
Algoritmo Genético para Produção de Conjuntos de Regras (GARP)

RICARDO SCACHETTI PEREIRA^{1*}
MARINEZ FERREIRA DE SIQUEIRA¹

¹ Centro de Referência em Informação Ambiental – CRIA, São Paulo, Brasil.

* e-mail: scachett@gmail.com

RESUMO

Neste trabalho é apresentado o uso de algoritmo genético GARP (*Genetic Algorithm for Rule Set Production*), com o objetivo de modelar a distribuição geográfica potencial de espécies. Tais algoritmos tentam encontrar relações não-aleatórias entre os dados de ocorrência da espécie e 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. Os conceitos em que os algoritmos genéticos se baseiam estão relacionados com os conceitos biológicos de evolução das espécies e de seleção natural. O GARP é um algoritmo genético projetado para criar e otimizar um conjunto de regras ou proposições lógicas. Este conjunto de regras é a base de inferência usada para modelar o nicho ecológico das espécies. A qualidade de cada regra presente no modelo é avaliada através de uma função de adaptação, que é calculada através da significância estatística obtida pela aplicação da regra ao conjunto de pontos de treinamento fornecidos ao algoritmo. Técnicas de modelagem de nicho ecológico são importantes, porque ainda é precário o nível de conhecimento existente sobre a distribuição geográfica de plantas tropicais, sendo a maioria das espécies representadas por poucos pontos de amostragem e, muitas vezes, esses pontos são georreferenciados de maneira muito imprecisa. Essas técnicas podem ser utilizadas para aumentar o conhecimento sobre a distribuição potencial dessas espécies, analisar possíveis rotas de distribuição de espécies invasoras e de vetores de doenças infecciosas, avaliar os melhores ambientes para re-introdução de espécies a seus hábitos naturais, entre outros. Neste trabalho foram utilizados dados de espécies nativas do Brasil para avaliar a eficácia dos algoritmos em prever a distribuição geográfica potencial de espécies tropicais.

ABSTRACT

This work presents the use of the genetic algorithm GARP (Genetic Algorithm for Rule Set Production) in modeling the potential geographic distribution of species. Such algorithms tend to find non aleatoric relations between species occurrence data and ecological/environmental data relevant to the species (such as temperature, precipitation, topography, type of soil, geology, among others) at the point where the species was found. The concepts that genetic algorithms are related to species biological evolution and natural selection. GARP is a genetic algorithm projected to create and optimize a number of rules or logical propositions. This set of rules is the base of inference used to model species ecological niche. The quality of each rule present in the model is

evaluated through an adaptation function that is calculated using statistical significance obtained through the application of the rule to the set of training points furnished by the algorithm. Ecological niche modeling techniques are important because the level of understanding about the geographic distribution of plants in tropics is still precarious, and most species are represented by very few sampling points that many times are georeferenced with little precision. Among many other applications, these techniques may be used to increase the knowledge about the potential distribution of species, analyze possible distribution routes of invasive species and of infectious disease vectors, and to evaluate the best environments to reintroduce species to their natural habitat. In this work, data on native species from Brazil was used to evaluate the algorithms' efficiency in predicting the potential geographic distribution of tropical species.

MODELAGEM DE NICHU 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 Hutchinson (1957) como sendo um conjunto de condições ecológicas com as quais as populações conseguem se manter, podendo ser representado por um espaço ecológico/ambiental multidimensional (Figura 1).

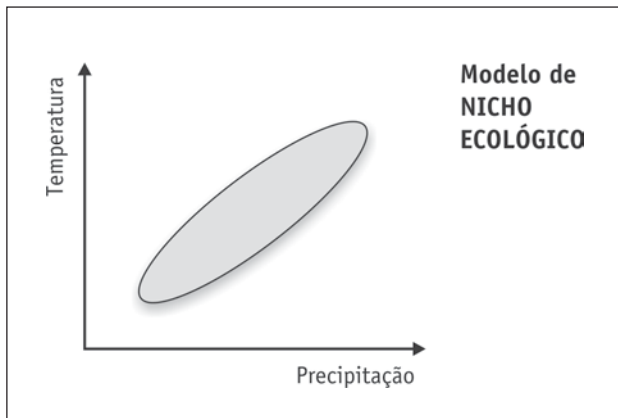


FIGURA 1 – Exemplo de espaço bidimensional definido como nicho ecológico de espécie, baseado em dois parâmetros ambientais (temperatura e precipitação).

