maioria dos aumentos altitudinais, basicamente dados em formas de cordilheiras e montanhas, é preciso pensar que ocorre a diminuição da área quase que diretamente proporcional à diminuição da temperatura, o que torna os dois fatores quase indissociáveis.

Klopfer e MacArthur (1961), sugeriram que "o mais importante fator que cause o aumento tropical no número de espécies de aves não é nem somente o aumento da complexidade de habitat, como também não é o aumento da especiação, mas sim o aumento da similaridade de espécies coexistentes, refletido por um deslocamento de caracteres reduzido, e portanto sobreposição de nicho." Esta é uma sugestão que está intimamente relacionada com outro mecanismo também sugeridos para explicar a alta diversidade de espécies nas regiões tropicais. De acordo com uma outra hipótese, o "empacotamento" de nichos nestas regiões seria efeito da alta energia disponível e adaptação evolutiva que as espécies sofreriam para que pudessem se especializar no aproveitamento de recursos. Assim, tanto a complexidade ambiental como alta especialização biológica permitiram que mais espécies coexistissem em uma mesma região, desde que houvesse alta disponibilidade energética.

De uma forma geral é também esperado, segundo a hipótese da variabilidade climática, que regiões com climas mais estáveis sejam mais favoráveis às taxas evolutivas, por meio de especializações finas e adaptações locais que causam nichos mais estreitos. Desta forma, espera-se que as espécies sejam mais flexíveis e generalistas nas regiões temperadas do que nos trópicos, já que estas últimas devem conviver com competidores, tornando-se cada vez mais especialistas.

Ao contrário, pode-se também esperar que exista um certo nível de distúrbio que maximize a diversidade. Isto funcionaria já que quando os distúrbios ocorrem com grande freqüência e força, algumas espécies se extinguirão se não tiverem altas taxas de crescimento. Entretanto, se não houver distúrbio algum, o sistema é levado ao equilíbrio competitivo, e espécies com baixas capacidades competitivas são também perdidas. Desta forma, se as taxas de crescimento das espécies forem mantidas igualmente baixa entre todas as espécies, o equilíbrio competitivo não é atingido, diminuindo os efeitos da competição (Krebs 2001). Entretanto esta hipótese tem se mantido até hoje no campo das conjecturas, e carece de evidências empíricas fortes. Não há um padrão claro de que ocorra uma rarefação de populações ao longo da latitude e com aumento em direção aos pólos, causado por eventos físicos como tempestades, secas, frio. Segundo esta hipótese, estes distúrbios repetidos criariam diferentes ambientes e uma sazonalidade capaz de abrigar diferentes espécies. Entretanto, não se observa este padrão tão claramente na natureza.

iv - Disponibilidade de Área

Desde muito tempo pensa-se que o tamanho da área disponível pode estar relacionado com o número de espécies que pode ali sem encontrado (Terborgh 1973). O próprio modelo de equilíbrio de biogeografia insular proposto por MacArtur e Wilson (1967), usa esta idéia intuitiva como pressuposto. Hoje o maior defensor desta hipótese é Michael Rosenzweig (ver Rosenzweig 1995).

A hipótese oferece um mecanismo, propondo que quanto maior a disponibilidade de área, maior seriam as áreas de distribuição geográfica das espécies, que por sua vez seriam menores na medida em que a diversidade aumenta (Rosenzweig 2003). Ainda, espécies com amplas distribuições geográficas teriam tanto altas taxas de especiação quanto baixas taxas de extinção. Isto por si só a oferece um mecanismo de ajuste por retro-alimentação negativa que poderia levar um estado de balanço dinâmico entre o tamanho da área e a diversidade.

Para que esta hipótese seja aceita, alguns pontos ainda carecem de evidências empíricas, como por exemplo o fato de que (i) o tamanho das distribuições geográficas exerce um papel importante na taxa de diversificação, (ii) que uma riqueza de espécies elevada em uma região acaba por gerar maior "empacotamento" das distribuições geográficas, tornando-as menores, e (iii) que espécies com áreas de distribuições geográficas mais restritas apresentam taxas de especiação menores.

Entretanto, a porção tropical das Américas tem uma área menor que a porção temperada, sendo isto também aplicável à Ásia. Assim, talvez para casos em grandes escalas espaciais até pudéssemos esperar uma relação inversa, já que nestes dois exemplos a porção tropical é bastante mais rica que a temperada. Na verdade nem toda porção tropical do planeta tem regiões maiores que as zonas frias ou temperadas, e talvez área não seja uma explicação geral para estes gradientes.

Existem evidências empíricas contraditórias para esta hipótese, mas ela por si só baseia-se em um mecanismo não tão direto quando as demais, sendo também de difícil teste. Como, senão por métodos de simulação computacional, poderemos medir e comparar o efeito de taxas de diversificação entre espécies, ao longo do tempo, para sabermos se o tamanho da distribuição geográfica está relacionado e por conseqüência poderia levar a uma variação na diversidade?

v - Restrições Geométricas e o Efeito do Domínio Médio

Entre os cinco grupos de hipóteses mais aceitas atualmente, a mais recente e sem dúvida mais controversa é a hipótese das restrições geométricas, que advoga a existência do chamado Efeito do Domínio Médio (*Mid-Domain Effect* - MDE) (Colwell & Lees 2000). Segundo esta hipótese, o gradiente no padrão espacial de diversidade é ocasionado por restrições geométricas de como (disponibilidade de área) e onde (posicionamento geográfico) as espécies localizam-se dentro das bordas do domínio (continente) no qual estão contidas. Assim, segundo esta hipótese, independentemente de todos os fatores evolutivos e ambientais, o padrão espacial esperado seria um aumento de diversidade das bordas em direção a um pico máximo no centro do continente, e isso se daria por efeitos puramente aleatórios criados pela sobreposição das áreas de distribuição das espécies em direção ao centro do domínio geográfico compartilhado por elas. Colwell & Lees

(2000) propuseram o MDE como um modelo nulo ao qual o padrão observado possa ser comparado, mas também advogam que muitos padrões observados são função do MDE, de modo que explicações baseadas em clima e história tornam-se supérfluas e menos parcimoniosas (ver Colwell *et al.* 2004, para uma revisão recente).

Entretanto, para uma ruptura com o pensamento majoritário entre os ecólogos e os biogeógrafos nos últimos 200 anos é de se esperar uma evidência que seja tão forte que a justifique. Entretanto, os valores de correlação entre a riqueza de espécies observada e o esperado de acordo com o modelo nulo para o padrão espacial de riqueza de espécies é em geral baixo (Diniz-Filho *et al.* 2002; Rangel e Diniz-Filho 2003; Rangel e Diniz-Filho 2004), especialmente aqueles que consideram contextos bidimensionais e não simplesmente variações ao longo de transectos latitudinais e/ou altitudinais (Zapata *et al.* 2002). Em um primeiro momento, poder-se-ia pensar que a rejeição do MDE indica simplesmente que as distribuições não estão posicionadas ao acaso, de modo que o ambiente (clima) determina os padrões de riqueza (Diniz-Filho *et al.* 2002). Por outro lado, Hawkins e Diniz-Filho (2002) e Zapata *et al.* (2003) mostraram também que o próprio pressuposto do modelo nulo, de que as distribuições podem ser aleatorizadas sobre o espaço, não pode ser assumido simultaneamente à ausência de efeitos ambientais, pois nessa situação (ausência de efeitos) essas distribuições se expandiriam até ocupar todo o continente, de modo que a riqueza seria constante ao longo do espaço geográfico. Mais recentemente, os proponentes e defensores desta hipótese argumentaram que ela requer tratamento especial nas análises estatísticas, sendo mais indicado seu uso como uma variável adicional em um modelo multivariado que inclua também outras variáveis climáticas e históricas, de forma que sua importância relativa poderá ser comparada o poder de predição do modelo com e sem ela (Colwell *et al.* 2004).

A despeito de sua utilidade como hipótese científica válida, e de todas as críticas que surgiram, o MDE trouxe à tona uma discussão importante: qual seria a expectativa nula (em sentido estatístico) para o padrão espacial de riqueza em grandes escalas? Esta é uma pergunta que ainda permanece sem resposta, e os próximos estudos talvez tentem respondê-la.

Apesar de toda discussão que o circula o MDE, este tem o mérito inspirar ecólogos e biogeógrafos a construir modelos que predizem um padrão espacial de biodiversidade. A partir de sua proposta original (Colwell & Lees 2000), alguns modelos já foram criados, como por exemplo o modelo de restrições geométricas de Jetz e Rahbek (2001), que aleatoriza a posição espacial da área de ocorrência da espécie, mantendo-o espacialmente coeso, com área total constante e restrito às bordas do continente. Posteriormente, ainda inspirados no MDE, porém afastando-se de seus pressupostos, Rangel e Diniz-Filho (2005) incorporaram ao modelo de restrição geométrica o padrão espacial do próprio ambiente, criando assim um modelo evolutivo no qual a tolerância fisiológica das espécies interage com a condição ambiental para determinar o tamanho da área de distribuição das espécies, sua posição no domínio. Isto tudo se dá enquanto ainda é possível variar e estudar a influência da "força" do ambiente, em relação à tolerância das espécies, na determinação da predição dos padrões espaciais de riqueza. Surpreendentemente o que encontraram foi que em condições de alta tolerância/baixa força ambiental, um padrão parabólico similar ao esperado pelo modelo nulo do MDE surge espontaneamente, o que sugere que este padrão seja não apenas uma expectativa nula, como também o esperado para condições em que as espécies podem ocupar grande parte dos biomas por terem tolerâncias fisiológicas altas, ou condições de homogeneidade ambiental ao longo de um único bioma.

vi - Desenvolvimentos Recentes

Mais recentemente, um novo mecanismo para explicar os padrões de riqueza de espécies em grandes escalas espaciais surgiu na literatura, no contexto dos modelos evolutivos. A chamada "conservação de nicho" (Wiens 2004, Wiens e Donoghue 2004) é a influência da inércia evolutiva sobre os caracteres das espécies que definem seu nicho ecológico. Assim, durante a evolução das espécies, e seu respectivo mecanismo de especiação, a maioria dos clados tendem a manter características peculiares mais ou menos constantes entre espécies. Desta forma, se isto de fato acontece, as espécies descendentes tendem a manterem-se não só ecologicamente, mas também espacialmente próximas das espécies ancestrais, já que se não há modificações suficientes para a evolução do nicho, os mesmos fatores ambientais que impediam a espécie ancestral de colonizar novas áreas e expandir sua distribuição geográfica, continuará por também impedir as espécies descendentes.

Desta forma, já que as regiões tropicais são mais estáveis ao longo do tempo, bem como também possuíram (e possuem) uma maior extensão geográfica, é possível admitir que muitos clados tenham surgido nas regiões tropicais. Então, se as espécies originaram-se tropicais e se especializaram a tais condições ambientais, e se a conservação de nicho é suficientemente forte, as espécies descendentes permaneceram, em sua maioria, também tropicais. Entretanto, ainda que lentamente, a evolução do nicho levou algumas espécies a colonizar regiões de clima temperado e/ou seco, porém muito tempo depois do surgimento do clado. Este mecanismo então poderia ter gerado os padrões de riqueza de espécies que hoje observamos na natureza, e é o foco do estudo do primeiro capítulo deste texto.

Por fim, processos evolutivos subjacentes aos padrões espaciais de riqueza de espécies ainda estão sobre discussão, e que falta aos estudos correlativos os fundamentos teóricos para explicar mecanisticamente os componentes evolutivos destes padrões. No capítulo primeiro deste texto, eu desenvolvi um modelo de simulação espacialmente explicito para avaliar, sob um enfoque de modelagem orientada por padrões, se o processo de conservação de nicho fornece uma explicação mecanistica para os padrões de riqueza de espécies. Meu modelo define o tamanho, forma e posição da distribuição geográfica das espécies sobre uma paisagem ambiental multivariada e heterogênea, através da simulação de processos evolutivos na qual a flutuação na tolerância ambiental das espécies cria fragmentação da distribuição geográfica e portanto regula os eventos de especiação e extinção. Eu também testei se um viés nas taxas líquidas de diversificação poderia aumentar a habilidade do modelo em predizer os padrões observados no mundo real. Executei modelagens no domínio da América do Sul, buscando o conjunto de parâmetros que maximizasse a correspondência entre os padrões de riqueza de espécies simulados e os observados para a riqueza de 3088 espécies de aves. O processo de conservação de nicho prediz padrões que muito se assemelham aos observados no mundo real, apesar de que o viés de diversificação aumenta o ajuste do modelo. Meu modelo de simulação permite a avaliação do papel de diferentes processos na explicação dos padrões espaciais de riqueza de espécies, relevando o enorme potencial da união entre ecologia e biogeografia histórica sob arcabouços teóricos e metodológicos.

Padrões Espaciais de Biodiversidade: Análise Estatística Espacial

Muito comumente ecólogos se deparam com a tarefa de mensurar e analisar variáveis distribuídas espacialmente. Entretanto, o fato das observações guardarem entre si uma relação espacial (distância ou proximidade, similaridade ou distinção) traz provavelmente conseqüências adicionais, tanto em âmbito metodológico quanto para a própria interpretação dos resultados estatísticos (Berry e Marble 1968; Cliff e Ord 1973; Cressie 1993; Dale et al. 2002; Diniz-Filho et al. 2003; Griffith 1987; Legendre 1993; Legendre e Fortin 1989; Legendre e Legendre 1998; Legendre et al. 2002; Lennon 2000; Moran 1950; Rossi et al. 1992; Sokal e Oden 1978a, b; Sokal et al. 1998b). Na verdade, a maior parte dos fenômenos naturais, e entre estes estão a maioria daqueles importantes em estudos ecológicos, tendem a se agrupar e possuir alguma estrutura espacial, fazendo com que as observações guardem relações estatísticas entre si (autocorrelação espacial) (Koenig 1999). Por isto, e outros motivos, existem os estudos cartográficos e a elaboração de mapas, que servem também para evidenciar o padrão espacial de uma variável, seja mostrando sua estrutura contínua e em gradiente, ou um padrão de agregação espacial em “manchas”.

Autocorrelação espacial pode ser definida, de forma geral, como a propriedade de certa variável em ser mais similar (autocorrelação positiva) ou distinta (autocorrelação negativa), dada certa distância espacial. Embora a escala do estudo influencie bastante conforme a variável, é possível encontrar autocorrelação espacial em virtualmente qualquer escala, desde micrômetros até distâncias continentais (Legendre e Fortin 1989; Legendre 1993). Entretanto, para a maioria dos testes de hipóteses clássicos, tais como os comumente usados por ecólogos em métodos estatísticos como correlações, análises de regressão ou análises de variância, é preciso que haja independência entre as observações (Sokal e Rohlf 1995; Zar 1999). Isto significa que a partir de uma ou de várias observações seja improvável predizer o valor de qualquer outra, independentemente de sua posição geográfica ou distância em relação às demais.

Mais especificamente, o número de graus de liberdade (ν) assumido pelo método estatístico em geral é baseado no número de observações (n) que o conjunto de dados possui, para que desta forma possa ser calculado a significância da estatística em questão (em geral o P). Este P expressa a significância da estatística calculada, ou seja, a probabilidade com que se espera que aquela estatística tenha sido obtida simplesmente ao acaso. P é calculado baseado em ν, que por sua vez usa n em sua estimativa, e assim, quanto maior for n, mais informação a respeito da população estatística será assumido existir, e desta forma, assumindo também que mais representativa a amostragem será da população. Na verdade o pressuposto lógico é que quanto mais conhecimento houver sobre o que se está medindo, mais confiança poderá ser colocada nos resultados trazidos pela análise. Desta forma, é preciso que cada uma das observações do conjunto amostral seja de fato independente, não guardando nenhuma relação entre si. Só assim realmente a confiança nestes resultados poderá ser aumentada conforme o número de observações, já que quanto mais observações existir, maior será o conhecimento real a respeito do que foi medido. Caso este independência não exista, os resultados estatísticos serão expressos com certo grau de contaminação "falaciosa".

O problema causado pela autocorrelação espacial é a deturpação dos valores de P, ou seja, na confiança atribuída aos testes de significância dos coeficientes estatísticos. Isto acontece, por exemplo, quando há uma dada quantidade de observações medidas no espaço, mas que, porém, guardem relações de similaridade ou dissimilaridade entre si. Desta forma, cada observação em si não adiciona conhecimento real à amostragem conforme o esperado, pois são redundantes, ou pseudo-réplicas, em relação à informação já existente (Hulbert 1984). Se assim for, por exemplo, é possível que observação pudesse ter sido, com certo grau de precisão, simplesmente estimada a partir do conhecimento já oferecido pelas demais observações. No caso do uso de um método estatístico convencional para analisar um conjunto de dados espacialmente autocorrelacionados, o que acontecerá é que ν será incorretamente estimado, pois não existe de fato aquele nível de conhecimento expresso por n. A estimativa incorreta de ν aumenta (infla) fatalmente a confiança estimada para a estatística (em geral em valores de P), porém de forma errônea (Diniz-Filho *et al.* 2002).

Os valores de P, por sua vez, são de vital importância na interpretação das estatísticas (Sokal e Rohlf 1995; Zar 1999). Eles expressam qual a probabilidade do resultado obtido ter ocorrido simplesmente ao acaso, ou seja, ser um resultado comum o suficiente para que não haja motivo para se rejeitar a hipótese nula. A presença de autocorrelação espacial em dados, quando analisados de forma comum e sem levar isto em consideração, faz com que seja mais provável a ocorrência de uma inferência estatística equivocada. O caso de inferência equivocada que a autocorrelação espacial aumenta a possibilidade de ocorrer é conhecido como Erro Tipo I, ou mais especificamente, o aumento da probabilidade de rejeição da hipótese nula quando, na verdade, ela deveria ter sido aceita. Como em geral o interesse dos cientistas, é rejeitar a hipótese nula e mostrar que de fato existe algum fenômeno biológico (ou ecológico) em atuação, a falsa estimativa do número de graus de liberdade pode por vezes convir com as expectativas e anseios pessoais, fazendo com que Erros Tipo I sejam muito comuns (Lennon 2000).

Já há muitos anos geógrafos, geólogos e também biólogos vêm estudando a autocorrelação, suas causas e conseqüências, e criando ferramentas estatísticas tanto para medi-la quanto para explorá-la analiticamente. Contudo, estas ferramentas têm chegado de forma muito deficitária até o público acadêmico geral, seja por falta de divulgação destas metodologias estatísticas em periódicos de grande penetração e de forma acessível, ou também pela falta de um programa computacional integrado (*software*) para os cálculos matemáticos (mas veja Sawada 1999).

Uma das atividades que desenvolvi ao longo do curso de mestrado em Ecologia e Evolução foi criar um programa computacional, de distribuição livre, que integra várias destas ferramentas estatísticas, principalmente aquelas destinadas à análise de dados de ecologia geográfica e/ou biogeografia (*surface pattern analysis*). Desta forma, é possível estudar corretamente o poder de explicação de cada uma das hipóteses mencionadas anteriormente para os padrões espaciais de riqueza de espécies, e fazer inferência sobre elas enquanto se leva o padrão de autocorrelação espacial em consideração.

