Marinez Ferreira de Siqueira

Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Doutor em Ciências da Engenharia Ambiental.

Orientador: Dra Giselda Durigan

São Carlos

2005

Dedico este trabalho aos meus pais,
Osny e Glorinha, pelo amor
incondicional que sempre me
dedicaram.

Agradecimentos

À Dra Giselda Durigan pela amizade de longa data, pelo companheirismo e pela maravilhosa orientação em todas as etapas do trabalho. Ao Dr. Luiz Eduardo Dias pelo aceite em me orientar até que a Giselda pudesse assumir oficialmente minha orientação.

Ao Prof. Dr. Andrew Townsend Peterson e Ricardo Scachetti, por terem me apresentado a modelagem como ferramenta para tratar os dados que eu estava levantando no campo e pelo valioso auxílio e orientação no desenvolvimento da modelagem.

Aos amigos Sidnei de Souza, Paulo de Marco Jr. e Mauro Enrique de Souza Munoz, pelo enorme auxílio no desenvolvimento das etapas de modelagem deste trabalho e na análise dos dados.

À FAPESP, que, através do Programa Biota, financiou o projeto *Viabilidade de Conservação dos Fragmentos de Cerrado do Estado de São Paulo,* no qual foram levantadas as espécies arbóreas de cerrado no estado de São Paulo utilizadas neste trabalho e o projeto *Sistema de Informação Distribuído para Coleções Biológicas: a Integração do Species Analyst e do SinBiota*, no qual foram desenvolvidas as técnicas de modelagem.

À Floresta Estadual de Assis, principalmente a Wilson Aparecido Contieri e Antônio Carlos Galvão de Melo pelos dados de vegetação natural da Bacia do Médio Paranapanema utilizados neste trabalho.

Ao CDVale (Centro de Desenvolvimento do Vale Paranapanema), principalmente a Hugo de Souza Dias, que forneceu os ambientais da Bacia do Médio Paranapanema.

Aos amigos do CRIA (Centro de Referência em Informação Ambiental), principalmente Dora Ann Lange Canhos e Ingrid Koch, pela leitura crítica do manuscrito e pelas valiosas contribuições feitas.

Aos amigos Giselda Durigan, Geraldo Antônio Daher Corrêa Franco e Edivaldo Furlan por toda a experiência que eles me passaram e também por toda a diversão na parte mais gostosa do trabalho que foi o tempo que passamos no campo, levantando os dados.

A todos os meus sinceros agradecimentos!

RESUMO

SIQUEIRA, M.F. (2005). *Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais*. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.

Atualmente vêm sendo desenvolvidas e utilizadas várias técnicas de modelagem de distribuição geográfica de espécies com os mais variados objetivos. Algumas dessas técnicas envolvem modelagem baseada em análise ambiental, nas quais os algoritmos procuram por condições ambientais semelhantes àquelas onde as espécies foram encontradas, resultando em áreas potenciais onde as condições ambientais seriam propícias ao desenvolvimento dessas espécies. O presente estudo trata do uso da modelagem preditiva de distribuição geográfica, através da utilização de algoritmo genético e algoritmo de distância, de espécies como ferramenta para a conservação de espécies vegetais, em três situações distintas: modelagem da distribuição do bioma cerrado no Estado de São Paulo; previsão da ocorrência de espécies arbóreas visando à restauração da cobertura vegetal na Bacia do Médio Paranapanema e modelagem da distribuição de espécies ameaçadas de extinção (*Byrsonima subterranea*). A metodologia empregada e os resultados obtidos foram considerados satisfatórios para a geração de modelos de distribuição geográfica de espécies vegetais, baseados em dados abióticos, para as regiões de estudo. A eficácia do modelo em predizer a ocorrência de espécies do cerrado é maior se forem utilizados apenas pontos de amostragem com fisionomias de cerrado, excluindo-se áreas de transição. Para minimizar problemas decorrentes da falta de convergência do algoritmo utilizado GARP (*Genetic Algorithm for Rule Set Production)*, foram gerados 100 modelos para cada espécie modelada. O uso de modelagem pode auxiliar no entendimento dos padrões de distribuição de um bioma ou ecossistema em uma análise regional e local.

Palavras-chave: algoritmos genéticos, modelagem de biodiversidade, espécies vegetais.

ABSTRACT

SIQUEIRA, M.F. (2005). *Using modeling based on fundamental niche in the evaluation of geographic distribution for plant species*. Ph.D. Thesis - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.

Many different techniques are being used today with different objectives to model the geographic distribution of species. Some of these techniques are based on environmental analysis where algorithms search for environmental conditions similar to those where the species were found, resulting in potential areas where the environmental conditions would be favorable for the development of these species. The present study is based on predictive modeling of the geographic distribution of species using genetic and distance algorithms as a tool for plant conservation in three different situations: modeling the Cerrado biome of the State of São Paulo; predicting the occurrence of tree species, aiming at the recuperation of the Médio Paranapanema watershed; and, modeling the distribution of endangered species (*Byrsonima subterranea*). The methodology used and the results obtained were considered satisfactory for the generation of geographic distribution models for plant species, based on abiotic data for the regions studied. The accuracy of the model in predicting the occurrence of species in the Cerrado is greater if only points that occur within the Cerrado and not in transition areas are used. In order to minimize problems due to the lack of convergence of the genetic algorithm GARP (*Genetic Algorithm for Rule Set Production*), 100 models were generated for each specie. The use of modeling tools can help understand the distribution patterns of a biome or ecosystem in a regional or local analysis.

Keywords: genetic algorithms, biodiversity modeling, plant species.

Sumário

Introdução Geral 1

A evolução da destruição da biodiversidade 1

O desenvolvimento da informática para a biodiversidade 2

O potencial da Modelagem 4

Modelagem da biodiversidade baseada no nicho fundamental das espécies 6

Objetivos 11

Objetivo geral 11

Objetivos específicos 11

Capítulo 1: Modelagem de espécies lenhosas para a região de cerrado no Estado de São Paulo 12

Introdução 12

Materiais e Métodos 19

Escolha das espécies 21

Parâmetros ambientais utilizados para a modelagem 22

Método de Modelagem 22

Resultados e Discussão 25

Testando a convergência do algoritmo 25

Avaliação da especificidade ambiental dos pontos de amostragem no resultado da modelagem para a área de Cerrado do Estado de São Paulo 27

Distribuição potencial do Bioma Cerrado no Estado de São Paulo. 31

Aplicações dos modelos resultantes 44

Conclusões 45

Capítulo 2: Modelagem de distribuição geográfica para espécies lenhosas da Bacia do Médio Paranapanema 46

Introdução 46

Objetivos 47

Material e Métodos 47

Área de estudo 47

Origem dos dados de ocorrência das espécies 53

Escolha das espécies 54

Pontos de amostragem 57

Método de modelagem 57

Parâmetros ambientais utilizados para a modelagem 57

Resultados e discussão 58

Teste dos diferentes tipos de bases de dados abióticos para distribuição potencial em uma escala local 58

Avaliação do efeito do número de amostras na porcentagem de acerto dos modelos 65

Modelagem de distribuição potencial de espécies vegetais para uma escala local..... 66

Influência de fatores ambientais na distribuição de espécies arbóreas 68
Aplicações de modelagem para auxiliar trabalhos de recuperação ambiental..... 71

Conclusões..... 74

Capítulo 3: Modelagem para espécies com distribuição restrita ou com poucos pontos de amostragem.....75

Introdução..... 75

Objetivos..... 76

Materiais e Métodos 76
Área de estudo 78

Desenvolvimento do algoritmo 78

Parâmetros ambientais utilizados na modelagem 79

Busca da espécie no campo 80

Método de modelagem..... 80

Resultados e discussão..... 80

Conclusões..... 85

Comentários gerais 86

Aplicação de modelagem para estudo de distribuição geográfica de espécies vegetais 86
Qualidade dos dados bióticos e abióticos..... 86

Compatibilidade de escalas 86

Referências Bibliográficas.....88

Anexo 1 - Algoritmos Genéticos.....95

Recombinação genética..... 97

Mutação..... 98

GARP 101

Anexo 2: Lista dos fragmentos de vegetação natural utilizados na modelagem das espécies da Bacia do Médio Paranapanema. 105

Lista de Figuras

Figura 1: Modelagem da biodiversidade baseada em dados abióticos..... 7

Figura 2: Esquema mostrando a relação entre as variáveis ambientais e os pontos de ocorrência da espécie. 8

Figura 3: Simulação dos dois possíveis erros associados ao modelo GARP, a omissão e a sobre-previsão..... 9

Figura 4: Evolução da devastação da vegetação natural do Estado de São Paulo..... 12

Figura 5: Distribuição do Bioma Cerrado no Estado de São Paulo em 1962 (BORGONOVI e CHIARINI 1965). 14

Figura 6: Estado atual da vegetação natural do Bioma Cerrado no Estado de São Paulo. Fonte: (KRONKA et al. 1998)..... 15

Figura 7: Fonte - Áreas prioritárias para conservação do Cerrado no Estado de São Paulo. PROBIO/SP – Secretaria do Meio Ambiente do Estado de São Paulo, 1997. . 16

Figura 8: Mapa da vegetação do Estado de São Paulo editado a partir do Mapa de Vegetação do Brasil segundo IBGE (1988) digitalizado por U.S. Geological Survey's (USGS) EROS Data Center, Sioux Falls, South Dakota com as localidades amostradas no projeto "Viabilidade de conservação dos remanescentes de Cerrado do Estado de São Paulo". 20

Figura 9: Mapa dos remanescentes de vegetação nativa do Estado de São Paulo segundo mapeamento do Instituto Florestal de São Paulo (KRONKA et al. 2003). 21

Figura 10: Resultado da modelagem da área potencial de distribuição para o mesmo conjunto de dados de ocorrência de *Byrsonima intermedia*. Cada mapa é o resultado de um conjunto da soma de 100 modelos gerados pelo GARP..... 26

Figura 11: Média das distâncias entre modelos baseados na ocorrência de *Byrsonima intermedia* em função do número de modelos somados. A linha representa o ajuste produzido por técnica de ajuste não-linear a um modelo de função potencia. A Equação da curva é $D = 0{,}441N^{-0{,}517}$; onde *D* é a distância e *N* o número de modelos. 27

Figura 12: Comparação entre as distribuições potenciais do Bioma Cerrado no Estado de São Paulo com base na ocorrência das 17 espécies apenas em áreas de cerrado (A) e em Cerrado e outros tipos de vegetação (B). O gradiente de cor representa o número de modelos somados (os 10 melhores modelos de 100 gerados para cada espécie), quanto mais escura a cor, mais modelos previram a ocorrência do conjunto de espécies analisado..... 29

Figura 13: Comparação entre a porcentagem de acerto de dois conjuntos de pontos de treino para a modelagem do bioma cerrado em São Paulo. Experimento A: localidades de amostragem contendo apenas áreas de cerrado. Experimento B: localidades de cerrado, floresta estacional e vegetação de transição entre as duas (região de ecótono cerrado/floresta). 30

Figura 14: Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (entre os 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP, a partir das 69 localidades exclusivas de Cerrado amostradas no escopo do projeto "Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo". 33

Figura 15: Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (dos 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP e a distribuição dos Remanecentes de Cerrado no Estado de São Paulo (Instituto Florestal de São Paulo 2004) com um recorte dos sub-tipos de solos nos quais as localidades amostradas de Cerrado ocorreram (localidades amostradas de Cerrado, Ecótono e Floresta Estacional Semidecidual no escopo do projeto "Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo"). 35

Figura 16: A - Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (dos 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP e a distribuição dos Remanecentes de Cerrado no Estado de São Paulo (Instituto Florestal de São Paulo 2004). B - Recorte das áreas de solo tipo Latossolos Vermelhos e Argissolos Vermelho Amarelos (localidades amostradas de Cerrado, Ecótono e Floresta Estacional Semidecidual no escopo do projeto "Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo"). 40

Figura 17A: Distribuição potencial de espécies de cerrado no Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos entre 100 gerados pelo GARP. 42

Figura 18: Localização geográfica da Bacia do Médio Paranapanema no Estado de São Paulo com os 63 pontos de amostragem da vegetação utilizados neste estudo. 49

Figura 19: Curva climática da Bacia do Médio Paranapanema (FUNDAG 1998)..... 49

Figura 20: Mapa da probabilidade de ocorrência de geada na Bacia do Médio Paranapanema e a distribuição das localidades amostradas na região. 50

Figura 21: Mapa Pedológico da Bacia do Médio Paranapanema. Fonte: Centro de Desenvolvimento Agropecuário do Médio Vale do Paranapanema – CDVale..... 51

Figura 22: Resultado da modelagem para a probabilidade de ocorrência de 28 espécies arbóreas, utilizando diferentes conjuntos de dados ambientais na Bacia do Médio Paranapanema, Estado de São Paulo. 60

Figura 23: Número de pontos de ocorrência de espécies e respectivas porcentagens de acerto entre os diferentes conjuntos de dados ambientais. **A-média**: média dos valores da porcentagem de acerto para os experimentos **A** e **A1**. **B-média**: média dos valores da porcentagem de acerto para os experimentos **B** e **B1**. 62

Figura 24: Comparação entre os valores de intercepto da regressão linear entre as porcentagens de acerto e número de pontos de teste para as diferentes bases de dados abióticos. Números iguais representam valores estatisticamente iguais a partir de um teste *a posteriori* de Tukey em um nível de 5% de significância..... 63

Figura 25: Comparação entre os valores de inclinação da regressão linear entre as porcentagens de acerto e número de pontos de teste para as diferentes bases de dados abióticos. Números iguais representam valores estatisticamente iguais a partir de um teste *a posteriori* de Tukey em um nível de 5% de significância..... 64

Figura 26: Efeito do número de pontos de treino sobre a média do erro e a variação entre cinco repetições do modelo de distribuição potencial de *Casearia sylvestris,* utilizando dados ambientais **D** (Worldclim – médias anuais). 65

Figura 27: Mapa da riqueza potencial para 28 espécies arbóreas da Bacia do Médio Paranapanema, produzido a partir da soma dos 10 melhores modelos (entre os 100 gerados para cada espécie) utilizando todos os registros (pontos de treino e pontos de teste) disponíveis na área de estudo. 67

Figura 28: Distribuição potencial de *Xylopia aromatica* para a Bacia do Médio Paranapanema e a probabilidade de ocorrência de geada na região..... 69

Figura 29: Mapas de distribuição potencial para quatro espécies arbóreas ocorrentes na Bacia do Médio Paranapanema. SG: espécies suscetíveis à geada. TG: espécies tolerantes à geada. Os pontos verdes representam os registros de ocorrência para cada espécie. 70

Figura 30: Cruzamento de dados entre o modelo de riqueza de espécies e as áreas de risco de erosão, uso do solo e remanescentes de vegetação nativa para a Bacia do Médio Paranapanema. As áreas em destaque A e B representam 2 pontos quaisquer da Bacia. 71

Figura 31: Lista das espécies com potencial de ocorrência no ponto A da Bacia do Médio Paranapanema, classificadas de acordo com o número decrescente de modelos que previram a ocorrência de cada espécie entre os 10 melhores modelos (dentre os 100 gerados) e a informação ambiental disponível para o ponto A. 72

Figura 32: Lista das espécies com potencial de ocorrência no ponto B da Bacia, classificadas de acordo com o número decrescente de modelos que previram a ocorrência de cada espécie entre os 10 melhores modelos (dentre os 100 gerados) e a informação ambiental disponível para o ponto B. 73

Figura 33: *Byrsonima subterranea* Brad. & Mark. (Malpighiaceae). Registro da planta no município de Assis, SP..... 77

