

Relatório sobre Modelagem no CRIA

Índice

1	Apresentação	1
2	Objetivo	1
3	Introdução	1
	Modelagem de nicho ecológico.....	1
	Algoritmos Genéticos	3
	3.1.1 Inspiração na Teoria da Evolução das Espécies e Terminologia	4
	3.1.2 Espaço de Busca de Soluções e as Superfícies de Otimização	4
	3.1.3 Operadores Heurísticos.....	5
	3.1.4 Genetic Algorithm for Rule-set Production (GARP).....	6
	Modelo de Nicho Ecológico no GARP	7
	3.1.5 Codificação das Regras no Algoritmo Genético.....	8
	3.1.6 Aplicação dos Operadores Heurísticos Sobre as Regras	8
	3.1.7 Seleção Natural no GARP e Início de uma Nova Iteração	9
4	Atividades Realizadas	10
	Vinda do Prof. A. Townsend Peterson	10
	4.1.1 Atividades Realizadas.....	10
	4.1.2 Pesquisa GARP	11
	4.1.3 Comentários.....	11
	Pesquisa em andamento	11
	Aquisição de novos software e de capacidade de armazenamento de dados	12
	4.1.4 ArcView Concurrent Master Lab Kit Pak – 8.2	12
	4.1.5 Armazenamento de dados.....	12
	Estudo e Análise dos Dados de Biodiversidade	12
	4.1.6 Dados biológicos	12
	4.1.7 Qualidade dos dados biológicos	14
	4.1.8 Dados Ambientais.....	14
	4.1.9 Qualidade dos dados ambientais.....	15
	4.1.10 Conversão de dados.....	15
	4.1.11 Interpolação de dados.....	15
5	Produtos	16
	Aplicações do GARP no Brasil.....	16
	5.1.1 Uso do GARP para prever a distribuição geográfica de espécies.....	16

5.1.2	Uso do GARP para prever algumas conseqüências de alterações climáticas futuras na distribuição geográfica de espécies.	17
5.1.3	Uso do GARP para prever o potencial de invasão de espécies exóticas.	18
	Publicações.....	19
	Palestras apresentadas	20
	Capacitação da Equipe	21
5.1.4	Curso de Geoprocessamento.....	21
5.1.5	Curso de Doutorado em Ciências da Engenharia Ambiental.....	21
5.1.6	Bolsista de pós-doutoramento - Fapesp.....	21
6	Principais Dificuldades Encontradas.....	21
7	Conclusões, Recomendações e Diretrizes Futuras.....	23
8	Bibliografia.....	24
	Anexo 1: Definições de Dados Ambientais.....	27
	Anexo 2: Curso de Geoprocessamento	31
	Projeto de curso: Fabrício Pavarin.....	32
	Projeto de curso: Alexandre Marino.....	32
	Projeto de curso: Rafael Fonseca.....	33

1 Apresentação

O nível de conhecimento sobre a distribuição geográfica de espécies tropicais é precário. A maioria das espécies é representada por poucos pontos de amostragem e o georeferenciamento, principalmente dos dados históricos disponíveis nas coleções biológicas, é muito impreciso.

O programa Biota/Fapesp introduziu a obrigatoriedade do uso do GPS (Global Positioning System) nas pesquisas de campo realizadas no âmbito do programa e tornou obrigatório o compartilhamento do resultado das pesquisas. O SinBiota¹ foi desenvolvido para armazenar e disponibilizar esses dados juntamente com a base cartográfica do Estado de São Paulo (produzida pelo Instituto Florestal). Tem-se ainda que o projeto speciesLink está desenvolvendo o catálogo virtual das coleções científicas do Estado de São Paulo.

A existência dessa nova base de informação a serviço da pesquisa torna factível estudar e desenvolver ferramentas que possam usar esta infra-estrutura para produzir informações para análise e tomadas de decisão.

O CRIA em parceria com pesquisadores de outras instituições vem estudando e desenvolvendo uma metodologia para análise espacial desses dados utilizando modelagem de nicho ecológico.

O relatório procura descrever os avanços e as principais dificuldades encontradas neste processo e também fornecer diretrizes futuras para essa linha de pesquisa dentro do CRIA.

2 Objetivo

O objetivo do trabalho é estudar, desenvolver e disseminar ferramentas de modelagem de distribuição geográfica adequadas para as espécies e para as condições ecológicas/ambientais brasileiras.

3 Introdução

A equipe do CRIA começou a se envolver com modelagem a partir da vinda do Professor Townsend Peterson ao Brasil em 1999 e a ida de um membro da equipe à Universidade de Kansas para trabalhar no desenvolvimento do GARP (Genetic Algorithm for Rule-set Production). A aprovação do projeto speciesLink deu ao CRIA a oportunidade e as condições necessárias para desenvolver esta linha de pesquisa no Brasil.

O GARP foi desenvolvido baseado nos conceitos de nicho ecológico e de algoritmos genéticos. Este algoritmo foi projetado para criar e otimizar um conjunto de regras ou proposições lógicas sendo desenvolvido originalmente por Stockwell e Noble (1992).

Modelagem de nicho ecológico

O processo de modelagem de nicho ecológico consiste em converter dados primários de ocorrência de espécies em mapas de distribuição geográfica indicando a provável presença ou ausência da espécie, neste caso, através da aplicação de algoritmo genético.

Estes modelos trabalham, na maioria dos casos, com o conceito de nicho ecológico fundamental da espécie. Este conceito foi definido por MacArthur (1972), como sendo um conjunto de condições ecológicas com as quais as populações conseguem se manter, que pode ser representado por um espaço ecológico/ambiental multidimensional (Figura 1).

¹ Sistema de Informação Biota/Fapesp: <http://sinbiota.cria.org.br>

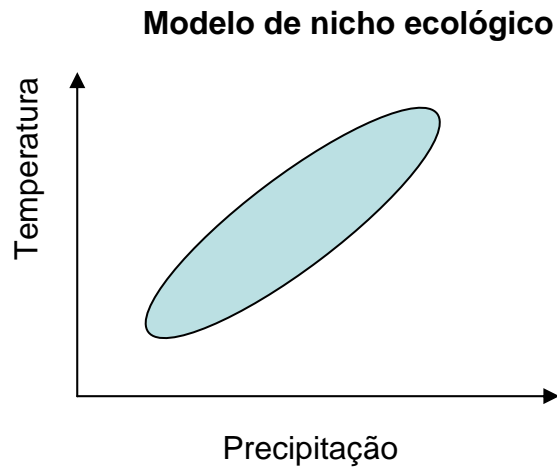


Figura 1. Exemplo de espaço bidimensional definido como nicho ecológico de espécie baseado em dois parâmetros ambientais (temperatura e precipitação).

Tais algoritmos tentam encontrar relações não-aleatórias entre os dados de ocorrência da espécie com os dados ecológico/ambientais relevantes para a espécie (tais como: temperatura, precipitação, topografia, tipo de solo, geologia, entre outros) no ponto onde a espécie foi registrada (Figuras 2 e 3).

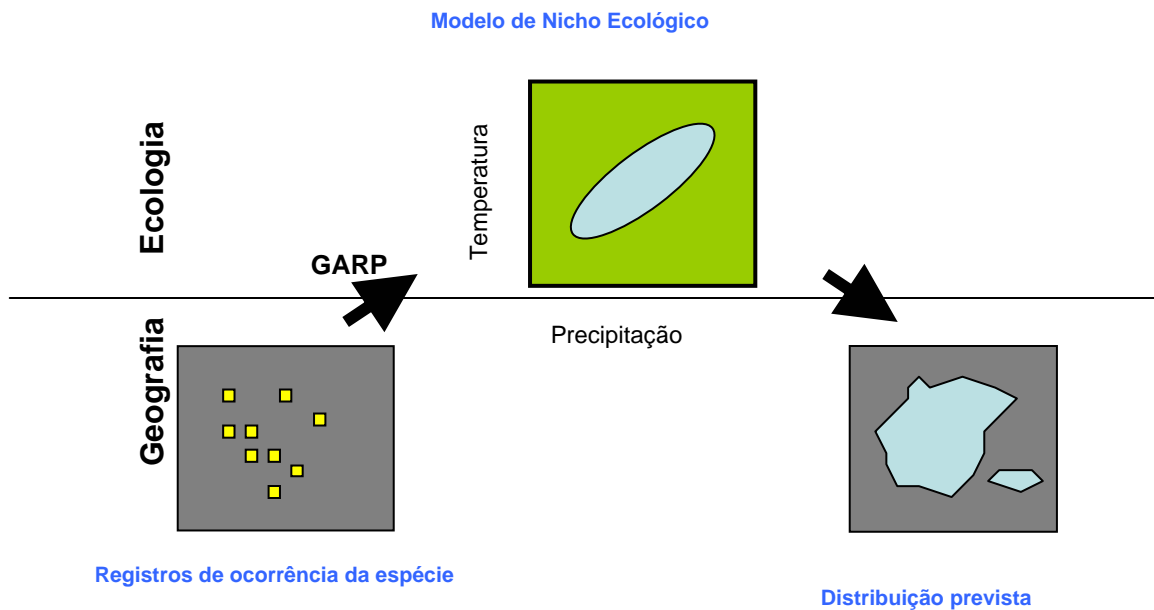


Figura 2. Esquema mostrando a relação entre os dados de campo e a previsão de distribuição geográfica através do processo de modelagem de nicho ecológico.

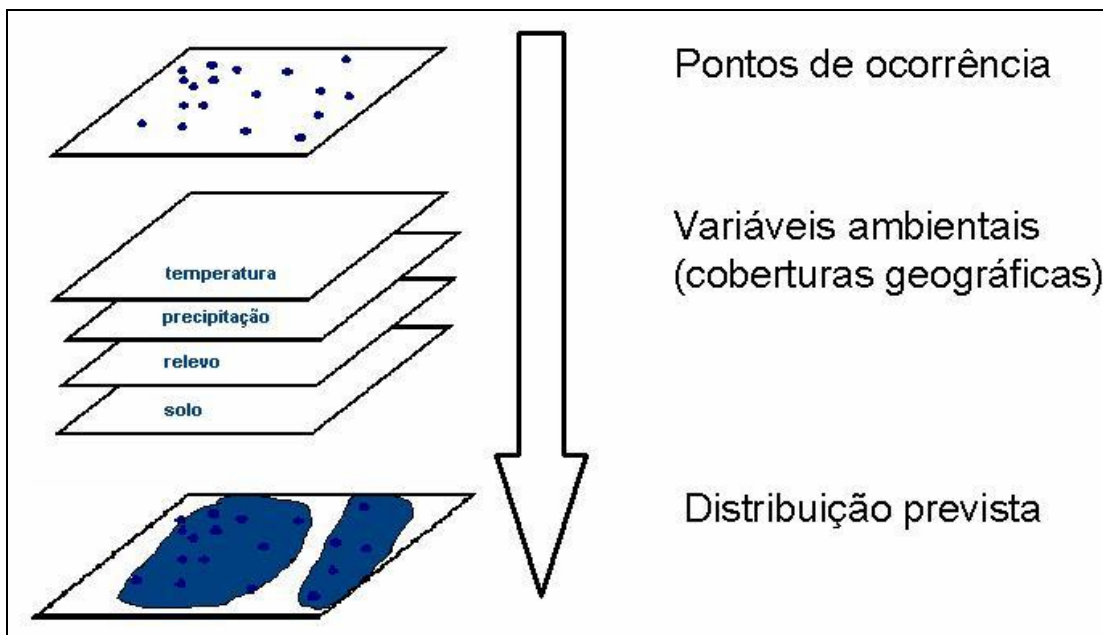


Figura 3. Esquema mostrando a relação entre as variáveis ambientais e os pontos de ocorrência da espécie.

Existem vários métodos na literatura que podem ser aplicados para se obter modelos de nicho ecológico a partir dos pontos de ocorrência de espécies, conforme é mostrado na figura 2. A pesquisa focou algoritmos genéticos com ênfase no GARP (Genetic Algorithm for Rule-Set Production).

Algoritmos Genéticos

Segundo Mitchel (1996), foi entre as décadas de 1950 e 60 que vários grupos de cientistas da computação começaram a estudar os sistemas evolucionários de maneira independente. Estes sistemas se baseavam na idéia de que a teoria da evolução das espécies (Darwin, 1859) poderia ser utilizada como ferramenta de otimização para problemas de engenharia.

Todos esses sistemas seguem a idéia central de que uma população de soluções candidatas para a resolução de um problema evoluía e seus indivíduos eram melhorados através da aplicação de operadores inspirados na variação genética e na seleção natural.

Ainda durante a década de 60, várias aplicações foram desenvolvidas usando este conceito, em problemas de engenharia aeronáutica (Rechenberg, 1965), em problemas de ciência da computação (Fogel *et. al.*, 1966), entre outras.

Foi no início da década de 70 que Holland e seu grupo de pesquisa começaram a estudar métodos para estudar o fenômeno da adaptação das espécies como ocorre na natureza e desenvolver estratégias para a utilização desses conceitos em sistemas computacionais.

Ao contrário dos outros pesquisadores da área que vinham aplicando a teoria da evolução a problemas específicos, Holland e seu grupo tinham como objetivo usar a teoria para desenvolver uma metodologia genérica de resolução de problemas. Como resultado de suas pesquisas, foram introduzidos os algoritmos genéticos (Holland, 1975). Nesta publicação, Holland apresentou o conceito dos algoritmos genéticos como uma abstração da evolução biológica e estabeleceu a base teórica para a adaptação de soluções através da variação genética e da seleção natural.