O programa computacional foi nomeado SAM, em abreviação de *Spatial Analysis in Macroecology*. Este foi escrito em linguagem Delphi 5, que é apropriada para desenvolvimento de aplicações de médio e grande porte, que possui uma grande versatilidade computacional e integração completa com as plataformas Windows. Esta linguagem computacional possui versões de livre distribuição, e permite a geração de arquivos de programas *stand-alone*, ou seja, que são independentes dos compiladores originais da linguagem-mãe. Isto facilita a distribuição do programa, já que ele será gratuito.

Para a divulgação de tal programa, foi escrito um artigo científico que anuncia a sua existência e vantagens. Este artigo é o capítulo dois desta dissertação, e já foi aceito para publicação depois de ser submetido à revista *Global Ecology and Biogeography* e ter passado pelo devido processo de revisão. A publicação do mesmo deve acontecer em algum momento do ano de 2006.

Na presente data (24 de janeiro de 2006), cópias eletrônicas do SAM já haviam sido distribuídas 136 vezes para cientistas e pesquisadores de todos os continentes do planeta, através do *site* oficial do SAM (www.ecoevol.ufg.br/sam), que está disponível na internet desde 25 de agosto de 2005. É também interessante ressaltar que o número de novos usuários cresce a cada dia, sendo atualmente, em média, cinco a cada dia.

Referências

Allen, A.P., Brown, J.H. e Gillolly, J.F. 2002. Global Biodiversity, Biochemical Kinetics, and the Energetic-Equivalence Rule. Science, 297: 1545-1548.

Brown, J.H., Gillolly, J.F., Allen, A.P., Savage, V.M. e West,G.B. 2004. Towards a Metabolic Theory of Ecology. Ecology, 85: 1771-1789.

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

Colwell, R.K., Rahbek, C., e Gotelli, N.J. 2004. The Mid-Domain Effect and Species Richness Patterns: What Have We Learned So Far? American Naturalist, 163: E-Paper.

Cressie, N. A. C. 1993. Statistics for Spatial Data. Revised Edition (Wiley series in probability and mathematical statistics. Applied probability and statistics section), Wiley, New York.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E.M. e Turner, J.R.G. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonimic richness. Ecology Letters, 7: 1121-1134.

Dale, M. R. T., Dixon, P., Fortin, M-J., Legendre, P., Myers, D. E. e Rosenberg, M. S. 2002. Conceptual and mathematical relationships among methods for spatial analysis. Ecography, 25: 558-577.

Davis, J. C. 1986. Statistics and Data Analysis in Geology. Second Edition, Wiley, New York.

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

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

Fotheringham, A. S., Brunsdon, C. and Charlton, M. 2002. Geographically Weighted Regression: the analysis of spatially varying relationships. Willey, New York

Hawkins, B.A. 2001. Ecology’s Oldest Pattern? Trends in Ecology & Evolution, 16: 470.

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: 3105-3117

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

Hawkins, B.A. e Diniz-Filho, J.A.F. 2003. 'Latitude' and geographic patterns in species richness. Ecography, 27: 268-272.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs, 54: 187-211.

Jetz, W., e Rahbek, C. 2001. Geometric constraints explain much of the species richness pattern in African birds. Procedings of National Academy of Science - USA, 98: 5661-5666.

Kaspari, M., Ward, P.S., Yuan, M. 2004. Energy gradients and the geographic distribution of local ant diversity. Oecologia, 140: 407-413.

Kerr, J.T. e Ostrovsky, M. 2003. From space to species: ecological applications for remote sensing. Trends in Ecology & Evolution, 18: 299-305.

Koening, W. D. 1999. Spatial autocorrelation of ecological phenomena. Trends in Ecology & Evolution, 14: 22-26.

Klopfer, P. H. e MacArthur, R. H. 1961. On the causes of tropical species diversity: niche overlap. American Naturalist, 95: 223-226.

Krebs, C. J. 2001. Ecology: The experimental analysis of distribution and abundance. 5th. ed. Addison-Wesley.

Kucharik, C. J., Foley, J.A., Delire, C., Fisher, V.A., Coe, M.T., Lenters, J., Young-Molling, C., Ramankutty, N., Norman, J.M., e Gower, S.T. 2000. Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance and vegetation structure. Global Biogeochemical Cycles 14: 795-825.

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

Lengendre, P. 1993. Spatial Autocorrelation: Trouble or New Paradigm? Ecology, 74: 1659-1673.

Legendre, P. e Fortin, M-J. 1989. Spatial pattern and ecological analysis. Vegetatio, 80: 107-138.

Legendre, P. e Legendre, L. 1998. Numerical Ecology. Second English Edition. Elsevier, New York.

Legendre, P. e Troussellier, M. 1988. Aquatic heterotrophic bacteria: modeling in the presence of spatial autocorrelation. Limnology and Oceanography 33: 1055–1067.

New, M.G., Hulme, M., e Jones, P.D. 1999. Representing 20th century space-time climate variability. I: Development of a 1961-1990 mean monthly terrestrial climatology. J. Climate. 12: 829-856.

NOAA. National Oceanic and Atmospheric Administration and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

Perry, J. N., Liebhold, A. M., Rosenberg, M. S., Dungan, J., Miriti, M., Jakomulska, A., e Citron-Pousty, S. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography, 25: 578-600.

Rahbek, C. e Graves, G.R. 2001. Multiscale assessment of patterns of avian species richness. Procedings of the National Academy of Science – USA, 98: 4534-4539.

Rangel, T.F.L.V.B e Diniz-Filho, J.A.F. 2003. Spatial patterns in species richness and the geometric constraint simulation model: a global analysis of mid-domain effect in Falconiformes. Acta Oecologica, 24: 203-207.

Rangel, T.F.L.V.B e Diniz-Filho, J.A.F. 2004. Worldwide patterns in species richness of Falconiformes: analytical null models, geometric constraints, and the mid-domain effect. Brazilian Journal of Biology, 64: 299-308.

Rangel, T.F.L.V.B e Diniz-Filho, J.A.F. 2005. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography, 28: 253-263.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos, 65: 514-527.

Sokal, R. R. e Rohlf, F. J. 1995. Biometry. Third Edition. Freeman, New York.

Whittaker, R.J., Willis, K.J., e Field, R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography, 28: 453-470.

Willig, M.R., Kaufman, D.M., Stevens, R.D. 2003. Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. Annual Review of Ecology and Systematics, 34: 273-309.

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

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

Capítulo 1

Article:

Species Richness and Evolutionary Niche Conservatism Hypothesis: a Spatial Pattern-Oriented Simulation Experiment

Thiago Fernando L. V. B. Rangel
t.rangel@terra.com.br or tf.rangel@gmail.com

and

José Alexandre Felizola Diniz-Filho
jafdiniz@uol.com.br or diniz@icb.ufg.br

Departamento de Biologia Geral, ICB, Universidade Federal de Goiás
CP 131, 74.001-970, Goiânia, GO, Brasil

Keywords
niche conservatism, spatial patterns, species richness, birds, South America, simulation model

Abstract

Evolutionary processes underlying the spatial patterns in species richness are still under discussion, and correlative studies lack the theoretical basis to explain mechanistically the evolutionary components of these patterns. In this paper, we developed a spatially explicit simulation model to evaluate, under a pattern-oriented modeling approach, if niche conservatism process provides a mechanistic explanation for patterns in species richness. Our model defines species' geographical range size, shape and placement over a multivariate heterogeneous environmental landscape by simulating an evolutionary process in which fluctuations in the species' tolerance to environment create geographic range fragmentation that regulates speciation and extinction. We also tested if a bias in net diversification rates could increase the model ability to predict real-world patterns. We ran the model on the South American domain, maximizing the correspondence between simulated and real-world patterns in richness of 3088 bird species. Niche conservatism process predicts patterns that closely resemble real-world ones, although a bias in diversification rates improved model fit. Our simulations allow evaluating the role of different processes explaining spatial patterns in species richness, revealing the enormous potential of the link between ecology and historical biogeography under integrated theoretical and methodological frameworks.

Introduction

As early as the 18th and 19th centuries, naturalists described and documented what we call today the latitudinal gradient in taxon diversity (species richness), which is a general global pattern of increase in species richness towards warm and wet tropical regions (Gaston 2000; Hawkins et al. 2003b; Hillebrand 2004). Initial hypotheses explaining this pattern were deduced solely by observing and describing nature and were based on merely intuitive correspondence between climatic and biological patterns (Hawkins 2001). However, even after 200 years of research in biogeography and ecology, the most common framework used in such investigations still relies on statistical measurements of the concordance between the spatial patterns in species richness and multiple environmental factors. Since there is generally a strong relationship between species richness and some of these environmental factors (e.g. Hawkins et al. 2003a, b), this approach has led researchers to conclude that environment (e.g. climate) is the main driver of species richness. Many variations of the environmental hypotheses, based on the effect of different environmental variables and on how these variables could affect organismal metabolism or population dynamics, have been developed to explain how environmental factors mechanistically drive species richness (see Allen et al. 2002; Currie et al. 2005; Ricklefs 2004; Rohde 1992; Whittaker et al. 2001; Willig et al. 2003). However, it is still difficult to understand the evolutionary components underlying these correlations between climate variation and species richness.

Surprisingly, climate-based and evolution-based hypotheses have developed since the early 1960’s (Ricklefs 2004) as almost independent sets of explanations. In general, evolutionary hypotheses invoke a latitudinal bias in diversification rates (see Cardillo et al. 2005, for a recent evaluation) driven by area availability (Rosenzweig 1995), time since the origin of lineages (Farrell and Mitter 1993; Latham and Ricklefs 1993; Ricklefs and Latham 1992), or climate factors (Rohde 1992, 1999; Wright 1983; Wright et al. 1993). This bias in diversification rates might have been created by a spatially patterned buffer against extinction (Hawkins et al. 2005, 2006) or by an acceleration of speciation rates due to metabolic activation (Bromham and Cardillo 2003; Rohde

1992). Both models are based on different net diversification rates for tropical vs. temperate regions (Cardillo et al. 2005). The most common approach to studying the evolutionary hypothesis is to try to decouple the effects of "ecological" (contemporary) and "evolutionary" (historical) components of the spatial patterns in species richness (e.g. Hawkins et al. 2003b; Diniz-Filho et al. 2004). However, as wisely highlighted by Wiens and Donoghue (2004) and Hawkins et al. (2005), historical biogeography and ecology have much to offer each other, and perhaps a better approach would be to join the two perspectives into a unified theoretical and analytical framework, capable of resolving the tangled and/or the interactive effects of both ecological and evolutionary processes affecting species richness.

However, most of the hypotheses developed in the last 200 years to explain spatial patterns in species richness are actually merely conjectures, and many of them are not testable and/or not falsifiable, or lack empirical supporting evidence (Currie et al. 1999, 2005; Rohde 1992; Willig et al. 2003). In an attempt to reduce the list of hypotheses, Jetz and Rahbek (2001) and Whittaker et al. (2001) suggested five general groups of "most plausible" explanations: (*i*) climate factors, (*ii*) area availability, (*iii*) habitat diversity, (*iv*) evolutionary time and (*v*) geometric constraints (see also Willig et al. 2003, for a slightly different set of plausible explanations). Furthermore, Currie et al. (2005) recently reviewed those hypotheses that invoke climatic factors to explain spatial patterns in species richness under a hypothetical-deductive Popperian approach, and concluded that the this hypotheses are still very misty, whereas biological mechanisms that link climate to species richness are still to be found.

Jackson and Overpeck (2000), and subsequently Ackerly (2003), developed a straightforward conceptual scheme, based on realized variation of environmental factors and ecological niche theory, which is useful to the purposes of this study (see below). Observing the fact that not all combinations of environmental conditions exist in the landscape at a particular time (i.e. there are no highly productive, cold places in South America; Fig. 1), and that sometimes environmental space may have "empty" sites, they coined the term *realized environment* to describe the environmental conditions that actually exist in nature at a given point in time. Also, a species' *fundamental niche* (Hutchinson 1957) is the set of all possible environmental conditions that permit survival and reproduction of individuals. However, Jackson and Overpeck (2000) also realized that a species' fundamental niche does not necessarily track the realized environment, and species may often have physiological tolerance to some environmental conditions that do not actually exist in a given moment (e.g. due to climate change, some previous environmental conditions may have ceased to exist). Thus, the subset of a species' fundamental niche that overlaps the realized environment at a given time is known as *potential niche*, which are the existing environmental conditions that are suitable for a given species. Also, since species usually have their geographical extent restricted by several factors other than their own fundamental niche, such as ecological interactions and effective dispersal barriers, the proportion of the potential niche that is indeed occupied by a species is known as its *realized niche* (Fig. 1; Hutchinson 1957; see Fig. 7 in Ackerly 2003; see Figs. 1 and 2 in Jackson and Overpeck 2000). Finally, Ackerly (2003) defined *environmental island* as "a habitat that is discontinuous in the realized environmental space from other habitats in the landscape". Here we broaden his concept to refer not only to the landscape characteristics in the environmental space, but to patches of differentiated environmental conditions across the geographical landscape. Thus, we use this term for habitats that have environmental conditions somehow distinct from their neighboring sites, such as a mountain or a desert (see Soberón and Peterson 2005 for a discussion of the importance of niche concept regarding predictive models of species' geographical distribution).

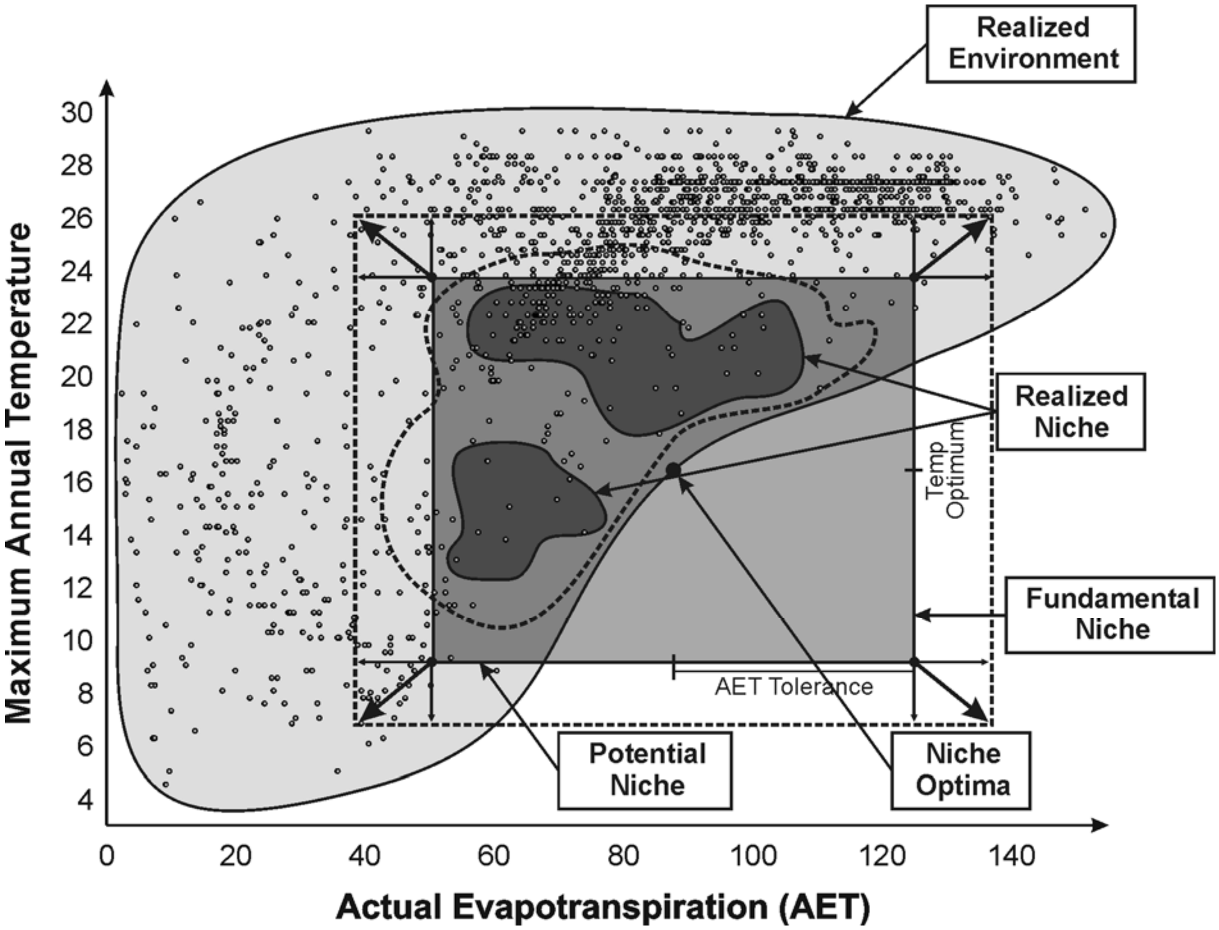

Figure 1: A diagram illustrating the 2-dimensional environmental space formed by Maximum Annual Temperature (MAT; Y axis) and Actual Evapotranspiration (AET; X axis). Dots scattered in the diagram are real-world measures of MAT and AET in South America. The largest, light gray polygon, which encompasses all dots, represents the realized environment for these two variables in South America. The dark square formed by solid lines in the center of the diagram represents the 2-dimensional fundamental niche of a hypothetical species. Notice that not all the fundamental niche overlaps the realized environment, and the portion that does overlap represents the potential niche. Also, due to the restriction of species' geographical range placement and spatial continuity, species do not always colonize all suitable regions, and the two small polygons within the fundamental niche represent the realized niche. Vectors at the corner of the fundamental niche indicate the direction when species' fundamental niche has fluctuated, and the dashed lines indicate the new fundamental niche. Following the fluctuation in species fundamental niche, the realized niche also expands and is then indicated by the dashed polygon.