Figura 34: Fisionomia de Cerrado aberto no qual foi encontrada a espécie Byrsonima subterranea no município de Avaré, SP..... 77

Figura 35: Localização da Bacia do Médio Paranapanema no Estado de São Paulo e do ponto original de ocorrência de *Byrsonima subterranea*..... 78

Figura 36: Método de modelagem de nicho fundamental, baseado em distância ambiental (distância euclidiana) para espécies com um único ponto de amostragem. 79

Figura 37: Modelo de distribuição potencial de *Byrsonima subterranea* baseado em um único registro de ocorrência (Estação Ecológica de Assis, Município de Assis, SP) utilizando dados climáticos e topográficos. O detalhe representa a Bacia do Médio Paranapanema e é apresentado na Figura 38..... 82

Figura 38: Resultado do trabalho de campo com a inclusão de mais seis registros de ocorrência para a espécie *Byrsonima subterranea* coletadas nas áreas de maior probabilidade de ocorrência da espécie associadas ao tipo de vegetação na região da Bacia do Médio Paranapanema. 83

Figura 39: Exemplo de superfície de otimização. Fonte: (OBITKO 1998). 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. 97

Figura 40: Exemplo de função de probabilidade na regressão logística..... 101

Lista de Tabelas

Tabela 1: Espécies utilizadas na modelagem da distribuição potencial do bioma Cerrado no Estado de São Paulo e o número de pontos de treino utilizados para gerar os modelos nos experimentos A e B. **vs** = verdadeiros; **fs** = falsos 24

Tabela 2: Análise de covariância para a porcentagem de acerto nos experimentos A e B e o efeito do número de pontos de teste. **gl trat** = graus de liberdade do tratamento; **gl erro =** graus de liberdade do erro; **P** = probabilidade do resultado ser dada ao acaso. 30

Tabela 3: Lista dos diferentes tipos de solos onde ocorreram as localidades amostradas de Cerrado (OLIVEIRA et al. 1999). 38

Tabela 4: Detalhamento da legenda para os 38 tipos de solos da Bacia do Médio Paranapanema. 51

Tabela 5: Espécies arbustivo-arbóreas utilizadas na modelagem para a Bacia do Médio Paranapanema, com as fisionomias em que ocorrem. Dados ecológicos, silviculturais e de fisionomia obtidos em Durigan et al. (2004ª) e Durigan et al. (2004b). **Cce** = campo cerrado; **Ce** = cerrado *s.s.* ; **CE** = cerradão; **FESxC** = região de transição entre a Floresta Estacional Semidecidual e o Cerrado; **MC** = Mata Ciliar; **FES** = Floresta Estadual Semidecidual; **A** = árvore; **a** = arbusto; **Pre** = preferencial de locais úmidos; **Ind** = Indiferente à umidade do solo; **Int** = intolerante a solos úmidos; **M** = moderado; **L** = lento; **R** = rápido; **SG** = suscetível à geada; **TG** = tolerante à geada e **T** = tolerante; **I** = intolerante. * *Xylopia aromatica* não compõe o conjunto dase 28 espécies utilizada na modelagem da Bacia porque não contem o número mínimo de pontos estabelecidos (30 pontos) para esse experimento, mas foi utilizada no experimento referente à análise dos fatores ambientais envolvidos na modelagem. .. 55

Tabela 6: Comparação entre as porcentagens de acerto total para os diferentes conjuntos de dados ambientais utilizados para modelar as espécies da Bacia do Médio Paranapanema. **A** e **A1**: Imagem de Satélite NDVI, duas repetições. **B** e **B1**: Worldclim médias mensais, duas repetições. **C**: Worldclim média anual com valores máximos, médios e mínimos. **D**: Worldclim médias anuais. 61

Tabela 7: Análise de covariância para o efeito do número de registros de ocorrência e diferentes conjuntos de dados ambientais sobre a porcentagem de acerto no modelo GARP para espécies arbóreas na Bacia do Médio Paranapanema. **gl trat** = graus de liberdade do tratamento; **gl erro =** graus de liberdade do erro; **P** = probabilidade do resultado ser dada ao acaso. 64

Introdução Geral

A evolução da destruição da biodiversidade

Vários estudos têm sido realizados para identificar as regiões com maior risco de perda de biodiversidade no planeta. Myers (1988) introduziu o conceito dos *hotspots* e dez anos depois Mittermeier et al. (1999) estabeleceram 25 *hotspots* no mundo, com base na existência de espécies endêmicas (espécies cuja distribuição é restrita a algum ecossistema específico) ou áreas onde 75% ou mais da vegetação original já tenha sido destruída. Dois outros critérios complementam as diretrizes para o estabelecimento de áreas prioritárias para conservação. São eles: o conceito de *Wilderness Areas* (áreas praticamente intactas com mais de 75% de sua vegetação original e com baixa densidade populacional) e países com os maiores índices de riqueza natural (MITTERMEIER et al. 1999).

O Brasil, de acordo com estes critérios, é o país com a maior diversidade biológica do mundo, detém a maior área verde praticamente intacta do planeta, (a Floresta Amazônica) e abriga dois grandes biomas considerados *hotspots* (Mata Atlântica e Cerrado), com grande ocorrência de espécies endêmicas associadas (MITTERMEIER et al. 1999).

O principal esforço para a definição de prioridades de conservação da biodiversidade no Brasil foi desenvolvido no escopo do projeto "Ações Prioritárias para a Conservação da Biodiversidade dos Biomas Brasileiros". Através deste projeto, coordenado pelo Ministério do Meio Ambiente, foram realizados vários workshops para discutir e definir prioridades de conservação para o Cerrado e Pantanal; Zona Costeira e Marinha; Floresta Amazônica; Floresta Atlântica e Campos Sulinos e Caatinga (Conservation International [1]).

Em termos vegetacionais o Estado de São Paulo abrange dois biomas, o Cerrado e a Mata Atlântica. Para o Cerrado foi realizado o Workshop “Ações Prioritárias para a Conservação da Biodiversidade do Cerrado e do Pantanal”,

[1] http://www.conservation.org.br/

realizado em 1998, e teve como principal resultado um Mapa de Áreas Prioritárias. Neste mapa foram identificadas 87 áreas prioritárias para a conservação da biodiversidade destes biomas no país, incluindo desde os cerrados na divisa do Paraná e São Paulo até os encraves de cerrado do Amapá e Roraima.

Apesar desses esforços, a situação atual do bioma Cerrado no Brasil é crítica, sendo que a área efetivamente protegida sob unidades de conservação é de menos de 3% da sua área total (MMA 1998). Essa devastação é particularmente grave nos Estados de São Paulo e Mato Grosso do Sul, que abrangem, junto com o Paraná, o limite sul da distribuição do Cerrado no Brasil.

Para a Mata Altântica, que teve sua área original diminuída para 7,3% da sua área original (SOS Mata Atlântica[2]), foi realizado em 1999 o *Workshop* "Ações e Áreas Prioritárias para a Conservação da Mata Atlântica e Campos Sulinos". O principal resultado foi a identificação de 184 áreas prioritárias para a conservação da biodiversidade da região.

A identificação dessas áreas é importante para direcionar investimentos e esforços na conservação e proteção desses Biomas ameaçados. As ações conservacionistas mais recomendadas pelos *workshops* foram a criação de Unidades de Conservação e o reconhecimento da necessidade de investimentos em projetos e programas de recuperação e recomposição de áreas degradadas envolvendo essas áreas. Dentro deste escopo, espécies do Cerrado e da Floresta Estacional Semidecidual, ocorrentes no Estado de São Paulo e na Bacia do Médio Paranapanema, foram alvo de estudos de modelagem no presente trabalho.

O desenvolvimento da informática para a biodiversidade

Devido à atual velocidade da devastação ambiental, é crescente a demanda por respostas rápidas que visem a solucionar problemas como a perda de biodiversidade, controle de espécies invasoras e de pragas agrícolas, previsão e controle de epidemias etc. Para que a implementação de estratégias

[2] http://www.sosmatatlantica.org.br/

adequadas para o desenvolvimento sustentável seja mais eficaz é necessário que sejam tomadas decisões baseadas na gestão do conhecimento. Para tanto, é necessário o desenvolvimento de processos que auxiliem o trabalho de produzir, trabalhar e usar a informação disponível. Estes processos envolvem a geração, o compartilhamento, a preservação e a utilização do conhecimento (FLORIANO 2004). Sendo assim, a adoção de estratégias adequadas poderá ser amplamente beneficiada com um acesso rápido e confiável aos dados disponíveis e com a utilização de ferramentas computacionais avançadas para a análise e visualização espacial desses dados, assim como para a construção de cenários de impacto e vulnerabilidade ambiental.

Na atual demanda por conhecimento, a geração e o uso de inovações em ciência e tecnologia são meios para a agregação de valor aos mais diversos produtos, tornando-se chave para a competitividade estratégica e o desenvolvimento social e econômico de uma região (CGECon[3]). Na última década, os avanços em tecnologia possibilitaram a criação de um ambiente cooperativo, envolvendo o compartilhamento de conhecimentos, sem precedentes. Além do desenvolvimento científico e tecnológico, as demandas das convenções e tratados internacionais sobre as questões ambientais estão também contribuindo para a mudança de paradigma na questão do acesso livre e compartilhado à informação sobre biodiversidade.

Estas novas demandas estão catalisando o desenvolvimento de uma nova área do conhecimento, a *informática para biodiversidade*. O *Global Biodiversity Information Facility - GBIF* (Infra-estrutura Global de Informação sobre Biodiversidade) é uma iniciativa internacional que tem como objetivo ampliar e consolidar o livre acesso à informação sobre biodiversidade na Internet. Esta iniciativa foi oficialmente estabelecida em 2001 e desde então vem incentivando o desenvolvimento e adoção de padrões e protocolos necessários para viabilizar a interação entre sistemas de informação.

O portal do “Instituto Virtual da Biodiversidade”, associado ao programa Biota/Fapesp, já possui cerca de 400 mil registros, disponíveis *on-line*, somando-se os dados relacionados a observações de coleta servidos pelo

[3] http://www.cgecon.mre.gov.br/

sistema *Sin*Biota (Sistema de Informação Ambiental do Biota)[4] aos dados de espécimes armazenados em coleções biológicas, servidos pela rede *species*Link (sistema distribuído de informação de dados primários de coleções científicas do Estado de São Paulo[5]).

No entanto, existem ainda grandes lacunas de conhecimento sobre a distribuição de espécies nos principais biomas do Brasil. O uso de ferramentas computacionais, como os Sistemas de Informação Geográfica (SIGs) e a modelagem de distribuição de espécies, pode auxiliar a direcionar a pesquisa de campo, identificar áreas de maior riqueza biológica, assim como delimitar áreas ricas em espécies ameaçadas ou endêmicas.

O potencial da Modelagem

Atualmente vêm sendo desenvolvidas e utilizadas várias técnicas de modelagem de distribuição geográfica de espécies com os mais variados objetivos. Algumas dessas técnicas envolvem modelagem baseada em análise ambiental, nas quais os algoritmos procuram por condições ambientais semelhantes àquelas onde as espécies foram encontradas, resultando em áreas potenciais onde as condições ambientais seriam propícias ao desenvolvimento dessas espécies. Esses métodos de modelagem podem ser divididos em dois tipos, os chamados modelos estatísticos e os modelos baseados em regras. No primeiro caso, temos os seguintes métodos:

- **Regressões múltiplas** (MR: *Multiple Regression)* (GUISAN e ZIMMERMANN 2000)
- **Modelos lineares** (GLM: *Generalized Linear Models)* (AUSTIN et al. 1994)
- **Modelos aditivos** (GAM: *Generalized Additive Models)* (YEE e MITCHELL 1991)
- **Árvores de classificação** (CTA: *Classification Tree Analysis)* (BREIMAN et al. 1984).

Como exemplo de modelos baseados em regras (NIX 1986) podemos citar:

---

[4] http://sinbiota.cria.org.br/
[5] http://splink.cria.org.br/

- **Algoritmos genéticos** (GA: *Genetic Algoritm*) um exemplo deste tipo de algoritmo é o GARP: Genetic Algorithm for Rule Set Production (STOCKWELL 1999; STOCKWELL e NOBLE 1992; STOCKWELL e PETERS 1999). Ver anexo 1 para maior detalhamento.
- **Redes neurais** (NN: *Neural Networks*) (OLDEN e JACKSON 2002; PEARSON et al. 2002).

Todos esses métodos buscam definir as limitações ambientais das espécies nas dimensões para as quais o modelo é desenvolvido. Assim, informações relacionadas aos pontos de coleta podem ser projetadas em espaço geográfico, e os algoritmos identificam locais com características ambientais similares na área de análise, indicando onde as espécies são potencialmente capazes, ou não, de manter populações viáveis (PETERSON e VIEGLAIS 2001).

Essas técnicas permitem identificar espécies que poderiam ser utilizadas em trabalhos de recuperação ambiental, avaliar o potencial de ameaça de espécies invasoras (PETERSON et al. 2003a; PETERSON et al. 2003b), avaliar o impacto das mudanças climáticas na biodiversidade (HUNTLEY et al. 1995; MAGANA et al. 1997; OBERHAUSER e PETERSON 2003; PETERSON et al. 2002a; SALA et al. 2000; SIQUEIRA e PETERSON 2003; THOMAS et al. 2004), estudar possíveis rotas de disseminação de doenças infecciosas (PETERSEN e ROEHRIG 2001; PETERSON et al. 2002b) e auxiliar na determinação de áreas prioritárias para conservação (BOJORQUEZ-TAPIA et al. 1995; CHEN e PETERSON 2002; EGBERT et al. 1998; ORTEGA-HUERTA e PETERSON 2004).

O uso de modelagem pode ainda indicar áreas de distribuição potencial para espécies em risco de extinção. Neste caso, o modelo resultante determina possíveis locais de ocorrência dessas espécies, aumentando a base de conhecimento a respeito da situação de risco em que se encontram. Esses resultados podem auxiliar na indicação de áreas onde existam as condições ambientais ideais para a sobrevivência das espécies analisadas, as quais podem ser recomendadas como áreas prioritárias para a conservação de espécies ou podem ser definidas como áreas potenciais para a reintrodução destas espécies.

O uso de técnicas de modelagem de distribuição geográfica de espécies é particularmente indicado em situações nas quais é preciso tomar decisões mas ainda não existe disponível um conjunto grande de informações, como é o caso de várias áreas dos principais biomas brasileiros, nas quais a coleta de dados ainda é bastante precária para o tamanho e riqueza biológica de cada região.

Técnicas de modelagem têm se mostrado de grande valia na avaliação da distribuição geográfica de espécies. Há, no entanto, uma série de questões e limitações intrínsecas a cada método utilizado. É preciso que sejam realizados testes e desenvolvidas técnicas de avaliação dos resultados para que o algoritmo a ser utilizado seja o mais coerente possível com a situação analisada. Também é importante frisar que a eficácia dessas técnicas é altamente dependente da qualidade dos dados biológicos e ambientais disponíveis.

Modelagem da biodiversidade baseada no nicho fundamental das espécies

O processo de modelagem consiste em converter dados primários de ocorrência de espécies em mapas de distribuição geográfica, indicando a provável presença ou ausência da espécie.

Estes modelos trabalham, na maioria dos casos, com o conceito de nicho ecológico da espécie. Este conceito foi definido por Hutchinson (1957), 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).