Tais algoritmos tentam encontrar relações não-aleatórias entre os dados de ocorrência da espécie e 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).

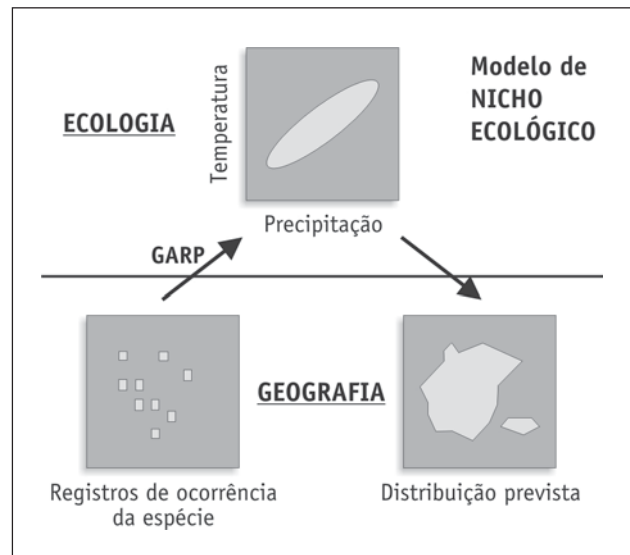


FIGURA 2 – Esquema mostrando a relação entre os dados de campo e a previsão de distribuição geográfica, através do processo de modelagem de nicho ecológico.

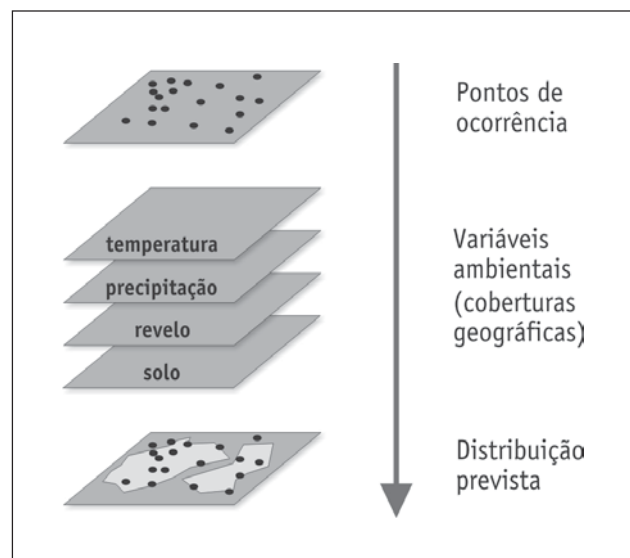


FIGURA 3 – Esquema mostrando a relação entre as variáveis ambientais e os pontos de ocorrência da espécie.

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, sendo que vários destes métodos são apresentados em outros artigos deste volume. No presente trabalho será apresentado em detalhe um destes métodos, chamado Algoritmo Genético para Produção de Conjuntos de Regras (GARP). Entretanto, antes de descrever o algoritmo GARP, é necessário apresentar uma breve descrição geral dos algoritmos genéticos.

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 seguiam 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 específicas 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.

Inspiração na teoria da evolução das espécies e terminologia

Grande parte dos conceitos em que os algoritmos genéticos se baseiam está relacionada 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 três 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.

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 bidimensional 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 .