Phylogenetic niche conservatism (Ackerly 2003; Holt 2003; Peterson et al. 1999) is a relatively new mechanism hypothesized to explain the effects of environmental fluctuation upon the biological evolution of a species' ecological niche, and the consequence for biogeograpical (Ackerly 2003; Wiens 2004; Wiens and Donoghue 2004; Wiens and Graham 2005) and community assembly patterns (Ackerly 2003; Webb et al. 2002). Indeed, niche conservatism is one of the mechanisms invoked by the proponents of stasis in the punctuated evolution debate (Gould 2002). According to this hypothesis, a pair of descendant species tend to resist a rapid adaptation to new environmental conditions at the time and place where they diverged, because a substantial proportion of their biological and physiological characteristics, which determine their fundamental ecological niches (Hutchinson 1957), is inherited from their common ancestral species, which in turn was adapted to different environmental conditions (Holt 1996; Holt and Gaines 1992; Wiens 2004). This leads to high similarity in phenotypic and ecological traits between closely related lineages, and suggests that there has been little adaptive biological modification since species' divergence from the common ancestor (e.g. Ricklefs and Latham 1992). In contrast, niche evolution refers to a change of ecological characteristics that define the niche during speciation events, which allows the descendant species to adapt rapidly to the new environmental conditions. Thus, it has been hypothesized that, if niche conservatism prevails over niche evolution, in regions of highly heterogeneous and/or asynchronously fluctuating environments, the diversification rate could be controlled by a process of range fragmentation and subsequent speciation (Wiens 2004; Wiens and Donoghue 2004). This might occur because environmental fluctuation can eventually fragment species geographical ranges into isolated populations (Ackerly 2003; Dynesius and Jansson 2000; Jackson and Overpeck 2000; Jansson and Dynesius 2002; Wiens 2004), due to species' inability to adapt to the changing environmental conditions in a continuous geographical area within the range. Eventually the reproductive isolation of those isolated populations would lead to allopatric speciation (see also Kirkpatrick and Barton 1997, Holt 2003), and possibly adaptive radiation (Gavrilets and Vose 2005). Akin to the process of environmental fluctuation, even weak evolutionary and/or ecological mechanisms (e.g. stabilizing selection and competition; Ackerly 2003) capable of producing changes in species' fundamental niche (e.g. environmental tolerance) can also lead to geographical range fragmentation, since a species' range is a biogeographical signature of the species' realized niche (Ackerly 2003; Brown et al. 1996; Holt 2003; Maurer 1999; Lomolino et al. 2005).

Niche conservatism may also have played an important role in determining the current latitudinal gradients in species richness (Wiens and Donoghue 2004; Wiens and Graham 2005). According to this hypothesis, most clades originated under a tropical climatic regime, which had, for most of Earth history, and still has, greater geographical extent and environmental stability than extra-tropical climates (Wallace 1878; Hawkins et al. 2005; Stephens and Wiens 2003; Wiens and Donoghue 2004). If descendant lineages tend to conserve ancestral niche characteristics (e.g. Ricklefs and Latham 1992), then species have been slowly dispersing towards extra-tropical regions, or are now extinct in those regions that are not tropical anymore (e.g. Hawkins et al. 2005, 2006). Such process would inevitably generate a higher richness in ancestral (tropical) regions.

To evaluate whether the niche conservatism hypothesis could be a potential mechanistic explanation for the spatial patterns in species richness, we developed a spatially explicit simulation model of niche evolution. The simulation approach is especially suitable for this study because it allows us to unravel and describe complex patterns (Grimm et al. 2005; Peck 2004; Winsberg 1999) that are simultaneously driven by multiple biological processes (Peck 2004) and independent agents (e.g. species and their geographical ranges; Grimm et al. 2005), in both spatial and temporal dimensions (Naylor et al. 1966). Furthermore, we used the simulation model as a ecologically controlled *quasi*-experiment (Peck 2004), in which some factors may be held constant in order to study the isolated and/or interactive effects of other ecological process on the emergent patterns (Peck 2004; Winsberg 2003). Our simulation model was also used as a predictive model, which offers a special advantage over correlative models because it allows predictions in units of species richness (number of species co-occurring at the same place and time), allowing a much more robust statistical inferences. As we previously pointed out, most analyses of broad-scale patterns of species richness are based on correlative approaches (e.g., Hawkins et al. 2003b), and only a few recent studies have used simulations to test alternative scenarios to distinguish among competing hypotheses, most of them in the context of the new geometric constraint models (e.g., Bokma et al.

2001; Colwell and Lees 2000; Davies et al. 2005; Jetz and Rahbek 2001; Rangel and Diniz-Filho 2005a, b).

Thus, we used real-world spatial patterns of South American bird species richness to guide our model parameterization and analyses (a “pattern-oriented modeling” approach; see Grimm et al. 2005). Our primary goal was to investigate whether the niche conservatism process (as described by Wiens 2004 and Wiens and Donoghue 2005), acting over broad-scale heterogeneous landscapes, could predict the real-world spatial patterns in species richness, as well as the shape of the range size frequency distribution. In addition to the niche conservatism mechanism, we also examined if a bias in diversification rates (see Cardillo et al. 2005; Davies et al. 2005), modeled as an extinction buffer linked to environmental productivity, could enhance model predictive power. We found that niche conservatism process can predict a significant portion of the variation in the the spatial patterns of bird richness in South America, and we conclude that niche conservatism can be considered as a plausible mechanistic explanation for broad-scale patterns in species richness.

Methods

We created a spatially explicit simulation model, using the framework of a Hutchinsonian *m*-dimensional niche hypervolume, to simultaneously distribute species in the ecological and geographical space. This “Niche Conservatism Model” (*NCM* hereafter) simulates the dynamics of species’ geographical range size, shape, and placement over a multivariate heterogeneous environmental landscape, by stochastically defining and dynamically modifying each species’ fundamental niche (Fig. 1). The basic idea of the model is to generate a species pool distributed in geographic space by an evolutionary process, in which fluctuations in the breadth of species’ tolerance to environmental factors (established by delimiting a species’ niche hypervolume) create geographic range fragmentation that, in turn, regulates speciation and extinction rates. Mode of speciation then is based on the niche conservatism process described by Wiens (2004). NCM can be used to understand which combinations of model parameters generate realistic patterns of species richness, range size frequency distribution and evolutionary components of diversification (see Wiens and Donoghue, 2004 and Wiens and Graham 2005). NCM uses a grid-based, discrete virtual space (a domain) modeled as a lattice, divided into n cells whose spatial location is defined by a unique set of two values (geographical coordinates). We used the South America domain, divided into 1º by 1º cells ($n = 1565$), as basis for our study. By assigning m values of real-world environmental variables ($e = 1, 2, \ldots, m$ – see below) to each grid cell of the South American domain, environmental spatial patterns are established.

Model Description

NCM starts by randomly placing the first species into a single cell (C). The geographical position (latitude and longitude) of this initial cell is recorded for use in later analysis (see below). The value of each environmental variable (C_e) of the initial cell becomes the "optimal" environmental condition ($O_e = C_e$, for $e = 1$ to m) for the initial species, thus defining the fundamental niche optima for that species. Next, symmetrical deviations to the optimal values, which we used as a surrogate for species' "physiological tolerance" (T_e, for $e = 1$ to m) to each environmental variable, are defined as half of the range (maximum minus minimum value) of the environmental variable across the whole domain (range_e), multiplied by a scaling coefficient randomly chosen from a positively truncated Gaussian probability distribution, with mean equal to 0 and standard deviation equal to σ ($T_e = (\text{range}_e / 2) \times \text{gauss}(0,\sigma)$, for $e = 1$ to m). This defines initial species' maximum feasible deviation from the optimal environmental values, and allows it to further occupy other cells with environmental condition within these upper and lower tolerance limits ($(O_e - T_e) \leq C_e \leq (O_e + T_e)$, for $e = 1$ to m), assuming spatial continuity in species' geographical range from the initial cell to all suitable neighboring cells. In other words, as stated by Liebig's law of the minimum (Krebs 2001), species' geographical range expansion is limited by that environmental variable in least amount beyond species' physiological tolerance. Thus, the larger the parameter σ is set, the broader the environmental tolerances the initial species will have and, consequently, the larger will be its geographic range size (r).

In this way, the size, shape, and placement of each species' geographic range are emergent consequences of species' environmental optima (defined by the geographical position of the species' initial cell position), environmental tolerances (which defines the breadth of the species' fundamental niche), and spatial pattern of the environmental gradients (which determine the extent of suitable areas around the species' initial position). Hence, in our model, the hypervolume formed by symmetrical tolerances around the species' environmental optima define the species' fundamental niche position and niche breadth (i.e. all environmental conditions in which the species could survive). However, as eventually some portion of species' fundamental niche may lie outside realized environment, only the potential niche can be converted into realized niche. Realized niche is always smaller and contained within the potential and fundamental niches, expressing those environmental conditions that species indeed occupy by means of their geographical range (Fig. 1).

In our model, time is measured in discrete steps, and at each step, each species' environmental tolerances are subjected to stochastic fluctuations. For computational simplicity, we assumed that environment factors remain constant over evolutionary time, and that species' environmental tolerances fluctuate stochastically. In practice, these two processes could be thought as interchangeable, and so we believe it acceptable to sacrifice realism (i.e., the environment fluctuates and species adapt) to decrease computational effort. Thus, with a given probability (P_c) each extant species has each of its environmental tolerances (T_e, for $e = 1$ to m) increased ($T_e = T_e \times (\text{beta}(1,\Omega) \times \mu)$), or decreased ($T_e = T_e \times ((1 / \text{beta}(1,\Omega)) \times \mu)$), symmetrically around the optimum, according to a value chosen from a Beta probability distribution, where Ω is a parameter that defines the shape of this distribution (Fig. 2), and μ is a scaling factor that defines the maximum possible change in species' tolerance (Fig. 1). Species' environmental tolerances, along each environmental variable, have independent oscillations, both in terms of magnitude and direction of change. This means that, for example, it is possible for a species' physiological tolerance to temperature to increase, whereas its tolerance to humidity decreases, each in different amounts. As a consequence of tolerance fluctuation, each species' geographical range may also fluctuate. Thus, after a tolerance fluctuation, a species will immediately become locally extinct in those cells with environmental conditions outside its new tolerance limits, such that the species' geographical range consequently contracts ($r_t > r_{t+1}$). On the other hand, if the new tolerance limits allow, species will colonize spatially adjacent suitable cells, and in this case species' geographical range will expand ($r_t < r_{t+1}$). Thus, fluctuations in fundamental niche usually affect geographical range size (r), and realized niche (See Figs. 1 and 3).

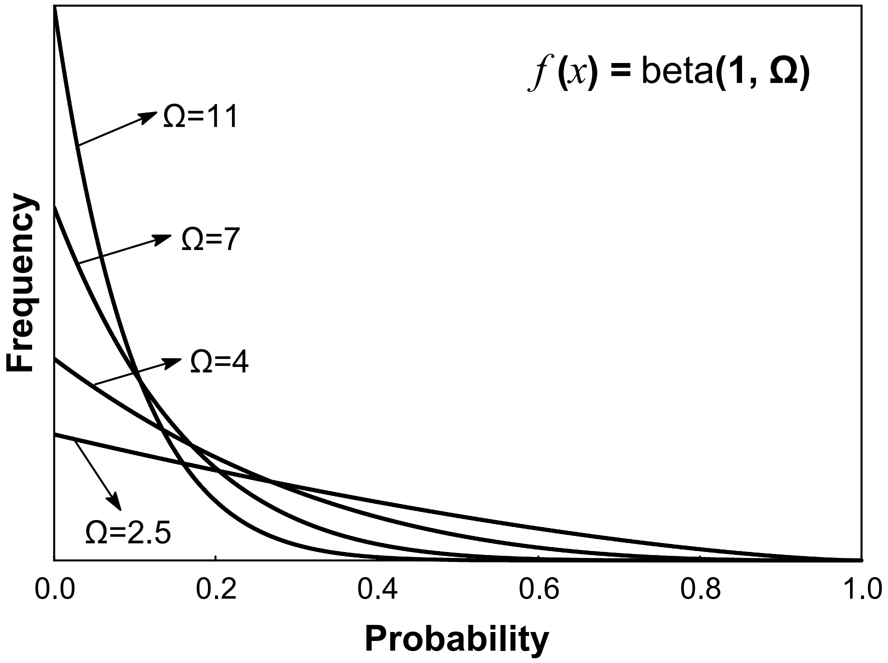

Figure 2: Density functions of the Beta probability distributions used to dynamically modify species' environmental tolerance for each environmental factor, according to the parameter Ω. The final magnitude of temporal fluctuation in a species' environmental tolerance is obtained by combining the parameters Ω and μ.

However, in our model, range plasticity (expansion, contraction, or both, in different regions) is not the only possible outcome of fluctuation in a species' tolerance. Besides waxing and waning, in some situations when species' tolerance to a given environmental factor shrinks, species' geographical range contracts and consequently it becomes spatially fragmented into smaller "local populations" (comprising isolated fragments of the original range, with different number of contiguous cells; Fig. 3; see also Fig. 1 in Wiens 2004). This occurs as a consequence of the heterogeneous nature of environmental landscapes, and because the breadth of species' environmental tolerance has become narrower, some areas within a species' geographical range may become unsuitable, leaving small populations isolated on "environmental islands" throughout the landscape. Nevertheless, not all fragmented populations are viable over time (Gomulkiewicz and Holt 1995; Hanski 1998), and in our model these local populations may go extinct as a function of their geographical range size (a surrogate for initial population abundance), according to the function $P_x = \alpha \times \exp(\beta \times r)$), where α and β are parameters that regulate the decrease in probability of extinction P_x as function of the range size (r) of the local population (Fig. 4).

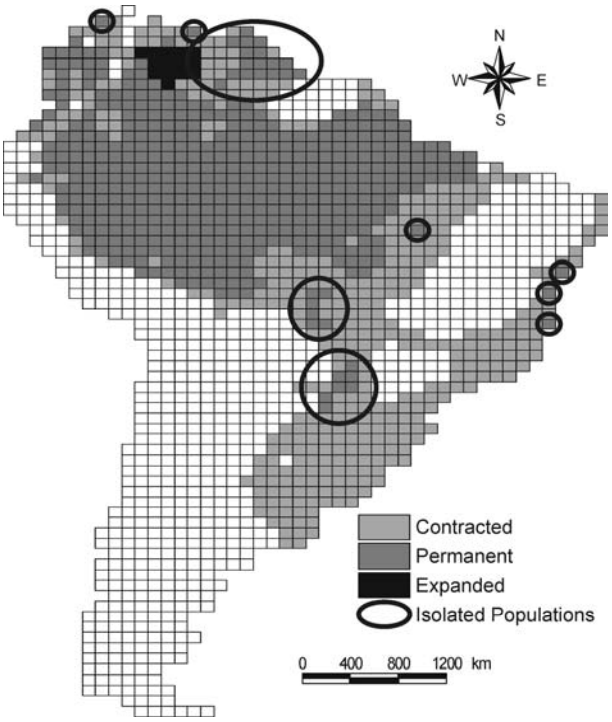

Figure 3: A hypothetical example of the behavior of a single species' geographical range when its environmental tolerances fluctuate. The light-gray cells (contracted) are those cells in which the species was present before tolerance fluctuation, but then became locally extinct due to a reduction in tolerance to some environmental factor. Medium-gray cells (permanent) are those cells that remained suitable even after the tolerance fluctuation, and thus are cells that species continues to occupy. Black cells (expanded) are those cells that the species did not occupy before the tolerance fluctuation, but that could be colonized after it due an expansion of the tolerance limit to some previous limiting environmental factor. Notice that the species' geographical range became fragmented into small, isolated populations with variable range sizes after tolerance fluctuation (medium-gray cells that are marked by circles). Each of these fragments may become a new species if survives to the extinction test, or may go extinct, according to functions shown in Fig. 4.

Each local population is next subjected to an “extinction test”. If it survives this test, we assume that the gene flow between populations was restricted for a time long enough that they diverge to become new (descendant) species, by a “classic” allopatric speciation process (Holt 2003; Lomolino et al. 2005). Because species are isolated, adaptive evolution shifts species niche optima along all environmental dimensions, such that each new species become specialized in the environmental conditions within its geographical range (i.e. the fragment of the ancestor’s range occupied by the incipient daughter species). Thus, no directional adaptive or neutral evolution happens in the model, except at the moment of species origination, when species shift their niche optima, in a pure Gouldian punctuationist fashion (Gould 2002). In this model, this shift is performed by defining the niche optima of each isolated population (both descendant and ancestor species) as the average value of each environmental variable within species’ geographical range ($O_e = \sum C_e/r$, for $e = 1$ to m). Also, breadth of descendant species’ environmental tolerances (DT) is inherited from the breadth of ancestor species’ environmental tolerance (AT), but increasing ($DT_e = AT_e \times (\text{beta}(1,\Omega) \times \mu)$, for $e = 1$ to m) or decreasing ($DT_e = AT_e \times ((1 / \text{beta}(1,\Omega)) \times \mu)$, for $e = 1$ to m) in some amount, creating an ‘heritability’ of niche dimensions and possibly range size. We assume here that physiological and life-history traits that, in last instance, determine species’ ability to adapt to various environments will have relatively high levels of phylogenetic conservatism. More importantly, a consequence of this process is that, because speciation processes and shifts in the niche optimum happened in an allopatric fashion over an spatially autocorrelated environments and, mainly, when there is a high similarity between tolerances of ancestral and descendant species, patterns in species richness will evolve under a process of high ‘heritability’ in niche space and, consequently, in range size and location (see Hunt et al. 2005; Webb and Gaston 2003, 2005). Thus, the parameter Ω regulates the shape of the Beta probability distribution, and the parameter μ determines the magnitude of the shift between descendant and ancestor and ancestral species’ tolerance. After the descendant species’ tolerance is defined, its geographical range spreads cohesively in the spatial dimension (at one and same time step), as previously described, to occupy contiguous and suitable cells. Thus, the spatial distance between ancestor's and descendant's geographical ranges tends to regulate the similarity of their niche optima, because spatial distance is usually correlated with environmental similarity (i.e. environmental variables are usually spatially autocorrelated). Also, the lower the parameter μ, the more similar the ancestor's and descendant's environmental tolerances (T) (niche conservatism), whereas the more μ increases, the more T differs between these two species (niche evolution). Therefore, the strength of niche conservatism over the evolutionary time is determined by the parameters Ω and μ, as well by the spatial distance between ancestor's and descendant's geographical range position.

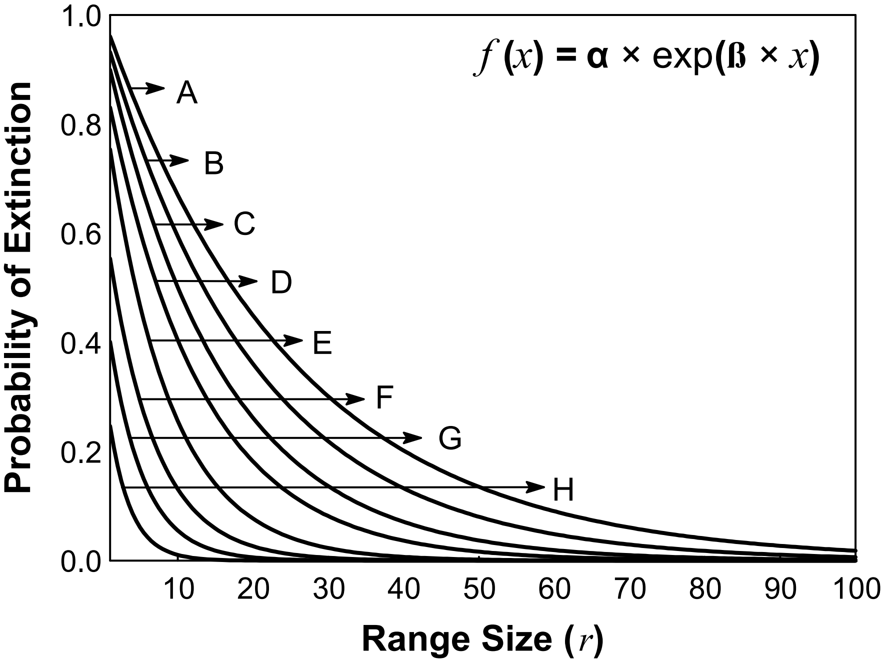

Figure 4: The density functions of extinction probabilities for each isolated local population, according to its geographical range size. Each line (A – H) is determined by the combination of the parameters α and β. A: α = 1.0 and β = -0.04; B: α = 0.98 and β = -0.05; C: α = 0.96 and β = -0.065; D: α = 0.9 and β = -0.08; E: α = 0.85 and β = -0.12; F: α = 0.65 and β = -0.16; G: α = 0.5 and β = -0.22; H: α = 0.35 and β = -0.35.