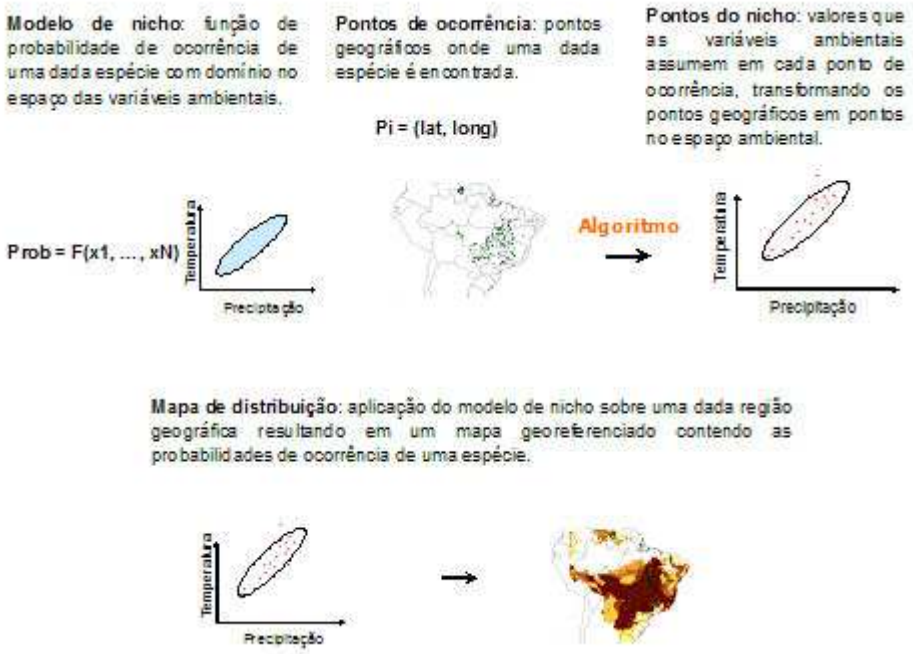


Figura 1: Modelagem da biodiversidade baseada em dados abióticos.

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 (Figura 2). Com base nesses dados são gerados mapas potenciais de distribuição geográfica das espécies (ANDERSON et al. 2002; CHEN e PETERSON 2002; GOLDBERG 1989; PETERSON 2001).

Neste trabalho será utilizada a definição de nicho fundamental em vez de nicho ecológico. O nicho fundamental consiste no conjunto de condições sob as quais a espécie pode existir na ausência de fatores bióticos limitantes, como competição, predação e parasitismo (ACIESP 1997). O uso deste tipo de abordagem encontra-se descrito no trabalho de Soberon e Peterson (2005).

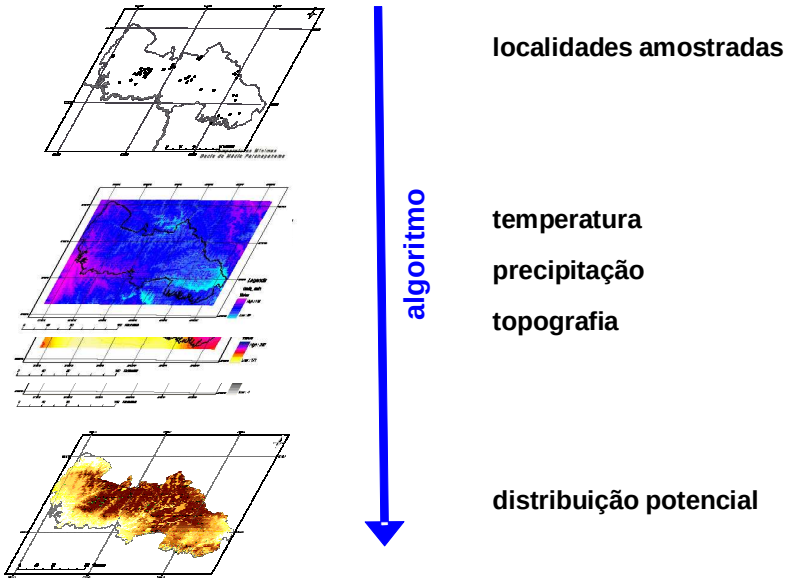


Figura 2: Esquema mostrando a relação entre as variáveis ambientais e os pontos de ocorrência da espécie.

Para a realização dos modelos apresentados neste estudo, será utilizado o algoritmo GARP - *Genetic Algorithm for Rule Set Production* (STOCKWELL e NOBLE 1992). No anexo 1 é apresentada uma descrição detalhada de como funcionam os algoritmos genéticos.

Em resumo, todo o processo descrito acima, utilizando o GARP, pode ser dividido em seis etapas:

1. Divisão do conjunto de pontos em *training data set* (para construir modelos) e *test data set* (para avaliação da eficácia do modelo).
2. Aplicação do algoritmo ao *training data set.*
3. Geração das regras
4. Avaliação da eficácia do modelo perguntando quando a omissão (quando a área ocupada não faz parte da previsão) e a sobre-previsão (quando a previsão inclui área não ocupada) (Figura 3) são significativamente *menores* que num modelo aleatório.

5. Seleção das melhores regras. Ao final de um número predeterminado de iterações, o conjunto de regras com melhor desempenho passa a representar o modelo do nicho fundamental da espécie.
6. Aplicação do modelo de volta ao espaço geográfico, indicando as regiões onde a espécie provavelmente estará presente ou ausente.

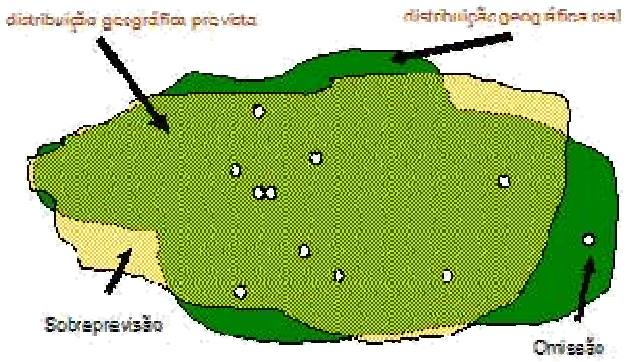


Figura 3: Simulação dos dois possíveis erros associados ao modelo GARP, a omissão e a sobre-previsão.

O presente estudo trata do uso da modelagem preditiva de distribuição geográfica de espécies como ferramenta para a conservação de espécies vegetais, em três situações distintas:

- Modelagem da distribuição do bioma cerrado no Estado de São Paulo.
- Previsão da ocorrência de espécies arbóreas visando à restauração da cobertura vegetal na bacia do Médio Paranapanema.
- Modelagem da distribuição de espécies ameaçadas de extinção (como é o caso da *Byrsonima subterranea*), visando à identificação de áreas com probabilidade de se encontrar a espécie, orientando assim trabalhos de campo e identificando áreas prioritárias para a conservação da espécie.

Cada um desses objetivos foi explorado em um capítulo separado ulizando as mesmas técnicas gerais de modelagem, mas com procedimentos e objetivos específicos próprios.

Este documento está, portanto, dividido em três capítulos:

<u>Capítulo 1</u>: Modelagem de espécies lenhosas de cerrado em todo o Estado de São Paulo

Capítulo 2: Modelagem de espécies arbóreas do Médio Paranapanema

Capítulo 3: Modelagem de espécies vegetais com distribuição restrita ou com poucos pontos de amostragem.

Objetivos

São apresentados a seguir os objetivos da pesquisa como um todo, envolvendo os três capítulos que se seguem:

Objetivo geral

Verificar a eficácia do uso de ferramentas de modelagem computacional e matemática para mapear a distribuição potencial de espécies vegetais.

Objetivos específicos

- Testar diferentes conjuntos de dados ambientais para otimizar o procedimento de modelagem.
- Predizer a distribuição geográfica potencial de algumas espécies do cerrado no Estado de São Paulo.
- Predizer a distribuição espacial para um conjunto de espécies arbóreas ocorrentes na Bacia do Médio Paranapanema, Estado de São Paulo.
- Identificar os parâmetros ambientais que melhor explicam a distribuição geográfica das espécies arbóreas em uma escala local (Bacia Hidrográfica).
- Desenvolver técnicas de modelagem para prever a ocorrência de espécies com poucos pontos de coleta.

Capítulo 1: Modelagem de espécies lenhosas para a região de cerrado no Estado de São Paulo

Introdução

Originalmente o Estado de São Paulo apresentava uma cobertura florestal correpondente a 81.8% do seu território (VICTOR 1975). O primeiro mapa da Figura 4 (Original) mostra, em verde, as áreas de florestas e, em branco, as áreas cobertas por vegetação de cerrado antes da ocupação humana. Os outros mapas da Figura 4 mostram a redução acelerada da cobertura florestal na qual o Cerrado e as áreas desmatadas são tratados igualmente, de modo que é impossível acompanhar a devastação desse bioma particularmente.

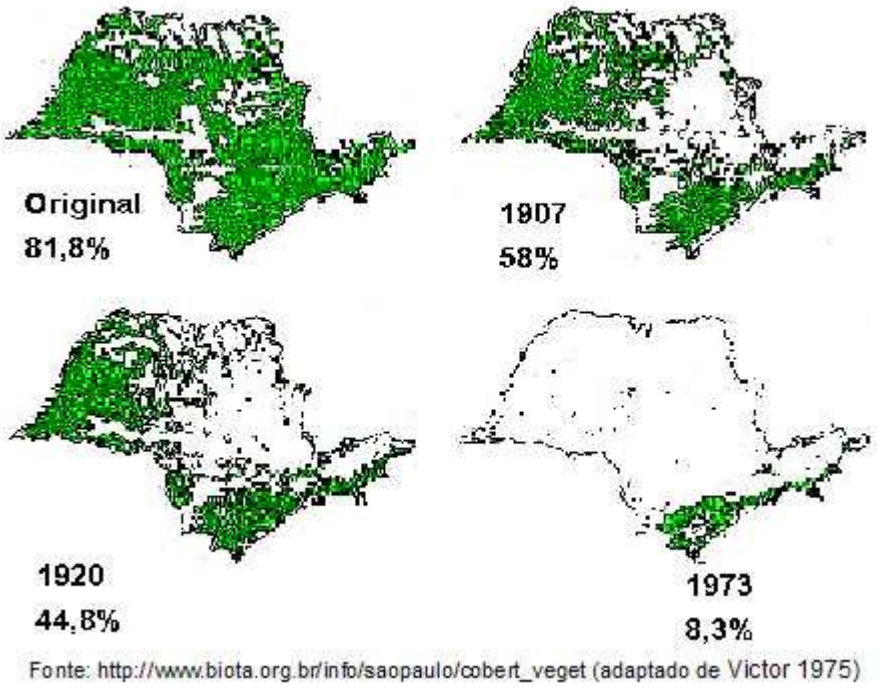


Figura 4: Evolução da devastação da vegetação natural do Estado de São Paulo.

Do ponto de vista de área total remanescente, a situação do Cerrado (sentido lato) no Estado é ainda mais devastadora do que a da Mata Atlântica. O Bioma Cerrado, em São Paulo, abrangia originalmente um quinto do Estado (EITEN 1970). Porém, naquela época, alguns autores comentaram a inadequação dos mapas de vegetação em seus trabalhos e experiências de campo. Eiten (1972) discute essas limitações e segundo o autor, nenhum mapa delimitava satisfatoriamente a área de cerrado no Brasil, especialmente em relação às áreas limítrofes, de contato com outros tipos de vegetação e isso era particularmente evidente no Estado de São Paulo. Ainda segundo Eiten (1972), o melhor mapeamento da época, apesar de excluir algumas áreas, foi o de Borgonovi e Chiarini (1965) que, no início da década de 60, realizaram um levantamento do Cerrado a partir de fotografias aéreas de 1962 (Figura 5). O resultado deste levantamento mostra a área de cerrado ocupando cerca de 15,4% do território paulista. No começo da década de 70, esta área já estava reduzida a cerca de 4,18%, (SERRA FILHO et al. 1974). Atualmente esta área é de menos de 1% da sua distribuição original (KRONKA et al. 1998), cobrindo apenas cerca de 23.790ha, de forma extremamente fragmentada (Figura 6).

Essa situação se agravou ainda mais quando o Cerrado foi excluído do decreto 750[6], o qual dispõe sobre o corte, a exploração e a supressão de vegetação primária ou nos estágios avançado e médio de regeneração de mata atlântica. Por conseqüência, a devastação do Cerrado no Estado atingiu níveis altíssimos. A velocidade da devastação levou à realização do *workshop* “Bases para Conservação e Uso Sustentável das Áreas de Cerrado do Estado de São Paulo” (SMA-SP 1997), com o objetivo de realizar um diagnóstico da situação atual das áreas de cerrado no Estado e propor diretrizes para a sua conservação. O mapa síntese desse workshop, indicando áreas consideradas prioritárias para a conservação do cerrado no Estado é apresentado na Figura 7.

---

[6] http://www2.ibama.gov.br/~misis/cnia/lema_texto/Decreto/DC00750-100293.htm

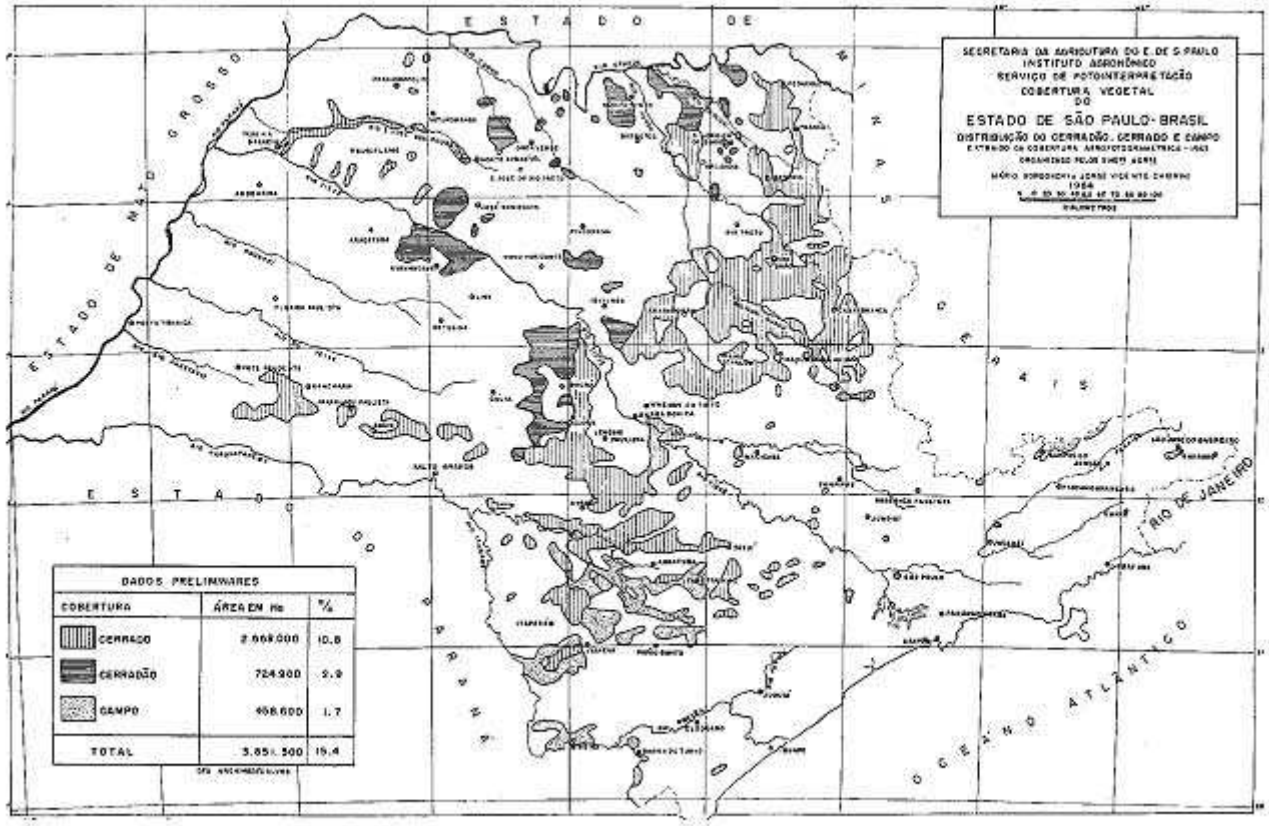


Figura 5: Distribuição do Bioma Cerrado no Estado de São Paulo em 1962 (BORGONOVI e CHIARINI 1965).