3.1.1 Inspiração na Teoria da Evolução das Espécies e Terminologia

Grande parte dos conceitos em que os algoritmos genéticos se baseiam estão relacionados com os conceitos biológicos da evolução das espécies e da seleção natural.

Para ilustrar como estes conceitos são empregados na resolução de um problema através de um algoritmo genético, considere o seguinte problema de minimização da função:

$$f(x, y, z) = x - 2y + 5z$$

no domínio dos números inteiros entre 0 e 255.

Uma solução válida para o problema é denominada um indivíduo. Um conjunto de indivíduos é considerado uma população. Neste caso os pontos: $p = (156, 18, 2)$ e $q = (4, 8, 100)$ são dois indivíduos e o conjunto dos dois pontos é uma população. Cada indivíduo tem seu próprio genoma que é representado por um ou mais cromossomos. No exemplo proposto, um cromossomo poderia ser representado pela concatenação das representações binárias das coordenadas dos pontos. Assim, os cromossomos P e Q (representando os pontos p e q) teriam a seguinte forma:

$$P = 100111000001001000000010$$

$$Q = 000001000000100001100100$$

Cada cromossomo é subdividido em genes que se encontram disponíveis em um número de alelos, e cada indivíduo pode ter presente em seu genoma uma combinação arbitrária de alelos. No caso do exemplo proposto, os cromossomos seriam subdivididos em 3 genes, um para cada variável da função, x , y e z . Os cromossomos P e Q teriam os alelos 156 e 4 para o gene x , respectivamente.

3.1.2 Espaço de Busca de Soluções e as Superfícies de Otimização

A idéia de se buscar uma solução ótima para um problema dentre um conjunto de soluções candidatas é um conceito bastante comum na Ciência da Computação. Este conjunto que contém todas as possíveis soluções para um problema é chamado de espaço de busca.

Cada solução pertencente ao espaço de busca está associada a um valor que define a qualidade da solução. Estes valores associados a cada solução, se expressados em um espaço multidimensional formam o que se chama de função ou superfície de otimização.

A Figura 4 mostra um exemplo de um espaço de soluções bidimensionais em que as diferentes soluções são mostradas no eixo x e os respectivos valores da função de otimização $f(x)$ são mostrados no eixo y .

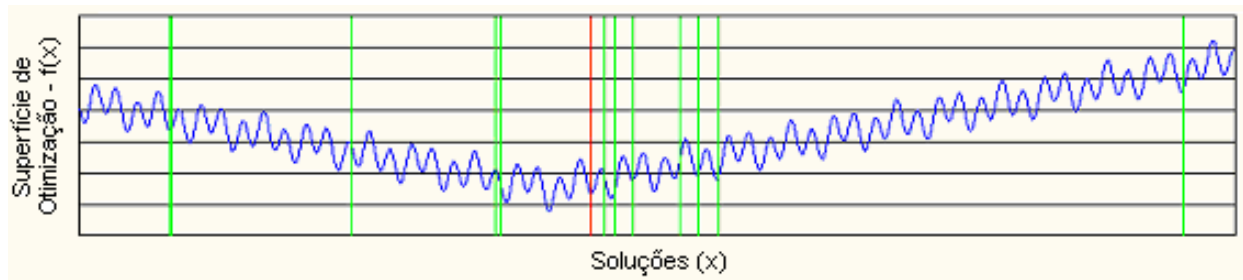


Figura 4. Exemplo de superfície de otimização. Fonte: Obitko, 1998

Na Figura 4, as barras verticais representam exemplos de soluções presentes no espaço de soluções para um problema arbitrário. A curva em azul representa a função de otimização.

No caso do exemplo da função de otimização proposta anteriormente, o espaço de busca seria a combinação de todos os valores possíveis das três variáveis em seus domínios e a

superfície de otimização consistiria na superfície tridimensional definida pelo valor função de otimização em todos os pontos do espaço de busca.

3.1.3 Operadores Heurísticos

Uma vez definida a maneira com que as soluções de um dado problema são codificadas, pode-se então aplicar um conjunto de operadores heurísticos para modificar soluções existentes, obtendo novas soluções para o problema.

Este processo de modificação de soluções existentes com objetivo de obter novas soluções é o mecanismo básico usado pelos algoritmos genéticos para explorar o espaço de soluções de um problema.

Os operadores heurísticos mais empregados são: a recombinação genética (crossing-over) e a mutação.

3.1.3.1 Recombinação genética

A operação de recombinação consiste na criação de novos cromossomos a partir da permutação de segmentos de dois outros cromossomos pais. Veja um exemplo de permutação entre os cromossomos P e Q definidos anteriormente na posição 12.

Antes da permutação tinha-se P e Q:

$$\begin{array}{lll} P = 100111000001-001000000010 & p = (156, 18, 2); & f(p) = 130 \\ Q = 000001000000-100001100100 & q = (4, 8, 100); & f(q) = 488 \end{array}$$

Após a permutação tem-se P' e Q':

$$\begin{array}{lll} P' = 100111000001-100001100100 & p' = (156, 24, 100); & f(p') = 158 \\ Q' = 000001000000-001000000010 & q' = (4, 2, 2); & f(q') = 10 \end{array}$$

Como resultado têm-se dois novos pontos do espaço de soluções, com valores distintos na superfície de otimização.

3.1.3.2 Mutação

A mutação consiste em realizar uma alteração do valor da seqüência de um cromossomo numa posição aleatória. Veja o exemplo de uma mutação no cromossomo P, na posição 17.

Antes da mutação tinha-se:

$$P = 100111000001<0>01000000010 \quad p = (156, 18, 2); \quad f(p) = 130$$

Após a mutação tem-se P':

$$P = 100111000001<1>01000000010 \quad p' = (156, 26, 2); \quad f(p') = 114$$

Da mesma forma que na recombinação, a mutação proporciona um mecanismo para o algoritmo explorar o espaço de soluções do problema.

3.1.3.3 Algoritmo Genético

De maneira geral, os algoritmos genéticos seguem a seguinte estrutura de execução:

1. **Início:** É gerada uma população inicial de n indivíduos;
2. **Adaptação:** Avalia-se o grau de adaptação de cada indivíduo da população através da função de adaptação $f(x)$. Os indivíduos menos aptos, isto é, aqueles que têm valor de $f(x)$ menor que um limiar v , morrem e são eliminados da população;

3. **Reprodução:** Os indivíduos mais adaptados, isto é, que têm valor de $f(x)$ maior que um limiar v , reproduzem-se e criam um novo conjunto de indivíduos. A reprodução se dá da seguinte forma:
 - a. **Seleção:** Dois indivíduos pais são selecionados a partir da população existente, de acordo com os valores de suas funções de adaptação $f(x)$;
 - b. **Recombinação:** Os cromossomos dos dois indivíduos são alinhados e recombinados para formar novos indivíduos;
 - c. **Mutação:** Cada novo indivíduo sofre mutação em posições aleatórias dos seus cromossomos segundo uma probabilidade de mutação pré-definida;
4. **Criação da Nova População:** Os indivíduos gerados na fase de reprodução são incorporados à população;
5. **Teste da Condição de Parada e Início de Nova Iteração:** Verifica-se se o número total de iterações chegou ao fim. Se sim, a execução do algoritmo é interrompida e o indivíduo mais apto da população atual é a solução do problema. Caso contrário, o algoritmo volta ao passo 2.

Os algoritmos genéticos processam as populações de cromossomos através das etapas descritas acima, substituindo uma população pela próxima. Os cromossomos são transferidos de uma geração para a próxima de acordo com o valor da função de adaptação, que consiste numa medida de quão boa é a solução representada pelo cromossomo (uma metáfora para o fenótipo).

Para mais detalhes sobre os algoritmos genéticos sugerem-se a leitura de Davis (1996), Mitchell (1996) e Goldberg (1989).

3.1.4 Genetic Algorithm for Rule-set Production (GARP)

O GARP é um algoritmo genético projetado para criar e otimizar um conjunto de regras ou proposições lógicas desenvolvido originalmente por Stockwell e Noble (1992). Este conjunto de regras é a base de inferência usada para modelar o nicho ecológico das espécies.

3.1.4.1 Tipos de Regras

As regras usadas no algoritmo apresentam o seguinte formato:

*Se **A** é verdadeiro então **B** é verdadeiro,*

onde **A** é a pré-condição e **B** é o resultado ou previsão da regra. Neste algoritmo, a pré-condição é uma expressão lógica que define uma região do espaço ecológico (através das variáveis ambientais) e o resultado é a previsão de presença ou ausência da espécie.

No GARP as regras podem ser de quatro tipos diferentes: regras atômicas, de faixa, de faixa negada e logit.

As **regras atômicas** são as regras mais simples usadas pelo algoritmo. As regras atômicas usam apenas um único valor de cada uma das variáveis ambientais na sua pré-condição. Um exemplo de regra atômica seria:

*Se **Temp = 23C e Elevação = 2000m** então **táxon está presente***

Ao serem criadas no modelo, cada variável ambiental recebe um valor aleatório.

As **regras de faixa** (ou range rules) são uma generalização das regras atômicas, em que as variáveis ambientais podem assumir valores dentro de uma faixa. Por exemplo:

*Se **Temp está entre 23 e 28C e Elevação está entre 2000 e 2700m**
então **táxon está presente***

Ao serem incluídas no modelo, estas regras recebem valores das faixas associados a cada variável ambiental através do algoritmo de previsão de distribuição de espécies chamado BIOCLIM (Nix, 1986).

As **regras de faixa negadas** consistem das regras de faixa em que a precondição da regra é negada. Por exemplo, uma regra de faixa negada teria a seguinte forma:

***Se Temp não está entre 23 e 28C e Elevação não está entre 2000 e 2700m
então táxon está presente***

As **regras logit** são uma adaptação dos modelos de regressão logística para o formato de regras lógicas. A regressão logística é uma forma de regressão em que o seu resultado é transformado em uma probabilidade. A regressão logística é definida pelas expressões abaixo.

$$p = 1 / [1 + e^{-y}]$$

$$y = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n$$

onde p é a probabilidade que determina se a regra deve ser aplicada (equivalente à expressão lógica dos outros tipos de regras), y é a variável dependente (presença ou ausência da espécie), x_i é a i -ésima variável independente (variável ambiental), c_i é o coeficiente que multiplica a i -ésima variável independente e c_0 é o termo constante ou de interceptação.

A função de probabilidade p tem a forma de uma curva em S como a apresentada na figura 5.

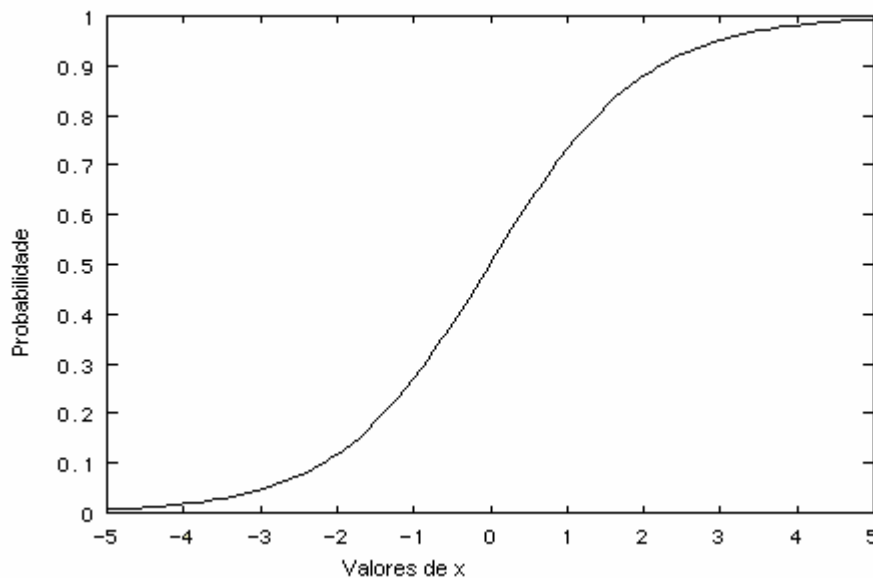


Figura 5. Exemplo de função de probabilidade na regressão logística.

Modelo de Nicho Ecológico no GARP

O algoritmo GARP define o modelo de nicho ecológico das espécies através de um conjunto de regras como as apresentadas anteriormente.

As regras presentes num modelo de nicho ecológico são aplicadas hierarquicamente, de forma que as primeiras regras do modelo se sobrepõem às últimas. Um exemplo possível de modelo do GARP seria:

Regra 1: Se **Temp está entre 5 e 10C e Elevação está entre 1000 e 2000m**
então **táxon está presente** (regra de faixa)

Regra 2: Se **Temp é de 18C e Elevação é 1500m**
então **táxon está ausente** (regra atômica)

Regra 3: Se $p = 1 / [1 + e^{-y}] > 0.75$, onde
 $y = 0.2 + \text{Temp} \times 0.45 + \text{Elevação} \times -0.12$
então **táxon está presente** (regra tipo logit)

3.1.5 Codificação das Regras no Algoritmo Genético

No GARP, cada regra é considerada como um indivíduo e o conjunto de regras é considerado uma população segundo a terminologia definida para os algoritmos genéticos.