Once the daughter species' geographical range spreads to occupy its realized niche in geographical space, this species is added to the species' pool. Each species in the pool becomes subjected to the same dynamical process of niche fluctuation, speciation and extinction, such that this process recurs iteratively for all new species. The simulation continues the diversification process (a balance between speciation and extinction) until a pre-defined number of species exists; this represents a single simulation. For each simulation, results of interest are (*i*) the resulting spatial patterns in species richness, (*ii*) the range size frequency distribution (RSFD), and (*iii*) the geographical position of the initial species (Table 1). The entire procedure is repeated (replicated simulations), and the results are averaged among replicates.

Table 2 – The state-variables of the model, and their respective notation.

Model's State-Variables	**Symbol**
Species' niche optimum for the environmental variable *e*	O_e
Width of species' environmental tolerance for the variable *e*	T_e
Species' geographical range size	r
Geographical position (latitude and longitude) of the origin of the initial species	(none)

Most historical hypotheses developed to explain patterns in species richness invoke the effects of processes that might create spatially patterned net diversification, by regional differences in speciation and/or extinction rates linked, for example, with time (Latham and Ricklefs 1993), climatic stability (Hawkins et al. 2005), or temperature (Rohde 1992). Following the same reasoning, we also tested if a bias in net diversification rates could increase the model's ability to predict real-world patterns in species richness, by hypothesizing that the probability of extinction of local populations could be buffered by real-world patterns of environmental productivity. The theoretical support for this hypothesis is that a population that settles in a very productive region has a higher carrying capacity compared to a population settling in a region of low productivity, all other factors being equal (Currie et al. 2005, Gotelli 1998; Krebs 2001). Thus, the carrying capacity of the former is expanded by environmental productivity, and if we assume that environments tend to be saturated, more abundant populations are less likely to go extinct as a result of environmental stochasticity (see Currie et al. [2005] for an explanation of the species-energy theory; Gotelli 1998).

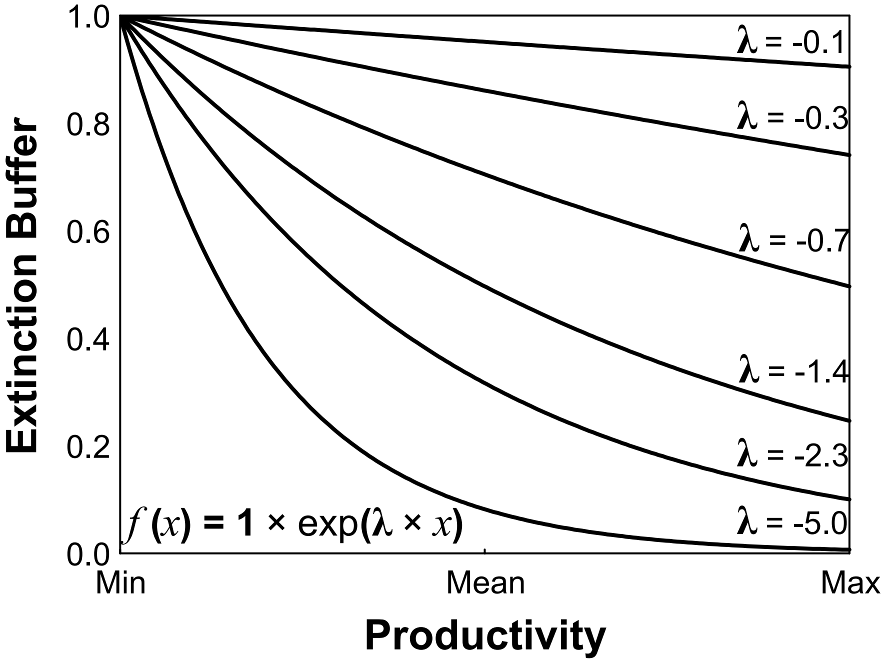

Figure 5: The functions of the proportion of reduction in the extinction probability due to the extinction buffer linked to environmental productivity. Each line is determined by the parameter λ. Thus, if the parameter λ = -5, and a species' geographical range is in the region of maximum productivity, the extinction probability will decline almost to zero. On the other hand, regardless of the parameter λ, if a species' geographical range is in a region of low productivity, extinction probability will not be much reduced.

To implement this extinction buffer in NCM, the previously defined probability of extinction (P_x) is rescaled (decreased) as a function of the environmental conditions of the region that the local population occupies, such that $P_x = P_x \times (1 \times \exp(\lambda \times (((\sum C_e/r) - \min(e)) / (\mathrm{range}(e)))$, where e is the variable describing environmental productivity (i.e. NPP or AET, see below), λ is a parameter that regulates the decrease in probability of extinction (Fig. 5) as function of the averaged value ($\sum C_e/r$) of e in those cells occupied by the local population, and r is the local population's range size. Also, $\min(e)$ and $\mathrm{range}(e)$ stand for the minimum and the range of values of the environmental variable e across the whole domain. It is noteworthy that the niche conservatism process is still the driver of speciation rates, although the extinction rate is now driven by a combined function of population range size and local environmental productivity (measured by AET, see below).

Thus, the entire algorithm of NCM can be summarized as follows:

1. A grid system is created, and environmental variables are assigned to each cell of the grid system in order to define landscape characteristics. Model parameters are set.
2. One cell of the grid is randomly chosen, and the values of the environmental variables assigned to that cell define the environmental optimum of the initial species in each environmental dimension.
3. For each environmental variable, a symmetrical environmental tolerance interval around the initial species' environmental optimum is stochastically defined. Environmental optima and tolerances define the species' fundamental niche.
4. The incipient (initial or recently originated) species expands its geographical range continuously to all contiguous, suitable cells, that is, with environmental conditions that lie within its lower and upper tolerance limits for all niche axes.
5. With a given probability, species may have their tolerance limits modified. If this happens, each species' environmental tolerance will be stochastically and symmetrically increased or decreased by some amount. When a given species has its tolerance limits modified, that species may eventually become locally extinct in cells that have environmental conditions outside the new tolerance limits, resulting in range contraction. On the other hand, according to step 4, a species' geographical range may expand into contiguous cells if the new tolerance limits allow;
6. If, during range contraction, a species' geographical range becomes fragmented into small isolated "local populations" (Fig. 2), each fragment may go extinct as function of its geographical range size (Fig. 4). Optionally, environmental productivity of the region where the fragment is settled may act as a buffer against extinction probability. If one (or many) of these local population does not go extinct, it speciates and becomes an independent, daughter species. In this case, the daughter species' optima are defined according to the environmental condition of the region it occupies. Finally, the daughter species' environmental tolerances are inherited from its ancestor, although tolerance breadth may remain conserved (little change) or evolve quickly

(greater change). The daughter species expands its geographical range (step 4) according to its tolerance limits;

7. Steps 5 and 6 recur until a previously defined number of species is reached, matching the number of species of some "target group";
8. When steps 2 - 7 have been repeated for a previously defined number of simulation replicates, and then the whole simulation process is completed. Results of averaged values among replicates are recorded.

Simulations

We used a pattern-oriented modeling approach (Grimm et al. 2005) to find the combination of parameters that maximizes the similarity between the predicted and real-world spatial patterns in species richness, for a total of 3088 South American birds. The digitized bird database (Rigdley et al. 2005) was processed using ESRI ArcView 3.1 scripts in order to record species' presence in the 1° by 1° grid cells (n = 1565) covering South America. Then, we calculated real-world spatial patterns in species richness by summing the number of species presences in each cell, and determined range size by summing the number of cells in which each species is present.

To evaluate the model's predictive power for real-world patterns, we explored the parameter space by performing a total of approximately 4000 simulations, with different parameter combinations. Each simulation was replicated 100 times, although previous analysis showed that results become stable at about 50 replicates. Each replicate simulates the biogeographical features of 3088 virtual species, such that model predictions are directly comparable with the real-world data. Previous analysis showed that model predictions are independent of the probability of tolerance fluctuation (P_c), because it affects only the pace of the diversification process. Thus, we fixed P_c at 0.5 for all further simulations. We performed simulations using different combinations of environmental variables (see below) in order to find the combination that maximizes the fit between real-world and predicted patterns in species richness. Also, we initially performed an exploration of parameter space by considering the isolated influence of the niche conservatism process (λ = nil), but then we activated the diversification bias linked with environmental productivity (λ ≠ nil) to evaluate if this additional process can enhance model's predictive power. We used a real-world measure of actual evapotranspiration (AET) as a surrogate for environmental productivity, a water-energy variable closely associated with plant productivity that has been considered one of the best predictors of species richness (see Hawkins et al. 2003a, b). The amplitude of exploration of parameter space and the best parameter combination are shown in Table 2.

Table 2 – The parameters of the model, their respective notation, values explored, and best sets.

Model's Parameters	**Symbol**	**Explored**	**Best Set (λ = nil)**	**Best Set (λ ≠ nil)**
Initial tolerance width (Gaussian distribution)	σ	0.25 – 1.25	0.5*	0.5*
Niche fluctuation and heritability (Beta distribution)	Ω	2.5 – 11	7	6*
	μ	5 – 15	8	11*
Extinction probability as function of species' range size	α	1 – 0.35	0.5	0.96
	β	-0.04 – -0.35	-0.22	-0.065
Extinction buffer as function of environmental productivity	λ	nil; -0.1 – -5	Nil	-2.3
Environmental Variables	$e=1,2,\ldots,m$	AET; PET; Humidity Temperature; Elevation	AET; Humidity; Temperature	AET; Humidity; Temperature
		Predictive Power (r_P)	0.814 ± 0.144	0.855 ± 0.083

Note: "Best Set" is combination of parameters under "pure" niche conservatism process (λ = nil) and niche conservatism associated with the diversification bias process (λ ≠ nil) that maximizes model predictive power. Asterisks indicate parameters that do not significantly influence the correlation between real-world and predicted spatial patterns in species richness.

We used eight real-world variables to describe the heterogeneous environmental conditions across South America: (*i*) AET, (*ii*) PET, (*iii*) minimum and (*iv*) maximum humidity, (*v*) minimum and (*vi*) maximum annual temperature and (*vii*) minimum and (*viii*) maximum altitude above sea level (Fig. 6; Kucharik et al. 2000; New et al. 1999; NOAA; Willmott and Kenji 2001), each one exerting equal control over simulated species. The parameter λ, which controls the strength of the extinction buffer as a function of environmental productivity, assumed seven different conditions (nil, and from -0.1 up to -5; Fig. 5). For the parameter σ, which regulates the width of initial species' environmental tolerances, we used five different values, varying from 0.25 up to 1.25. The parameters Ω and μ, which jointly regulate the niche fluctuation and heritability processes, assumed four (from 2.5 up to 11; Fig. 2) and three different values (5, 10 and 15), respectively. The parameters α and β, which jointly regulate the strength of extinction probability as function of range size, assumed eight different combinations (α varying from 1 down to 0.35, and β varying from -0.04 down to -0.35; Fig. 4).

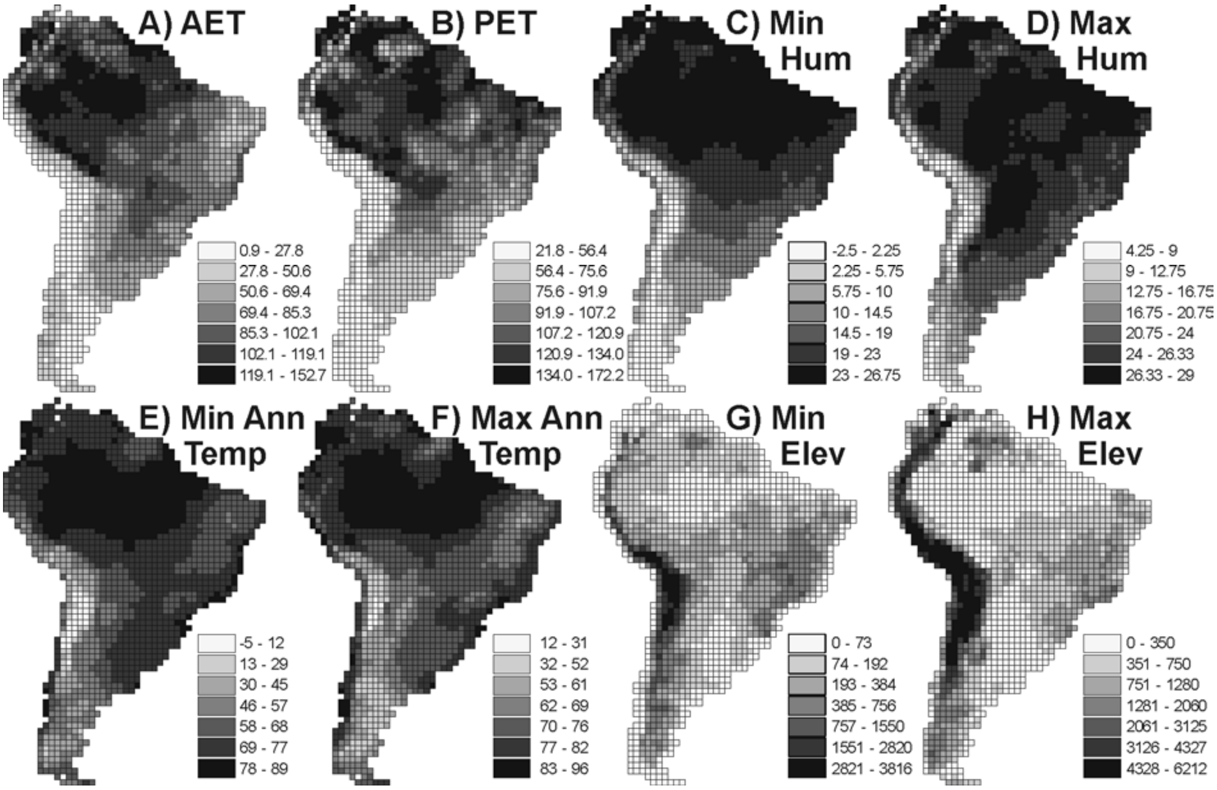

Figure 6: Spatial patterns of eight environmental variables in South America that were used to define landscape and climate characteristics in the Niche Conservatism Model. (*A*) AET is actual evapotranspiration; (*B*) PET is potential evapotranspiration; (*C*) Min Hum and; (*D*) Max Hum are minimum and maximum annual humidity, (*E*) Min Ann Temp; and (*F*) Max Ann Temp are minimum and maximum annual temperature, (*G*) Min Elev and; (*H*) Max Elev are minimum and maximum altitude above sea level.

Data Analysis

Because of the complexity of model and wide range in parameter variation, we proceeded following a sequential approach in which the most promising scenarios were selected for closer examination, following the pattern-oriented modeling strategy. We used Pearson's correlation coefficient (r_P) between predicted and real-world spatial patterns in species richness to find the parameter combination that maximized model predictive power. However, since broad scale patterns in species richness are usually strongly spatially autocorrelated (Diniz-Filho et al. 2003), we performed the significance test of r_P using the number of degrees of freedom estimated by Dutilleul's (1989) method, which accounts for spatial autocorrelation in both variables, and discarded any insignificant predictions. In addition, an Analysis of Variance (ANOVA; Sokal and Rohlf, 1995) was used to evaluate the influence of each model parameter on the correlation coefficient (r_P, after z-transformation) calculated between real-world and predicted spatial patterns in species richness.

Next, we selected for a more detailed analysis the results of the 10 best simulations under the "pure" niche conservatism process, and the 10 best for niche conservatism with the diversification bias. We used simultaneous autoregressive models (SARs) to find the spatially unbiased intercept and slope values on the regression of real-world against predicted spatial pattern in species richness (we performed this spatial analysis only for the best 20 models of NCM due to the time investment involved). SAR is a form of generalized least-squares regression in which the spatial structure of the data is directly incorporated into the model. Thus, SAR coefficients (β) can be found by solving

$$\beta = (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}^{-1} \mathbf{Y}$$

where **X** is the matrix of predictor variables, **Y** is the vector of response variable, and **C** is the covariance matrix defined by

$$\mathbf{C} = \sigma^2 \, [(\mathbf{I} - \rho\mathbf{W})^T]^{-1} \, [\mathbf{I} - \rho\mathbf{W}]^{-1}$$

where σ^2 is the variance of the OLS residuals and **I** is an n x n identity matrix, ρ is an autoregressive parameter to be estimated, and **W** is neighbor weight matrix (Rangel et al. 2006). SAR allowed us to find the regression coefficients, taking into account the strong spatial autocorrelation in both response and predictor variables (Cressie 1993; Tognelli and Kelt 2004; Lichstein et al. 2002). Thus, since our model yields predictions in "real units" of species richness, we used the values of a spatially coefficients and their respective 95% confidence intervals (CI), to find the parameter combination that best approaches a regression slope of 1.0 and intercept of 0.0. All spatial statistics were carried out on SAM v1.1 (Rangel et al. 2006), freely available at www.ecoevol.ufg.br/sam.

We also report and compared the shapes of the real-world and simulated range size frequency distribution (RSFD) for best-fitting model, in order to verify the realism of these predicted patterns.

Results

Bird species richness is strongly spatially patterned over South America, although the patterns are very complex (Fig. 7 *A*). In fact, it is clear that some specific regions, such as Andean mountains, Amazon basin and Atlantic Forest are highly species rich, and these regions are generally located in a NW-SE axis of South America. On the other hand, there are also species poor regions, such as Atacama desert and Brazilian Caatinga, which are aligned in a NE-SW direction. Also, the range size frequency distribution of birds in South America is strongly right-skewed (Fig. 8 *A*), a pattern commonly observed in nature (Gaston 2003).