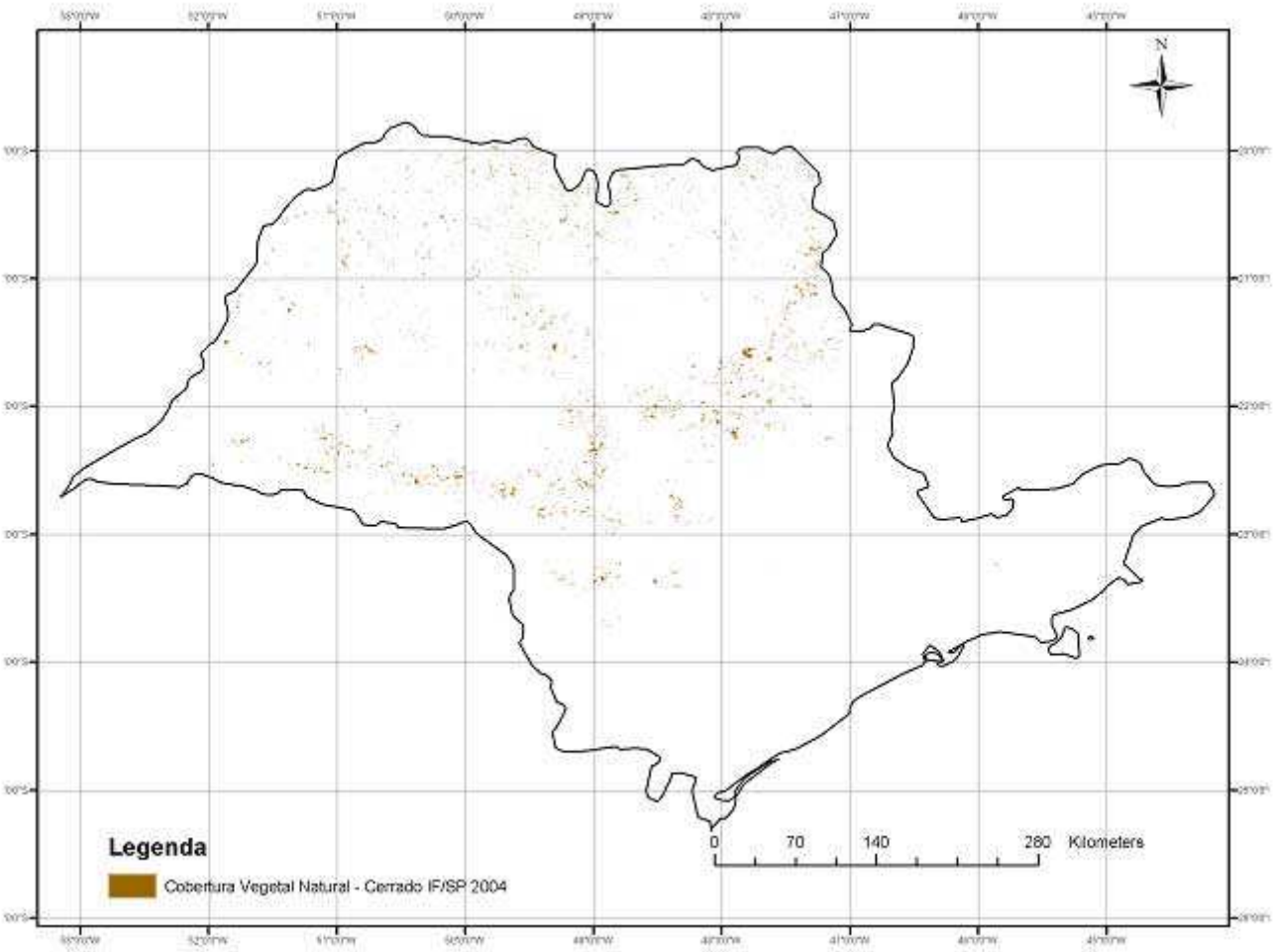


Figura 6: Estado atual da vegetação natural do Bioma Cerrado no Estado de São Paulo. Fonte: (KRONKA et al. 1998).

Figura 7: Fonte - Áreas prioritárias para conservação do Cerrado no Estado de São Paulo. PROBIO/SP – Secretaria do Meio Ambiente do Estado de São Paulo, 1997.

As áreas prioritárias geradas neste *workshop* foram alvo de estudo pelo projeto "Viabilidade de conservação dos remanescentes de Cerrado do Estado de São Paulo") dentro do Programa Biota/Fapesp[7] e os resultados do levantamento de flora deste projeto encontram-se disponíveis no site do SinBiota[8]. Estes dados serviram de base para a publicação recente de alguns artigos sobre padrões fitogeográficos do cerrado paulista. Durigan et al. (2003a) analisaram a flora arbórea de 202 remanescentes de cerrado ocorrentes em São Paulo e em Estados vizinhos. Neste trabalho, a distribuição das áreas geradas por análise multivariada sugere que as relações de similaridade se estabelecem em primeiro lugar em função do clima e em segundo lugar, em função da fertilidade dos solos.

A vegetação de cerrado é decorrente de vários fatores, sendo os mais importantes: estação seca definida, fertilidade e drenagem do solo, regime de

[7] http://www.biota.org.br/

fogo, e as flutuações climáticas do Quaternário (OLIVEIRA-FILHO e RATTER 2002).

Dentre esses, o solo é hoje considerado um dos fatores mais importantes para a distribuição de espécies e a variação florística no cerrado (RATTER e DARGIE 1992).

As principais classes de solos na região dos cerrados são os latossolos podzólicos, cambissolos, areia quartzosas, litólicos, laterita hidromórfica e gleis. Esses tipos de solos apresentam baixa capacidade de armazenamento de água, ou seja, são bastante lixiviados e, conseqüentemente, possuem baixa fertilidade natural, acidez elevada e altos níveis de alumínio trocável (VARGAS e HUNGRIA 1997).

Em termos geomorfológicos, a região central do Brasil é um exemplo clássico de evolução policíclica da paisagem, ou seja, é constituída de formas jovens de superfícies oriundas do Pleistoceno e remanescentes de superfícies muito mais antigas (MOTTA et al. 2002).

Vários autores analisaram as questões relativas às condições climáticas e à vegetação durante o período Quaternário no Brasil Central (BEHLING 1998, 2003; LEDRU et al. 1998; SALGADO-LABOURIAU 1998; SALGADO-LABOURIAU et al. 1998; SALGADO-LABOURIAU et al. 1997).

Estes estudos envolvendo o passado têm confirmado que as florestas e o cerrado mostraram sucessivas expansões e retrações, seguindo as flutuações climáticas do Quaternário (LEDRU 2002).

Segundo Behling (1998), a área de cerrado no sudeste do Brasil foi bem maior no começo do Holoceno (cerca de 10.000 anos A.P. - antes do presente) do que no final deste período. As ilhas de cerrado encontradas hoje no meio da floresta estacional parecem indicar uma distribuição anterior do cerrado bem maior.

Mais especificamente, há cerca de 32.000 anos A.P., o clima era quente e úmido. Os registros mais antigos de pólen de cerrado são dessa época (LEDRU 2002). De 30.000 a 26.000 A.P. o clima tornou-se mais frio, mas ainda úmido. De 19.000 a 7.000 anos A.P. houve uma época mais seca, com a presença de desertificação em várias áreas, com predomínio de formações de

---

[8] http://sinbiota.cria.org.br/

cerrado aberto. Cerca de 7.000 anos A.P. a umidade começou a aumentar, o que levou a uma expansão de florestas de galeria e da floresta estacional. De 4.000 anos A.P. até o presente o cerrado arbóreo e a floresta estacional chegaram aos níveis atuais de abrangência (LEDRU et al. 1998). Os encraves de cerrado presentes hoje na Amazônia e em São Paulo podem ser resultantes destas expansões do passado (LEDRU 2002).

Para algumas áreas a vegetação de cerrado foi substituída pela floresta semidecidual, mas isso somente no último milênio, sugerindo que o fim do Holoceno foi um período mais úmido (BEHLING 2003). Segundo o autor, por volta de 7.560 anos A.P. as florestas de galeria se expandiram nos vales, demostrando uma maior umidade. Porém, entre 7.560 e 6.000 anos A.P., houve um retorno às condições mais secas e a savana voltou a se expandir, mas entre 6.000 e 2180 anos A.P., os vales foram novamente cobertos por florestas semideciduais, mas nas regiões mais elevadas o cerrado aberto continuou prevalecendo. Essa situação mudou entre 2.180 a 600 anos e a vegetação de cerrado aberto nos topos mudou para cerrado mais fechado, refletindo as condições mais úmidas do período. E finalmente, apenas há 600 anos, a floresta semidecídua vem se expandindo nesta região (BEHLING 2003).

Contudo, desde então, apesar dos níveis de umidade terem aumentado, as florestas não retomaram sua distribuição original em relação ao Cerrado. Isso parece estar ligado a fatores como a utilização de fogo, cujo uso recorrente, assim como o empobrecimento do solo, associado a períodos de seca, tendem a impedir a sucessão de espécies florestais (OLIVEIRA-FILHO e RATTER 2002). A distribuição atual do cerrado e da floresta estacional no Estado de São Paulo apresenta um padrão de mosaico bastante evidente, sendo impossível, muitas vezes, determinar onde uma formação termina e a outra começa.

Visando a fortalecer estratégias de conservação, diante do atual Estado de fragmentação deste Bioma no Estado, foram aplicadas neste estudo técnicas de análise espacial e ferramentas de modelagem que pudessem predizer a potencial distribuição do Bioma Cerrado no Estado, baseadas em parâmetros ambientais como o clima e o meio físico do presente. Com base em dados biológicos (pontos de ocorrência de espécies) e dados ambientais (mapas

temáticos), foram aplicados algoritmos com o objetivo de modelar o nicho fundamental e a área potencial de ocorrência dessas espécies (veja a introdução geral).

No presente capítulo pretende-se responder às seguintes perguntas:

- É possível usar modelagem para mapear as áreas potenciais de distribuição atual do cerrado no Estado de São Paulo?
- A mistura de diferentes biomas na modelagem afeta a eficácia do modelo em predizer os pontos de teste?
- Para o conjunto de dados disponível, qual o grau de incerteza que a falta de convergência do algoritmo genético acarreta nos modelos resultantes?

Materiais e Métodos

Para a realização deste estudo, foram amostradas 86 áreas, originalmente mapeadas como Cerrado (KRONKA et al. 1998), localizadas em 40 municípios do Estado, tendo sido registradas 548 espécies lenhosas, das quais 17 foram selecionadas para a realização de modelagem. O levantamento dessas 86 áreas foi realizado dentro do escopo do projeto “Viabilidade de conservação dos remanescentes de Cerrado do Estado de São Paulo”, cujo objetivo foi realizar um levantamento da flora e do estado de conservação dos remanescentes, em áreas não protegidas por Unidades de Conservação, dentro do Estado. Adotou-se, para o levantamento da flora neste projeto, um método de inventário florístico expedito, no qual a suficiência de amostragem baseia-se em uma curva de espécies x tempo. O método, inicialmente utilizado por Ratter et al. (2000), encontra-se descrito em Durigan et al. (2003b), que apresentam a análise dos padrões fitogeográficos encontrados para o cerrado em São Paulo, resultado do referido projeto.

As análises desenvolvidas neste capítulo tratam da distribuição do cerrado no Estado de São Paulo. São apresentadas na Figura 8 os tipos de vegetação ocorrentes no Estado, assim como as áreas de contato entre os diferentes tipos, segundo o IBGE (1988) e as localidades amostradas no projeto. A Figura 9 apresenta os remanescentes de vegetação nativa segundo Kronka et al. (2003).

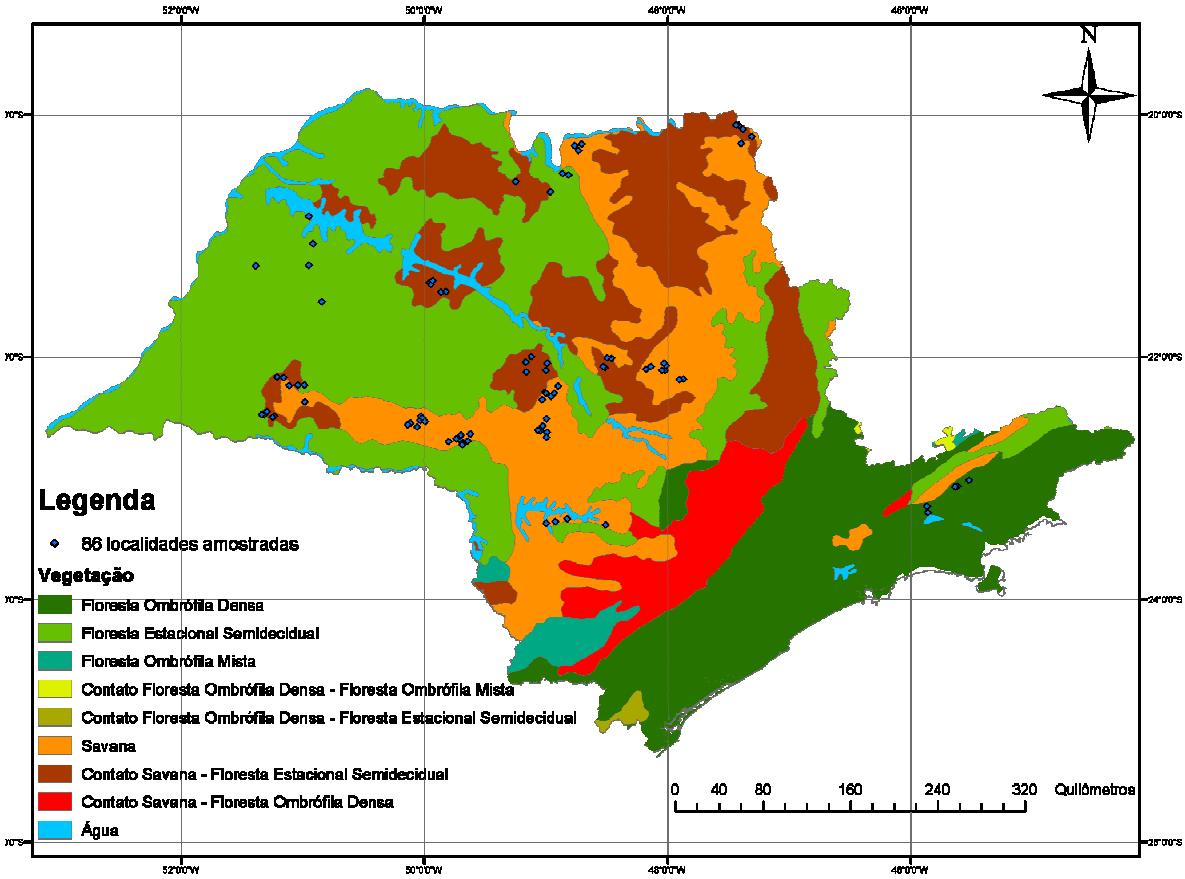


Figura 8: Mapa da vegetação do Estado de São Paulo editado a partir do Mapa de Vegetação do Brasil segundo IBGE (1988) digitalizado por U.S. Geological Survey's (USGS) EROS Data Center, Sioux Falls, South Dakota[9] com as localidades amostradas no projeto "Viabilidade de conservação dos remanescentes de Cerrado do Estado de São Paulo".