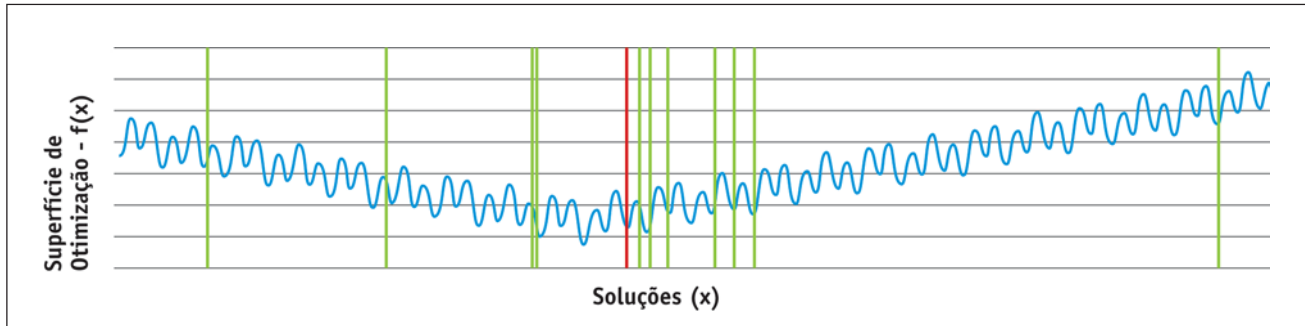


FIGURA 4 – Exemplo de superfície de otimização. Fonte: Obitko, 1998.

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 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.

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.

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

Antes da permutação tinha-se P e Q:

$P = 100111000001-001000000010$ $p = (156, 18, 2);$
 $f(p) = 130$

$Q = 000001000000-100001100100$ $q = (4, 8, 100);$
 $f(q) = 488$

Após a permutação tem-se P' e Q':

$P' = 100111000001-100001100100$ $p' = (156, 24, 100);$
 $f(p') = 158$

$Q' = 000001000000-001000000010$ $q' = (4, 2, 2);$
 $f(q') = 10$

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

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:

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

Após a mutação tem-se P':

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

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

Estrutura de execução dos algoritmos genéticos

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 , se reproduzem 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 interação: 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 Goldberg (1989), Davis (1996) e Mitchell (1996).

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 e foi desenvolvido originalmente por Stockwell e Noble (1992). Este conjunto de regras é a base de inferência usada para modelar o nicho ecológico das espécies.

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. Elas usam um único valor de cada uma das variáveis ambientais na sua pré-condição. Um exemplo de regra atômica seria:

Se **Temperatura = 23°C** e **Elevação = 2000 m**, então o 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, sendo que as variáveis ambientais podem assumir valores dentro de uma faixa. Por exemplo:

Se **Temperatura está entre 23° e 28°C** e **Elevação está entre 2000 e 2700 m**, então o 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 **Temperatura não está entre 23° e 28°C** e **Elevação não está entre 2000 e 2700 m**, então o 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 a seguir.

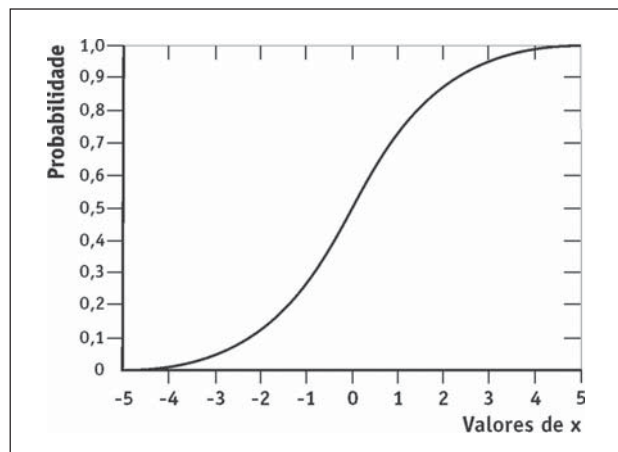


FIGURA 5 – Exemplo de função de probabilidade na regressão logística.