Internamente as regras são codificadas através das faixas de valores ou coeficientes relativos às variáveis ambientais e também ao valor da previsão da regra. As mesmas regras apresentadas acima seriam codificadas como cromossomos da seguinte forma:

Regra	Tipo da Regra	Temp: Primeiro Fator	Temp: Segundo Fator	Elevação: Primeiro Fator	Elevação: Segundo Fator	Presença ou Ausência	Função de Adaptação
1	Faixa	5	10	1000	2000	P	50%
2	Atômica	18	25*	1500	3000*	A	12%
3	Logit	0.45	0.17*	-0.12	0.02*	P	95%

Os coeficientes das variáveis ambientais (representados nas colunas 3 a 6) correspondem aos genes que compõem os cromossomos. No caso das regras de faixa (e negadas) existem dois valores que definem a faixa para cada variável ambiental, e por isso cada variável é representada por dois genes. A previsão das regras (coluna "Presença e Ausência") também é codificada como um gene, podendo sofrer alterações durante a execução do algoritmo.

Note que as regras do tipo logit e atômica não utilizam o segundo fator das variáveis. Estas posições (marcadas com um asterisco - *) são ocupadas por valores irrelevantes para estas regras, mas que são importantes nas operações de mutação e recombinação.

A qualidade de cada regra presente no modelo é avaliada através de uma função de adaptação (última coluna da tabela anterior), que é calculado através da significância estatística obtida pela aplicação da regra ao conjunto de pontos de treinamento fornecidos ao algoritmo.

3.1.6 Aplicação dos Operadores Heurísticos Sobre as Regras

Durante a execução do algoritmo as regras são modificadas aleatoriamente através de operadores heurísticos de recombinação e mutação.

Por exemplo, uma operação de mutação sobre a regra 1, na posição do segundo gene (segundo fator da variável Temp) seria realizada da seguinte forma:

Situação	Regra	Temp: Primeiro Fator	Temp: Segundo Fator	Elevação: Primeiro Fator	Elevação: Segundo Fator	Presença ou Ausência	Função de Adaptação
Antes:	1	5	10	1000	2000	P	50%
Depois:	4	5	17	1000	2000	P	67%

Note que a mutação criou uma nova regra, que quando aplicada aos pontos de treinamento, obtém um valor maior na função de adaptação, devido à mudança realizada em um de seus genes.

Já uma operação de recombinação sobre a posição do segundo gene resultaria em:

Situação	Regra	Temp: Primeiro Fator	Temp: Segundo Fator	Elevação: Primeiro Fator	Elevação: Segundo Fator	Presença ou Ausência	Função de Adaptação
Antes:	1	5	10	1000	2000	P	50%
	2	18	25	1500	3000	A	12%
Depois:	5	5	25	1500	3000	A	5%
	6	18	10	1000	2000	P	97%

Note que os novos cromossomos gerados a partir da recombinação e da mutação são diferentes dos cromossomos originais e por isso apresentam valores de adaptação diferentes. Estes valores de adaptação resultantes podem ser tanto melhores quanto piores que os valores dos cromossomos originais.

3.1.7 Seleção Natural no GARP e Início de uma Nova Iteração

Após a criação de novos cromossomos e inclusão destes na população existente, é executada uma operação de seleção natural. Nesta operação aqueles cromossomos que têm valor da função de adaptação abaixo de um certo limiar pré-definido são eliminados da população.

O GARP processa os conjuntos de regras, gerando novos cromossomos, aplicando os operadores heurísticos a eles e em seguida executando a seleção natural, excluindo-se as regras menos eficientes sucessivamente.

Quando um número predeterminado de iterações é atingido, o algoritmo é encerrado. É criado como resultado um conjunto de regras a partir dos indivíduos sobreviventes. Este conjunto de regras representa o modelo de nicho ecológico da espécie.

Este modelo é aplicado de volta ao espaço geográfico, indicando as regiões onde a espécie está provavelmente presente ou ausente.

Diferentemente da concepção original dos algoritmos genéticos em que a solução para o problema é representada apenas pelo indivíduo mais apto da população, o GARP considera toda a população (conjunto de regras) como solução para o problema de modelagem.

Mais detalhes do processo utilizado pelo GARP estão disponíveis na literatura (Stockwell 1999, Stockwell & Peters 1999).

Testes anteriores mostraram a habilidade do modelo em prever a distribuição de espécies em várias regiões (Peterson & Cohoon 1999, Peterson *et al.* 1999, Peterson 2001, Anderson *et al.* 2002a, b, Fera & Peterson 2002, Peterson *et al.* 2002a, Peterson *et al.* 2002c, Peterson *et al.* 2002d, Stockwell & Peterson 2002a, b, Anderson *et al.* 2003, Peterson *et al.* Submitted, Peterson *et al.* 2003 e Peterson & Viegas, 2001).

Atualmente o CRIA está estudando novas abordagens para uso do GARP, envolvendo testes e validação dos modelos resultantes e também uso em diferentes escalas de trabalho, e também vem estudando a utilização de outras abordagens de modelagem. Este estudo será abordado em outro documento que norteará a continuidade dessa atividade no CRIA.

A seguir é apresentado um detalhamento das atividades realizadas, dos produtos gerados e das principais dificuldades encontradas nesse período.

4 Atividades Realizadas

Vinda do Prof. A. Townsend Peterson

O Dr. Peterson veio ao Brasil participar do desenvolvimento do projeto "Sistema de Informação Distribuído para Coleções Biológicas: A Integração do Species Analyst e SinBiota", processo **02/01298-8**. Seu trabalho teve como foco o desenvolvimento de tecnologia e aplicativos de modelagem do nicho ecológico de espécies. A vinda desse especialista foi decisiva para o desenvolvimento dos trabalhos e para o envolvimento da equipe com a temática.

4.1.1 Atividades Realizadas

4.1.1.1 Curso de Treinamento - Métodos e Ferramentas para Modelagem Preditiva de Espécies

O curso, oferecido em parceria com a Embrapa Informática Agropecuária (CNPTIA/EMBRAPA), contou com a participação de 35 pesquisadores². Os seguintes tópicos foram abordados:

- Disponibilidade de dados;
- Teoria de modelagem; e,
- Exemplos práticos e aplicações.

A preparação prévia dos participantes possibilitou o desenvolvimento de discussões de alto nível sobre o tema apresentado. As discussões tiveram como foco a teoria relacionada com as idéias apresentadas, e o potencial para aplicações do método em diferentes situações.

Durante o curso foram feitos diversos contatos que possibilitaram o cumprimento da segunda atividade prevista "Estudos de aplicação da modelagem de espécies, em conjunto com pesquisadores de várias instituições de pesquisa brasileiras".

4.1.1.2 Visitas Técnicas

As seguintes visitas técnicas foram realizadas durante a estadia do Professor Peterson no Brasil:

- Visita à USP – reunião com Jeffrey Shaw para discutir pesquisas colaborativas no campo da Leishmaniose;
- Visita a SUCEN – para discutir pesquisa colaborativa em questões de saúde pública, com a Dra. Vera Camargo Neves e o superintendente da SUCEN, Dr. Luiz Jacintho da Silva;

² Informações sobre o curso estão disponíveis no endereço www.cria.org.br/eventos/mfmpe/

- Museu de Zoologia, USP – apresentação de palestra e discussão a respeito de pesquisa colaborativa com o Professor Naércio Menezes;
- Museu Nacional, RJ – apresentação de palestra e discussão a respeito de pesquisa colaborativa com Paulo Buckup.

Como resultado dessas visitas, tiveram início vários dos projetos listados a seguir.

4.1.1.3 Pesquisa Colaborativa

O Professor Peterson trabalhou em vários projetos durante sua estadia no CRIA. Vários destes projetos resultaram em publicações juntamente com membros da equipe do CRIA. Maiores detalhes podem ser obtidos no item Publicações deste relatório.

4.1.2 Pesquisa GARP

Durante a estadia do Professor Townsend Peterson (U. Kansas), várias melhorias do algoritmo foram discutidas e analisadas. Foi implementado um procedimento de filtragem de modelos que otimiza o tempo de computação do algoritmo nos melhores modelos gerados. Cálculos preliminares sugerem que este processo irá economizar de 50 a 75% do tempo de computação dos modelos, e cerca de 90% do esforço humano envolvido no processo.

4.1.3 Comentários

A visita do Professor Townsend Peterson ao Brasil foi muito importante para o desenvolvimento do projeto de pesquisa em modelagem. O professor Peterson contribuiu para a disseminação das idéias e tecnologias de acesso e integração de dados de biodiversidade, e também de aplicações de metodologias de previsão de nicho ecológico de espécies a casos brasileiros. Iniciou também um processo de pesquisa colaborativa, mostrando a importância de compartilhar dados e experiências.

Como resultado dos contatos com pesquisadores brasileiros, principalmente do Estado de São Paulo, foram criadas condições para iniciar o processo de disseminação das tecnologias envolvidas no projeto e na capacitação da comunidade científica local no uso dessas tecnologias em pesquisa.

Pesquisa em andamento

- Modelagem de espécies de distribuição geográfica restrita baseada em similaridade ambiental. Marinez Siqueira (CRIA), Giselda Durigan (IF-Assis), Ricardo S. Pereira (CRIA) e A. Townsend Peterson (U. Kansas)
- Modelagem de espécies de plantas arbóreas da Bacia do Médio Paranapanema - Marinez F. de Siqueira (CRIA), Giselda Durigan (IF-ASSIS), Wilson Aparecido Contieri (IF-Assis), Ricardo S. Pereira (CRIA) e A. Townsend Peterson (U. Kansas).
- Determinação de áreas para estudos de controle biológico do ácaro vermelho do tomate - Rafael Luís Fonseca (CRIA), Ricardo Scachetti Pereira (CRIA), Imelda Peixoto Furtado (ESALQ), Dr. Gilberto José de Moraes (ESALQ).
- Modelos de nicho potencial de espécies vegetais dispersas pela megafauna: os efeitos da perda de dispersores. Rafael Luís Fonseca (CRIA), Dr. Mauro Galetti (UNESP – Rio Claro).
- Sensibilidade do GARP a dimensionalidade dos dados: um teste usando análise de componentes principais (PCA). Rafael Luís Fonseca (CRIA), Ricardo

Scachetti Pereira (CRIA), Rafael Luís Galdini Raimundo (Unicamp) , Dr. Thomas M. Lewinsohn (Unicamp).

- Distribuição geográfica de duas espécies crípticas de Tomoplagia (Diptera: Tephritidae): condicionantes em macroescala. Aluana Gonçalves de Abreu (Unicamp), Marcio Uehara-Prado(Unicamp) e Rafael Luís Fonseca (CRIA).
- Modelagem de distribuição geográfica com espécies de Apocynaceae – detecção de centros de diversidade no Brasil. Ingrid Koch (CRIA).
- Biogeografia das espécies neotropicais de *Rauvolfia* L. (Apocynaceae). Ingrid Koch (CRIA).
- Redes Neurais Artificiais. As RNAs são amplamente utilizadas para a classificação supervisionada no reconhecimento de padrões. Para testar se um ponto (uma condição) ambiental pertence à classe que define o nicho ecológico de uma dada espécie o ponto será utilizado como entrada da RNA e sua saída indicará a pertinência ou não ao nicho. Mauro Munhoz (CRIA).

Aquisição de novos software e de capacidade de armazenamento de dados

4.1.4 ArcView Concurrent Master Lab Kit Pak – 8.2

Este pacote conta com 25 licenças, dessas, 8 estão em uso sendo 7 em PCs e uma em notebook.

4.1.5 Armazenamento de dados

Hoje a modelagem conta com um espaço em um servidor PowerEdge 6600 com Linux RedHat instalado. Esse servidor contém quatro discos com capacidade de 36Gb cada um. A utilização da área modelagem hoje é de 19 Gb aproximadamente.

Estudo e Análise dos Dados de Biodiversidade

4.1.6 Dados biológicos

Os dados referentes a registros de ocorrência de espécies são compostos, basicamente, pelo nome da espécie, as coordenadas (latitude e longitude) onde a espécie foi encontrada, a precisão destas coordenadas (desejável) e a data da coleta/observação. Associadas a estas informações ainda pode-se ter informações referentes à precisão e grau de confiabilidade dos dados.

Estes dados são geralmente extraídos dos acervos de coleções biológicas (herbários, museus, etc) e de observações realizadas por especialistas (como no programa Biota). Aqui se vê a importância de projetos como o *speciesLink* que engendram toda a complexidade inerente à obtenção destes dados, permitindo que com apenas uma consulta a um portal web sejam obtidos dados resultantes dos esforços distribuídos por várias coleções tanto do Brasil como do mundo.

Hoje o CRIA conta com um conjunto de dados bastante razoável sobre registros de ocorrência de espécies nativas. A maior parte desses dados é composto por dados qualitativos, ou seja, dados sobre a presença da espécie em um determinado ponto no espaço. Com base nesses dados é possível produzir informação sobre como essas espécies estão distribuídas localmente, regionalmente ou ainda continentalmente. Além disso, pretende-se utilizar modelagem para responder outras perguntas como: quais fatores

ambientais são responsáveis pela distribuição geográfica das espécies? É possível utilizar modelagem para buscar espécies no campo? É possível utilizar modelagem para auxiliar a tomada de decisão para conservação de espécies? É possível utilizar modelagem para auxiliar a tomada de decisão quanto ao potencial de invasão de espécies exóticas?