[9] http://www.whrc.org/southamerica/LBAData/Braz_Veg_IBGE.htm

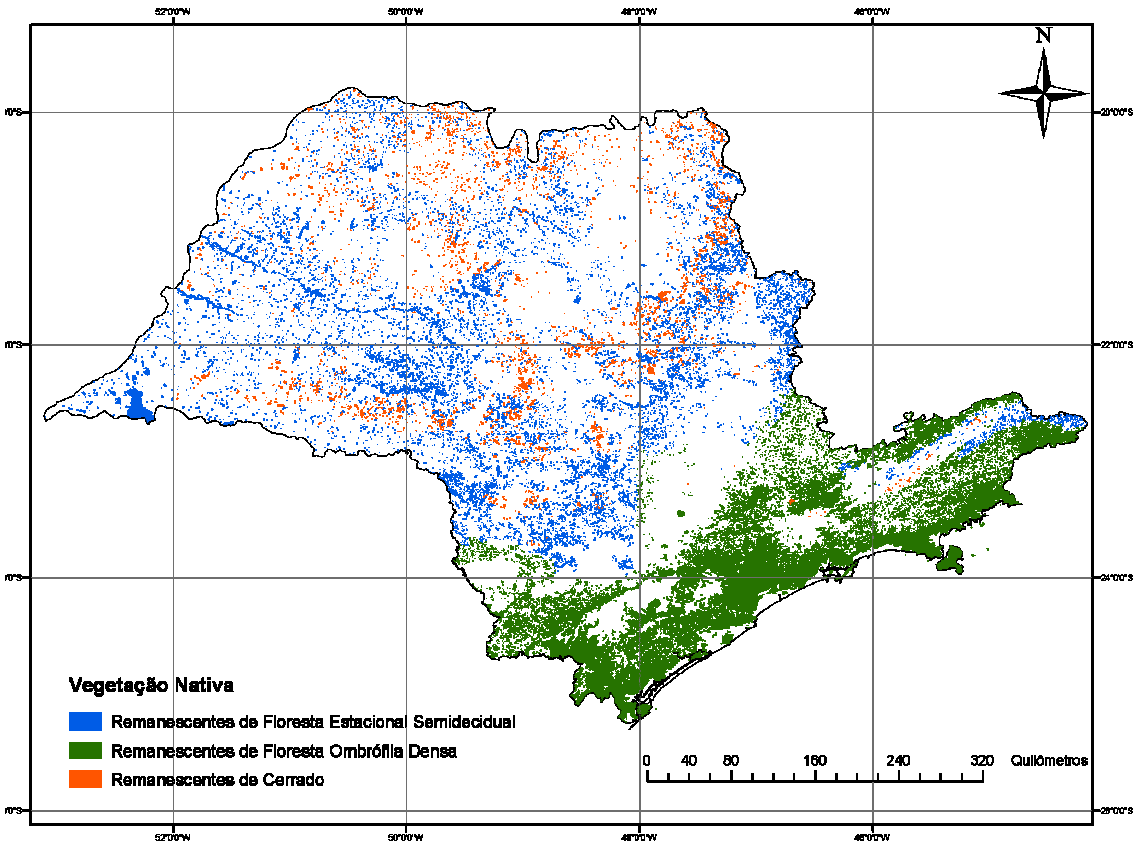


Figura 9: Mapa dos remanescentes de vegetação nativa do Estado de São Paulo segundo mapeamento do Instituto Florestal de São Paulo (KRONKA et al. 2003).

Escolha das espécies

Para a realização da modelagem, foram escolhidas 17 espécies (Tabela 1) por apresentarem os maiores conjuntos de pontos disponíveis ou seja, registradas em 20 ou mais localidades diferentes dentre as 86 áreas levantadas e por serem espécies características do cerrado (DURIGAN et al. 2004b). A escolha deste limite de corte (20 localidades) seguiu as recomendações para o algoritmo utilizado GARP *Genetic Algorithm for Rule Set Production* (STOCKWELL e NOBLE 1992) para se conseguir um bom resultado com a modelagem.

Embora sejam típicas de cerrado, essas espécies são eventualmente registradas em áreas ecotonais ou de contato entre o Cerrado e a Floresta Estacional Semidecidual.

Parâmetros ambientais utilizados para a modelagem

- Dados climáticos*: Worldclim* version 1.1 *Global Cimate Surface* [10]. Resolução 0.83km.
  - ⇒ precipitação mensal (12 camadas).
  - ⇒ temperaturas máximas mensais (12 camadas).
  - ⇒ temperaturas médias mensais (12 camadas).
  - ⇒ temperatura mínimas mensais (12 camadas).
- Dados topográficos: U.S. *Geological Surveys* [11]. Resolução de 1km.
  - ⇒ elevação (1 camada).
  - ⇒ relevo (1 camada).
  - ⇒ aspecto do terreno (1 camada).
  - ⇒ capacidade de retenção de umidade (1 camada).
- Mapa de Solos do Estado de São Paulo - Instituto Agronômico de Campinas. Escala 1:500.000 (1 camada) (OLIVEIRA et al. 1999).

O mapeamento pedológico apresenta os dados categorizados e não contínuos. Quando os dados são divididos em classes, os valores originais são perdidos, o que para efeito de modelagem é bastante limitante, mas, devido à grande importância dos solos na distribuição de espécies vegetais, esta variável foi incluída nas análises de modelagem deste estudo.

Método de Modelagem

Neste trabalho foi usado o algoritmo genético GARP (*Genetic Algorithm for Rule Set Production*) (STOCKWELL e NOBLE 1992). Este algoritmo permite prever a distribuição geográfica potencial a partir de modelos de nicho fundamental de espécies (veja a introdução geral e anexo 1, para maiores detalhes). Os resultados da modelagem foram visualizados e analisados através de um Sistema de Informação Geográfica (ArcView 8.2 - ESRI).

Para efeito de comparação da eficácia do modelo em predizer áreas de cerrado no Estado de São Paulo, foram realizados dois experimentos.

<u>Experimento A</u>: modelagem utlizando apenas pontos de ocorrência das espécies de cerrado, em fragmentos de Cerrado.

---

[10] http://biogeo.berkeley.edu/worldclim/worldclim.htm

Experimento B: modelagem utilizando pontos de ocorrência das mesmas espécies de cerrado presentes também em áreas ecotonais e bordas de áreas florestais, além das áreas de Cerrado.

Para realizar os experimentos, os conjuntos de dados biológicos disponíveis foram divididos em 50% para pontos de treino (Tabela 1) e 50% para pontos de teste para cada espécie modelada. Este teste é externo e não deve ser confundido com a divisão de pontos descrito na Introdução Geral que é interno, ou seja, realizado pelo algoritmo (ver Anexo 1). No caso do teste externo, os pontos de teste são utilizados apenas para calcular a porcentagem de acerto de cada espécie, não participando portanto da geração do modelo.

Este teste é feito através da comparação célula a célula entre a matriz resultante da soma de modelos e a planilha de dados de campo para a área de estudo. Esta comparação mostra o quanto o modelo acertou ou errou na previsão. A contagem é feita a partir dos números de teste da seguinte forma:

verdadeiro: quando o modelo e a planilha de campo dizem que a espécie ocorre ou quando ambas dizem que não ocorre.

falso: quando o modelo diz que a espécie ocorre e a planilha de campo diz não, ou vice-versa.

n/a: quando o campo diz que a espécie ocorre, mas trata-se de um ponto de treino (ou seja, utilizado na modelagem). Neste caso este dado não é contabilizado, não se aplica.

A porcentagem de acerto dá uma idéia de quanto o modelo é capaz de acertar a previsão de ocorrência (presença) ou de não ocorrência (ausência) para cada espécie. Considerando que todas as espécies utilizadas são bastante abundantes, ou seja, ocorrem com vários indivíduos no campo, foi considerada ausência da espécie quando nenhum indivíduo da espécie foi encontrado no fragmento aplicando-se o método expedito de levantamento botânico.

Na Tabela 1 são apresentadas as 17 espécies lenhosas utilizadas na modelagem, os números de treino e as porcentagens de acerto para cada espécie.

---

[11] http://edcdaac.usgs.gov/gtopo30/hydro/

Tabela 1: Espécies utilizadas na modelagem da distribuição potencial do bioma Cerrado no Estado de São Paulo e o número de pontos de treino utilizados para gerar os modelos nos experimentos A e B. **vs** = verdadeiros; **fs** = falsos

		Experimento A					**Experimento B**				
Especies	Familias	**% vs**	% fs	N vs	N fs	**N treino**	**% vs**	% fs	N vs	N fs	**N treino**
Byrsonima intermedia A. Juss.	MALPIGHIACEAE	86	13	33	5	33	81	18	39	9	38
Gochnatia barrosii Cabrera	ASTERACEAE	83	16	30	6	32	79	20	38	10	38
Stryphnodendron obovatum Benth.	MIMOSACEAE	84	15	33	6	29	80	19	42	10	34
Schefflera vinosa (Cham. & Schltdl.) Frodin	ARALIACEAE	82	17	33	7	29	69	30	39	17	30
Anadenanthera falcata (Benth.) Speg.	MIMOSACEAE	79	20	35	9	25	70	29	40	17	29
Aegiphila lhotszkiana L.	VERBENACEAE	59	40	26	18	26	53	46	31	27	28
Brosimum gaudichaudii Trécul.	MORACEAE	65	34	29	15	25	58	41	34	24	28
Dimorphandra mollis Benth.	CAESALPINIACEAE	60	40	27	18	24	51	48	31	29	26
Duguetia furfuracea (A. St. -Hil.) Benth. & Hook f.	ANNONACEAE	60	40	27	18	23	52	47	32	29	26
Miconia stenostachya DC.	MELASTOMATACEAE	65	34	30	16	25	60	40	36	24	26
Qualea grandiflora Mart.	VOCHYSIACEAE	60	39	28	18	24	51	48	31	29	26
Annona coriacea Mart.	ANNONACEAE	52	47	24	22	22	43	56	27	35	25
Campomanesia adamantium (Cambess.) O. Berg	MYRTACEAE	54	45	25	21	24	45	54	28	33	25
Caryocar brasiliense Cambess.	CARYOCARACEAE	56	43	26	20	23	49	50	30	31	25
Ouratea spectabilis (Mart.) Engl.	OCHNACEAE	54	45	25	21	24	44	55	27	34	25
Tocoyena formosa (Cham. & Schltdl.) K. Schum.	RUBIACEAE	56	43	25	19	25	45	54	28	33	25
Annona dioica A. St. -Hil.	ANNONACEAE	65	34	30	16	22	59	40	37	25	24

Para um melhor resultado são gerados vários modelos. No presente estudo foram gerados 100 modelos por espécie, dos quais foram escolhidos os dez melhores modelos (ANDERSON et al. 2003) que, somados (sobrepostos utilizando-se um SIG), compuseram o modelo de distribuição potencial de cada espécie. Posteriormente os modelos gerados para cada espécie foram somados para produzir os mapas com a distribuição potencial das áreas de Cerrado.

Resultados e Discussão

Uma grande crítica ao uso de algoritmos genéticos é a falta de convergência dos resultados gerados, ou seja, cada vez que o algoritmo é gerado, com o mesmo conjunto de dados, o resultado pode ser diferente. Para saber se ocorreu esse efeito, neste conjunto de dados, foi realizado um teste a partir dos pontos de treino de *Byrsonima intermedia* para os mesmos parâmetros ambientais. Esta espécie foi escolhida por representar o melhor conjunto de dados disponível, dentre as 17 espécies analisadas, tanto em relação ao número de pontos de amostragem como na porcentagem de acerto (Tabela 1).

Testando a convergência do algoritmo

A questão aqui é saber se a falta de convergência, que é uma característica intrínsica de algoritmos genéticos, produz realmente modelos muito distantes (não semelhantes) para o mesmo conjunto de dados (bióticos e abióticos). Cada modelo gerado é comparado, célula a célula, com outro. Somando-se os modelos, é feita nova comparação, novamente célula a célula, para se calcular o quanto um modelo único, ou um modelo resultante de soma, é distante do outro. O resultado desta comparação é apresentado nas Figuras 10 e 11.

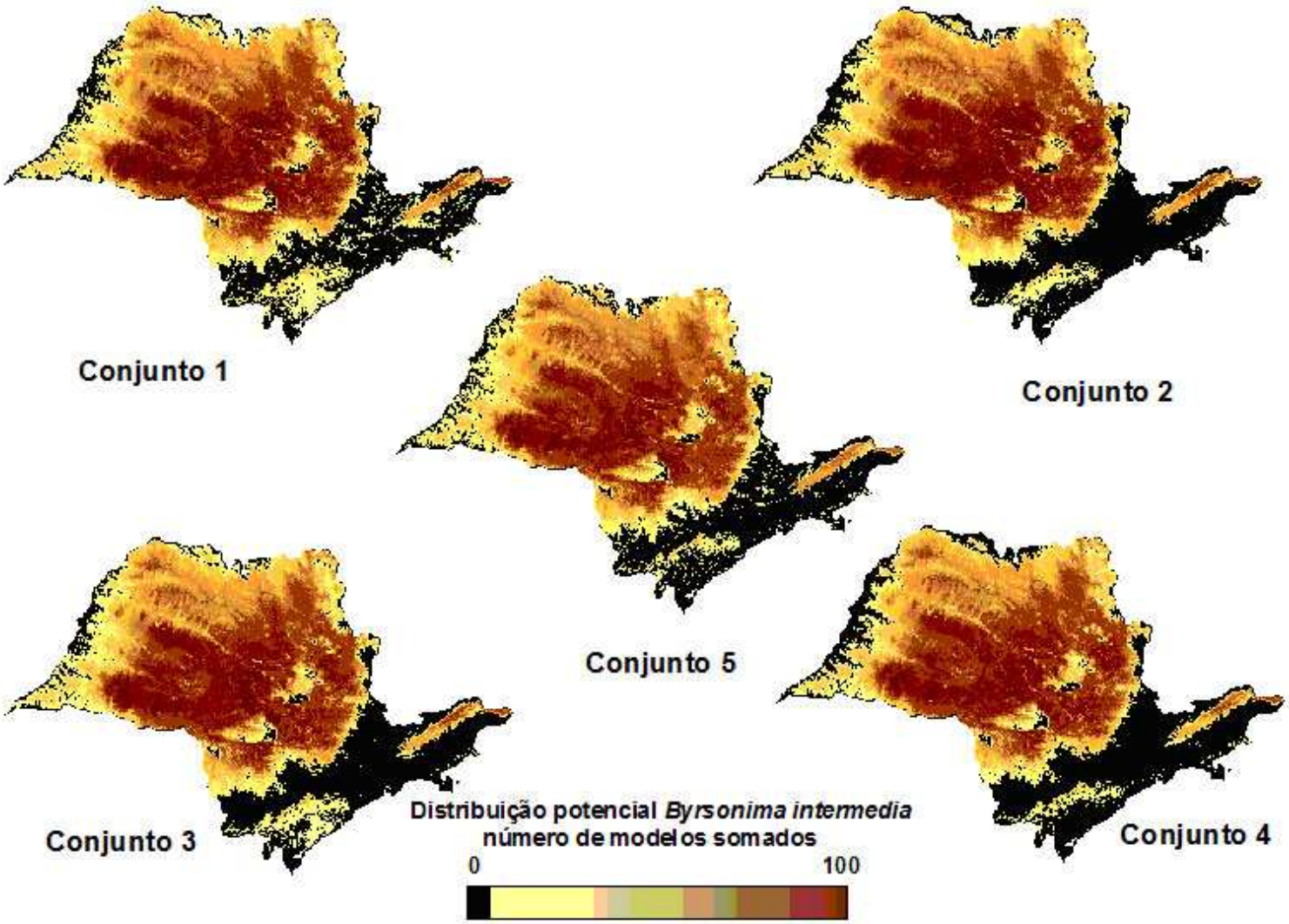


Figura 10: Resultado da modelagem da área potencial de distribuição para o mesmo conjunto de dados de ocorrência de *Byrsonima intermedia*. Cada mapa é o resultado de um conjunto da soma de 100 modelos gerados pelo GARP.