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 **Temperatura está entre 5° e 10°C** e **Elevação está entre 1000 e 2000 m**, então o táxon está presente (regra de faixa).

REGRA 2 – Se **Temperatura é de 18°C** e **Elevação é 1500 m**, então o táxon está ausente (regra atômica).

REGRA 3 – Se $p = 1 / [1 + e^{-y}] > 0,75$, onde $y = 0,2 + \text{Temperatura} \times 0,45 + \text{Elevação} \times -0,12$, então o táxon está presente (regra tipo logit).

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 como na Tabela 1.

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 ou 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), que é calculada através da significância estatística obtida pela aplicação da regra ao conjunto de pontos de treinamento fornecidos ao algoritmo.

Aplicação dos operadores heurísticos sobre as regras

Durante a execução do algoritmo, as regras são modificadas aleatoriamente através de operadores heurís-

TABELA 1

Regra	Tipo da Regra	Temperatura: Primeiro Fator	Temperatura: 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%

TABELA 2

Situação	Regra	Temperatura: Primeiro Fator	Temperatura: 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%

TABELA 3

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	0%
	6	18	10	1000	2000	P	97%

ticos de recombinação e mutação. Por exemplo, uma operação de mutação sobre a regra 1, na posição do segundo gene (segundo fator da variável Temperatura), seria realizada como na Tabela 2.

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 como na Tabela 3.

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.

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 provavelmente a espécie está 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 está disponível 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; Peterson & Viegas, 2001; Anderson *et al.*, 2002a, b; Fera & Peterson, 2002; Peterson *et al.*, 2002a, b, c; Stockwell & Peterson 2002a, b; Anderson *et al.*, 2003; Peterson *et al.*, submetido; Peterson *et al.*, 2003).

APLICAÇÕES DO GARP NO BRASIL

O GARP vem sendo utilizado para estabelecer previsões sobre a distribuição geográfica de espécies para condições tropicais. A seguir, mostramos alguns exemplos do que vem sendo desenvolvido no Brasil, com base no uso do GARP.

Uso do GARP para prever a distribuição geográfica de espécies

Atualmente ainda é precário o nível de conhecimento existente sobre a distribuição geográfica de plantas tropicais, sendo a maioria das espécies representadas por poucos pontos de amostragem e, muitas vezes, esses pontos são georreferenciados 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, dados não publicados). 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). 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.

Os dados provenientes de outras bases de dados que não as disponíveis em Ratter *et al.* (2001), pontos representados por quadrados na Figura 6, não participaram da modelagem, mas ilustram, em alguns casos, a qualidade do modelo gerado pelo GARP. Em outros casos (como o ponto isolado no canto superior esquerdo da figura, localizado no Equador) mostram que o modelo não foi capaz de prever aquela região como de provável ocorrência da espécie. A causa disso, neste caso, muito provavelmente é resultante da ausência de pontos no conjunto de dados de treinamento que representassem as condições ambientais presentes naquela região.