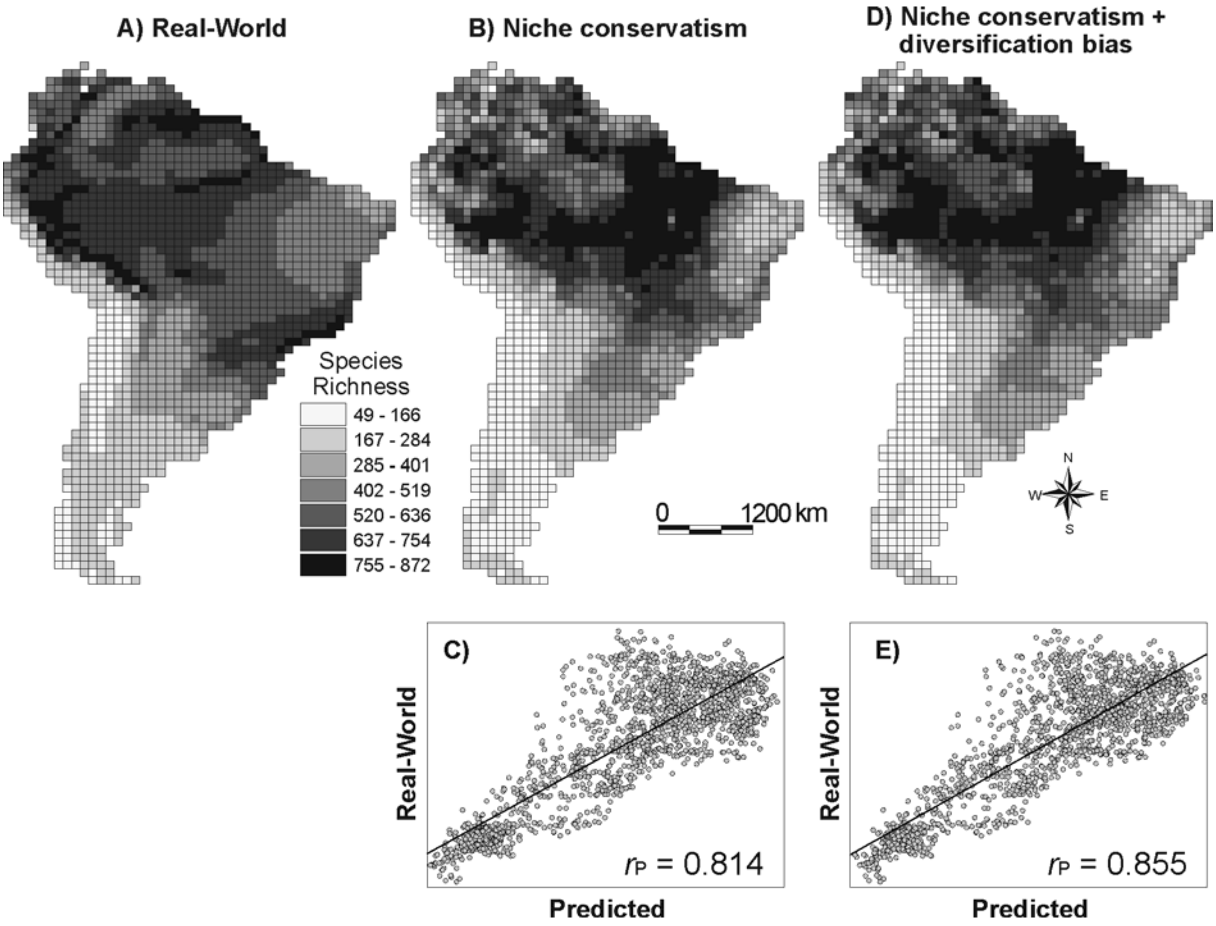

Figure 7: (*A*) Real-world spatial patterns in South American bird species richness. (*B*) Predicted spatial patterns in South American bird species richness, under "pure" niche conservatism process. (*C*) Relationship and regression fit between the patterns in the maps (*A*) and (*B*). (*D*) Predicted spatial patterns in South American bird species richness, under the niche conservatism and diversification bias processes. (*E*) Relationship and regression fit between the patterns in the maps (*A*) and (*D*).

We found that NCM is very robust to small variations in model parameters, which indicates that the basic processes underlying the model can capture and predict the real-world spatial patterns in species richness. It is possible to observe this robustness by the frequency distribution of correlation coefficients obtained by all explored parameter combinations (Fig. 9 *A* and *B*). Despite the wide range of parameter variation (see Table 1), which creates a large variation in the values of the correlation coefficient between predicted and real-world patterns, NCM is clearly a good predictor of species richness, showing that niche conservatism process is capable of capturing the essentials of real-world spatial patterns in species richness. Correlation values (r_P) for the "pure" niche conservatism process (λ = nil) ranged between r_P = 0.01 and 0.849 (Fig. 9 *A*), whereas the niche conservatism process associated with the diversification bias ($\lambda \neq$ nil) yielded strong correlations, ranging between $r_P = 0.02$ and 0.855 (Fig. 9 *B*). Considering all eight environmental factors we tested, the set of variables that maximized model predictive power, regardless the combination of all other parameters, includes AET, minimum and maximum humidity, and minimum and maximum annual temperature. Thus, we further report results of simulations that used this set of environmental variables.

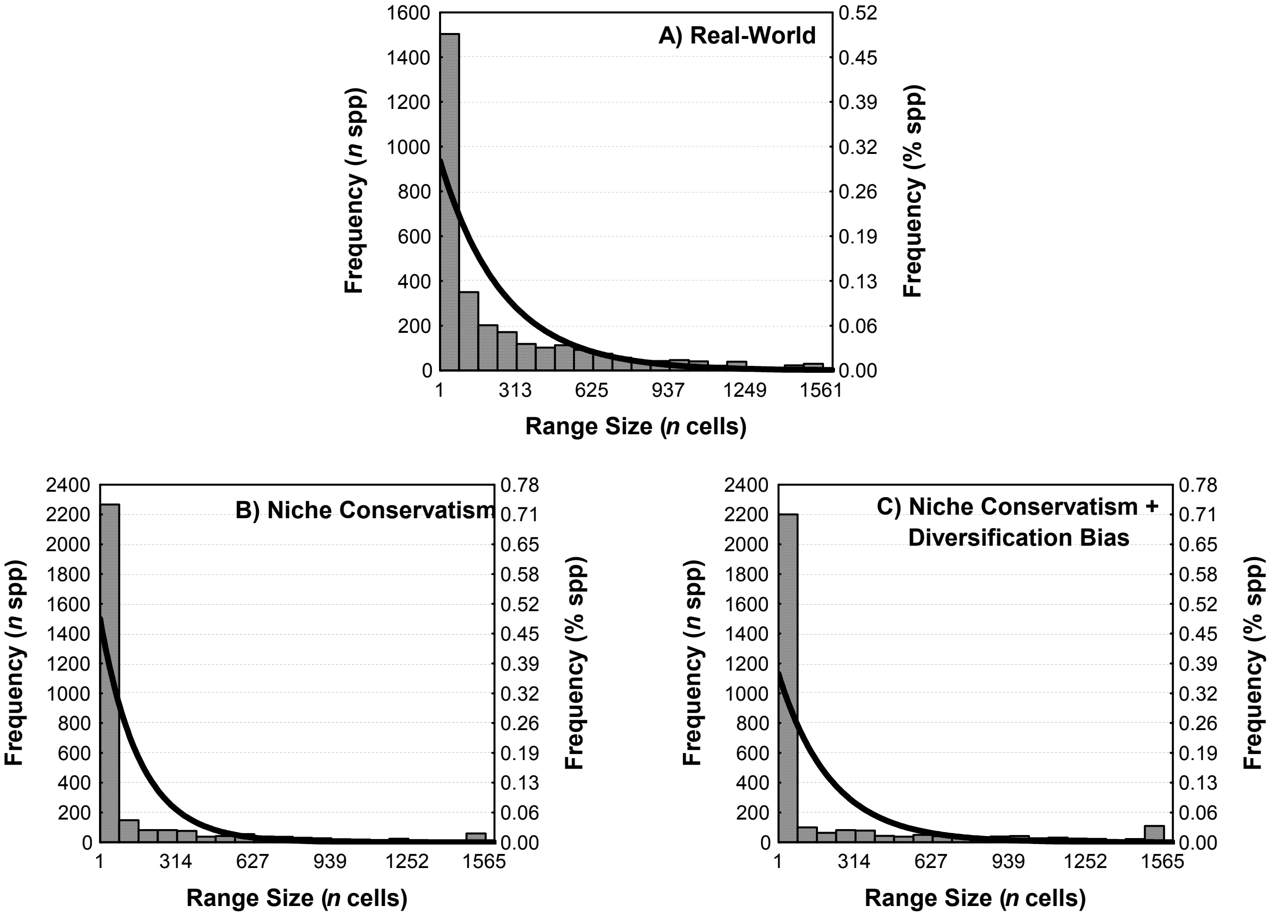

Figure 8: Range size frequency distribution of real world (*A*) and predicted species richness according to the best simulation models. Range distribution in (*B*) is a prediction under the "pure" niche conservatism process, whereas (*C*) is the distribution obtained under the niche conservatism and diversification bias processes.

Niche conservatism process alone (λ = nil)

We first consider the scenario in which the niche conservatism process acts alone, without the bias in diversification rates through the extinction buffer (λ = nil). In this case, the factor with most influence on the correlation (r_P) between real-world and predicted spatial patterns in species richness is the strength of extinction probability as function of geographical range size (parameters α and β). Indeed, when there is a low probability of extinction, even for very small range fragments (curves F, G and H of Fig. 4), the fit between predicted and real-world patterns increases significantly (F-ratio = 372.824; $P < 0.001$).

The following most influential factors are those that regulate the modification of species' environmental tolerances and niche heritability processes (parameter Ω, F-ratio = 17.937; $P < 0.001$; parameter μ, F-ratio = 26.919; $P < 0.001$). Our results showed that at intermediate levels of niche conservatism ($\Omega \sim 7$ and $\mu \sim 8$), the correlation between predicted and real-world patterns increases. This is a balanced situation in which niche evolution is sufficient to allow some lineages to adapt and disperse towards environmentally distinct regions, but niche conservatism maintains most of descendant species close to the core diversification region.

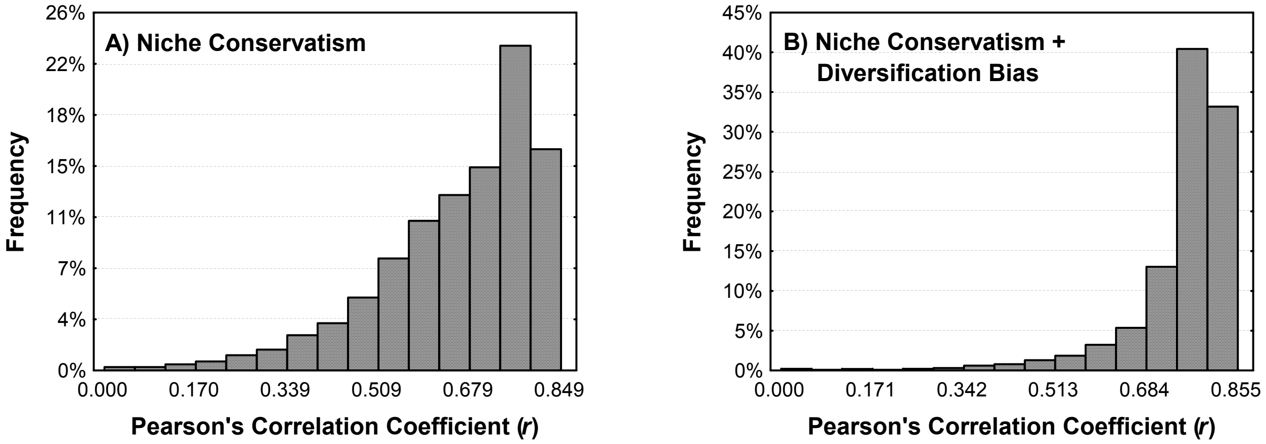

Figure 9: Histogram illustrating the frequency of values of Pearson's correlation coefficient obtained between real-world and predicted patterns, under the (*A*) "pure" niche conservatism process, and (*B*) niche conservatism and diversification bias processes, for all parameter combinations we simulated.

The final significant influential factor is the latitudinal position of the initial species (F-ratio = 12.008; $P < 0.001$), such that when the initial species arises in the tropics, the diversification process starts there, and thus the area of origin becomes the most species-rich region. Therefore, niche conservatism exerts a strong force, preventing most species from reaching extra-tropical regions, and thus a latitudinal gradient in species richness arises, increasing model predictive power.

Thus, these results suggest that when the extinction probability is very low ($\alpha = 0.5$ and $\beta = -0.22$; curve G of the Fig. 4), niche conservatism is reasonably strong ($\Omega = 7$ on Fig. 2, and $\mu = 8$), and the whole lineage has originated in the tropics, the model has a strong predictive power of real-world spatial patterns in bird species richness (mean $r_P = 0.824$, std. dev. = 0.144; Fig. 9 *A*). This predictive power can be illustrated comparing the maps illustrated in Fig. 7 *A* with the one in Fig. 7 *B*, which show that the "pure" niche conservatism process may create a latitudinal gradient in species richness reasonably similar to the observed in the real-world (for the relationship between these patterns see the Fig. 7 *C*). Although NCM fails to predict the high bird richness along Andean mountains and Atlantic Forest, its predictions reasonably match the low species richness along the NE-SW axis (Fig. 7 *B*).

Considering the predictions yielded by the best 10 parameter combinations (all with $r_P > 0.810$), they provide not only good predictions of the shape of spatial patterns in species richness, but also generate very accurate predictions in terms of magnitude of variation in species richness. The regression coefficients of the 10 best models, spatially corrected by SAR, show that the mean (± standard deviation) intercept value is -4.404 ± 34.663, whereas mean slope value is 1.067 ± 0.07. Furthermore, the 95% CI of the intercept of nine models, out of the 10 best models, bracket 0.0, whereas, the 95% CI of the slope of seven model, out of 10 models, bracket 1.0.

The real-world and predicted range size frequency distributions are clearly similar (Figs. 8 *A* and *B*). Moreover, like the real-world RSFD, the vast majority of simulated species are small-ranged, whereas very few species have wide geographical ranges.

Niche conservatism and diversification bias processes ($\lambda \neq$ nil)

When we activate a net diversification bias driven by environmental productivity, the overall model predictive power increases (mean $r_P = 0.855$, std. dev. = 0.083; Fig. 10 *B*), although it also changes the relative importance of the model parameters. Extinction pressure as function of geographical range size remains the most important parameter (F-ratio = 65.522; $P < 0.001$), although, in contrast to when there is no extinction buffer (λ = nil), the higher is the probability of extinction (curves B and C of the Fig. 4), the higher the predictive power of the model.

The latitudinal position of the initial species now becomes now the second most significant factor explaining correlations between real-world and simulated patterns (F-ratio = 43.584; $P < 0.001$), such that when the lineage arise in a tropical region, model predictive power increases. Next, the diversification bias, modeled as an extinction buffer (parameter λ) becomes significantly influential (F-ratio = 18.062; $P < 0.001$), such that strong extinction buffers ($\lambda < -2.3$; Fig. 5) increase the predictive power of the model. However, in contrast to the "pure" niche conservatism model, the niche heritability parameter, which regulates niche conservatism/evolution processes, becomes not significant (F-ratio = 0.04 and $P = 0.990$ for Ω, and F-ratio = 1.246 and $P = 0.273$ for μ), being responsible only for causing range fragmentation and consequently speciation events, but not for driving spatial patterns in species richness.

Thus, these analyses reveal that when extinction pressure is strong ($\alpha = 0.5$ and $\beta = -0.22$; curve B of the Fig. 4), which is strongly buffered in highly productive regions ($\lambda < -2.0$), and the oldest ancestor's lineage originated in a tropical region, the model becomes a better predictor than when the "pure" niche conservatism process acts alone (Fig. 9 *B*). The spatial patterns in species richness predicted by the niche conservatism and diversification bias processes (Fig. 7 *D*) look similar to the observed in real-world patterns ($r_P = 0.855$; Fig. 7 *E*), although differences from the prediction yielded under the "pure" niche conservatism process (Fig. 7 *B*) are not so clear.

Although the net diversification bias slightly improves model predictive power (Fig. 7 *A* and *B*), it decreases model accuracy. In fact, considering the 10 best parameter combination (all with $r_P > 0.828$), the mean (± standard deviation) of the intercept calculated by SAR is 26.289 ± 27.222, whereas for the slope is 0.957 ± 0.042, respectively. Out of these 10 best models, seven of them have 95% CI for the intercept that bracket 0.0, and eight for slope that bracket 1.0.

Adding the net diversification bias does not change the predicted pattern of the range size frequency distribution (Fig. 8 *A* and *C*). It remains strongly right-skewed, and closely resembles real-world patterns (Fig. 8 *A*).

Discussion

This study has yielded two important results: (*i*) a niche conservatism process seems to be sufficient to generate broad-scale spatial patterns in species richness that closely resemble the real-world contemporary ones, although (*ii*) it is also probable that diversification bias linked with environmental factors has also contributed to the final pattern. These findings are in accordance with recent theoretical and empirical advances in ecology and biogeography (e.g. Cardillo et al. 2005; Hawkins 2005, 2006; Holt 1996; Holt and Gaines 1992; Qian and Ricklefs 2004; Prinzing et al. 2001; Ricklefs and Latham 1992; Stephens and Wiens 2003; Svenning 2003; Wiens 2004; Wiens and Donoghue 2004; Wiens and Graham 2005).

Assuming climatic niche specialization of the most ancestral species of a lineage, and niche heritability over evolutionary processes, descendant species tend to have similar environmental optima and tolerances. In our model, this will occur because of spatial autocorrelation in environments and because descendant species will appear within ancestral distributions. We also assume that geographic isolation and the climatic fluctuations causing “environmental islands” will generate a quick adaptation of the new species, shifting its optimum to a new environmental centroid within its range (see Kirkpatrick and Barton 1997). Over evolutionary time, climate factors may change the environmental landscape in which the species are embedded (Ackerly 2003; Dynesius and Jansson 2000; Jackson and Overpeck 2000; Jansson and Dynesius 2002), and if species fail to rapidly adapt to the new environmental conditions to which they are subjected (by expanding or shifting environmental tolerances), the climate shift may eventually lead to range fragmentation and subsequent allopatric speciation (Wiens 2004). The failure of species to adapt to environmental oscillations within its geographical range is known as niche conservatism (Wiens 2004), and besides range fragmentation, it has been also suggested that this process may have contributed to shaping the contemporary latitudinal gradients in species richness (Wiens and Donoghue 2004; Wiens and Graham 2005). Since most lineages may have originated and were adapted to tropical climatic regimes, species may either have become extinct in regions in which the climate changed from tropical to temperate (Hawkins et al. 2006), or may have failed to adapt to non-tropical climatic regimes (Wiens and Donoghue 2004). In this study, we used a computer simulation model to evaluate if the niche heritability process may contribute to range fragmentation and to real-world spatial patterns in species richness. By modeling the niche heritability process as variable parameter, in which it is possible to simulate conditions of strong niche conservatism or fast niche evolution, our model has shown that niche conservatism in an ancestral species may promote allopatric speciation, because species' geographical ranges are subjected to fragmentation if species fail to adapt to climate dynamics. Also, when we simulate a condition in which a lineage has originated in the tropics, and niche heritability is subjected to a relatively strong evolutionary conservatism process, our simulation model yields predictions that closely match the real-world spatial patterns in species richness.