Temos hoje no CRIA (tabela1) os seguintes dados georeferenciados disponíveis para se trabalhar com modelagem.

4.1.6.1 Tabela 1: Dados de biodiversidade disponíveis no CRIA:

Dados	Nº de espécies distintas	Nº de registros únicos de espécies e coordenadas	Fonte	Disponibilidade de uso
SinBiota	5928	24127	Sistema de Informação Ambiental - BIOTA/FAPESP	Dados disponíveis na Internet.
Bacia do Médio Paranapanema	478	4023	Estação Ecológica de Assis – IF/SP	Dados disponíveis mediante a realização de parceria com a EEAssis/IF-SP
Cerrado de São Paulo	552	8745	Projeto Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo – Programa Biota Fapesp	Dados disponíveis no SinBiota ou através de parceria com EEAssis e o CRIA.
Coleções Biológicas – Projeto SpeciesLink – CRIA				
AcariDZSJRP	232	379	Coleção de Ácaros do Departamento de Zoologia e Botânica - IBILCE/UNESP.	Dados disponíveis na Internet.
AcariESALQ	1036	206	Coleção de Ácaros do Departamento de Entomologia, Fitopatologia e Zoologia - LEF/ESALQ.	Dados disponíveis na Internet.
DZSJRP	470	1690	Coleção de Peixes do Departamento de Zoologia e Botânica - IBILCE/UNESP	Dados disponíveis na Internet.
LIRP	653	2146	Coleção de Peixes do Laboratório de Ictiologia de Ribeirão Preto - FFCLRP/USP	Dados disponíveis na Internet.
SPF	1769	524	Herbário do Departamento de Botânica - IB/USP	Dados disponíveis na Internet.
UEC	1428	379	Herbário da Universidade Estadual de Campinas - IB/UNICAMP	Dados disponíveis na Internet.
Total	---	42219	Centro de Referência em Informação Ambiental	Parcialmente disponível na Internet.

Os dados acima relacionados disponíveis na Internet foram coletados a partir do sistema *speciesLink*³ na data de 10/06/2003. Para os dados de coleções biológicas e para o SinBiota, o termo “registro de espécie” significa registros cujo valor do campo espécie é “não nulo”, ou seja, estão sendo também considerados como registros identificados até gênero (registros com "sp" e demais variantes).

³ <http://splink.cria.org.br>

4.1.7 Qualidade dos dados biológicos

Os maiores problemas com os dados de biodiversidade para a modelagem referem-se à atualização da nomenclatura e à qualidade do seu georeferenciamento. Como vimos na tabela anterior, grande parte dos dados disponíveis hoje são dados de coleções biológicas. Esses dados compõem um vasto conjunto de informação, provendo não só informação atual sobre essas espécies, mas também informação histórica sobre vários séculos passados (Chapman & Busby 1994). Porém, muitos destes dados têm pouca informação geográfica associada, muitas vezes não passando de uma descrição genérica da localidade onde foram coletados (Chapman & Milne 1998), ou seja, o georeferenciamento nesses casos não é muito acurado, especialmente em se tratando de dados históricos (Chapman 1999). Para tentar minimizar este tipo de problema, existem atualmente alguns recursos, como é o caso do uso de Gazetteers (lista de localidades associada a informações geográficas, como latitude e longitude). Atualmente, existem vários gazetteers disponíveis na WEB. Quando provindos de Instituições fidedignas, estas listas com coordenadas geográficas são extremamente úteis para georeferenciar espécimens coletados em uma época em que aparelhos de GPS não eram disponíveis. É importante, portanto, desenvolver sistemas de georeferenciamento automático e de checagem da nomenclatura para auxiliar as coleções em seu processo de validação a atualização dos dados de seus acervos. Neste contexto, foi desenvolvido um projeto denominado “BioGeoMancer: Georeferenciamento e validação automática para coleções biológicas”, que está em processo de avaliação na Fapesp (processo número 02/08379-4).

4.1.8 Dados Ambientais

No contexto da modelagem de distribuição de espécies, os dados ambientais dizem respeito às informações acerca do clima (temperatura, precipitação, vento, incidência de geadas etc) e do meio físico (topografia, aspecto, solo, geologia, uso da terra, cobertura vegetal etc) da região de estudo. A obtenção destes dados é feita através de cartografia, imagens de satélite, de fotos aéreas ou das leituras de estações de dados. Os dados brutos são refinados através de técnicas de interpolação (no caso de estações) ou classificação/agrupamento (no caso de imagens). Como resultado desse refinamento são gerados mapas ambientais que são armazenados em forma de *grid* e que servirão de entrada para os algoritmos responsáveis por gerar os modelos. Temos hoje (tabela 2) os seguintes dados ambientais disponíveis.

4.1.8.1 Tabela 2: Dados ambientais disponíveis no CRIA

Dados	Metadados	Fonte
Bacia do Médio Paranapanema – solos, geologia, uso do solo, declividade, temperatura (mínimas, médias e máximas mensais), precipitação (médias mensais), geada, risco e adequação de uso de solo.	Horizontal coordinate system Projected coordinate system name: WGS_1984_UTM_Zone_22S Geographic coordinate system name: GCS_Corrego_Alegre	O mapa básico que gerou o mapa de solos foi extraído de trabalhos em escalas 1:100.000, do IAC, Quadriculas de Assis, Maracáí, Marília e mapas municipais de 1:50.000 de vários municípios. Além disto, também foi usada a base do Radam, não publicada, e algumas amarrações de campo.
SinBiota (cobertura vegetal, UCs, municípios etc)	Horizontal coordinate system Projected coordinate system name: World_polyconic Geographic coordinate system name: SAD69	Programa Biota/FAPESP – Instituto Florestal
World Data (clima e topografia)	Geographic coordinate system name: GCS_WGS_1984	Universidade de Kansas

Os demais dados ambientais disponíveis no CRIA (Flora Neotropical; Bacia do Piracicaba; IPCC entre outros) ainda serão analisados do ponto de vista de metadados e fonte.

4.1.9 Qualidade dos dados ambientais

A utilização de dados ambientais seria simples se houvesse um padrão de armazenamento e de projeção utilizadas. Como formatos de armazenamento pode-se citar: Arc/Info, ERDAS, GEO-TIFF, DEM, Raster, etc. O problema da padronização da projeção é um pouco mais complexo, pois dependendo da extensão, do formato e da localização da área que se deseja representar, diferentes projeções são mais indicadas. O intuito dessa parte do relatório é mostrar o cuidado que se deve ter quando se trabalha com coordenadas geográficas (latitude e longitude) de pontos. Para isso serão abordados alguns conceitos de Cartografia (Anexo 1) que serão simplificados para que seja facilitada a compreensão em seu sentido mais amplo. Parte da equipe do CRIA está fazendo um curso em geoprocessamento para poder auxiliar a equipe de informação nessa área.

São necessários alguns cuidados para se trabalhar com a transformação de sistemas de coordenadas utilizadas pelos usuários. O uso de diferentes procedimentos de transformação gera diferenças nos valores das coordenadas transformadas. As conseqüências dessas diferenças são percebidas em problemas de fusão de dados de múltiplas fontes. Duas bases de dados podem estar sob o mesmo sistema de referência, mas terem sido transformadas de outros sistemas utilizando diferentes procedimentos. Isso pode fazer com que a mesma representação de uma feição seja transformada em duas representações diferentes (i.e., com diferença nas coordenadas dos seus pontos). Para se evitar este tipo de problema, cada base de dados deve ter a devida documentação técnica dos dados e dos processamentos aplicados a eles.

4.1.10 Conversão de dados

Os dados ambientais comumente são armazenados em arquivos do tipo *grid*. Porém cada aplicativo proprietário tende a utilizar um formato próprio para os arquivos *grid* que geram. Este foi o caso para os dados climáticos da bacia do médio Paranapanema cujos dados originais estão no formato Idrisi. Para permitir sua utilização e cruzamento com outros dados já utilizados pela equipe fez-se necessária a criação de um filtro conversor do formato Idrisi para o ArcInfo ASCII-Grid. O formato destino foi escolhido devido à familiaridade da equipe com o software ArcView que utiliza este formato e que não consegue ler o formato Idrisi.

O programa responsável pela conversão entre os dois formatos foi desenvolvido pela equipe do CRIA em linguagem C++ através da implementação de uma classe capaz de ler o formato Idrisi para a memória, fornecer uma interface para a manipulação destes dados e ainda gravá-los em um arquivo no formato ArcInfo ASCII-Grid. Desta forma pode-se realizar a conversão entre os formatos e ainda permitir a leitura e utilização do formato Idrisi por programas futuros.

4.1.11 Interpolação de dados

Devido à dificuldade que o CRIA vem encontrando para localizar e adquirir bases ambientais para o Brasil e mesmo para o estado de São Paulo foi desenvolvido um estudo para gerar mapas baseados em dados brutos, que no caso de dados climáticos, são mais fáceis de conseguir.

Esta seção é dedicada à exposição desse estudo sobre o processo de interpolação dos dados de estações climáticas para a obtenção de mapas ambientais, desde a entrada dos dados brutos gerados pelas estações, passando pela interpolação em si e finalizando com a gravação dos arquivos contendo os mapas ambientais.

Para o estudo foram utilizados os dados do “Arquivos de Série Pluviometria Histórica” disponibilizados pela Rede Nacional de Agrometeorologia⁴. Os arquivos estão no modo texto e contêm os valores diários de precipitação para uma estação pluviométrica, além de seu nome e de sua localização (latitude, longitude e altitude).

O processo de interpolação foi realizado com a ajuda do programa *Grass* (Geographic Resources Analysis Support System)⁵. Este programa foi escolhido por ser um dos GIS *open source* mais utilizado. Este programa permite que se entre com uma lista de estações com suas coordenadas e sua medição, e se obtenha um mapa (grid) que representa a superfície resultante da interpolação das medições.

Tendo os dados brutos das estações e definido o programa de interpolação, foi necessário criar um outro programa capaz de ler os dados brutos e gerar as listas de estações no formato aceito pelo *Grass*.

Foram gerados mapas contendo informações sobre as médias do total de precipitação no período chuvoso (dezembro, janeiro e fevereiro) e no período de seca (junho, julho e agosto) para o estado de São Paulo no período de 1952 a 1995. Estes mapas foram visualmente comparados com os fornecidos pela RNA e pareceram satisfatórios, porém ainda sem testes devidamente conduzidos para que possam ser utilizados como fonte de informação ambiental para modelagem. Porém, muitas dificuldades foram encontradas para a geração de mapas ambientais a partir das medições de estações climáticas, então, aconselha-se que a geração de mapas ambientais seja realizada por agências especializadas no assunto, ou em parceria com elas através da proposta de um projeto exclusivo para a geração de mapas ambientais ou ainda através da compra de bases prontas para o uso, em formato digital.

5 Produtos

Aplicações do GARP no Brasil

O GARP vem sendo utilizado para estabelecer previsões sobre a distribuição geográfica de espécies para condições tropicais. A seguir apresentamos alguns exemplos, a título ilustrativo, do que vem sendo desenvolvido no Brasil.

5.1.1 Uso do GARP para prever a distribuição geográfica de espécies

Ainda é precário o nível de conhecimento sobre a distribuição geográfica de plantas tropicais. A maioria das espécies é representada por poucos pontos de amostragem e muitas vezes, esses pontos são georeferenciados de maneira muito imprecisa.

Como um primeiro teste para se avaliar a eficácia do GARP em prever a distribuição geográfica de espécies tropicais (figura 6), foi selecionado um conjunto de dados, provenientes do *Projeto de Cooperação Técnica Conservação e Manejo da Biodiversidade do Bioma Cerrado – EMBRAPA Cerrados – UnB – Ibama/DFID e RBGE/Reino Unido* para gerar os modelos (Koch *et. al.* Submitted). Foi escolhido este conjunto de dados por se tratar de uma base de dados bastante completa e precisa referente às espécies arbóreas de cerrado (Ratter *et. al.* 2001).

Este conjunto foi dividido em dois subconjuntos de dados independentes, um utilizado para construir os modelos e prever a distribuição geográfica das espécies (dados de treinamento) e outro utilizado para testar a previsão de distribuição (dados de teste).

⁴ <http://masrv54.agricultura.gov.br/rna>

⁵ <http://grass.itc.it/>

Posteriormente, devem-se empregar análises estatísticas para validar se o resultado obtido é significativo estatisticamente.

Com base em análises deste tipo é possível avaliarmos se a modelagem de nicho ecológico é uma ferramenta robusta para auxiliar no entendimento da distribuição geográfica de espécies de plantas tropicais. A figura 6 mostra um exemplo de resultado do uso de GARP para prever a distribuição geográfica de espécies arbóreas de cerrado.