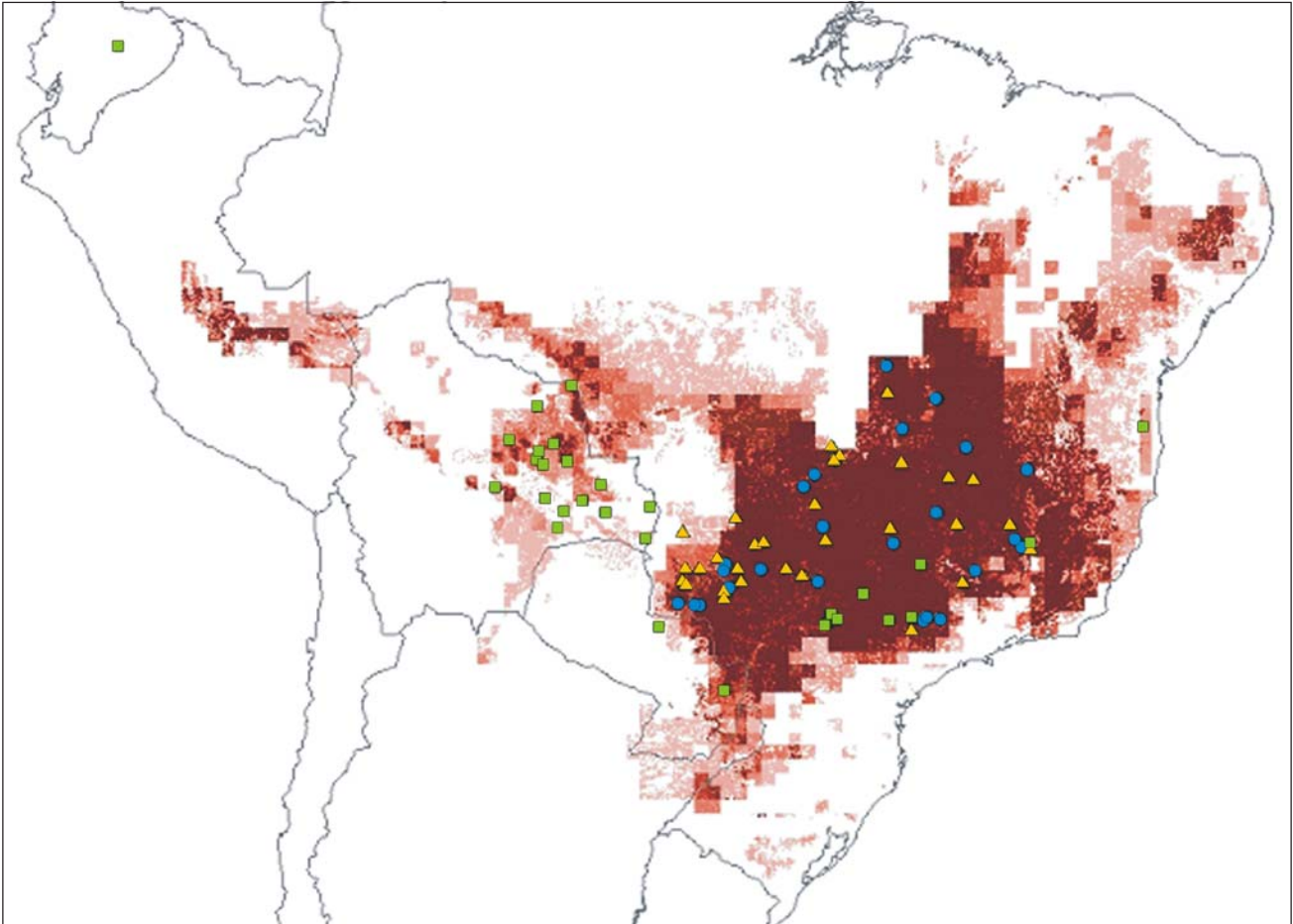


FIGURA 6 – Distribuição prevista para *Terminalia argentea*. Os círculos azuis representam os pontos usados na confecção do modelo (dados de treinamento); os triângulos amarelos são os pontos usados para testar o modelo (dados de teste); e os quadrados verdes são registros dessa espécie provenientes de outros conjuntos de dados (New York Botanical Garden, Biota/Fapesp e Missouri Botanical Garden).