Actually, climate change dynamics exert a stronger influence upon species' geographical ranges than fluctuations in the breadth of species' ecological niche do (Ackerly 2003; Jackson and Overpeck 2000). The fact is that species range boundaries in nature fluctuate greatly (Ackerly 2003; Gaston 2003; Holt 2003; Kirkpatrick and Barton 1997; MacArthur 1972), either due to environmental changes, or due to niche fluctuation caused by evolutionary, ecological, or demographic forces. Certainly, while the niche fluctuation affects only one species, a significant shift in climate factors would affect many species simultaneously, each one in a different fashion (Dynesius and Jansson 2000; Jackson and Overpeck 2000; Jansson and Dynesius 2002). Although we do think that the main source of geographical range dynamics might be environmental fluctuations, fluctuations in species' environmental tolerances throughout evolutionary time can also cause the same effect, although in slightly different time scales (Spicer and Gaston 1999), or at least can buffer the magnitude and direction of the biogeographical consequences of climate dynamics. Over long time scales, evolutionary changes in the breadth of ecological niches inevitably exert some impact over species' ranges (Holt 2003). Our reasons to implement niche fluctuation instead of environmental fluctuations are, on the one hand, the lack of availability of accurate multivariate data on time scale of climate change. On the other hand, the computational effort involved in simultaneously modifying several environmental variables and re-computing species geographical ranges across large spatial scales is probably beyond current computational capabilities. Nonetheless, although this feature of our model may seem not too realistic at first glance, our principal results and interpretations are not artifacts of this specific feature of model design, and we believe that any implementation of the niche conservatism process based on climatic changes acting over broad geographical scales would yield similar results.

One may argue that there is an inherent contradiction in modeling the effect of niche conservatism process by stochastically modifying the breadth of species' ecological niche. However, it is important to stress that, in our model, species' ecological niche modification is a surrogate for fluctuation of environmental factors and the generation of "environmental islands" (Ackerly 2003). Also, in our model, the environmental optimum on each niche axis of a species' ecological niche does not evolve (except for the niche optima shifts of daughter species at the moment of origination), but only the breadth of species' environmental tolerances are subjected to stochastic fluctuations. Finally, magnitude of change in the niche fluctuation process is modeled as a variable parameter, which can be regulated within a continuous spectrum from strong niche conservatism to fast niche evolution. Thus, under a pattern oriented modeling approach, we explored parameter space to find that magnitude of niche conservatism that maximizes model predictive power in relation to real-world patterns in species richness.

A parallel explanation for spatial patterns in species richness invokes a hypothetical diversification bias acting over evolutionary time. It has been suggested that this diversification bias could be consequence of a high speciation rates driven by environmental factors (Cardillo et al. 2005; Rohde 1992), or increased extinction rates driven by climate change (Diniz-Filho et al. 2004; Hawkins et al. 2005, 2006). In our model, we implemented a net diversification bias driven by high environmental productivity, which is hypothesized to buffer extinction rates. The reasoned biological mechanism behind this process is that high environmental productivity increases the carrying capacity of the environment for a species' population, and thus tends to reduce extinction probability (Currie et al. 2005). Assuming speciation rate remains constant, a lower extinction rate increases the net diversification rate. In fact, model predictive power increased when we simulated the joint effect of niche conservatism and diversification bias, because species tended to accumulate in highly productive tropical regions. Indeed, when the net diversification bias was activated, it becomes a very important model parameter. We believe that other process creating increased speciation in the tropics, or increased extinction outside the tropics, would have the same effect on the predicted spatial patterns in species richness, and thus could also be considered as possible explanations for real-world patterns.

As expected, the environmental variables used in the NCM model can explain much of real-world spatial patterns in species richness. Using all eight environmental variables as predictors of real-world spatial patterns in species in a multiple regression analysis, the calculated r^2 was equal to 0.756, a magnitude of explanation close to levels usually found in most correlative studies (e.g. Hawkins et al. 2003b). Thus, one could argue that the predictive power of our model is actually indirectly driven by the environmental variables that we used to characterize landscape patterns in South America, and that NCM could therefore not be interpreted as an independent predictor. However, it is important to highlight that, although useful, correlative studies lack the theoretical basis to explain mechanistically how these environmental variables drive spatial patterns in species richness (Currie et al. 1999, 2005). In contrast, our model introduces a strong mechanistic framework, and its predictions, which were analyzed under a robust statistical methodology, proved to be quite accurate in reproducing overall broad-scale patterns in species richness (although it failed to predict patterns of extreme species richness along the Andes or in Atlantic rainforest).

More importantly, the predicted spatial patterns under a “pure” niche conservatism process are also highly correlated with the environmental variables used in the model, and a multiple regression analyses yielded an $r^2 = 0.836$, slightly higher than the one obtained with the real dataset. However, it is important to emphasize that, under this condition of “pure” niche conservatism, the model does not use any kind of qualitative assessment of environmental variables controlling diversification (i.e., there is not a higher chance of speciation or extinction depending of the species’ position along the environmental space). This high correlation can be thought as an indirect effect of evolutionary history, and not as a direct effect of climate driving species richness (see below). Thus, it leads us to the conceptual distinction between cause and correlation (see Shipley 2002) and can be more clearly shown when we consider that the environmental condition of the region where the lineage has originated is very important to define the shape of the spatial patterns in species richness, because it defines the optimal environmental condition of the initial species, and by niche conservatism process, this optimal condition tends to be conserved between closely related species. This means that, if one lineage originates at the cold, dry, and low productivity region of Atacama Desert, for example, and the niche conservatism process is sufficiently strong, it is very likely that the model will predict this region as very rich in species. Indeed, the region of origin of the lineage is a very important factor driving model predictions, as highlighted by Wiens and Donoghue (2004).

However, it is noteworthy that the “pure” niche conservatism process cannot be considered absolutely random with respect to ancestral niche optima and evolution. Although speciation can occur with the same probability in any portion of species’ geographic range, in our model, extinction would be more frequent at regions of extreme environmental conditions (e.g. very high or very low temperature). For example, if a species has its fundamental niche optimum at the highest condition of a given environmental variable (i.e. productivity), and this environmental variable oscillates, decreasing its productivity level all other the domain, the species' realized niche (as expressed in its range size) would be more frequently affected due to "habitat loss". On the other hand, another species with niche optimum at intermediate environmental conditions would be able to track its habitat preferences across the geographical space, regardless of the direction of the shift of the environmental variable. In our model, which does not operate through environmental shifts, the same process is also applicable, because the potential niche of species with niche optima at extreme environmental conditions is subjected to greater reduction due to fluctuations in the breadth of the fundamental niche. Actually, it is the effect of boundary constraint acting over a species' potential niche. Thus, it is clear that the final geographical patterns in species richness would be also dependent not only on environmental shifts and the dynamics of niche evolution, but also on the particulars of the ancestral condition.

In contrast with the "pure" niche conservatism process, when the net diversification bias is activated, then the model indeed makes qualitative assessment of the characteristics of a given environmental condition to the macroevolutionary dynamics (i.e., rates of extinction and thus, net speciation, are related to position in niche space), and tends to buffer extinction in highly productive regions. Thus, this process favors an increase in net diversification rate in more productive regions, which then leads to a rapid, local species accumulation. Needless to say, because species richness will generally be well correlated with environmental productivity (e.g. AET, PET and NPP), adding this diversification bias to the model increases the predictive power of original environmental variables in relation to simulated patterns. In fact, using a multiple regression analysis to evaluate the agreement between the predicted richness patterns and the environmental variables used in the model, 84.5% of the variation in the predicted richness patterns is explained by the environmental variables.

In this study, our model has failed to explain the high bird richness at the Andean mountains. However, we believe that it is due to the coarse spatial resolution of our grid system, and that simulations of the niche conservatism process over finer grid cells could lead to better fits between predicted and real-world species richness within the Andes. This would be due to better description of local characteristics of the region, which would magnify the isolation and the origin of “environmental islands”. Additionally, other complex evolutionary and ecological factors may act specifically in this region to create conditions especially conducive to speciation. Further studies are necessary to focus the origin of extreme species richness in Andes using simulation models. High richness in the Atlantic coast (i.e., associated with Atlantic forest) was also not well predicted by simulations, although it is more difficult to find a reason why this happened.

Our simulation set was based on randomly selecting the map position where the lineages originate, in order to study the effect of niche specialization to different initial environmental conditions in the origin of the clade. However, under a more realistic “biogeographical” modeling approach, one might use the niche conservatism process to model the spatial patterns in species richness for a small lineage or clade that has a well known center of origin. Also, by comparing environmental optima and tolerances of phylogenetically related species within the selected group, or disjunct lineages with intercontinental distributions, it is also possible to estimate the magnitude of evolutionary divergence (or niche conservatism) from a common ancestor, which would lead to a more specific test of our simulation model (Ackerly 2003; Peterson et al. 1999; Ricklefs and Latham 1992). Thus, after estimating a specific position of the center of origin, and the magnitude of niche conservatism, one can evaluate the correlations between predicted and real-world spatial patterns in species richness. We expect that model parameterization based on real-world phylogenetic and biogeographical data may decrease the need for exploring a wide scope of parameters, and also further improve model predictive power.

Finally, further simulation studies could implement more realistic models, including habitat choice, dispersal patterns through time (Holt 2003), species’ interaction and consequences of niche overlap (Case and Taper 2000), gene flow within and between range fragments (Kirkpatrick and Barton 1997), population and metapopulation dynamics (Hanski 1998), genetic basis of speciation (e.g. Gavrilets and Vose 2005; Gomulkiewicz and Holt 1995;), and dynamics of climate change and its effects on species' geographical ranges (Ackerly 2003; Dynesius and Jansson 2000; Jackson and Overpeck 2000; Jansson and Dynesius 2002). However, this would lead to much more complex models and it is possible that the results of such models would become quite difficult to interpret and to evaluate with respect to the relative importance and role of model parameters. Although it is difficult, at the present time, to evaluate this particular problem of model complexity, hopefully our NCM model falls well within the "Medawar zone" (Grimm et al. 2005), in which model design includes some sacrifice of biological realism for the sake of general interpretation of model results, which can be projected in different scales and hierarchical levels, instead of very specific and "realistic" models, which require cumbersome analytical procedures.

Our "Niche Conservatism Model" is an attempt to merge, in a spatially explicit context, several widely accepted evolutionary, biogeographical and ecological mechanisms, together with recently developed ideas on niche evolution. NCM is a simulation model designed to understand the evolution of spatial patterns in species richness, under niche conservatism. We found that it is very likely that niche conservatism process has played a role in shaping these patterns, and indeed our model has strong and accurate predictive power of real-world spatial patterns of South American birds. However, diversification bias, linked to environmental factors, seems to be also important, and gives more realism to predicted patterns. We do not claim that the validity of our results and interpretations are based on the materiality of our study (Sismondo 1999), but on the close connection between our model and (*i*) the underlying scientific theory that supports this study, (*ii*) the accuracy and fidelity of the model-building techniques we employed, (*iii*) the real-world database on geographical distribution of birds, (*iv*) the pattern-oriented approach used to parameterize the model, (*v*) the stability and robustness of our model in relation to parameter variation, (*vi*) the strength of the statistical inference used here, and most important (*vii*) the power of the model to predict real-world patterns. Finally, we agree with Wiens and Donoghue (2004) that ecology and historical biogeography have much to offer each other, we and hope this paper will motivate researchers to study spatial patterns in species richness under an integrated theoretical and methodological framework.

Acknowledgements

We thank to Rob Colwell and Mau Bini for the careful revision and the useful suggestions on earlier versions of the manuscript. The first author appreciated the “coffee-break” theoretical discussion on biogeography and evolutionary biology with Mau Bini and Cleiber Vieira, which somehow influenced the development of the model design. Biological data were provided by NatureServe in collaboration with Robert Ridgely, James Zook, TNC, CI, WWF, and Environment Canada. Simulations were performed on Dell® Precision® 450 series Workstations provided by CNPq/SECTEC-GO (PRONEX Project nº 23234156). Both authors have been continuously supported by several CNPq fellowships.

References

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. International Journal of Plant Science 164:S165-S184.

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

Bokma, F., Bokma, J. and Mönkkönen, M. 2001 Random process and geographic species richness patterns: why so few species in the north? Ecography 24:43-49.

Bromham, L. and Cardillo, M. 2003 Testing the link between the latitudinal gradient in species richness and rates of molecular evolution. Journal of Evolutionary Biology 16: 200-207.

Brown, J. H., Stevens, G. C. and Kaufman, D. M. 1996. The geographic range size: size, shape, boundaries, and internal structure. Annual Reviews of Ecology and Systematics 27:597-623.

Cardillo, M., Orme, C. D. L. and Owens, I. P. F. 2005. Testing for latitudinal bias in diversification rates: an example using new world birds. Ecology 86:2278-2287.

Case, T. J. and Taper, M. L. 2000. Interespecific competition, environmental gradients, gene flow, and the coevolution of species' borders. American Naturalist 155:583-605.

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

Cressie, N. A. C. 1993. Statistics for Spatial Data. Revised Edition (Wiley series in probability and mathematical statistics. Applied probability and statistics section), Wiley, New York.

Currie, D. J., Francis, A. P. and Kerr, J. T. 1999. Some general propositions about the study of spatial patterns of species richness. Ecoscience 6:392-399.

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guégan, J-F., Hawkins, B. A., Kaufman, D. M., Kerr, J. T., Oberdorff, T., O'Brien, E. and Turner, J. R. G. 2005. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecology Letters 7:1121-1134.

Davies, T. J., Grenyer, R. and Gittleman, J. L. 2005. Phylogeny can make the mid-domain effect an inappropriate null model. Biology Letters doi: 10.1098/rsbl.2005.297.

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

Diniz-Filho, J. A. F., Rangel, T. F. L. V. B. and Hawkins, B. A. 2004. A test of multiple hypotheses for the species richness gradient of South American owls. Oecologia 140:633-638.

Dynesius, M. and Jansson, R. 2000. Evolutionary consequences of changes in species' geographical distribution driven by Milankovitch climate oscillations. Proceedings of the National Academy of Science USA 97:9115-9120.

Dutilleul, P. 1993. Modifying the t test for assessing the correlation between two spatial processes. Biometrics 49:305-314.

Farrell, B. D. and Mitter, C. 1993. Phylogenetic determinants of insect/plant community diversity. In: Species diversity in ecological communities: Historical and geographical perspectives (eds. Ricklefs, R. E. and Schluter, D.) University of Chicago Press, Chicago.

Gaston, K. J. 2000 Global patterns in biodiversity. Nature 405:220-227.

Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University Press, Oxford.

Gavrilets, S. and Vose, A. 2005. Dynamic patterns of adaptive radiation. Proceedings of the National Academy of Science USA 102:18040-18045.

Gomulkiewicz R. and Holt, R. D. 1995. When does evolution by natural selection prevent extinction? Evolution 49:201-207.

Gotelli, N. J. 1998. A primer of Ecology. Sinauer, Sunderland.

Gould, S. J. 2002. The structure of evolutionary theory. Harvard University Press. Cambridge.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H-H., Weiner, J., Wiegand, T. and DeAngelis, D. L. 2005. Pattern-Oriented Modeling of Agent-Based Complex systems: Lessons from Ecology. Science 310:987-991.

Hanski, I. 1998. Metapopulation dynamics. Nature 396:41-49.

Hawkins, B. A. 2001. Ecology's oldest pattern? Trends in Ecology and Evolution 16:470.

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

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

Hawkins, B. A., Diniz-Filho, J. A. F. and Soeller, S. A. 2005. Water links the historical and contemporary components of the Australian bird diversity gradient. Journal of Biogeography 32:1035-1042.

Hawkins, B. A., Diniz-Filho, J. A. F., Jaramillo, C. A. and Soeller, S. A. 2006. Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. Journal of Biogeography 33:in press.

Hillebrand, H. 2004. On the Generality of the Latitudinal Diversity Gradient. TheAmerican Naturalist 163:192-211.

Holt, R. D. 1996. Demographic constrains in evolution: towards unifying the evolutionary theories of senescence and niche conservatism. Evolutionary Ecology 10:1-11.

Holt, R. D. 2003. On the evolutionary ecology of species' ranges. Evolutionary Ecology Research 5:159-178.

Holt, R. D. and Gaines, M. S. 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolutionary Ecology 6:433-447.

Hunt, G., Roy, K., and Jablonski, D. 2005. Species-level heritability reaffirmed: a comment on "on the heritability of geographic range sizes". American Naturalist 166:129-135.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symp. Quant. Biol 22:415-427.

Jackson, S.T. and Overpeck, J.T. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology 26:S194-S220.

Jansson, R. and Dynesius, M. 2002. The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. Annual Reviews of Ecology and Systematics 33:741-777.

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

Kirkpatrick, M. and Barton, N. H. 1997. Evolution of a species' range. American Naturalist 150:1-23.

Kucharik, C. J., Foley, J.A., Delire, C., Fisher, V.A., Coe, M.T., Lenters, J., Young-Molling, C., Ramankutty, N., Norman, J.M., e Gower, S.T. 2000. Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance and vegetation structure. Global Biogeochemical Cycles 14:795-825.

Krebs, C. J. 2001. Ecology: The experimental analysis of distribution and abundance. 5th. ed. Addison-Wesley.

Latham, R. E. and Ricklefs, R. E. 1993. Continental comparisons of temperate-zone tree species diversity. In: Species diversity in ecological communities: Historical and geographical perspectives (eds. Ricklefs, R. E. and Schluter, D.) University of Chicago Press, Chicago.

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

Lomolino, M. V., Riddle, B. R. and Brown, J. H. 2005. Biogeography. 3rd ed. Sinauer, Sunderland.

Maurer, B. A. 1999. Untangling Ecological Complexity: The Macroscopic Perspective. Univ. Chicago Press. Chicago.

McArthur R. H. 1972. Geographical Ecology. Harper & Row, New York.

Naylor, T. H., Balintfly, J. L., Burdick, D. S. and Chu, K. 1966. Computer Simulation Techniques. Wiley & Sons, New York.

New, M.G., Hulme, M., e Jones, P.D. 1999. Representing 20th century space-time climate variability. I: Development of a 1961-1990 mean monthly terrestrial climatology. Journal of Climate. 12:829-856.

NOAA. National Oceanic and Atmospheric Administration and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

Peck, S. L. 2004. Simulation as experiment: a philosophical reassessment for biological modeling. Trends in Ecology and Evolution 19:530-534.

Quian, H. and Ricklefs, R. E. 2004. Geographical distribution and ecological conservatism of disjuct genera of vascular plants in eastern Asia and eastern North America. Journal of Ecology 92:253-265.

Peterson, A. T., Soberón, J. and Sánchez-Cordero, V. 1999. Conservatism of Ecological Niches in Evolutionary Time. Science 285:1265-1267.