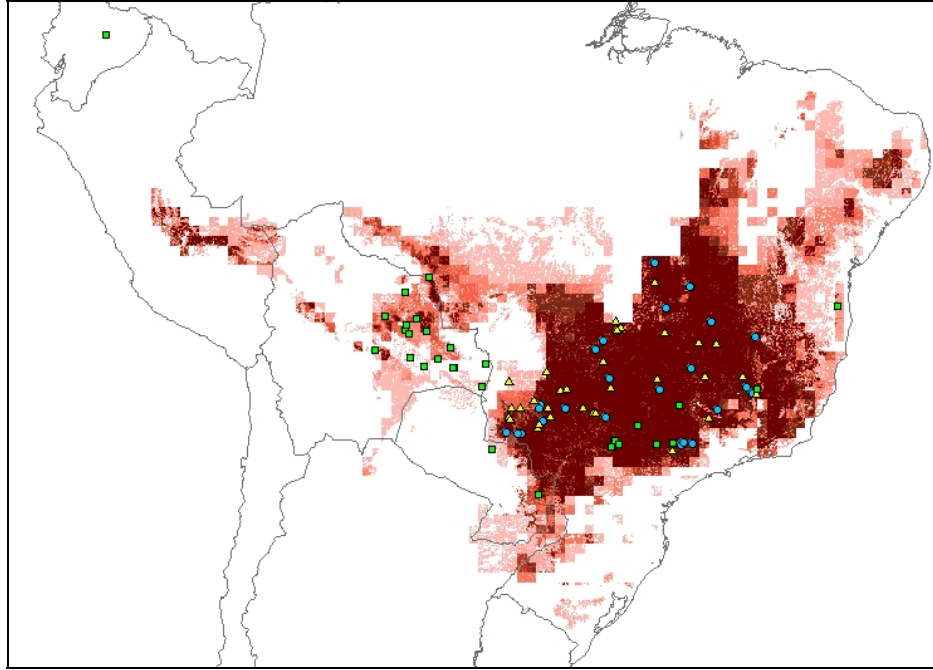


Figura 6. Distribuição prevista para *Terminalia argentea*.

Os círculos azuis representam os pontos usados na confecção do modelo (dados de treinamento), os triângulos amarelos são os pontos usados para testar o modelo (dados de teste) e os quadrados verdes são registros dessa espécie provenientes de outros conjuntos de dados (New York Botanical Garden, Biota/Fapesp e Missouri Botanical Garden).

5.1.2 Uso do GARP para prever algumas conseqüências de alterações climáticas futuras na distribuição geográfica de espécies.

A figura 7 mostra as conseqüências de alterações climáticas para três espécies arbóreas de cerrado baseado em dois cenários futuros de alteração climática (Siqueira & Peterson, Submitted). Em vermelho temos a área prevista de distribuição da espécie no presente, em azul temos a área habitável em 2055 baseado no cenário HHGSDX50 (cenário mais otimista) e em rosa temos a área habitável em 2055 baseado no cenário HHGGAX50 (cenário mais pessimista).

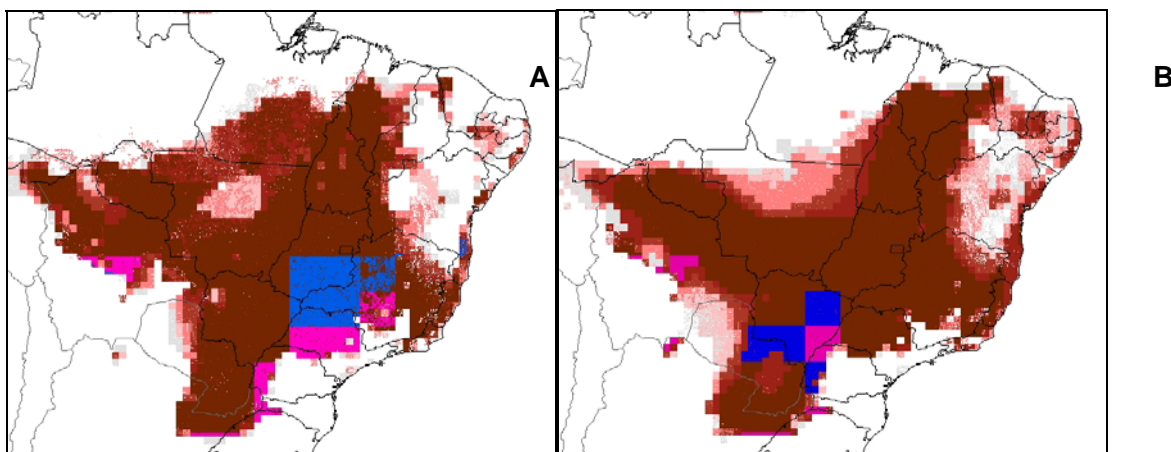


Figura 7. Exemplo de previsão de distribuição atual (1961-1990) em vermelho escuro e futura (2055) em azul para o cenário HHGSDX50 e em rosa para o cenário HHGGAX50, para espécies arbóreas de cerrado: **A** - *Rudgea viburnoides* (Cham.) Benth. (Rubiaceae) e **B** - *Qualea parviflora* Mart. (Vochysiaceae).

Os modelos de circulação usados foram provenientes do HadCM2 (Carson, 1999). Foi utilizado um cenário mais otimista e outro mais pessimista em relação ao modo como o clima poderia se alterar nos próximos 50 anos usando cenários Hadley HHGSDX50 e HHGGAX50 respectivamente. O cenário HHGSDX50 baseia-se em um aumento de 0,5%/ano de CO₂ e o cenário HHGGAX50 baseia-se em um aumento de 1%/ano de CO₂ na atmosfera, prevendo um aquecimento global e redução de precipitação em várias regiões. Como um exemplo sobre o que representam essas alterações climáticas futuras foi feito um cálculo para o México, cuja conclusão foi que essas alterações ocasionariam um aumento de 1,6 a 2,5°C na temperatura média e uma diminuição de 70 a 130 mm na precipitação média (Peterson *et. al.*, 2002b).

Como resultado podemos observar uma redução de área na distribuição geográfica, que varia de intensidade dependendo da espécie, para algumas espécies arbóreas de cerrado.

5.1.3 Uso do GARP para prever o potencial de invasão de espécies exóticas.

A possibilidade de invasão de espécies exóticas em outras regiões é uma questão de relevância conservacionista e econômica. É possível utilizar o GARP para se tentar prever o comportamento de uma espécie potencialmente invasora do seguinte modo. Modela-se o nicho ecológico de uma espécie na sua área de distribuição geográfica nativa e posteriormente projetar essa modelagem, através da projeção do nicho potencial de ocupação dessa espécie, em outras áreas de interesse.

O exemplo a seguir (Peterson *et. al.* 2003) ilustra o caso da possibilidade de invasão da América do Sul por *Homalodisca coagulata* (inseto vetor de transmissão da bactéria *Xylella fastidiosa*) dado o potencial de seu impacto como um vetor para agentes fitopatológicos. Foram desenvolvidos modelos de nicho ecológico para a espécie na sua área de distribuição geográfica nativa, no sudeste dos Estados Unidos.

A capacidade de previsão dos modelos foi testada tanto na área de distribuição nativa quanto em projeções no estado da Califórnia, onde a espécie tem estado presente por muito tempo como uma espécie invasora. Em ambos os casos, os testes indicaram alta significância estatística para as previsões. Projeções dos mesmos modelos para a América do Sul indicaram poucas chances de invasão ao sudeste do Brasil, local em que doenças em culturas de cítricos são preocupantes. Entretanto, todos os modelos são concordantes

ao prever alto risco de estabelecimento da espécie na região onde se encontram as vinícolas do norte da Argentina e no extremo sul do Brasil.

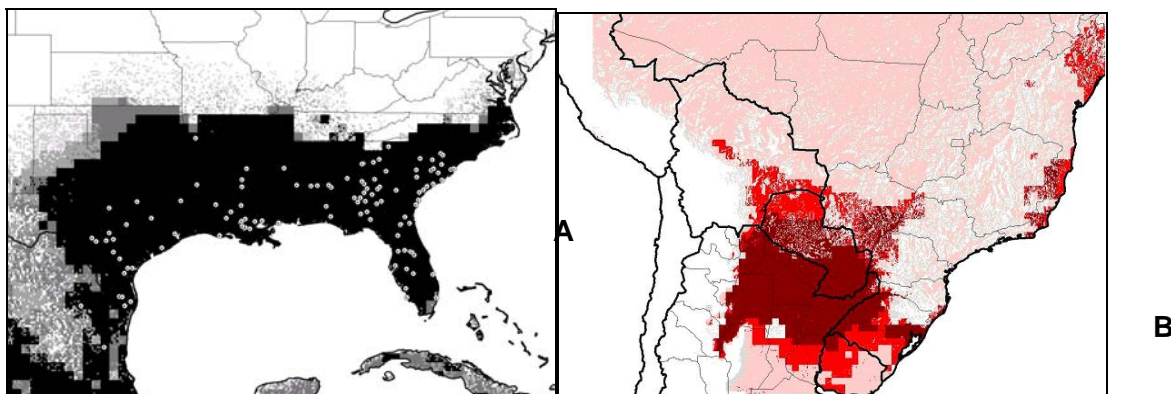


Figura 8. Previsão de invasão de *Homalodisca coagulata* na América do Sul. **A** – previsão de distribuição nativa da espécie. **B** – projeção do nicho ecológico potencial de ocupação da espécie na América do Sul (modificado de Peterson *et. al.* 2003).

Publicações

- Modelagem GARP da distribuição nativa e exótica da planta invasora *Chromolaena odorata* (Asteraceae). Rafael L. G. Raimundo (Unicamp), Rafael L. Fonseca (CRIA), Ricardo S. Pereira (CRIA), A. T. Peterson (U. Kansas), Thomas M. Lewinsohn (Unicamp). Submetido para "Journal of Applied Ecology".
- Predicting the potential of invasion of two *Crotalaria* species (Fabaceae) in Conservation Units in Brazil. Rafael Luís Fonseca (CRIA), Paulo Guimarães Jr. (Unicamp), Sérgio R. Morbiolo (Herbário Unicamp), Ricardo Scachetti Pereira (CRIA), Townsend Peterson (KUNHM-BRC). Manuscrito finalizado.
- Modelagem do potencial invasivo de *Lantana camara* L. (Verbenaceae) em ecorregiões e unidades de conservação tropicais. Sérgio R. Morbiolo (Herbário Unicamp) e Rafael Luís Fonseca (CRIA). Submetido para "Biota Neotropica".
- Characterizing geographic distributions of tropical woody plant species via ecological niche modeling. Ingrid Koch (Unicamp), Marinez Ferreira de Siqueira (CRIA), A. T. Peterson (U. Kansas). Status da publicação: Submetido para "Global Ecology and Biogeography".
- Global Climate Change Consequences for Cerrado Tree Species Distribution. Marinez Ferreira de Siqueira (CRIA) & A. T. Peterson (U. Kansas). Submetido para "Biota Neotropica".
- Avaliação do potencial de invasão de *Homalodisca coagulata* na Califórnia e no Brasil. A. T. Peterson (U. Kansas), Ricardo Scachetti Pereira (CRIA), Daniel A. Kluza (U. Kansas). *Biota Neotropicalica* 3(1): <http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003>
- Detectando problemas de identificação em conjuntos de dados sobre biodiversidade baseado em modelagem de nichos ecológicos. A. T. Peterson (U. Kansas), Ingrid Koch (Unicamp), Ricardo Scachetti Pereira (CRIA), Adolfo G. Navarro-Sigüenza (UNAM, Mexico). Submetido para "Diversity and Distributions".
- Detection of errors in biodiversity data: Collectors' itineraries flag mislabeled specimens. A. T. Peterson (U. Kansas), Adolfo G. Navarro-Sigüenza (UNAM,

México), Ricardo Scachetti Pereira (CRIA, Brasil). No prelo. "Bulletin of the British Ornithologists' Club".

- Distribuição de vetores da Leishmaniose cutânea em São Paulo. Vera Camargo Neves (SUCEN), A. Townsend Peterson (U. Kansas), Ricardo Scachetti Pereira (CRIA). Submetido para "Sociedade Brasileira de Medicina Tropical".
- Distribuição da capivara na bacia do Rio Piracicaba no Estado de São Paulo. Kátia Ferraz (ESALQ, Piracicaba), Ricardo Scachetti Pereira (CRIA), A. T. Peterson (U. Kansas). Aceito para publicação nos anais do IALE2003 (Simpósio sobre landscape ecology).

Palestras apresentadas

- **Workshop – Modelagem de Biodiversidade**

Local: Belém Data: 10-12 de Fevereiro de 2003

GEOMA - Rede Temática de Pesquisa em Modelagem Ambiental da Amazônia - MCT

Participantes do CRIA: Ricardo Scachetti Pereira e Marinez Ferreira de Siqueira

Apresentação da palestra: Algoritmos genéticos - GARP

- **A Informática como Ferramenta para Conservação da Biodiversidade.**

II Semana da Biologia UFSCar - 24 a 29 de março de 2003

Universidade Federal de São Carlos – São Carlos – SP

Apresentação: Rafael Luís Fonseca

- **Modelagem Preditiva de Distribuição de Espécies**

I Curso de Introdução ao Sistema de Informação do Programa Biota/Fapesp (SinBiota)

13 e 14 de fevereiro de 2003 - Instituto de Biologia – Unicamp

Apresentação: Rafael Luís Fonseca

- **Modelagem do potencial invasivo de *Lantana camara* L. (Verbenaceae)**

Apresentação oral no III Segundo Simpósio do Programa Biota/Fapesp

26 a 28 de novembro de 2002 – Universidade Federal de São Carlos – SP

Apresentação: Rafael Luís Fonseca e Sérgio Rodrigues Morbiolo

- **Modelagem Preditiva de Distribuição Geográfica de Espécies**

Participação em Aula de Biogeografia – Prof. Dr. Thomas Michael Lewinsohn

Curso de Geografia

06 de novembro de 2002 – Instituto de Biologia – Unicamp

Apresentação: Rafael Luís Fonseca

- **Modelagem Preditiva de Distribuição Geográfica de Espécies**

Participação em Aula de Ecologia Básica - Prof. Dr. Paulo Sergio oliveira

Curso de Ciências Biológicas

18 de outubro de 2002 – Instituto de Biologia – Unicamp

Apresentação: Rafael Luís Fonseca

Capacitação da Equipe

5.1.4 Curso de Geoprocessamento

Curso de Geoprocessamento 2003 modalidade Extensão Universitária, oferecido pelo Instituto de Computação da Unicamp.