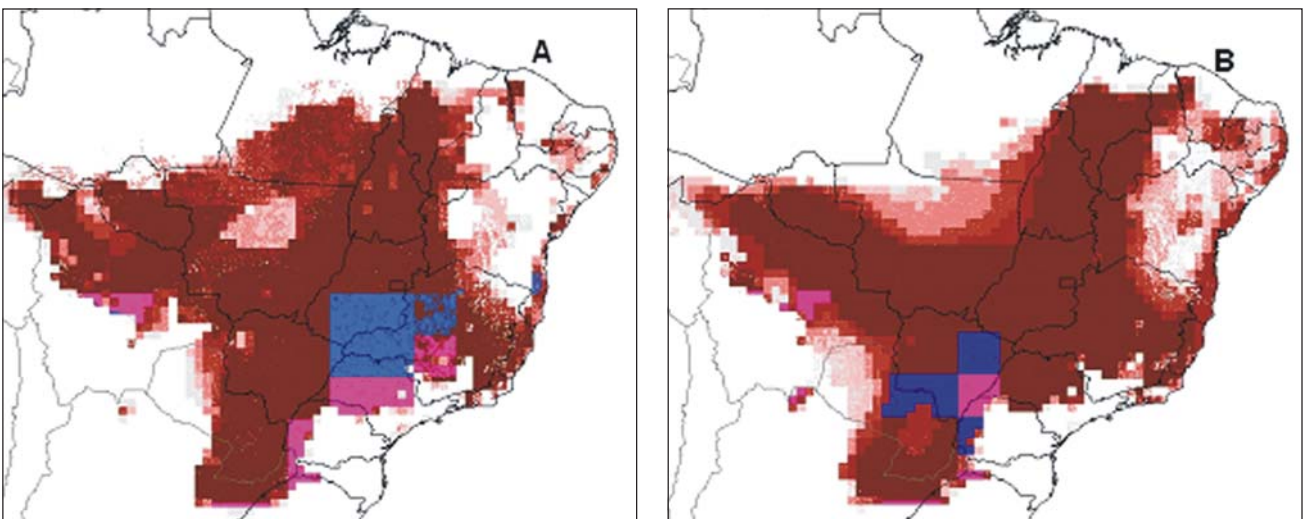


FIGURA 7 – Exemplo de previsão de distribuição atual (1961-1990) em vermelho escuro e futura (2055), em azul, para o cenário HHGSDX50, e em rosa para o cenário HHGGAX50, para espécies arbóreas de cerrado: A - *Rudgea viburnoides* (Cham.) Benth. (Rubiaceae) e B - *Qualea parviflora* Mart. (Vochoysiaceae).

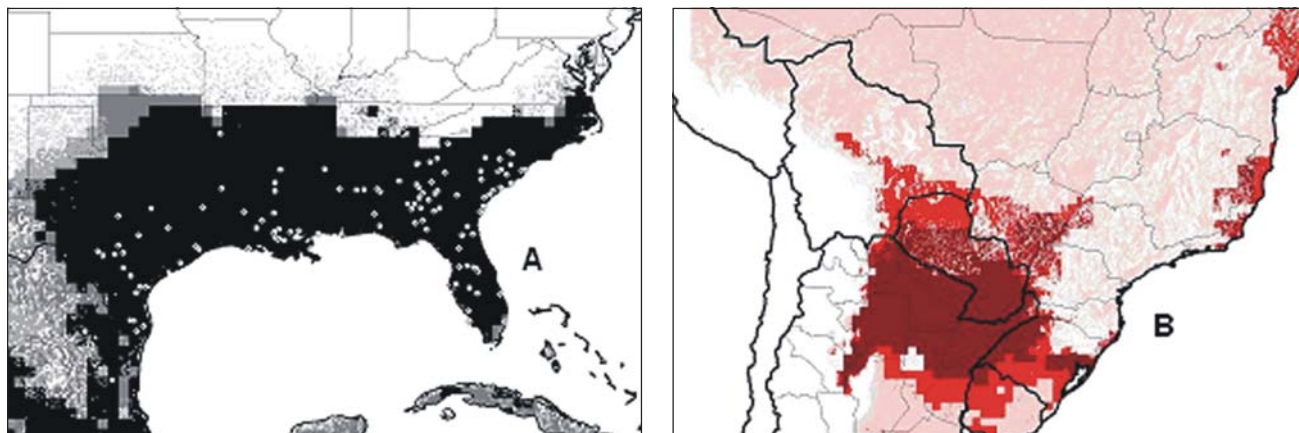


FIGURA 8 – Previsão de invasão de *Homalodisca coagulata* na América do Sul. A - previsão de distribuição nativa da espécie no sudoeste dos Estados Unidos. B - projeção do nicho ecológico potencial de ocupação da espécie na América do Sul (modificado de Peterson *et. al.*, 2003).