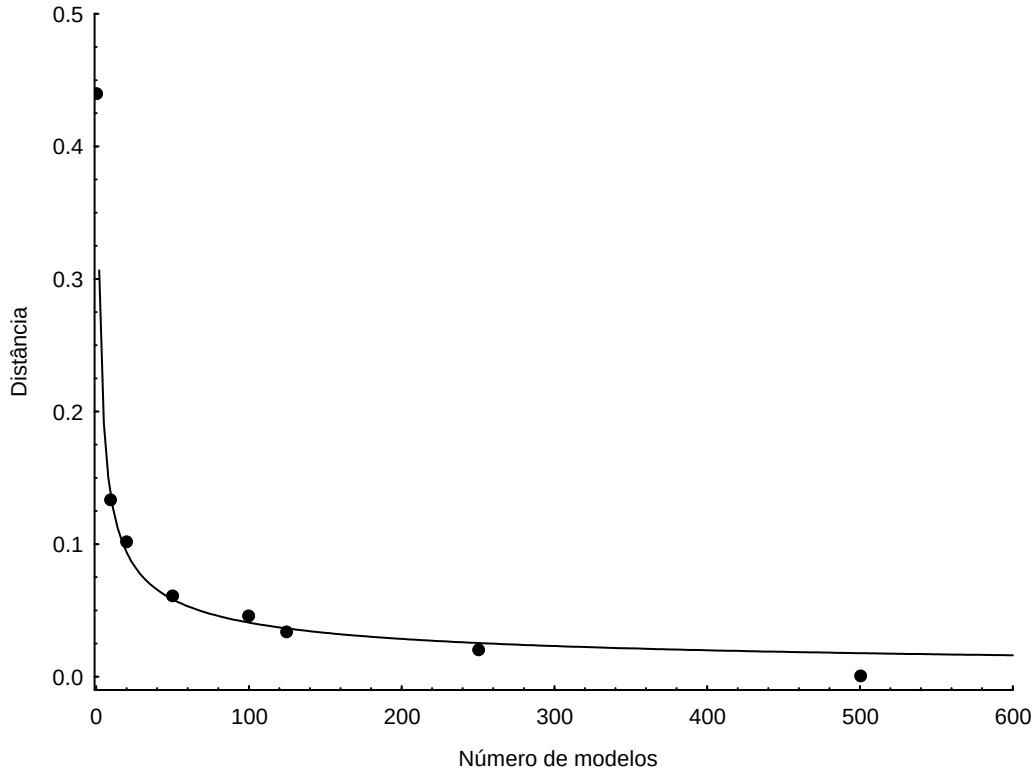


Figura 11: Média das distâncias entre modelos baseados na ocorrência de *Byrsonima intermedia* em função do número de modelos somados. A linha representa o ajuste produzido por técnica de ajuste não-linear a um modelo de função potencia. A Equação da curva é $D = 0{,}441N^{-0{,}517}$; onde *D* é a distância e *N* o número de modelos.

O resultado mostra que os modelos resultantes para os cinco conjuntos de dados, gerados a partir da soma de 100 modelos, são bastante semelhantes (Figura 12). Além disso, considerou-se mais eficiente a utilização de 100 modelos, uma vez que rodar mais modelos exigiria um tempo de processamento bem maior sem que isso trouxesse um resultado significativamente melhor para os modelos somados resultantes (Figura 13).

Avaliação da especificidade ambiental dos pontos de amostragem no resultado da modelagem para a área de Cerrado do Estado de São Paulo

Na figura 14 é apresentado o resultado da modelagem obtida nos dois experimentos. A modelagem utilizando apenas localidades de cerrado (experimento A) mostrou uma porcentagem de acerto maior que a modelagem realizada a partir de pontos com fisionomias de outros biomas (experimento B), apesar do maior número de pontos deste segundo experimento (Figura 12 e

Tabela 1). Observando-se as fisionomias presentes em cada fragmento verifica-se que as áreas com baixa porcentagem de acerto são predominantemente compostas por fragmentos exclusivos de ecótono ou de Floresta Estacional Semidecidual. Por outro lado, nos fragmentos que apresentaram as maiores porcentagens de acerto há predomínio de fisionomias exclusivas de Cerrado.

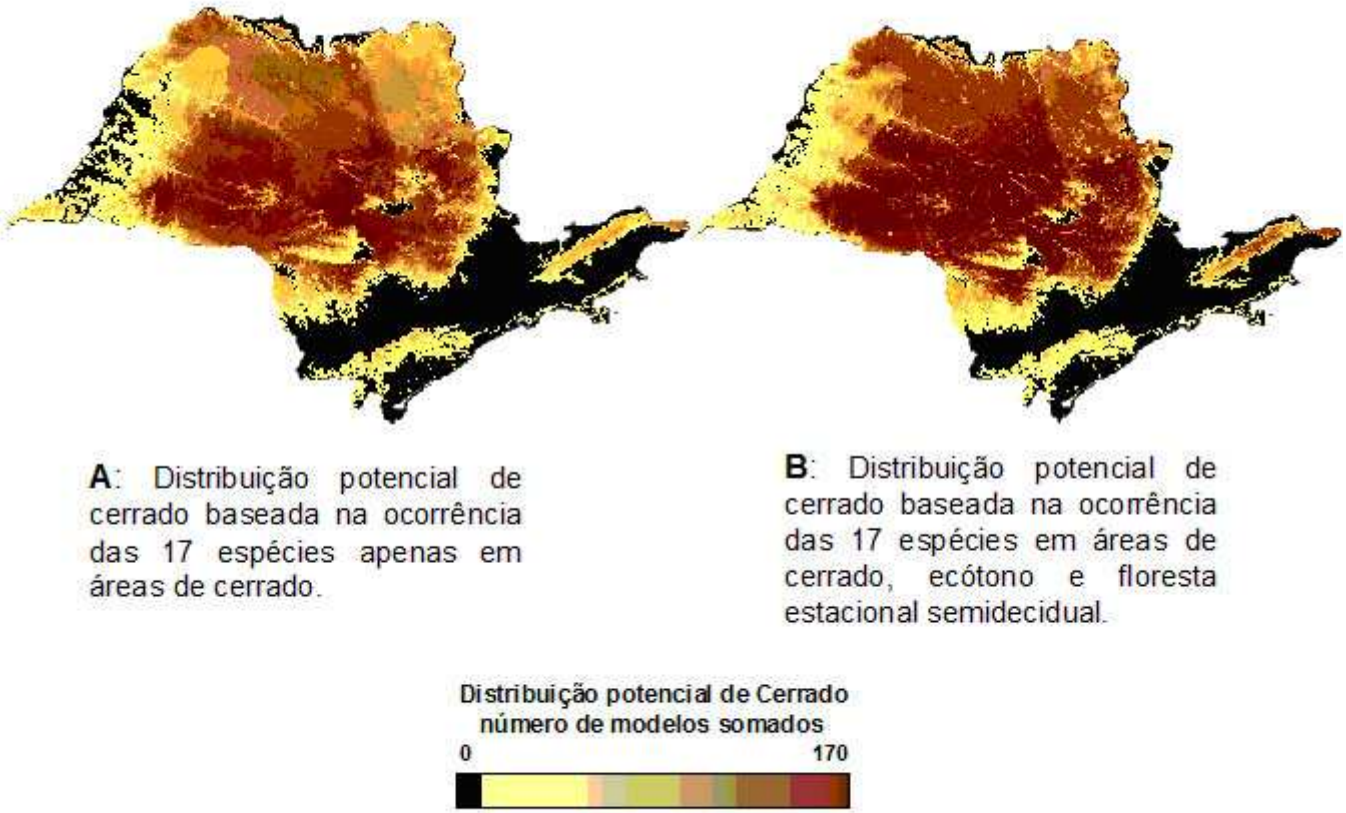


Figura 12: Comparação entre as distribuições potenciais do Bioma Cerrado no Estado de São Paulo com base na ocorrência das 17 espécies apenas em áreas de cerrado (A) e em Cerrado e outros tipos de vegetação (B). O gradiente de cor representa o número de modelos somados (os 10 melhores modelos de 100 gerados para cada espécie), quanto mais escura a cor, mais modelos previram a ocorrência do conjunto de espécies analisado.

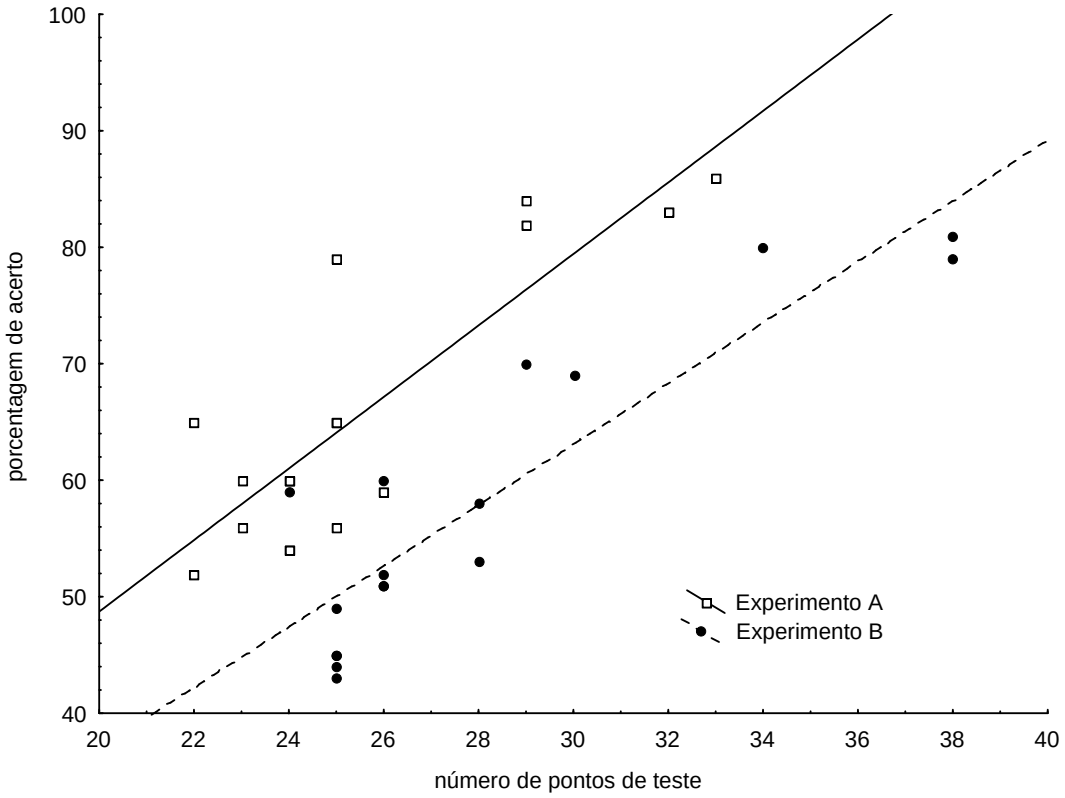


Figura 13: Comparação entre a porcentagem de acerto de dois conjuntos de pontos de treino para a modelagem do bioma cerrado em São Paulo. Experimento A: localidades de amostragem contendo apenas áreas de cerrado. Experimento B: localidades de cerrado, floresta estacional e vegetação de transição entre as duas (região de ecótono cerrado/floresta).

Para um mesmo número de pontos de treino e teste, o experimento A foi sempre o que apresentou maior porcentagem de acerto (Figura 13), para todas as espécies. Essas diferenças não são explicadas pelo acaso (Tabela 2) e representam um efeito consistente da escolha dos pontos sobre a eficácia do modelo.

Tabela 2: Análise de covariância para a porcentagem de acerto nos experimentos A e B e o efeito do número de pontos de teste. **gl trat** = graus de liberdade do tratamento; **gl erro =** graus de liberdade do erro; **P** = probabilidade do resultado ser dada ao acaso.

Comparação	**F**	**gl trat**	**gl erro**	**P**
Diferença de inclinação	0.566	1	30	0.458
Efeito do tipo de experimento	40.783	1	30	**<0.001**
Efeito do número de amostras	92.257	1	30	**<0.001**

Esses resultados mostram que a eficácia do modelo aumenta se forem utilizadas localidades em que o bioma é único, sem fisionomias de biomas vizinhos. Quando as espécies têm distribuição mais ampla, ocorrendo também em bordas de fragmentos pertencentes a outras formações vegetais, essas ocorrências podem diminuir a porcentagem de acerto do modelo resultante. Isso provavelmente deve estar relacionado ao número desigual de pontos provenientes de diferentes biomas. No caso do experimento A foram incluídas 69 áreas, sendo 66 áreas de Cerrado e três áreas contendo fisionomias tanto de Cerrado, como de Ecótono e Floresta Estacional. No experimento B foram incluídas todas as 86 áreas, sendo 66 áreas com vegetação típica de cerrado, 11 áreas com vegetação ecotonal, seis áreas de Floresta Estacional Semidecidual e as três áreas contendo as fisionomias de Cerrado, Ecótono e Floresta Estacional. Ou seja, no segundo caso (experimento B) o algoritmo tem poucos pontos de amostragem de floresta estacional e ecótono em comparação com as áreas de Cerrado. Portanto, é de se esperar uma capacidade mais baixa na porcentagem de acerto do modelo.

Por outro lado, se o que se busca é um modelo de distribuição de determinada espécie mais inclusivo, com o objetivo de definir os locais onde a espécie tem condições de sobreviver, então é aconselhável utilizar todos os registros disponíveis de ocorrência daquela espécie, dentro da área de estudo.

Distribuição potencial do Bioma Cerrado no Estado de São Paulo.

A distribuição potencial do Cerrado gerada pela modelagem resulta em uma área bem maior que a área mapeada pelo IBGE (1988) (Figura 14). Isso também era verdadeiro no início dos anos 70, como sugere a revisão de Eiten (1972). A subestimativa da distribuição do cerrado foi particularmente severa em relação à região oeste do Estado. O autor cita especificamente a presença de importantes áreas de Cerrado na região de Paraguaçu Paulista e Rancharia (sudoeste do Estado), no nordeste de São José do Rio Preto, na região do Rio Grande (norte do Estado), em várias áreas de cerradão no noroeste do Estado e também muitas áreas menores dispersas na região oeste que não eram incorporadas nos mapeamentos de Cerrado. Ainda considerando o Estado de

São Paulo, para Eiten (1972) o melhor mapeamento do Cerrado foi feito por Borgonovi e Chiarini (1965) (Figura 5), mas mesmo esses autores deixaram de incluir algumas áreas, novamente as áreas do nordeste de São José do Rio Preto e uma área grande em São Carlos.