Participantes: Alexandre Marino, Fabrício Pavarin e Rafael Fonseca.

O curso é oferecido de janeiro a dezembro de 2003, nas instalações do Instituto de Computação da Unicamp. O programa do curso é composto por 16 disciplinas totalizando 360 horas. Maiores detalhes são apresentados no anexo 2.

5.1.5 Curso de Doutorado em Ciências da Engenharia Ambiental

Marinez Ferreira de Siqueira, Departamento de Hidráulica e Saneamento. PPG – Ciências da Engenharia Ambiental. Escola de Engenharia de São Carlos – Universidade de São Paulo.

Título: Uso de modelagem de nicho ecológico na avaliação do padrão de distribuição geográfica de espécies arbóreas.

Neste trabalho pretende-se utilizar modelagem para conhecer a distribuição observada e esperada para um conjunto de espécies (aquelas registradas em 30 ou mais localidades diferentes) arbóreas para o Cerrado e para a Floresta Estacional Semidecidual ocorrentes no estado de São Paulo. Para algumas espécies, a serem escolhidas de acordo com alguns critérios pré-estabelecidos, pretende-se fazer a modelagem de nicho ecológico para conhecer as áreas onde essas espécies potencialmente poderiam ocorrer e verificar a possibilidade desse tipo de abordagem auxiliar o trabalho de busca dessas espécies no campo, para uma área de estudo previamente definida. Para alguns casos podem também ser aplicados modelos de alteração climática para observar qual seria a tendência de resposta da vegetação, ou da espécie analisada, mediante alguns cenários futuros de mudanças climáticas.

5.1.6 Bolsista de pós-doutoramento - Fapesp

Dra Ingrid Koch

Parte deste projeto é relacionado à modelagem de nicho ecológico para previsão da distribuição de espécies a partir de coleções de herbário. Com a finalidade de se capacitar para este trabalho, a bolsista realizou um curso para o uso do programa GARP com o Dr. A. Townsend Peterson e interagiu com o pesquisador nos meses de junho e julho de 2002. Tem trabalhado com a equipe do CRIA, realizando experimentos desde então. Como a pesquisadora é taxonomista e tem trabalhado com a família Apocynaceae, os experimentos têm sido realizados com espécies desta família. Atualmente, tem tido contato com outros programas SIG, como o FloraMap e o Diva por intermédio do Dr. Arthur Chapman, pesquisador visitante vinculado ao CRIA.

6 Principais Dificuldades Encontradas

- Problemas com o georeferenciamento dos pontos de coletas. Quando começamos a utilizar bases de dados georeferenciadas, foi muito comum nos depararmos com problemas do tipo: pontos que caem no mar (em caso de espécies terrestres) ou fora da máscara, ou seja, fora da área de estudo. No geral esses erros são devidos a inconsistências nas coordenadas geográficas (erros ou diferenças no datum). Pode também haver erro na máscara utilizada. Por exemplo, no caso das pequenas ilhas

da América Central, a máscara utilizada não incluía determinadas ilhas devido à falta de informações ambientais para essas ilhas.

- Conversão de dados: geralmente cada aplicativo proprietário tende a utilizar um formato próprio para os arquivos grid que geram. Quando isso ocorre, a utilização desses dados por outro aplicativo fica bastante complicada. Para resolver este problema no CRIA foi necessária a criação de um filtro conversor de formatos (Ildrisi para o ArcInfo ASCII-Grid). O programa responsável pela conversão entre os dois formatos foi implementado em linguagem C++.
- Criação de novos conjuntos de dados (*datasets*): quando se utilizou o Garp para modelar espécies distribuídas em ilhas da América Central, surgiram problemas relacionados à definição das informações climáticas (0,5°). O programa considera apenas um ponto em cada "GRID" e, como as ilhas são relativamente pequenas, os pontos de coleta são muito próximos, poucos pontos foram considerados na análise. Além disso, muitas vezes as áreas costeiras não possuem informações ambientais vinculadas a elas, bem como muitas das ilhas menores. A solução encontrada foi buscar dados climáticos com uma melhor definição e criar novos *datasets*. Quando criados os novos *datasets*, foi possível notar que, em algumas situações, havia um deslocamento do GRID em relação ao GRID original, e por isso, os pontos apareciam fora da máscara. Foi necessário corrigir as coordenadas nos arquivos ASCII originais e refazer os *datasets* para fazer nova modelagem. Este problema de deslocamento dos GRIDS está relacionado ao Script que corta os *datasets* e ocorre quando se re-amostra o tamanho das células.
- Uso do SIG: a troca de versão do ArcView vem acarretando vários problemas para as pessoas que trabalham com modelagem no CRIA. A versão 8.2 lançada no mercado vem apresentando problemas *bugs* que desconhecíamos na versão 3.2. Isso têm acarretado atrasos no desenvolvimento dos projetos em andamento. Estes problemas estão sendo avaliados por usuários do software e pelo departamento de suporte do CRIA. A GEMPI - Gestão Empresarial & Informática Ltda (representante do software no Brasil) está sendo informada e consultada para ajudar na solução do problema.
- Aquisição de novos dados ambientais: estamos encontrando bastante dificuldade na localização e aquisição de bases ambientais, em formato digital, em escalas mais adequadas (maior detalhe) para o Brasil e para o estado de São Paulo.
- Interpolação de dados: as seguintes dificuldades foram encontradas para a geração de mapas ambientais a partir das medições de estações climáticas.
 - Há grande dificuldade de se obter as medições históricas das estações. Como exemplo, a própria RNA (Rede Nacional de Agrometeorologia) só fornece séries históricas de precipitação, além de haver vários arquivos corrompidos, como a maioria das estações do Amazonas e todas as do Acre.
 - Os dados das estações precisam ser corrigidos. Isto porque há um elevado número de inconsistências, tais como duas estações na mesma localidade, localidades dadas apenas com graus e não graus, minutos e segundos, o que leva a uma grande incerteza na hora de realizar a interpolação.
 - Deficiência de dados do entorno. Para gerar um bom mapa ambiental de uma certa região é necessário utilizar medições em pontos que estejam próximos à fronteira, mas fora da região. Portanto, se quisermos um mapa ambiental do Brasil, temos que utilizar estações em outros países.
 - A realização de todo o processo dispense bastante esforço e tempo. Isto porque a adequação dos dados das estações para o formato a ser utilizado pelo

interpolador é dispendioso e dificilmente pode ser reaproveitado para outros tipos de estação, pois não parece haver um formato padrão.

- Impossibilidade de adquirir novos pontos que possam melhorar o mapa. Para se obter mapas com qualidade deve-se fazer uma análise criteriosa das distribuições das estações. Por exemplo, regiões onde há maior variação devem ser cobertas por mais estações. No entanto, a decisão de adicionar estações não é de competência do Cria.

7 Conclusões, Recomendações e Diretrizes Futuras

Para dar continuidade à linha de pesquisa em modelagem recomenda-se realizar as seguintes atividades.

1. Continuar fomentando o desenvolvimento de parcerias para que não só tenhamos acesso a mais e melhores dados biológicos, mas também para que tenhamos mais respaldo científico quanto à análise biogeográfica, para diferentes grupos taxonômicos, dos resultados gerados.
2. Estudo de ferramentas para verificar a qualidade e melhorar a confiabilidade de dados biológicos, como por exemplo, o uso de *Gazetteers*.
3. Desenvolvimento do algoritmo GARP em forma de biblioteca implementada na linguagem C++.
4. Desenvolvimento de outros tipos de modelagem: além do uso do GARP é recomendável que o CRIA conheça, teste e avalie outros tipos de modelagem disponíveis hoje para que possamos ampliar e melhorar nossa área de atuação nesse ramo.
5. Desenvolvimento de um ambiente para o desenvolvimento de algoritmos para a modelagem da distribuição de espécies. O ambiente de experimentação proposto permitirá o desenvolvimento, teste e validação de algoritmos utilizados para a previsão da distribuição de espécies e também a geração dos modelos de distribuição. Uma consequência da utilização deste ambiente para o desenvolvimento de algoritmos de modelagem é a modularização do algoritmo através de uma interface padrão. Esta característica permite a fácil transposição do algoritmo para outros sistemas, permitindo a fácil reprodução do processo de modelagem por parte de vários grupos de pesquisa.
6. Em relação à qualidade de dados ambientais, é necessária atenção ao procedimento de transformação de coordenadas pelos usuários. O uso de diferentes procedimentos de transformação gera diferenças nos valores das coordenadas transformadas. As consequências dessas diferenças são percebidas em problemas de fusão de dados de múltiplas fontes, por exemplo.
7. É aconselhada a aquisição de bases ambientais em escalas mais detalhadas, em formato digital, para uso em projetos de modelagem cuja abordagem seja de teste de eficácia do modelo no campo e validação estatística do modelo. Atualmente estamos em processo de compra do Mapa Pedológico do Estado de São Paulo (escala 1:500000) - IAC, formato digital (ArcView), georeferenciado em coordenadas geográficas. Contém os limites (poligonizados) das unidades de mapeamento de solo, com os respectivos identificadores e hidrografia.
8. Ainda sobre bases ambientais, é necessário incluir a devida documentação técnica dos dados e dos processamentos aplicados a eles. Duas bases de dados podem estar sob o mesmo sistema de referência, mas terem sido transformadas de outros sistemas utilizando diferentes procedimentos. Isso pode fazer com que uma mesma

representação de uma feição seja transformada em duas representações diferentes (i.e., com diferença nas coordenadas dos seus pontos).

9. Em relação aos dados ambientais, referente ao tópico interpolação de dados, aconselha-se que a geração de mapas ambientais seja realizada por agências especializadas no assunto, ou em parceria com elas através da proposta de um projeto exclusivo para a geração de mapas ambientais ou ainda através da compra de bases prontas para o uso, em formato digital.
10. Implementar estudo para a possibilidade de desenvolvimento de novas bases ambientais, ou adequação das que já existem, às necessidades do CRIA.
11. Estudo e implementação de técnicas para validação de modelos.

8 Bibliografia

- Anderson, R. P., D. Lew, & A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162:211-232.
- Anderson, R. P., M. Laverde, & A. T. Peterson. 2002a. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. *Global Ecology and Biogeography* 11:131-141.
- Anderson, R. P., M. Laverde, & A. T. Peterson. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 93:3-16.
- Carson, D. J. 1999. Climate modelling: Achievements and prospects. *Quarterly Journal of the Royal Meteorological Society* 125:1-27.
- Chapman, A.D. 1999. Quality Control and Validation of Point-Sourced Environmental Resource Data pp. 409-418 in Lowell, K. and Jatón, A. (eds). *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. Chelsea, MI: Ann Arbor Press.
- Chapman, A.D. and Busby, J.R. 1994. Linking Species Information to Climate Modeling for Continental Biodiversity Inventory and Monitoring pp. 177-195 in Miller, R.I. (ed). *Mapping the Diversity of Nature*. London: Chapman and Hall.
- Chapman, A.D. and Milne, D.J. 1998. *The Impact of Global Warming on the Distribution of Selected Australian Plant and Animal Species in relation to Soils and Vegetation*. Canberra: Environment Australia.
- Darwin, C. R. 1859. *On the Origin of Species by Means of Natural Selection*. John Murray, London.
- Davis, L. 1996. *Handbook of Genetic Algorithms*. International . Thomson Computer Press, Boston, Massachusetts, 1996.
- Feria, T. P., & A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. *Diversity and Distributions* 8:49-56.
- Fogel, L. J., Owens, A. J. & Walsh, M. J. 1966. *Artificial Intelligence through Simulated Evolution*. Wiley.
- Goldberg, D. Editor. 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1989.