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, baseadas em dois cenários futuros de alteração climática (Siqueira & Peterson, 2003). Em vermelho, temos a área prevista de distribuição da espécie no presente. Em azul, temos a área habitável em 2055, baseada no cenário HHGSDX50 (cenário mais otimista). Em rosa temos a área habitável em 2055, baseada no cenário HHGGAX50 (cenário mais pessimista).

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.*, 2002d). Como resultado, podemos observar uma redução de área na distribuição geográfica, que varia de intensidade dependendo da espécie, para algumas espécies arbóreas de cerrado.

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 projeta-se 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 a 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 (Figura 8a).

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 8b).

REFERÊNCIAS BIBLIOGRÁFICAS

- 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.
- 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.
- Carson, D.J. 1999. Climate modelling: Achievements and prospects. *Quarterly Journal of the Royal Meteorological Society* 125: 1-27.
- Darwin, C.R. 1859. *On the origin of species by means of natural selection*. John Murray, London.
- Davis, L. 1996. *Handbook of genetic algorithms*. Thomson Computer Press, Boston, Massachusetts.
- 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., A.J. Owens & M.J. Walsh. 1966. *Artificial intelligence through simulated evolution*. Wiley, New York.
- Goldberg, D. 1989. *Genetic algorithms in search, optimization and machine learning*. Addison. Wesley Publishing Company, Inc., Reading, Massachusetts.
- Hutchinson, R.E. 1957. Concluding remarks. *Cold Spring Harbor. Symposia on Quantitative Biology*. 22: 415-427.
- Holland, J. 1975. *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, Michigan.
- Mitchell, M. 1996. *An introduction to genetic algorithms*. MIT Press, Cambridge, MA.
- Nix, H.A. 1986. A biogeographic analysis of Australian Elapid snakes. In: R. Longmore (ed). *Atlas of Australian Elapid Snakes*. Australian Flora and Fauna Series 8: 4-15.
- Obitko, M. 1998. *Genetic Algorithms*. Disponível em <http://www.professor.webizu.org/ga/> (acessado em agosto de 2007).
- 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. & D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling. *BioScience* 51:363-371.
- 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: 27-32.
- Peterson, A.T., V. Sanchez-Cordero, C.B. Beard & J.M. Ramsey. 2002b. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases* 8: 662-667.
- Peterson, A.T., D.R.B. Stockwell & D.A. Kluza. 2002c. Distributional prediction based on ecological niche modeling of primary occurrence data. In: M. Scott (ed). *Predicting species occurrences: Issues of scale and accuracy*. pp. 617-623. Island Press, Washington, D.C.
- Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R.H. Buddemeier & D.R.B. Stockwell. 2002d. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416: 626-629.
- Peterson, A.T., M. Papes & D.A. Kluza. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*. 51: 863-868.
- 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 Neotropica* 3. Disponível em: <http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003> (acessado em julho de 2007).
- Ratter, J.A., S. Bridgewater & J.F. Ribeiro. 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.
- Siqueira, M.F. & A.T. Peterson. 2003. Consequences of global climate change for geographic distributions of cerrado tree species. *Biota Neotropica* 3: 1-14.
- Stockwell, D.R.B. 1999. Genetic algorithms II. In: A. H. Fielding (ed). *Machine learning methods for ecological applications*. pp. 123-144. Kluwer Academic Publishers, Boston.
- Stockwell, D.R.B. & I.R. Noble, 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. & A.T. Peterson, 2002a. Controlling bias in biodiversity data. In: J. M. Scott (ed). *Predicting species occurrences: Issues of scale and accuracy*. pp. 537-546. 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 Systems* 13: 143-158.