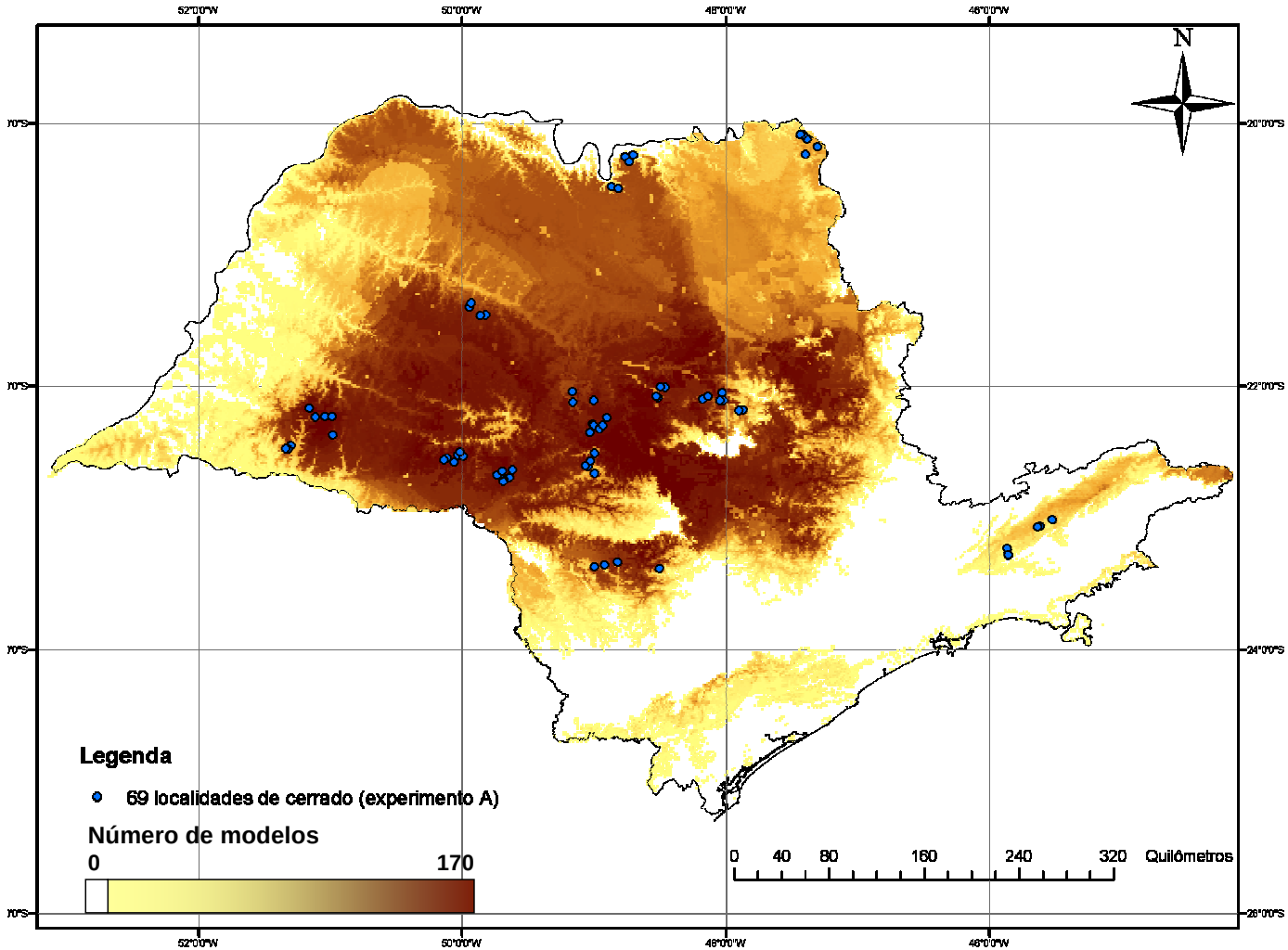


Figura 14: Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (entre os 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP, a partir das 69 localidades exclusivas de Cerrado amostradas no escopo do projeto “Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo”.

Tanto o modelo gerado, como o recente mapeamento do Instituto Florestal apresentados nas Figuras 8, 9 e 14, corroboram a idéia de subamostragem, mostrando uma área de cerrado potencialmente maior. Os fragmentos atuais de cerrado podem ser encontrados tanto nas áreas mapeadas como Floresta Estacional Semidecidual como nas áreas de vegetação ecotonal.

O que ocorre na prática é que é bastante difícil estabelecer os limites entre o Cerrado e a Floresta Estacional no Estado. A região ecotonal é bastante descontínua, formando um mosaico de vegetação, ocorrendo fragmentos de Cerrado e fragmentos de Floresta Estacional e extensas áreas de vegetação de transição. Segundo Oliveira-Filho e Ratter (2002) onde os solos são mais férteis, a vegetação climax é definitivamente floresta mesofítica (incluindo florestas decíduas e semidecíduas), no caso de São Paulo, floresta estacional semidecidual.

Para se observar a influência do solo no padrão de distribuição do Cerrado, a Figura 15 apresenta uma comparação entre o mapa potencial de distribuição de Cerrado no Estado, os remanescentes de vegetação do Cerrado (Instituto Florestal de São Paulo 2004) e um recorte da distribuição potencial de Cerrado baseado nos sub-tipos de solos nos quais as localidades amostradas se encontram. As 69 localidades amostradas apresentaram 19 diferentes sub-tipos de solo (Tabela 3). A união desses 19 sub-tipos de solos é apresentada na Figura 15 como um recorte formando uma potencial área de distribuição de Cerrado.

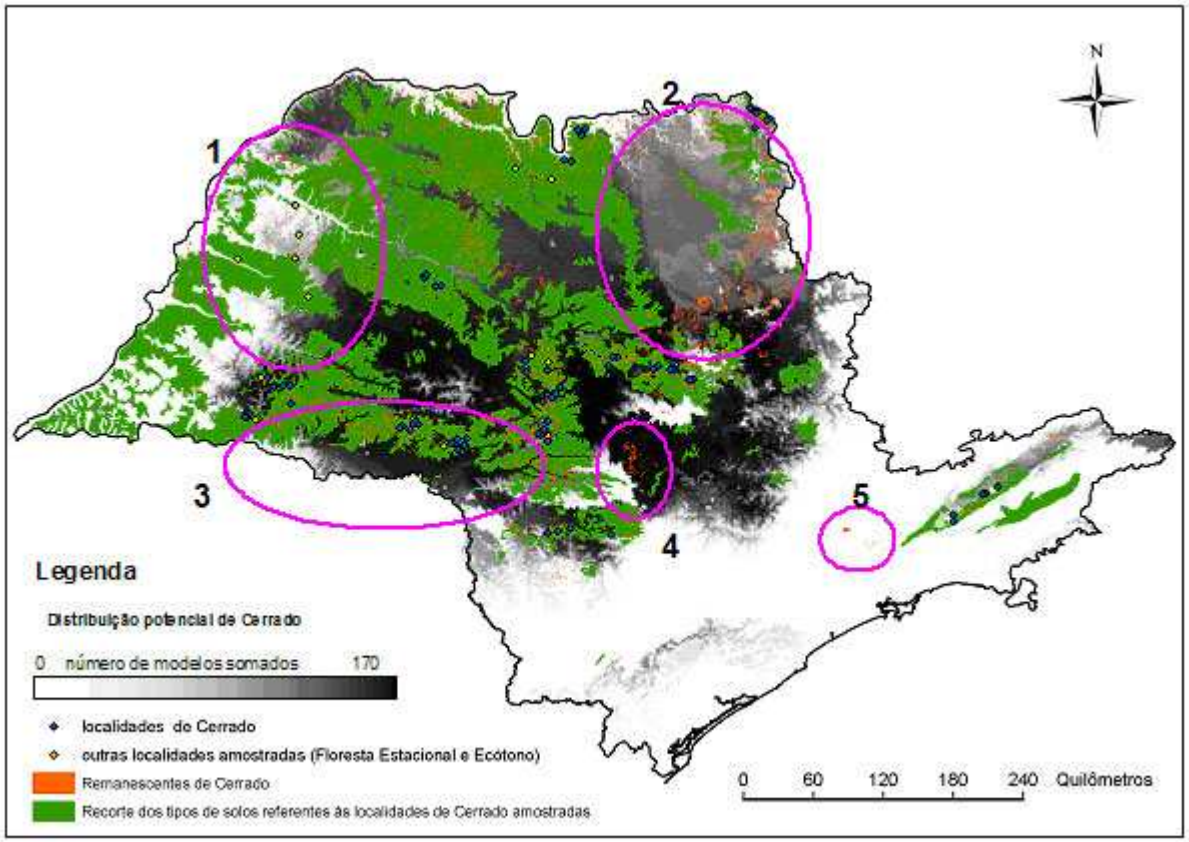


Figura 15: Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (dos 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP e a distribuição dos Remanecentes de Cerrado no Estado de São Paulo (Instituto Florestal de São Paulo 2004) com um recorte dos sub-tipos de solos nos quais as localidades amostradas de Cerrado ocorreram (localidades amostradas de Cerrado, Ecótono e Floresta Estacional Semidecidual no escopo do projeto “Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo”).

As áreas demarcadas na Figura 15 evidenciam as seguintes situações:

1. Dentro desta área, tanto o recorte de solos como o modelo de distribuição potencial de Cerrado (embora com baixa probabilidade) mostram algumas regiões com ocorrência potencial de Cerrado, enquanto as localidades amostradas são de Floresta Estacional Semidecidual ou Ecótono. Já os remanescentes de Cerrado estão dispersos nas duas regiões.
2. O recorte de solos é pouco representado enquanto o modelo prevê a ocorrência de Cerrado para esta região que abriga vários remanescentes de Cerrado.
3. O recorte de solos exclui a área ao sul enquanto o modelo prevê a distribuição de Cerrado para esta região, na qual não estão presentes os remanescentes de Cerrado.
4. Esta área apresenta praticamente ausência de recorte de solos e uma alta probabilidade de ocorrência de Cerrado indicada pelo modelo, na qual estão presentes vários remanescentes de Cerrado.
5. Esta área apresenta alguns remanescentes de Cerrado mas nem o modelo, nem o recorte de solos foram capazes de prever estas ocorrências.

Esses resultados parecem evidenciar alguns pontos, tais como:

A área 1 parece abranger um mosaico de vegetação, com áreas de transição entre o Cerrado e a Floresta Estacional. As áreas 2 e 4 evidenciam uma melhor capacidade de previsão do modelo em relação ao recorte de solos.

A área 3 evidencia uma melhor capacidade do recorte de solos em predizer a não ocorrência de Cerrado em relação ao modelo. A razão disso parece ser que apesar das condições climáticas e topográficas serem adequadas para a ocorrência de Cerrado, a região apresenta Latossolo Roxo e Vermelho Escuro, Eutroficos e Distróficos, textura argilosa ((OLIVEIRA et al. 1999) (maiores detalhes sobre os solos dessa região são apresentados no Capítulo 2), que são terras bastante férteis, portanto, sendo mais propícias ao desenvolvimento de florestas. Os fragmentos de Cerrado ocorrentes dentro da área 5 pertencem a dois municípios, Franco da Rocha e Guarulhos, ambos com solos do tipo Podzólicos Vermelho Amarelo e Latossolos Vermelho Amarelos, que são os principais tipos de solos para ocorrência de Cerrado, segundo a Tabela 3; porém, são de sub-tipos diferentes daqueles apresentados pelas localidades amostradas. Sendo assim, o recorte de solos não foi amplo o bastante para incluir essas regiões.

Em uma tentativa de observar o quão inclusivo pode ser o recorte de solos, foi elaborado um recorte mais genérico (Figura 16), incluindo todos os sub-tipos de solos do tipo LATOSSOLOS VERMELHOS e Podzólicos Vermelho-Amarelos, os mais importantes para

distribuição do Cerrado (Tabela 3). O resultado é um recorte de solos ainda mais inclusivo (Figura 16) que o modelo gerado (Figura 14).

Tabela 3: Lista dos diferentes tipos de solos onde ocorreram as localidades amostradas de Cerrado (OLIVEIRA et al. 1999).

Tipos de Solos	**Código da legenda**	**Legenda dos sub-tipos de solos**	**Nº de fragmentos por tipo de solo**
LATOSSOLOS VERMELHOS (LV) Latossolos Roxos + Latossolos Vermelho-Escuros	**LV78**	Distróficos A moderado textura média, relevo plano + ARGISSOLOS VERMELHO-AMARELOS e VERMELHOS ambos Eutróficos e Distróficos A moderado textura arenosa/média e média, relevo suave ondulado.	19
	LV45	Distróficos A moderado textura média, relevo plano e suave ondulado.	12
ARGISSOLOS VERMELHO-AMARELOS (PVA) Podzólicos Vermelho-Amarelos	**PVA10**	Eutróficos + ARGISSOLOS VERMELHOS Distróficos e Eutróficos ambos textura arenosa/média e média, relevo suave ondulado + LATOSSOLOS VERMELHOS Distróficos text. média relevo. plano todos A moderado.	12
LATOSSOLOS VERMELHOS (LV) Latossolos Roxos + Latossolos Vermelho-Escuros	**LV56**	Distróficos + LATOSSOLOS VERMELHOS-AMARELOS Distróficos ambos A moderado text. média relevo. Plano e suave ondulado.	4
LATOSSOLOS VERMELHO-AMARELOS (LVA) Latossolos Vermelho-Amarelos	**LVA13**	Distróficos textura média + NEOSSOLOS QUARTZARÊNICOS Órticos distróficos ambos A moderado, relevo suave ondulado e plano.	4
LATOSSOLOS VERMELHOS (LV) Latossolos Roxos + Latossolos Vermelho-Escuros	**LV48**	Distróficos textura argilosa, relevo suave ondulado + CAMBISSOLOS HÁPLICOS Distróficos textura média e argilosa, relevo ondulado ambos A moderado.	3
LATOSSOLOS VERMELHO-AMARELOS (LVA) Latossolos Vermelho-Amarelos	**LVA40**	Distróficos, relevo suave ondulado e plano + ARGISSOLOS VERMELHO-AMARELOS Distróficos latossólicos, relevo suave ondulado ambos A moderado textura. argilosa.	2
ARGISSOLOS VERMELHO-AMARELOS (PVA) Podzólicos Vermelho-Amarelos	**PVA65**	Distróficos textura média/argilosa, relevo forte ondulado e ondulado + ARGISSOLOS VERMELHO-AMARELOS Distróficos latossólicos textura argilosa, relevo ondulado ambos A moderado.	2
LATOSSOLOS AMARELOS (LA)	**LA6**	Distróficos, relevo suave ondulado e plano + ARGISSOLOS VERMELHO-AMARELOS Distróficos latossólicos, relevo suave ondulado ambos A moderado textura argilosa.	1
LATOSSOLOS VERMELHOS (LV) Latossolos Roxos + Latossolos Vermelho-Escuros	**LV74**	Distroférricos textura argilosa e Distróficos textura média + LATOSSOLOS VERMELHO-AMARELOS Distróficos textura média + LATOSSOLOS VERMELHOS Distroférricos textura argilosa todos A moderado, relevo suave ondulado.	1