- Holland, J. 1975. *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, Michigan, 1975.
- <http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003>
- Koch, I., M. F. Siqueira, & A. T. Peterson. Submitted. Characterizing geographic distributions of tropical woody plant species via ecological niche modeling. *Global Ecology and Biogeography*.
- MacArthur, R. 1972. *Geographical Ecology*. Princeton Univ. Press, Princeton, NJ.
- Mitchell, M. 1996. *An Introduction To Genetic Algorithms*. MIT Press, 1996.
- Nix, H.A., 1986. A biogeographic analysis of Australian Elapid snakes, in Longmore, R. (ed.) *Atlas of Australian Elapid Snakes. Australian Flora and Fauna Series 8*: 4-15.
- Obitko, M. 1998. Genetic Algorithms. <http://cs.felk.cvut.cz/~xobitko/ga/>
- Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor* 103:599-605.
- Peterson, A. T., & K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling* 117:159-164.
- Peterson, A. T., & Vieglais, D.A. 2001. Predicting species invasions using ecological niche modeling. *BioScience* 51:363-371.
- Peterson, A. T., D. R. B. Stockwell, & D. A. Kluza. 2002d. Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617-623 in J. M. Scott, editor. *Predicting species occurrences: Issues of scale and accuracy*. Island Press, Washington, D.C.
- Peterson, A. T., J. Soberon, & V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265-1267.
- Peterson, A. T., L. G. Ball, & K. C. Cohoon. 2002a. Predicting distributions of tropical birds. *Ibis* 144:e27-e32.
- Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R. H. Buddemeier, & D. R. B. Stockwell. 2002b. Future projections for Mexican faunas under global climate change scenarios. *NATURE* 416:626-629.
- Peterson, A. T., M. Papes, & D. A. Kluza. Submitted. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*.
- Peterson, A. T., R. Scachetti-Pereira, & D. A. Kluza. 2003. Assessment of invasive potential of the glassy-winged sharpshooter *Homalodisca coagulata* in California and South America. *Biota Neotrópica* 3(1):
- Peterson, A. T., V. Sanchez-Cordero, C. B. Beard, & J. M. Ramsey. 2002c. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases* 8:662-667.
- Ratter, J. A., Bridgewater, S. & Ribeiro, J. F. 2001. Espécies lenhosas da fitofisionomia cerrado sentido amplo em 170 localidades do bioma cerrado. *Boletim do Herbário Ezechias Paulo Heringer, Brasília* 7: 5-112.

- Rechenberg, I. 1965. Cybernetic Solution Path of An Experimental Problem, Ministry of Aviation, Royal Aircraft Establishment, United Kingdom, 1965.
- Siqueira, M. F. & A. T. Peterson. Submitted. Consequences of Global Climate Change for Geographic Distributions of Cerrado Tree Species. *Biota Neotropica*.
- Stockwell, D. R. B. & Noble, I. R., 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* 33: 385-390.
- Stockwell, D. R. B. 1999. Genetic algorithms II. Pages 123-144 in A. H. Fielding, editor. Machine learning methods for ecological applications. Kluwer Academic Publishers, Boston.
- Stockwell, D. R. B., & A. T. Peterson. 2002a. Controlling bias in biodiversity data. Pages 537-546 in J. M. Scott, editor. Predicting species occurrences: Issues of scale and accuracy. Island Press, Washington, D.C.
- Stockwell, D. R. B., & A. T. Peterson. 2002b. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148:1-13
- Stockwell, D. R. B., & D. P. Peters. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science* 1999;13:143-58.

Anexo 1: Definições de Dados Ambientais

Parte das informações disponíveis neste anexo foi retirada dos seguintes sites da Internet:

www.dpi.inpe.br/~julio/cartogeo/geoprocessamento.pdf

www.geotrack.hpg.com.br/cart.pdf

www2.prudente.unesp.br/dcartog/galo/pdf/2002_propagacao.pdf

www.spg.com.br/informacoes/artigos_gps/125.HTM

Mapas: é a representação do globo terrestre, ou de trechos de sua superfície, sobre um plano, indicando fronteiras políticas, características físicas, localização de cidades e outras informações geográficas, sócio-políticas ou econômicas. Os mapas, normalmente, não tem caráter técnico ou científico especializado, servindo somente para fins ilustrativos ou culturais e exibindo suas informações por meio de cores e símbolos. Ou seja, mapas têm finalidade ilustrativa, como por exemplo, um "Mapa Turístico". Às vezes, nem se quer tem sistema de coordenadas, e a escala é aproximada.

Cartas: trata-se também de uma representação da superfície terrestre sobre um plano, mas são especialmente traçadas para serem usadas em navegação ou outra atividade técnica ou científica. As Cartas permitem medições precisas de distâncias e direções (azimutes), podendo inclusive ser temáticas (Carta topográfica, gravimétrica, geológica, etc).

Maiores informações podem ser obtidas no site da CPRM - Definições de documentos cartográficos - <http://www.cprm.gov.br/didote/p0135.htm>

Coordenadas: o mapa serve apenas para dar uma idéia do terreno, mas também para identificar pontos dentro dele. Para isso, os pontos do mapa podem ser referenciados por suas coordenadas cartesianas. As coordenadas podem ser angulares (graus, minutos e segundos) ou métricas.

Latitude: no sistema de coordenadas angulares, o ângulo "vertical" entre o equador e o paralelo que passa sobre o ponto, é chamado de latitude. Se o ponto está ao Norte do Equador, tem latitude positiva. Se estiver ao Sul do Equador, tem latitude negativa.

Longitude: é a distância angular entre o meridiano de Greenwich e o meridiano que passa sobre o ponto visado. Imaginando-se o planisfério onde a Inglaterra ocupa o centro do mapa, o que estiver à oeste de Greenwich, tem longitude negativa e o que estiver à Leste tem longitude positiva.

Portanto, quase todo o território brasileiro tem coordenadas duplamente negativas. Por exemplo, um ponto em Brasília com coordenadas (-15° 48' 11.5" -48° 03'57.2") pode ser registrado como (15S 48' 11.5" 48W 03' 57.2") ou (15° 48' 11.5" S 48° 03'57.2" W) ou ainda (-15.8030 -48.0530) em graus decimais. É um ponto a 15 graus ao sul do equador, e a 48 graus a oeste de Greenwich.

Escala: é a relação que expressa a diferença de grandeza entre as feições no terreno e como elas aparecem no mapa. A escala pode ser gráfica ou numérica. A escala gráfica tem a aparência de uma régua que mostra o tamanho no terreno de um segmento de reta no mapa. É sempre uma fração que tem: o número "1" como numerador, indicando uma unidade de comprimento no mapa (Ex: cm, mm, polegada). Um número muito maior que 1 como denominador, indicando quantas unidades no terreno equivalem a uma unidade no mapa. Assim, uma escala 1:100.000 indica que: 1cm no mapa equivale a 100.000cm ou 1km no terreno.

Projeções: os mapas são representações em papel (portanto em um plano) de trechos da superfície da terra que é aproximadamente esférica. Para fazer a representação de uma

superfície curva em um plano, os cartógrafos fazem uma projeção, ou seja, fazem a correspondência entre cada ponto da superfície da terra a um ponto sobre o mapa. A projeção cartográfica mais comum é a de Mercator (Transversa de Mercator). Claro que no século XXI não se faz mais projeções diretamente do globo. Os mapas são feitos hoje com levantamentos aerofotogramétricos, imagens de radar, e mais recentemente por satélite. A diferença é que a projeção se faz em uma área pequena, normalmente a área coberta por um mosaico de fotos aéreas já corrigidas e ajustadas, o que reduz muito as distorções.

Projeção UTM - "Universal Transverse Mercator": o mapeamento sistemático, ou seja, detalhado e completo do Brasil é feito usando a projeção UTM nas escalas (1:250.000, 1:100.000, 1:50.000). Cartas detalhadas do Exército / IBGE / Outros.

Para se obter um mapa em Projeção de Mercator, que é um tipo de projeção cilíndrica, a idéia é a seguinte: Envolver o globo terrestre por um cilindro. Este cilindro que tem o mesmo diâmetro da Terra, e que é tangente a um ponto conveniente, não necessariamente no Equador. Depois de projetados os pontos, perpendicularmente à superfície do cilindro, este é desenrolado.

Como característica desta projeção, teremos os paralelos projetados como retas paralelas, com distância entre elas cada vez menor, à medida que distanciamos do Equador. Teremos também um meridiano central que é uma linha reta, e os demais meridianos ligeiramente curvos.

Na projeção UTM, a Terra foi dividida em 60 fusos (meridianos), criando setores (fatias) de 6 graus de extensão. Da mesma forma, foi dividido em 30 outros setores no sentido dos paralelos criando "retângulos esféricos" de 6x6 graus. Cada cilindro é chamado de zona, e esta zona recebe um nome formado por uma letra e um número. (Ex: L23). O cilindro de projeção é tangente ao centro deste "retângulo esférico" e recebe com pouca distorção os pontos nele projetados.

Outra característica da Projeção UTM, é que as coordenadas são métricas. Ou seja, os pontos têm coordenadas (X,Y) cartesianas em metros, em relação a um ponto de origem. Isso facilita calcular a distância, e também a extrair as coordenadas métricas de um ponto no mapa usando uma régua.

Elipsóide: a Terra está aproximada de um elipsóide de revolução, que consiste em um sólido, gerado pela rotação de uma elipse em torno do eixo dos pólos (eixo menor). Nas determinações geodésicas, esse elipsóide é utilizado como superfície de referência.

Estudos geodésicos apresentam valores diferentes para os elementos do elipsóide, medidos nos vários pontos da Terra. Assim, cada região deve adotar como referência o elipsóide mais indicado. Por diferentes razões técnicas, históricas, políticas e econômicas, cada país tem utilizado um determinado elipsóide, que em geral é diferente dos demais, pois antes da utilização de satélites era muito difícil determinar com exatidão o centro da terra para que se pudesse derivar um datum padrão para o mundo. Atualmente, o padrão mais aceito é o que leva o nome WGS84 (World Geodetic System, 1984). Toda carta traz na legenda o elipsóide de referência.

Um sistema geodésico consta de um elipsóide (superfície de referência), de um conjunto de parâmetros matemáticos que definem a amarração do elipsóide à superfície da Terra e de um conjunto de pontos da superfície. O conjunto formado pelo elipsóide e os parâmetros é conhecido como Datum.

No Brasil, inicialmente, adotou-se o elipsóide de Hayford, recomendado na conferência de Madri de 1924, que utiliza o Datum Córrego Alegre. Suas dimensões foram consideradas as mais convenientes para a América do Sul.

Atualmente, o Sistema Geodésico Brasileiro utiliza o elipsóide da União Astronômica Internacional de Geodésia, que passou a se chamar Elipsóide de Referência Internacional, homologado pela UGGI (União Geodésica e Geofísica Internacional) em 1967. Posteriormente este mesmo elipsóide foi utilizado para o Sistema Geodésico Sul-Americano, ficando então conhecido como South American Datum de 1969 - SAD-69.

No site da Universidade do Colorado pode-se ver alguns dos Elipsóides de Referência - http://www.colorado.edu/geography/gcraft/notes/datum/datum_f.html

A razão para existirem tantos elipsóides, é que um elipsóide pode aproximar bem a superfície da terra em um determinado ponto, mas pode ficar meio distante em outros locais.

Datum: utilizado como referência em toda a cartografia mundial para elaboração de mapas e na recuperação das informações nele contidas, o Datum é uma característica comum nas cartas cartográficas. Ele pode ser entendido como um modelo matemático que aproxima a forma da terra e permite cálculos como posição e área a serem levantadas, de forma consistente e precisa. Para a definição do Datum escolhe-se um ponto central em relação à área de sua abrangência. Existem vários tipos de Datum, como:

Local: Datum geodésico utilizado na cobertura de países ou regiões, escolhido de forma a minimizar as distâncias entre o geóide e o elipsóide de referência, numa determinada zona de interesse. Exemplos: SAD-69, Córrego Alegre, NAD-83, Indian.

Global: Datum geodésico utilizado na cobertura geral do globo, escolhido de forma a fazer coincidir o centro de massa da Terra com o centro do elipsóide de referência, e o eixo da Terra com o eixo menor do elipsóide, procurando assim minimizar, globalmente, as diferenças entre este e o geóide. Os sistemas globais de posicionamento utilizam um Datum global. Exemplos: WGS-84, SIRGAS (Sistema de Referência Geocêntrico para as Américas).

Geodésico: Conjunto dos parâmetros que constituem a referência de um determinado sistema de coordenadas geográficas, e que inclui a definição do elipsóide de referência e a sua posição relativamente ao globo terrestre.

Buscando ainda sua definição, o Datum é uma superfície de referência para controle horizontal e vertical de pontos em um sistema de referência cartográfico. Uma superfície de referência (Datum horizontal) consiste em cinco valores: latitude, longitude de um ponto inicial, azimute de uma linha que parte deste ponto e duas constantes necessárias para definir o elipsóide de referência. Assim, forma-se a base para o cálculo dos levantamentos de controle horizontal no qual considera-se a curvatura da Terra. A superfície de nível (Datum vertical) refere-se às altitudes.

Qualquer carta tem uma referência a um Datum e geralmente é apresentado em sua legenda. As linhas de latitude e longitude numa carta ou mapa são referenciados a um específico Datum. Se comparar as coordenadas de GPS a uma carta ou outra referência, o mapa Datum selecionado deve ser o mesmo do da carta.

Sistema Geodésico Brasileiro - SGB:

O Sistema Geodésico Brasileiro foi regulamentado pelas instruções reguladoras das Normas Técnicas da Cartografia Nacional e instituído pelo Decreto número 89.317 de 20 de junho de 1984.

Estas normas estabelecem o Datum SAD-69 como datum oficial a ser utilizado em toda e qualquer representação cartográfica do território nacional. O SAD-69 foi adotado para uniformizar a rede geodésica de controle básico continental. Porém, boa parte da cartografia do Brasil ainda existente está referida ao Datum anterior ao SAD-69. E ainda existem mapas feitos até com Datum locais.

Após uma visão geral sobre todos esse conceitos podemos mostrar alguns exemplos reais e mensurar alguns possíveis erros:

Córrego Alegre - MG

Latitude: 19° 45' 41.34" S

Longitude: 48° 06' 07.08" W

SAD-69

Latitude: 19° 45' 41.6527" S

Longitude: 48° 06' 04.0639" W

Azimute de Uberaba: 271° 30' 04.05"

As coordenadas geográficas, na verdade geodésicas, variam menos que 60m entre SAD-69 e Córrego Alegre e menos de 100m entre SAD-69 e WGS84, no território brasileiro.