Prinzing, A., Durka, W., Klotz, S. and Brandl, R. 2001. The niche of higher plants: evidence for phylogenetic conservatism. Proceedings of the Royal Society, London, B 268:2383-2389.

Rangel, T. F. L. V. B. and Diniz-Filho, J. A. F. 2005a. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography 28:253-263.

Rangel, T. F. L. V. B. and Diniz-Filho, J. A. F. 2005b. 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. and Bini, L. M. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. Global Ecology and Biogeography 15:in press.

Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters 7:1-15.

Ricklefs, R. E. and Latham, R. E. 1992. Intercontinental correlation of geographical ranges suggests stasis in ecological traits of relict genera of temperate perennial herbs. American Naturalist 139:1305-1321.

Ridgely, R. S., Allnutt, T. F., Brooks, T., McNicol, D. K., Mehlman, D. W., Young, B. E. and Zook, J. R. 2005. Digital Distribution Maps of the Birds of the Western Hemisphere, version 2.1. NatureServe, Arlington, Virginia, USA. Available at www.natureserve.org/getData/

Rosenzwig, M. L. 1995. Species diversity in time and space. Cambridge University Press, Cambridge.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65:514-527.

Rohde, K. 1999. Latitudinal gradients in species diversity and Rapoport's rule revisited: a review of recent work and what can parasites teach us about the causes of the gradients? Ecography 22:593-613.

Shipley, B. 2002. Cause and Correlation in Biology. Cambridge University Press, Cambridge.

Sismondo, S. 1999. Models, simulations and their objects. Science in Context 12:247-260.

Spicer, J. I. and Gaston, K. J. 1999. Physiological diversity and its ecological implication. Blackwell Publishing, Oxford.

Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics 2:1-10.

Sokal, R. R. and Rohlf, F. J. 1995. Biometry, 3rd ed. Freeman.

Stephens, P. R. and Wiens, J. J. 2003. Explaining species richness from continents to communities: the time-for-speciation effect in Emydid turtles. American Naturalist 161:112-128.

Svenning, J. -C. 2003. Deterministic plio-pleistocene extinctions in the European cool-temperature tree flora. Ecology Letters 6:646-653.

Tognelli, M. F. and Kelt, D. A. 2004. Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. Ecography 27:427-436.

Wallace, A. R. 1878. Tropical nature and other essays. MacMillan.

Webb, C. O., Ackerly, D. D., McPeek, M. A. and Donoghue, M. J. 2002. Phylogenies and community ecology. Annual Reviews of Ecology and Systematics 33:475-505.

Webb, T. J. and Gaston, K. J. 2003. On the heritability of geographic range sizes. American Naturalist 161:553-566.

Webb, T. J. and Gaston, K. J. 2005. Heritability of geographic range sizes revisited: a reply to Hunt et al. American Naturalist 166:136-143.

Whittaker, R. J., Willis, K. J. and Field, R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453-470.

Wiens, J. J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution 58:193-197.

Wiens, J. J. and Donoghue, M. J. 2004. Historical biogeography, ecology and species richness. Trends in Ecology and Evolution 19:639-644.

Wiens, J. J. and Graham, C. H. 2005. Niche Conservatism: integrating evolution, ecology, and conservation biology. Annual Reviews of Ecology and Systematics 36:519-539.

Willig, M. R., Kaufman, D. M. and Stevens, R. D. 2003. Latitudinal gradients of biodiversity: pattern, process, scale and synthesis. Annual Reviews of Ecology and Systematics 34:273-309.

Willmott, C. J. and Kenji, M. 2001. Terrestrial Water Budget Data Archive: Monthly Time Series (1950-1999). Available at www.sage.wisc.edu/atlas/

Winsberg, E. 1999. Sanctioning models: the epistemology of a simulation. Science in Context 12:247-260.

Winsberg, E. 2003. Simulated experiments: methodology for a virtual world. Philosophy of Science 70:105-125.

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

Wright, D. H., Currie, D. J. and Maurer, B. A. 1993. Energy supply and patterns of species richness on local and regional scales. In: Species diversity in ecological communities: Historical and geographical perspectives (eds. Ricklefs, R. E. and Schluter, D.) University of Chicago Press, Chicago.

Capítulo 2

Este capítulo foi submetido e aceito para publicação em 2006
na *Global Ecology and Biogeography* (GEB 2005-0093)

Ecological Soundings

Towards an integrated computational tool for spatial analysis in macroecology and biogeography

Thiago Fernando L. V. B. Rangel[*],
José Alexandre Felizola Diniz-Filho
Luis Mauricio Bini

Departamento de Biologia Geral, ICB, Universidade Federal de Goiás
CP 131, 74.001-970, Goiânia, GO, Brasil

[*] Author for correspondence; e-mail: t.rangel@terra.com.br

Running head: Spatial Analysis in Macroecology

Keywords: spatial autocorrelation, spatial statistics, modeling, macroecology, biogeography, statistical package, software

Abstract

Because most macroecological and biodiversity data are spatially autocorrelated, special tools for describing spatial structures and dealing with hypothesis-testing are usually required. Unfortunately, most of these methods have not been available in a single statistical package. Consequently, using these tools is still a challenge for most ecologists and biogeographers. In this paper, we present SAM (Spatial Analysis in Macroecology), a new, easy-to-use, freeware package for spatial analysis in macroecology and biogeography. Through an intuitive, fully graphical interface, this package allows the user to describe spatial patterns in variables and provides an explicit spatial framework for standard techniques of regression and correlation. Moran's *I* autocorrelation coefficient can be calculated based on a range of matrices describing spatial relationships, for original variables as well as for residuals of regression models, which can also include filtering components (obtained by standard Trend Surface Analysis or by Principal Coordinates of Neighbor Matrices). SAM also offers tools for correcting the number of degrees of freedom when calculating the significance of correlation coefficients. Explicit spatial modeling using several forms of autoregression and generalized least squares models are also available. We believe this new tool will provide researchers with the basic statistical tools to resolve autocorrelation problems and, simultaneously, to explore spatial components in macroecological and biogeographical data. Although the program was designed primarily for the applications in macroecology and biogeography, most of SAM's statistical tools will be useful for all kinds of surface pattern spatial analysis. The program is freely available at www.ecoevol.ufg.br/sam (permanent URL at http://purl.oclc.org/sam/).

"The process of preparing programs for a digital computer is especially attractive, not only because it can be economically and scientifically rewarding, but also because it can be an aesthetic experience much like composing poetry or music."

Donald E. Knuth

Introduction

Ecologists recognize that nearly all macroecological and biodiversity data show strong spatial patterns, driven by spatially structured biological processes, and consequently are often spatially autocorrelated. Following Legendre (1993), spatial autocorrelation may be defined as "the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) that expected for randomly associated pairs of observations". The endogenous or exogenous causes of spatial structure or autocorrelation in ecological and biogeographical data are a function not only of how spatially dynamic processes drive observed variables such as abundance, richness, endemism, body size and range size, but also depend on how the data are collected by spatial sampling schemes (Diniz-Filho *et al*., 2003; Fortin & Dale, 2005).

Spatially autocorrelated datasets present both a potential statistical problem and an opportunity to recognize the importance and understand the causes of the spatial structure in ecology (Legendre, 1993; Diniz-Filho *et al*., 2003). Accordingly, ecologists and biogeographers now dealing with spatially autocorrelated datasets may act in two different ways: (i) they ignore it or brush it aside (jeopardizing publication of their papers in the better mainstream journals) or (ii) they incorporate realistic assumptions about spatial structure in their analyses and endeavor to understand the underlying spatial processes.

Some idea of the emerging importance of spatial autocorrelation in the worldwide scientific literature can be gleaned using a bibliometric approach. The Thomson Institute (ISI Web of Science) bibliographic database (1945 to June, 20 2005) was used to identify all papers indexed that used the terms "spatial" and "autocorrelation". We found a total of 2284 studies distributed in 82 subject categories. After 1990, a clear increase in the number of papers that used the terms "spatial" and "autocorrelation" in the title or in the abstract was detected in nearly all subject categories (Figure 1). This indicates the interest in spatial autocorrelation by different research fields. However, the predominance of papers on spatial autocorrelation recently published in ecological journals is notable, indicating that, at least with respect to this issue, ecologists produce scientific knowledge and not just consume it (Peters, 1991).

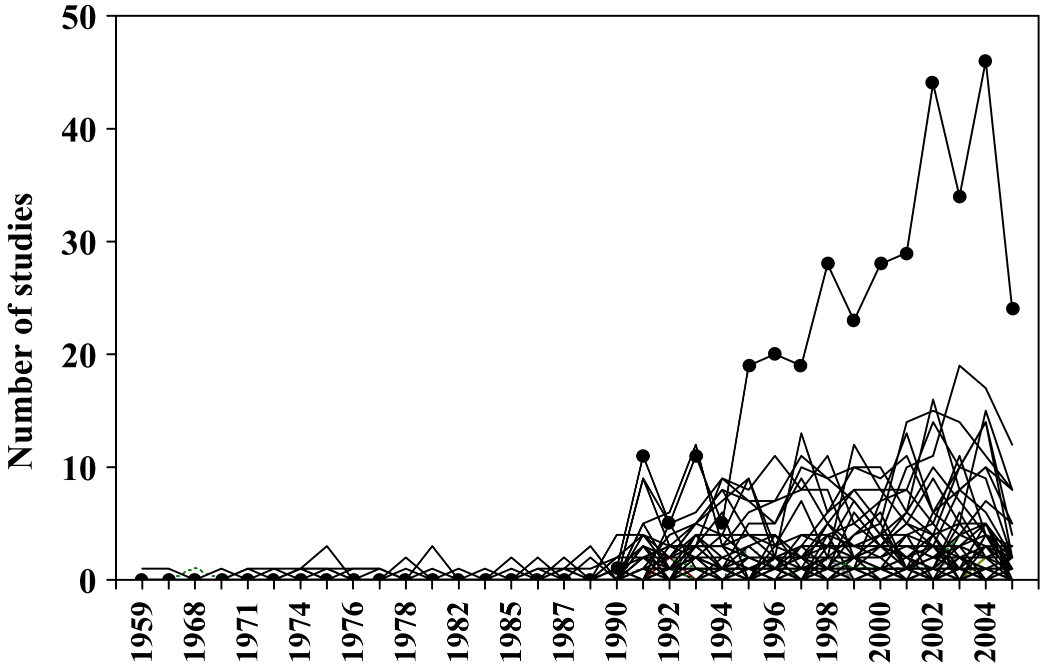

Figure 1: Temporal trends in the number of papers that consider spatial autocorrelation, distributed among 82 subject categories distinguished by ISI. Filled circles indicate the number of studies published in ecological journals, whereas other lines indicate trends in all other categories, showing that that spatial autocorrelation is now a major issue in ecology and biogeography.

Given the increasing importance of spatial autocorrelation analysis in ecology and biogeography, it is important to stress that there are still difficulties in applying these methods routinely, since they are usually not available in most commercial statistical packages. This was one of the principal problems identified by the participants of the workshop on statistical methods for spatial analysis in ecology, sponsored by the National Center for Ecological Analysis and Synthesis (NCEAS). As stated by Liebhold & Gurevitch (2002), "Ultimately it would be desirable to develop software packages…that are capable of computing a full array of spatial statistics". But for most researchers in our field, understanding these complex spatial analyses and modeling approaches is still a challenge. We believe that a key step in improving the understanding of spatial issues in ecology and biogeography is to provide comprehensive, user-friendly and freely available software, together with a useful help file. In this paper, we introduce a computational program named *Spatial Analysis in Macroecology* (SAM), a new software application for managing spatial analyses that was designed with the particular needs of macroecologists and biogeographers in mind. We also highlight how these analyses interact with some current issues that are of interest for macroecologists and biogeographers.

Exploring and identifying the spatial structure of the data

The first step in every spatial analysis, as in any other statistical procedure, is an exhaustive exploratory data analysis (EDA). Conducting an EDA facilitates immensely the visualization of "hidden" patterns in a (usually) large-sized macroecological dataset (see Rossi *et al.*, 1992 for a discussion of EDA within the framework of spatial analysis). The EDA module of SAM provides five analytical methods to: (i) compute basic and descriptive statistics, (ii) visualize the data (distribution and relationships) by means of graphs (2D and 3D scatterplots, histograms and maps), (iii) create/edit a connectivity matrix using several, alternative criteria (e.g. Delaunay triangulation, Gabriel, Minimum Spanning Tree; Sokal & Oden, 1978a; see more below), (iv) transform the data (log, square-root, etc), and (v) reduce data dimensionality (principal component analysis). This last approach may be particularly useful when analyzing highly correlated variables, such as climatic data at broad scales, reducing the problem of multicolinearity in multiple regression (Philippi, 1993; Graham 2003). Also, SAM has two special modules that handle presence-absence data in matrices of species X spatial unit (e.g. cells in a $1^o \times 1^o$ grid), allowing the calculation of richness values using different combinations of species based on macroecological criteria, such as body size, range size, habitat type or evolutionary age (Marquet's *et al.*, 2004 deconstructive approach; see also Cardillo, 2002; Jetz & Rahbek, 2002; Bini *et al.*, 2004; Rahbek, 2005; Hawkins *et al.*, 2005). For example, a researcher may be interested in calculating spatial pattern in species richness for small-bodied and large-bodied species (Hillebrand & Azovsky, 2001), or computing the average body size of several species in each spatial unit.

SAM includes different modules to test for spatial autocorrelation in a variable (e.g. species richness) and also to distinguish the type of spatial structure (e.g. clines, patches, etc; see Legendre & Fortin, 1989). Following Rossi *et al*., (1992), these methods (spatial continuity measures) quantify the relationship between the value of a variable in one spatial unit and the value of the same variable in other spatial units.

Because Moran's *I* coefficient is the most commonly used statistic for autocorrelation analyses in macroecological and biogeographical studies (see Bagley & Fox, 2000; Diniz Filho *et al*., 2003 and Bini *et al*., 2004 for recent applications of this coefficient), and because it is statistically robust (Tiefelsdorf, 2000), it is the primary statistic for describing spatial structure in SAM, although semi-variance is also available. Moran's *I* can be calculated for multiple distance classes, established using a variety of different criteria, allowing the generation of a graph relating autocorrelation coefficients to the spatial distances among sampling units, called a spatial correlogram (Sokal & Oden, 1978a,b). If there is only one variable in the dataset, several ecological issues can be evaluated with a correlogram. Minimally, the analysis of a correlogram furnishes a description of the spatial pattern in the data (Legendre & Fortin, 1989). As we will show below, this is also an important tool for evaluating if a fundamental assumption (independence of residuals) of general linear models holds or not.

Three issues regarding the description of the spatial structure in the data using autocorrelation coefficients have been inadequately discussed in ecological and biogeographical literature. First, it is possible to estimate Moran's *I* coefficient considering various types of geographical connections (criteria for connecting geographical localities). As indicated by Sokal & Oden (1978a), the criteria for considering a pair of sampling units as connected or not connected depend on the hypothesis being tested (e.g. two sampling units in a stream network, despite the geographical proximity, might be considered unconnected; see Ganio *et al*., 2005; Peres-Neto, 2004). For this reason, SAM offers the opportunity to create an ad hoc connectivity matrix that indicates the relationships among sampling units by consideration of the hypothesis under study (e.g. to take into account the presence of an ecological barrier, or to match migration routes or dispersion flows). Generally, connections are treated as some function of geographical proximity. As variations on this approach, SAM offers five standard methods for defining connections among the sampling units: Delaunay Triangulation, Gabriel Criterion, Relative Neighbourhood, Minimum Spanning Tree and Distance Criterion (Legendre & Legendre, 1998; Fortin & Dale, 2005).

Second, it is also possible to run a LISA (Local Indicator of Spatial Autocorrelation) analysis (Anselin, 1995; Sokal *et al*., 1998) in SAM, an overlooked method in macroecology. LISA can be used to measure the contribution of each sampling unit to the overall (global) level of spatial autocorrelation (Cocu *et al*., 2005). Finally, it is possible to test the statistical significance of Moran's *I* using randomization (Monte Carlo) (see Manly, 1997; Tiefelsdorf, 2000), which is a reliable way to assess statistical significance, especially for small sample sizes (Sokal & Oden, 1978 a,b).

At this point it is worthwhile to mention another very useful feature implemented in SAM. Most spatial analyses use a matrix of distances among pairs of sampling units to describe the spatial relationships in the data, usually assuming a planar surface (Euclidean distance). However, for broad spatial scales (e.g.: continents or large domains), which are common in macroecology and biogeography, the calculation of planar distances may bias the spatial relationship among sampling units because of the curvature of the Earth. For this reason, SAM allows the user to compute geodesic surface distance among pairs of sampling points, with an accuracy of about 50 meters, assuming not only that the earth is a approximately spherical, but also taking into account the actual polar flattening of the Earth and the equatorial bulge. Geodesic surface distances may be used for all spatial analyses in SAM, but only if geographic coordinates are measured in decimal degrees of latitude and longitude.

SAM provides two basic ways to describe and control for spatial structure in the data under the general concept of spatial filtering. The first is the well-known trend surface analysis (TSA), which has been extensively discussed elsewhere (Wartenberg, 1985; Davies, 1986). SAM allows the automatic calculation of TSA polynomial expansions up to 6th order. However, a highly recommended alternative is Principal Coordinates of Neighbor Matrices (PCNM; see Borcard & Legendre, 2002; Borcard *et al.*, 2004; Diniz-Filho & Bini, 2005), because of the efficiency of this method in partitioning variation between spatial and environmental components (Borcard *et al.*, 2004). Further, the importance of scale in detecting the magnitude and direction of relationships among variables is well known, and PCNM can deal effectively with this issue (Whittaker *et al.*, 2001; Rahbek, 2005).

Spatial filters obtained by TSA or PCNM can be used in different ways, depending on how spatial patterns are taken into account. One approach is to use them to remove all spatial structure from the data and work only with (residual) non-spatial components to evaluate, for example, the effect of predictors on richness. In this example, applying this approach would be appropriate if broad-scale spatial processes did not contain, or could not reliably reveal, information regarding causal process associated with richness, due to confounding effects of intrinsic and extrinsic processes affecting this variable. Alternatively, these filters can be treated as candidate explanatory variables together with other, environmental predictors. With this approach, the effects of environmental predictors are evaluated as partial effects, taking space into account explicitly (see below). These two different approaches may produce different results, depending on the collinearity between predictors and space.

A different subject in the analysis of spatial pattern is the identification of patches and regions in space (Fortin & Dale, 2005). This approach may be especially necessary in a broad or multiple scale spatial analysis, when several ecological processes may be independently driving different regions of an observed spatial pattern. However, the likelihood of detecting these ecological processes depends on our ability to delineate ecological patches, boundaries, edges or ecotones (Fortin, 1994; Fortin & Drapeau, 1995; Fortin *et al.* 2000; Oden *et al*., 1993), and also on the ecological processes under investigation, the sampling design and the employed analytical methodology. To help macroecologists and biogeographers detect edges, two edge detection algorithms, called triangulation-wombling and lattice-wombling (Fagan *et al.*, 2003; Fortin & Drapeau, 1995) are available in SAM. The difference between these two methods concerns rules to join adjacent sampling units, although, in both methods, a region of with spatial discontinuity is detected by the steepness and direction of the slope of the plane formed by a set of joint sampling units (for details, see Fortin & Dale, 2005).