A primeira conclusão é a necessidade de atenção ao procedimento de transformação de coordenadas pelos usuários. O uso de diferentes procedimentos de transformação gera diferenças nos valores das coordenadas transformadas.

As conseqüências dessas diferenças são percebidas em problemas de fusão de dados de múltiplas fontes, por exemplo. Duas bases de dados podem estar sob o mesmo sistema de referência, mas terem sido transformadas de outros sistemas utilizando diferentes procedimentos. Isso pode fazer com que a mesma representação de uma feição seja transformada em duas representações diferentes (i.e., com diferença nas coordenadas dos seus pontos). Portanto a segunda conclusão é a necessidade da devida documentação técnica dos dados e dos processamentos aplicados a eles.

Anexo 2: Curso de Geoprocessamento

Curso de Geoprocessamento 2003 modalidade Extensão Universitária, oferecido pelo Instituto de Computação da Unicamp.

Participantes: Alexandre Marino, Fabrício Pavarin e Rafael Fonseca.

O programa do curso é composto por 16 disciplinas, que variam de 8 a 40 horas, oferecidas em aulas semanais aos sábados em tempo integral e em determinadas noites de sextas-feiras (18 às 23h), totalizando 360 horas. As disciplinas estão estruturadas numa seqüência lógica e de complementaridade. O curso é oferecido de janeiro a dezembro de 2003, nas instalações do Instituto de Computação da Unicamp.

Descrição dos módulos:

- Introdução ao Geoprocessamento - A importância de informações georreferenciadas, principais sistemas de geoprocessamento, Sistemas de Informações Geográficas, aplicações de geoprocessamento (8 horas).
- Cartografia Digital Básica - Conceitos básicos fundamentais que permitem utilizar, de um modo mais adequado e confiável, produtos cartográficos como mapas, imagens de satélites, fotografias aéreas (24 horas).
- Pré-processamento de Dados Espaciais - Dados analógicos vs dados digitais, modelos de dados digitais, estrutura de dados vetoriais, sistema CAD, técnicas de digitalização, exercícios práticos (24 horas).
- Introdução ao Sensoriamento Remoto - Conceitos radiométricos básicos, órbitas e sensores, comportamento espectral de alvos, pré-processamento de imagens de satélites (16 horas).
- Processamento de Imagens de Sensoriamento Remoto - Imagens digitais, histogramas, realce, contraste, filtragem, imagens de radar, classificação e exatidão de classificação, correções (geométrica, topográfica), composição, interpretação, uso e aplicações de imagens de satélite (24 horas).
- Introdução a Banco de Dados para Aplicações Geográficas - Banco de dados e sistemas de arquivos, gerenciadores e arquitetura de banco de dados, utilização em aplicações geográficas, modelagem de dados (MER), projeto lógico de dados, processamento de consultas em banco de dados (linguagem SQL) (16 horas).
- Engenharia de Software para Aplicações Geográficas - Paradigma da engenharia de software, ciclo de vida de um software geográfico, técnicas e ferramentas de levantamento de requisitos do usuário e do sistema, método de análise e projetos de sistemas, planejamento e execução de testes, documentação e manutenção de aplicações (24 horas).
- Sistemas de Informação Geográfica (SIG) - Introdução, arquitetura de um SIG, modelos de dados, superfícies contínuas em SIG, análise espacial, aplicações básicas (40 horas).
- Bancos de Dados Espaciais (BD) - Modelagem e especificação BD, BD geográficos para aplicações ambientais e urbanas, estruturação de dados geográficos em BD, BD temáticos, textuais e de imagens, processamento e otimização de consultas espaciais (28 horas).
- Redes de Computadores e Sistemas Distribuídos para Geoprocessamento - Noções básicas de redes, sistemas distribuídos, Internet para gerenciamento de dados

geográficos, protocolo de redes, Ethernet, token ring, ATM, TCP/IP, servidores Web, transmissão de dados via Web, segurança, estudos de caso (24 horas).

- Sistemas Espaciais de Apoio à Decisão (AD) - Modelagem de sistemas vs AD, inteligência artificial e AD, interfaces e visualização em sistemas de AD, sistemas cooperativos, sistemas de AD em grupo, tomada de decisão em geoprocessamento, análise multicritério (24 horas).
- Fundamentos de Geoestatística - Estatística descritiva, funções aleatórias, estacionariedade, conceito de variável regionalizada, semivariograma, estimação e krigeagem (16 horas).
- Monitoramento Ambiental (MA) - Conceitos de planejamento ambiental, geoprocessamento e MA, base de dados para MA, modelagem de dados do meio físico para MA, MA em bacias hidrográficas, zoneamento ambiental, Meio Físico e planos diretores municipais e estudos de caso (16 horas).
- Projeto de SIGs para Infraestrutura de Redes Urbanas - Caracterização de aplicações de redes de concessionárias de serviços públicos, municipalidade, mapeamento urbano básico, características e modelagem, redes de infra-estruturas, conversão de dados (16 horas).
- Projeto de Sistemas Georeferenciados - Especificação e projeto de ferramentas e técnicas de geoprocessamento combinadas com metodologias e tecnologias da computação, para soluções de problemas em aplicações geográficas (24 horas).
- Implementação de Sistemas Georeferenciados - Desenvolvimento e implementação de projetos em geoprocessamento a partir de problemas específicos em aplicações geográficas (36 horas).

Projeto de curso: Fabrício Pavarin

Com base em levantamentos realizados na bacia do Médio Paranapanema existem hoje disponíveis informações biológicas e ecológicas para 481 espécies arbóreas que ocorrem naturalmente na região. A Estação Ecológica de Assis liderou um trabalho de levantamento de dados no campo em 83 diferentes localidades da região. Porém, esses dados sozinhos, apesar de serem uma fonte riquíssima de informação, sem uma ferramenta de análise e de integração de dados, não são suficientes para a tomada de decisão quanto à escolha sobre quais espécies usar e em qual situação. Neste caso, uma solução seria desenvolver ferramentas, utilizando técnicas de geoprocessamento, modelagem espacial e a integração de dados através de sistemas de informação georreferenciadas (SIG), para gerar uma metodologia aplicável em casos de identificação de áreas prioritárias para recuperação ambiental e conservação de espécies nativas. Esta metodologia servirá de base para orientar futuros trabalhos não só de identificação de áreas para recuperação como também dará apoio para a escolha das espécies mais adequadas, de acordo com suas necessidades ecológicas, para diferentes áreas da região de estudo.

Projeto de curso: Alexandre Marino

Título: Ferramenta de classificação da qualidade da água da Bacia do Rio Piracicaba (PCJ) na WEB

Objetivo: O presente projeto tem por objetivo a elaboração de uma ferramenta SIG na WEB com o intuito de facilitar a interpretação de qualidade da água de alguns rios da bacia em questão, numa estrutura acessível a todos os tipos de usuários, possibilitando sua utilização

como instrumento para o conhecimento, o planejamento e o gerenciamento dos recursos hídricos.

Descrição do Produto: A bacia dos rios Piracicaba, Capivari e Jundiá localiza-se entre as coordenadas geográficas 22°00' e 23°00' latitude Sul e 46°00' e 48°00' longitude Oeste; correspondendo a uma superfície de 12.450 km². Desta bacia, são retirados cerca de 2,3Hm³ (2.300.000m³) de água por dia, para uso na área urbana (37%), uso em usinas de açúcar e álcool (16,5%), uso em outras indústrias (32,7%) e uso agropecuário (13,8%). A produção de água durante a estiagem fica em situação crítica para o abastecimento público das cidades, das bacias e de mais 5 municípios de outras bacias. Adicionalmente a essa demanda, há também a reversão de 31m³/s, pelo sistema Cantareira, para o abastecimento de cerca de 55% da população da Região Metropolitana de São Paulo. Para agravar a situação, somente 12% dos esgotos domésticos recebem tratamento e, cerca de 75% das cargas industriais são removidas, lançando uma carga remanescente nos nossos rios e cerca de 106 tDBO/dia referente aos esgotos industriais e 133 tDBO/dia referente aos esgotos domésticos. Este projeto tem como proposta criar uma ferramenta na WEB para exibir um mapa com os principais rios da bacia coloridos de acordo com a qualidade da água, partindo dos valores dos parâmetros da água medidos em diversos pontos do rio em um determinado espaço de tempo. Um índice de classificação será adotado para facilitar a interpretação das informações de qualidade de água de forma abrangente e útil, para especialistas ou não. As cores servirão para mostrar a qualidade da água em questão, classificando-a em ótima, boa, aceitável, ruim ou péssima. Esta ferramenta terá uma estrutura acessível a todos os tipos de usuários, possibilitando sua utilização como instrumento para o conhecimento, o planejamento e o gerenciamento dos recursos hídricos.

Projeto de curso: Rafael Fonseca

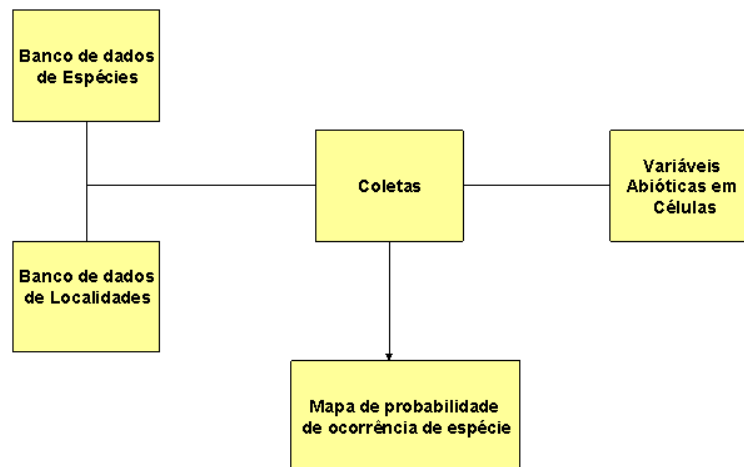
Título: Sistema de Análise de Similaridade de Áreas

Orientador: Professor Dr. Jurandir Zullo Junior (Cepagri/Unicamp)

Uma espécie pode habitar uma determinada região se possuir adaptações necessárias às condições abióticas do local. Algumas espécies encontram condições ótimas e, na ausência de inimigos naturais, tornam-se pragas. Identificar áreas cujas condições climáticas sejam similares a da região invadida pela praga, possibilita um estudo de campo direcionado para busca de inimigos naturais. O Sistema de Análise de Similaridade de Áreas visa identificar similaridade ambiental de áreas e criar um mapa de probabilidade de ocorrência de espécies. A distribuição poderá ser associada com o tipo de dispersão que a planta possui (síndrome de dispersão) e com as variáveis ambientais da localidade da coleta. Também permitirá visualizar as coletas em uma escala temporal.

A relação entre coleta e variáveis abióticas permite compreender a distribuição de espécies. Essa compreensão possibilita inferir em quais regiões há maior probabilidade de ocorrência de espécies, auxilia na elaboração de inventários, no estudo de espécies raras e nos programas de controle biológico. O controle biológico é dado pela busca de inimigos naturais para uma determinada praga. Essa busca pode ser otimizada se for direcionada para locais similares (clima, topografia, solos etc) a regiões de invasão da praga. As regiões indicadas pelo Sistema de Análise de Similaridade de Áreas apenas demonstram que, se for encontrado um inimigo natural na área prevista, ele terá grande potencial de ser um bom agente de controle.

As acamadas ambientais serão divididas em células, pequenos polígonos que possuem igual valor de temperatura, precipitação etc. Essas quadrículas serão definidas por um ponto central. Quando se insere uma coleta, o sistema realizará uma associação entre os pontos de coletas e as células. Essa configuração permite criar buscas de variáveis, cálculos estatísticos e os resultados podem ser visualizados graficamente.



Sistema de Análise de Similaridade de Áreas

Casos de Uso

1. Recomendação de locais prioritários para estudos de controle biológico

Considerando-se a importância da produção agrícola do Brasil em sua economia e sua presença significativa nas exportações, tornam-se essenciais estudos para aumentar a produção, diminuir gastos na lavoura e melhorar a qualidade dos produtos. Atualmente, esse controle da lavoura é realizado com a utilização em massa de agrotóxicos e outros produtos químicos que além de influenciarem na qualidade do produto final, encarecem a produção, pois muitos desses componentes são importados. A utilização de controle biológico é uma alternativa positiva para os agricultores brasileiros. Utilizar um inimigo natural para controlar a praga, ao mesmo tempo em que reduz ou elimina a utilização de agrotóxico, agrega uma qualidade ao produto.

2. Identificação de áreas de riscos

Muitas doenças possuem vetores de transmissão. Através de uma análise de ocorrência de vetores, podem-se buscar áreas similares onde ele possa vir a ocorrer, impedindo uma nova epidemia em uma outra área similar.

3. Estudos de espécies endêmicas

Espécies endêmicas são muito específicas quanto ao seu habitat ótimo, isto é, são tolerantes a poucas variações no ambiente. O sistema pode auxiliar na busca de áreas onde a espécie possa vir a ocorrer bem como auxiliar a reintrodução dessa espécie em áreas nativas.