Modeling and hypothesis testing

The most frequently discussed issue in the ecological literature regarding spatial autocorrelation is the inflation of Type I errors in significance tests of correlation and regression analyses (Legendre, 1993; Diniz-Filho *et al*., 2003 and references therein). If two (or more) variables are each strongly spatially autocorrelated, spatial units close in geographic space are partially redundant with respect to the information they provide about the relationships between variables. In other words, in the presence of spatial autocorrelation, the number of degrees of freedom is overestimated and, consequently, confidence intervals are much narrower than they should be. This may cause an error in judging the statistical significance under a null hypothesis. Thus, the non-independence of data caused by spatial structure can lead these analyses to be quite liberal, and thus, even variables that are in fact weakly correlated will appear to yield significant coefficients due to the confounding effects of space. Significance inflation is quite important when trying to understand the effect of different predictors on a response variable, such as which environmental factors give better explanations for spatial variation in species richness. Such analyses are beyond the simple description of spatial variation previously described, and taking the spatial dimension into account usually improves the ability to model the spatial variation and understand the causal factors underlying it.

The simplest solution for testing a correlation coefficient in the presence of autocorrelation is to adjust the number of degrees of freedom, an approach developed by Clifford *et al.* (1989) and Dutilleul (1993) (see Legendre *et al.*, 2002 and Hawkins *et al.*, 2005 for discussions and applications). SAM provides two estimators to calculate the geographically effective number of degrees of freedom, both using spatial correlograms of the raw variables to be correlated. Although they usually provide quite similar results, they have different computational requirements (because Dutilleul's approach is more computationally intensive, it will demand more time for large matrices).

For a simple spatial modeling, SAM provides tools for ordinary least squares (OLS) regression, with three special features: (i) the evaluation and mapping of spatial structure in model residuals (which may reveal the need for explicit spatial modeling), (ii) partial regression analysis, using up to a 6th order polynomial expansion of geographical coordinates and (iii) the calculation of the Akaike information criterion (AIC) (Burnham & Anderson, 2002; Johnson & Omland, 2004), allowing an easy and powerful comparative evaluation of model fit when competing hypothesis are confronted with data. AIC as computed by SAM is based on the sum of squares of residuals and provides a approximation of AIC based on likelihood under a normal distribution of error terms (Mangel & Hilborne, 1997).

When strong autocorrelation is found in model residuals, alternative modeling strategies are available. The first is to include filters (TSA or PCNM), as previously discussed, along with the predictive variables, to minimize residual autocorrelation (see Diniz-Filho & Bini, 2005). However, SAM also allows the fitting of explicit spatial regression models to data. Three forms of autoregression models (ARM) are available.

The first set of routines allows estimation of 'lagged-models' (see Haining, 1990; 2002), which are based on fitting a pure autoregressive model that describes the spatial structure of only response variable **Y**, given by

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{e}$$

where ρ is the autoregression parameter, and the matrix **W** contains neighbor weights (w_{ij}), indicating the relationships among spatial units. The elements w_{ij} can be given by the connectivity matrices previously discussed or as an inverse power function of geographic distances (d_{ij}), given by functions of the form $w_{ij} = 1 / d_{ij}^{\alpha}$, where α is an additional parameter that regulates the relationship and that usually improves the performance of the model (Davis, 1986).

Because ARM allows a description of the spatial structure in data, it could just as well have been offered in within the 'Structure' section of SAM. It appears, instead, in the 'Modeling' section as the basis of more complex spatial regression models that can be used to evaluate the effects of predictors on the response variable, by adding additional terms (see Haining 1990; 2002). The first option for added terms assumes that the autoregressive process modeled by ARM occurs only in the response variable (lagged-response model), and thus includes a term for the spatial autocorrelation in **Y**, as in ARM (above), but also includes the standard term for the predictors in OLS. The model then becomes

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X}\beta + \mathbf{e}$$

where β is a vector representing the slopes associated with the predictors in the original predictor matrix **X**. Alternatively, autocorrelation can affect both response and predictor variables (lagged-predictor model, or Durbin econometric model – see Anselin, 1988). In this case yet another term must appear in the model, which now takes the form

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X}\beta + \mathbf{WX}\gamma + \mathbf{e}$$

where γ are the autoregressive parameters of each of the predictors. Note that in this more complex model there is an autoregression parameter for each predictor. In all cases, AIC and residual spatial autocorrelation can be used to choose among alternative models, which can also be generated using different α values, as discussed for the simple ARM. These last two models are in fact fitted by working with the residuals **e** of the pure ARM, described above, in a standard OLS regression. For this reason, these models can be also interpreted as techniques that filter the effect of space (see Haining, 1990; Anselin, 2002). For the lagged-response model, the ARM residuals of the response variable are regressed against the original predictors, using OLS, whereas in the lagged-predictor model the ARM residuals of the response variable is regressed against ARM residuals of each predictor variable (Haining, 1990, p. 347).

Another explicit way to take autocorrelation into account in a regression is by changing the estimator of the vector of slopes (β) by applying a generalized least-squares (GLS) model that incorporates spatial structure directly into model residuals (see Selmi & Boulinier, 2001; Hawkins & Diniz-Filho, 2002; Evans *et al.*, 2005). This vector is given by

$$\beta = (\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X})^{-1}\,\mathbf{X}^T\,\mathbf{C}^{-1}\mathbf{Y}$$

where **C** is the covariance among residuals. In the standard OLS model, it is assumed that $\mathbf{C} = \mathbf{I}\sigma^2$, (**I** is an identity matrix, and σ^2 is the variance of the residuals) so that residuals are independent. However, it is possible to model the values of **C** using a semi-variogram, fitted by exponential, powered-exponential, Gaussian, spherical, hole-effect, Matérn, truncated-linear or pure-nugget models (Legendre & Legendre, 1998; Banerjee *et al.*, 2004). Fitting these models usually requires an iterative process, and SAM allows the user to manually fit the semi-variograms and decide among them based on visual inspection or on their explanatory ability. After fitting the semi-variograms, regression slopes of GLS are obtained — in effect, a two-step evaluation (modeling OLS residuals then using their spatial structure to approximate GLS estimation). This approach is sometimes called 'kriging regression' (Cressie, 1993).

The GLS framework also allows the direct computation of other forms of spatial regressions that have been recently used in ecology and macroecology (e.g., Lichstein et al., 2002; Tognelli & Kelt, 2004; Dark, 2004), such as Simultaneous and Conditional Autoregression Models (SAR and CAR, respectively), by computing the matrix **C** with different formats (see Haining, 1990; Cressie, 1992). For SAR, the covariance matrix among residuals is given by

$$\mathbf{C} = \sigma^2\,[(\mathbf{I} - \rho\mathbf{W})^T]^{-1}\,[\mathbf{I} - \rho\mathbf{W}]^{-1}$$

where σ^2 is the variance of the residuals and **I** is an *n* x *n* identity matrix. For CAR, this matrix is given as

$$\mathbf{C} = [(\sigma^2\, \mathbf{W}_{i+})\mathbf{I}]\, (\mathbf{I} - \rho\mathbf{W})^{-1}$$

SAM also computes the moving average model, where the covariance matrix among residuals is given by

$$\mathbf{C} = \sigma^2\, [(\mathbf{I} + \rho\mathbf{W})(\mathbf{I} + \rho\mathbf{W})]$$

See Wall (2004) for the conditions under which these three models can be expressed, as described above, as functions only of autoregressive parameter ρ and neighbor weight matrix **W**. Note that, in these models, the matrix of neighbor weights (**W**) may also be computed as an inverse power function of geographic distances among sampling units ($w_{ij} = 1 / d_{ij}^{\alpha}$).

The GLS is then a regression in which the spatial component is explicitly modeled in the residual terms, defined by the fitted semi-variogram. Thus, these residuals contain a strong spatial component, which must be decomposed into spatially-structured residuals and a pure error term using Cholesky decomposition (see Cressie, 1993, p. 202; Haining, 2002, p. 333). This error vector **e**, or noise component, is given as

$$\mathbf{e} = \mathbf{L}^{-1}\, (\mathbf{Y} - \mathbf{X}\beta)$$

where β is the vector of estimated slopes and $\mathbf{L}\mathbf{L}^{T} = \mathbf{C}$, so that the **L** matrix can be obtained by the Cholesky decomposition of the covariance among residuals. In the GLS, this error term then is a function of the model used to fit the semi-variogram, whereas in SAR, CAR and MA models the error term is a function of the autoregressive parameter according to the functions defined above. The effectiveness of GLS-based models, in terms of taking autocorrelation into account, can be judged by absence of spatial structures in this error term. These properties can be visualized in SAM by spatial correlograms of residual and error terms in the GLS-based routines.

The r^2 due to explanatory variables for all these GLS-based spatial models is obtained using Nagelkerke's (1991) general formulation for coefficients of determination, given as

$$r^2 = 1 - \mathrm{e}^{-2/n(l_{\mathrm{A}} - l_0)}$$

where n is number of spatial units, l_{A} is the log-likelihood of the model, and l_0 is log-likelihood of the null model fitted with only the intercept (Lichstein *et al.* 2002). The 'full r^2' of the model, which incorporates the joint effects of the spatial component and the predictors, is given as the complement of the squared linear correlation coefficient between response and the error term, $r^2 = 1 - \text{Pearson}\,[\mathbf{Y}; \mathbf{e}]^2$.

A full comparison of these various spatial regression models in macroecology and biogeography is still lacking, primarily with respect to understanding how choosing each of them will affect the relative importance of predictors driving biodiversity at different spatial scales (see Diniz-Filho *et al.*, 2003; Tognelli & Kelt, 2004; Ferrer-Castan & Vetaas, 2005). This is not just a statistical issue, since the choice among different predictors may provide support for alternative, and sometimes competing, biogeographical hypothesis (Hawkins *et al.*, 2003; Currie *et al.*, 2005). Hopefully, this new software will help researchers to use more spatial models and to perform more accurate hypothesis testing and data exploration and, thus, allow a deeper understanding of complex ecological and biogeographical patterns and processes.

Computational features and software availability

SAM is able to open input files in any of three data formats: Excel, dBASE and ASCII (tab-delimited). Analytical results from all analyses appear in text windows in SAM modules and may be exported by Windows's copy-paste tool. Any new variables created by SAM during analysis (e.g. regression residuals, principal components, spatial filters, etc.) may be saved in the original data file or exported from SAM to another file in any permissible format.

In principle, SAM has no computational limits, but practical limits are set by the computational power of the machine that runs it. The constraint on working with very large datasets (above 3,000 cases) is probably not on opening the matrix, but on the time and memory needed to perform some of the analyses, which require varying amounts of memory and CPU time. The time needed to compute the more computationally demanding statistics in SAM depends on the computer, the dataset size and the operation of interest. Our experience shows that, running most currently available PCs (about 2.5 GHz, 256 Mb of RAM), SAM can easily handle matrices of up to

1,000 cases for all routines (albeit with variable computer times). The Dell Precision 450 Workstation (3.5 GHz, 4 GB RAM) on which SAM was developed was able to run easily about 5,000 cases for all routines.

SAM is freeware, and the present version is less than 3.0 MB. Researchers interested in using SAM may download it from the official web site: www.ecoevol.ufg.br/sam (permanent URL at http://purl.oclc.org/sam/). Three complete, real sample datasets are distributed together with the SAM application file (birds of South America, birds and mammals of the Brazilian cerrado, birds and mammals of the Western Hemisphere), and these datasets may be helpful for new users as trial examples. Prior to downloading SAM, users are required to provide name, institution, country and e-mail address, so that a user list can be compiled to determine how many researches are interested and using the package, to justify the grant used to build the software, and to allow us to contact users about future versions.

Acknowledgments

We thank (alphabetically) Miguel Araújo, Daniel Borcard, Rob Colwell, Nick Gotelli, Brad Hawkins, Holger Kreft, Matt Talluto, and about 20 initial users in two SAM courses (Universidade Federal de Goiás - Brazil, and Universidad de Alcalá - Spain), and other 70 users since first SAM release (June 2005), for suggestions on the software. Rob Colwell also kindly revised and suggested improvements on the text. We are especially grateful to Daniel Borcard and Ole Vetaas for the insightful comments and suggestions on initial version of the manuscript. SAM was developed on a DELL Precision 450 Series Workstation provided by CNPq/SECTEC-GO (PRONEX project no. 23234156). The authors received graduate and productivity fellowships from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

References

Anselin, L. (1988) *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Boston.

Anselin, L. (1995) Local indicators of spatial association – LISA. *Geographical Analysis*, **27**, 93-115.

Anselin, L. (2002) Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models. *Agricultural Economics*, **27**, 247-267.

Badgley, C. & Fox, D.L. (2000) Ecological biogeography of North American mammals: species density and ecological structure in relation to environmental gradients. *Journal of Biogeography*, **27**, 1437-1467.

Banerjee, S., Carlin, B.P. & Gelfand, A.E. (2004) *Hierarchical modelin and analysis for spatial data*. Chapman & Hall / CRC, Boca Raton.

Bini, L.M., Diniz-Filho, J.A.F. & Hawkins, B.A. (2004) Macroecological explanations for differences in species richness gradients: a canonical analysis of South American birds. *Journal of Biogeography*, **31**, 1819-1827.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51-68.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826-1832.

Burnham, K.P & Anderson, D.R. (2002) Model selection and multimodel inference. A practical information – theoretical approach. Springer, New York.

Cardillo, M. (2002) Body size and latitudinal gradients in regional diversity of New World birds. *Global Ecology and Biogeography*, **11**, 59-65.

Clifford, P., Richardson, S. & Hemon, D. (1989) Assessing the significance of the correlation between 2 spatial processes. *Biometrics*, **45**, 123-134.

Cocu, N., Harrington, R., Hulle, M. & Rounsevell, M.D.A. (2005) Spatial autocorrelation as a tool for identifying the geographical patterns of aphid annual abundance. *Agricultural and Forest Entomology*, **7**, 31-43

Cresci, N.A.C. (1992) *Statistics for Spatial Data*. Wiley-Interscience Publications, New York.

Currie, D.J., Mittelbach, G.C., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E.M. & Turner, J.R. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121-1134.

Dark, S.J. (2004) The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diversity and Distributions*, **10**, 1-9.

Davis, J. C. 1986. *Statistics and Data Analysis in Geology*. Second Edition, Wiley, New York.

Diniz-Filho, J.A.F. & Bini, L.M. (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14, 177-185.

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

Dutilleul, P. (1993) Modifying the t-test for assessing the correlation between 2 spatial processes. *Biometrics*, **49**, 305-314.

Evans, K.L., Warren, P.H., Gaston, K.J. (2005) Species-energy relationship at macroecological scale: a review of the mechanisms. *Biological Reviews*, **80**, 1-25.

Fagan, W.F., Fortin, M.-J., & Soycan, C. (2003) Integrating Edge Detection and Dynamic Modeling in Quantitative Analysis of Ecological Boundaries. *BioScience*, **53**, 730-738.

Ferrer-Castan, D. & Vetaas, O.R. (2005) Pteridophyte richness, climate and topography in the Iberian Peninsula: comparing spatial and nonspatial models of richness patterns. *Global Ecology and Biogeography*, **14**, 155-165.

Fortin, M.-J. (1994) Edge-detection algorithms for 2-dimensional ecological data. *Ecology*, **75**, 956-965.

Fortin, M.-J. & Drapeau, P. (1995) Delineation of ecological boundaries - comparison of approaches and significance tests. *Oikos*, **72**, 323-332.

Fortin, M-J. & Dale, M.R.T. (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.

Fortin, M.-J., Olson, R.J., Ferson, S., Iverson, L., Hunsaker, C., Edwards, G., Levine, D., Butera, K., & Klemas, V. (2000) Issues related to the detection of boundaries. *Landscape Ecology*, **15**, 453-466.

Ganio, L.M., Torgersen, C.E. & Gresswell, R.E. (2005) A geostatistical approach for describing patterns in stream networks. *Frontiers in Ecology and the Environment*, **3**, 138-144.

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

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

Haining, R. (2002) *Spatial data analysis.* Cambridge University Press, Cambridge.

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

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

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**, 1035-1042.

Hillebrand, H. & Azovsky, A.I. (2001) Body size determines the strength of the latitudinal diversity gradient. *Ecography*, **24**, 251-256.

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

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

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

Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J., Hohn, M. & Myers, D. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, **25**, 601-615.

Legendre, P. & Fortin, M.J. (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107-138.

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

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

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

Mangel, R. & Hilborne, M. (1997) *The ecological detective: confronting models with data.* Princeton University Press, Princeton.

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, London.

Marquet, P. A., Fernández, M., Navarrete, S.A. and Valdovinos, C. (2004) Diversity Emerging: Toward a Deconstruction of Biodiversity Patterns. *Frontiers of Biogeography: New Directions in the Geography of Nature*. (ed. by M.V. Lomolino and L.R. Heaney), pp. 191-210. Sinauer, Sunderland.

Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. *Biometrika,* **78,** 691-692.

Oden, N.L., Sokal, R.R., Fortin, M.J. & Goebl, H. (1993) Categorical wobbling - detecting regions of significant change in spatially located categorical variables. *Geographical Analysis*, **25**, 315-336.

Peres-Neto, P.R. (2004) Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, **140**, 352-360.

Peters, R.H. (1991) *A critique for ecology*. Cambridge University Press, Cambridge.

Philipp, T.E. (1993) Multiple regression: herbivory. *Design and analysis of ecological experiments.* (ed. by Scheiner, S. M. & J. Gurevitch). pp. 183-210, Chapman & Hall, New York.

Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology letters*, **8**, 224-239.

Rossi, R.E., Mulla, D.J., Journel, A.G. & Franz, E.H. (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological monographs*, **62**, 277-314.

Selmi, S. & Boulinier, T. (2001) Ecological biogeography of Southern Ocean islands: The importance of considering spatial issues. *American Naturalist*, **158**, 426-437.

Sokal, R. R. & Oden, N. L. (1978a) Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of Linnean Society,* **10**, 199-228.

Sokal, R.R. & Oden, N.L. (1978b) Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of Linnean Society,* **10**, 229-249.

Sokal, R.R., Oden, N.L. & Thomson, B.A. (1998) Local spatial autocorrelation in biological variables. *Biological Journal of the Linnean Society*, **65**, 41-62.

Tiefelsdorf, M. (2000) Modelling Spatial Processes: the identification and analysis of spatial relationships in regression residuals by means of Morans' I. Cambridge University Press, Cambridge.

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

Wall, M.M. (2004) A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121: 311-324.

Wartenberg, D. (1985) Canonical trend-surface analysis - a method for describing geographic patterns. *Systematic Zoology*, **34**, 259-279.

Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453-470.