	LV8	Eutroférricos e Distroférricos textura argilosa + LATOSSOLOS VERMELHOS Distróficos todos A moderado, relevo suave ondulado + NEOSSOLOS LITÓLICOS Eutróficos e Distróficos A moderado e chernozêmico textura argilosa, relevo ondulado.	1
LATOSSOLOS VERMELHO-AMARELOS (LVA) Latossolos Vermelho-Amarelos	**LVA12**	Distróficos text. média + NEOSSOLOS QUARTZARÊNICOS Órticos distróficos ambos A moderado relevo. suave ondulado.	1
	LVA15	Distróficos text. argilosa + CAMBISSOLOS HÁPLICOS Distróficos text. argilosa e média ambos A moderado e proeminente relevo. forte ondulado.	1
	LVA31	Distróficos A moderado e proeminente textura média + LATOSSOLOS VERMELHOS Distróficos A moderado textura média e argilosa ambos, relevo suave ondulado.	1
	LVA4	Distróficos A moderado textura média, relevo suave ondulado.	1
	LVA51	Distróficos + LATOSSOLOS VERMELHOS Distróficos ambos textura média + NEOSSOLOS QUARTZARÊNICOS órticos distróficos todos A moderado, relevo suave ondulado.	1
NEOSSOLOS LITÓLICOS (RL) Solos Litólicos	**RL11**	Distróficos + CAMBISSOLOS HÁPLICOS Distróficos ambos A moderado textura média fase pedregosa, relevo forte ondulado e montanhoso.	1
NEOSSOLOS QUARTZARÊNICOS (RQ) Arelevoas Quartzosas, Arelevoas Quartzosas Hidromórficas	**RQ1**	órticos Distróficos A moderado, relevo suave ondulado e plano.	1
	RQ2	órticos + LATOSSOLOS VERMELHO-AMARELOS textura média ambos Distróficos A moderado, relevo suave ondulado.	1
Total			19

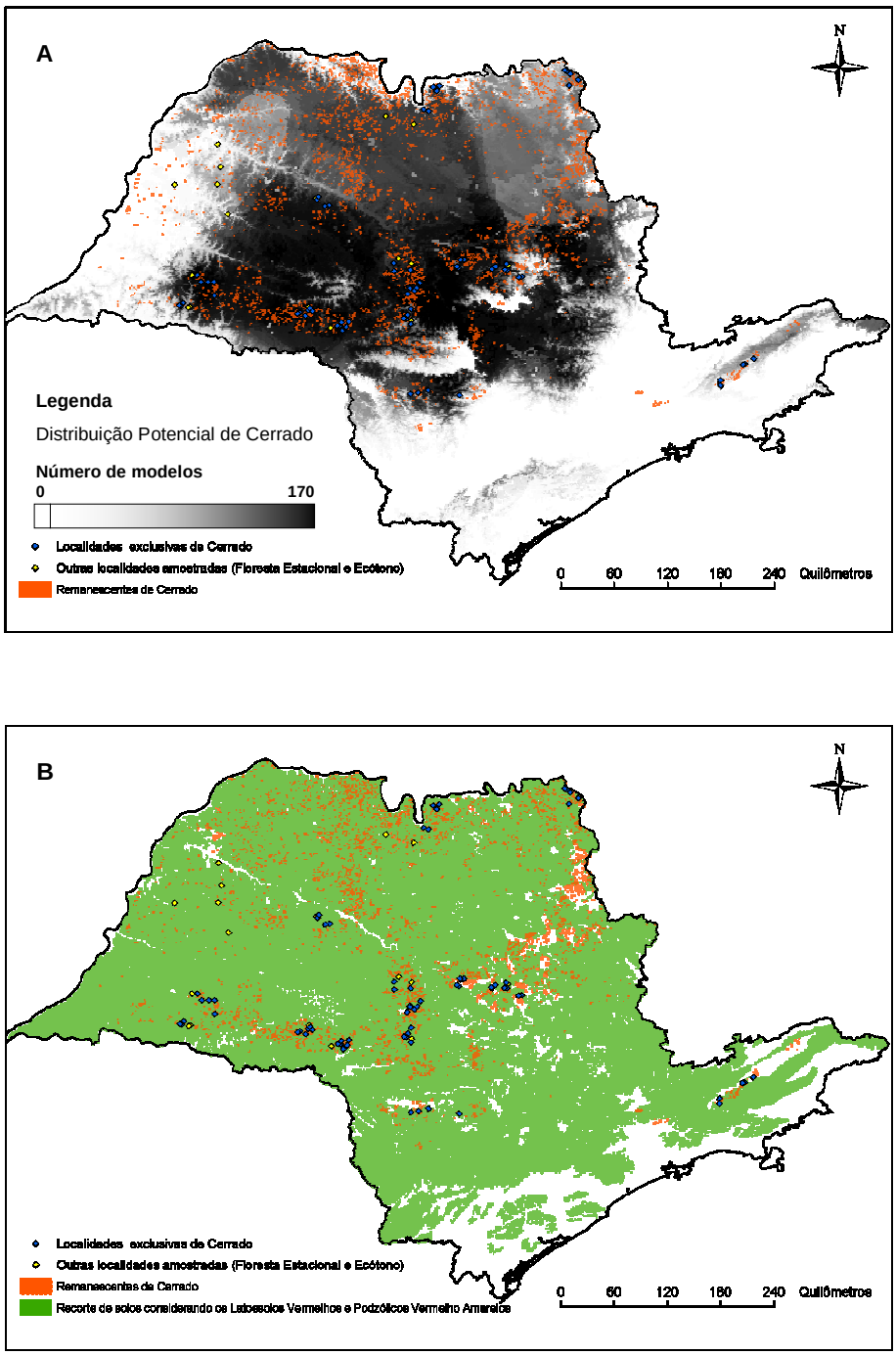


Figura 16: A - Distribuição potencial do Bioma Cerrado para o Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos (dos 100 gerados) para 17 espécies de Cerrado, utilizando o algoritmo GARP e a distribuição dos Remanecentes de Cerrado no Estado de São Paulo (Instituto Florestal de São Paulo 2004). B - Recorte das áreas de solo tipo Latossolos Vermelhos e Argissolos Vermelho Amarelos (localidades amostradas de Cerrado, Ecótono e Floresta Estacional Semidecidual no escopo do projeto "Viabilidade de Conservação dos remanescentes de Cerrado do Estado de São Paulo").

Esses resultados sugerem que para a escala geográfica analisada, dados climáticos, topográficos e dados de solos (mesmo categorizados) geram um bom modelo de distribuição do Bioma Cerrado, mas ainda assim os modelos gerados apresentam problemas, ou seja, o modelo não garante sua precisão em todas as situações. Por outro lado, apenas um recorte de sub-tipos de solos pode deixar de incluir áreas importantes e a generalização de tipos de solos gera um mapa com maiores áreas de sobreprevisão (quando o modelo diz que o elemento modelado deve ocorrer mas que na realidade não ocorre) se comparado com o modelo. Talvez, com a inclusão de dados contínuos de solos (riqueza, granulometria, capacidade de absorção de água etc.) na modelagem, a distinção entre a distribuição do Cerrado e da Floresta Estacional Semidecidual possa ser melhor mapeada pelas técnicas utilizadas.

Em relação às 17 espécies modeladas, com base nos resultados obtidos (Figuras 17A, 17B e 17C), verifica-se que há diferentes especificidades na distribuição de cada espécie. Esse resultado corrobora aqueles obtidos por Durigan et al. (2003b), que encontraram padrões fitogeográficos distintos para o Estado. Aparentemente, algumas espécies não são sensíveis a variações ambientais dentro do Estado (*Brosimum gaudichaudi, Anadenanthera falcata, Byrsonima intermedia, Duguetia furfuracea*), enquanto outras são restritas a determinadas regiões (*Aegiphila lhotzkyana, Annona dioica, Campomanesia adamantium, Miconia stenostachya* e *Caryocar brasiliesis*).

A junção de todas as 17 espécies em um único modelo resulta em um mapa que é, provavelmente, muito próximo da distribuição original do cerrado em São Paulo (Figura 14).

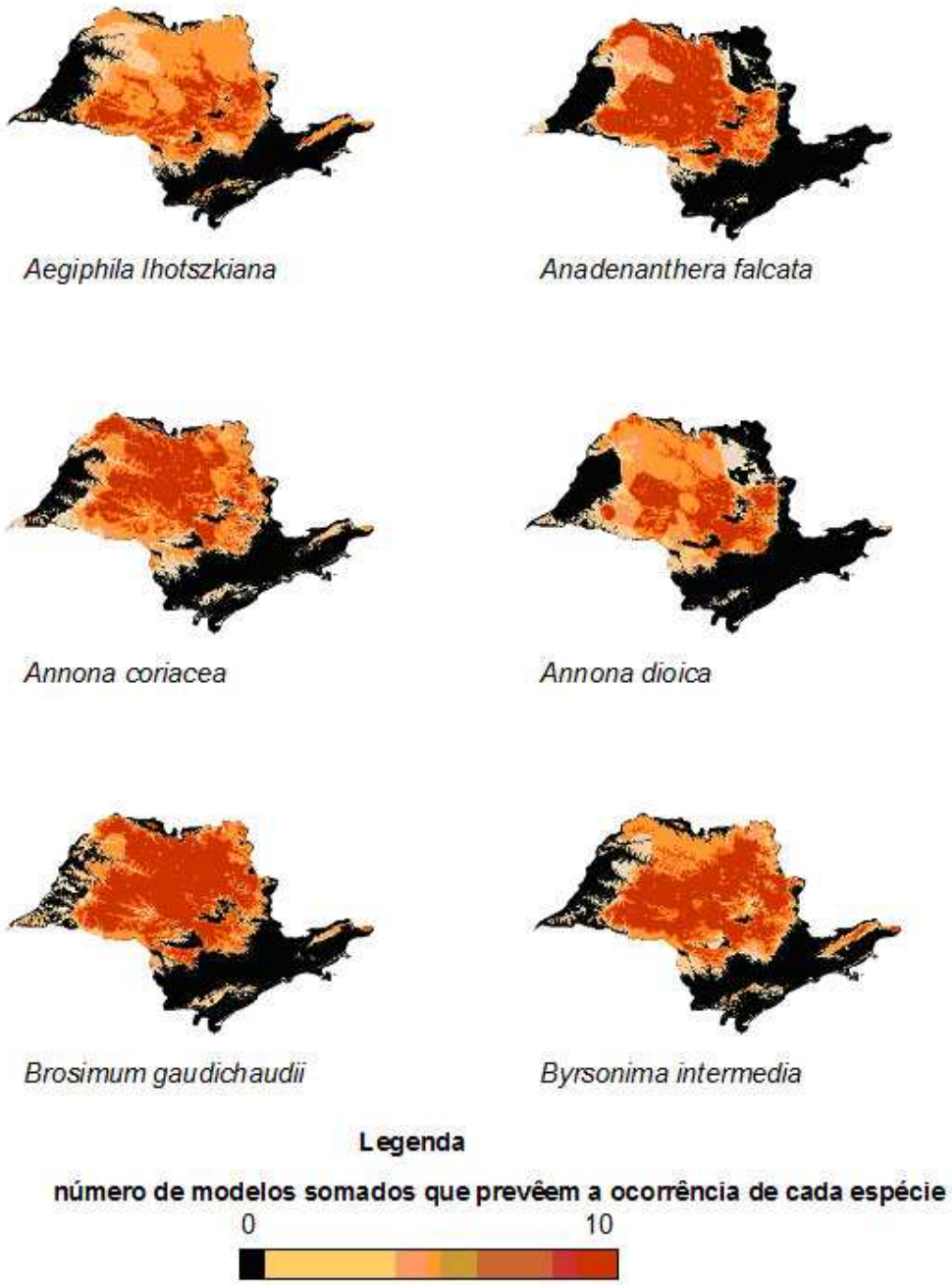


Figura 17A: Distribuição potencial de espécies de cerrado no Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos entre 100 gerados pelo GARP.

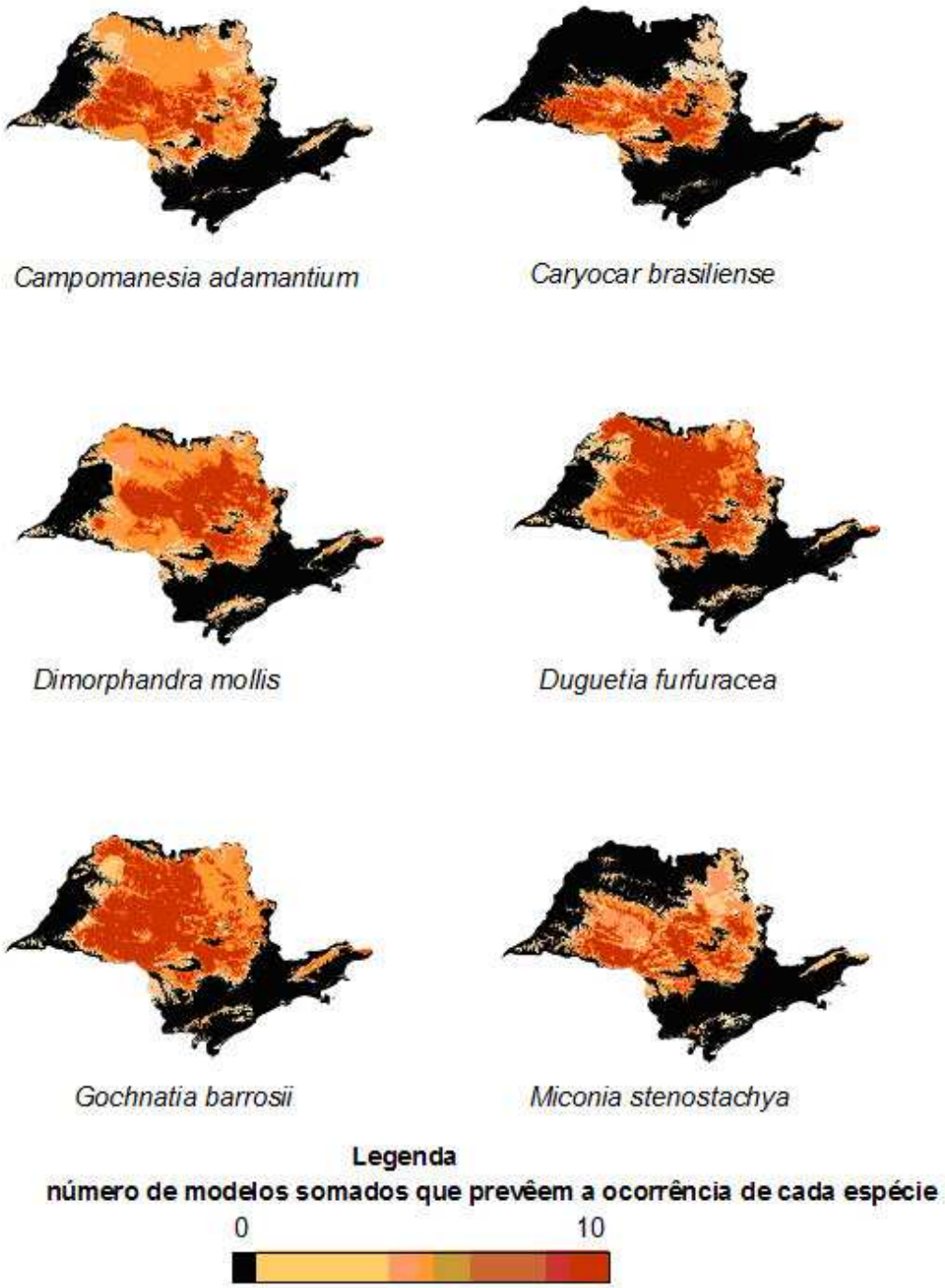


Figura 17B: Distribuição potencial de espécies de cerrado no Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos entre 100 gerados pelo GARP.

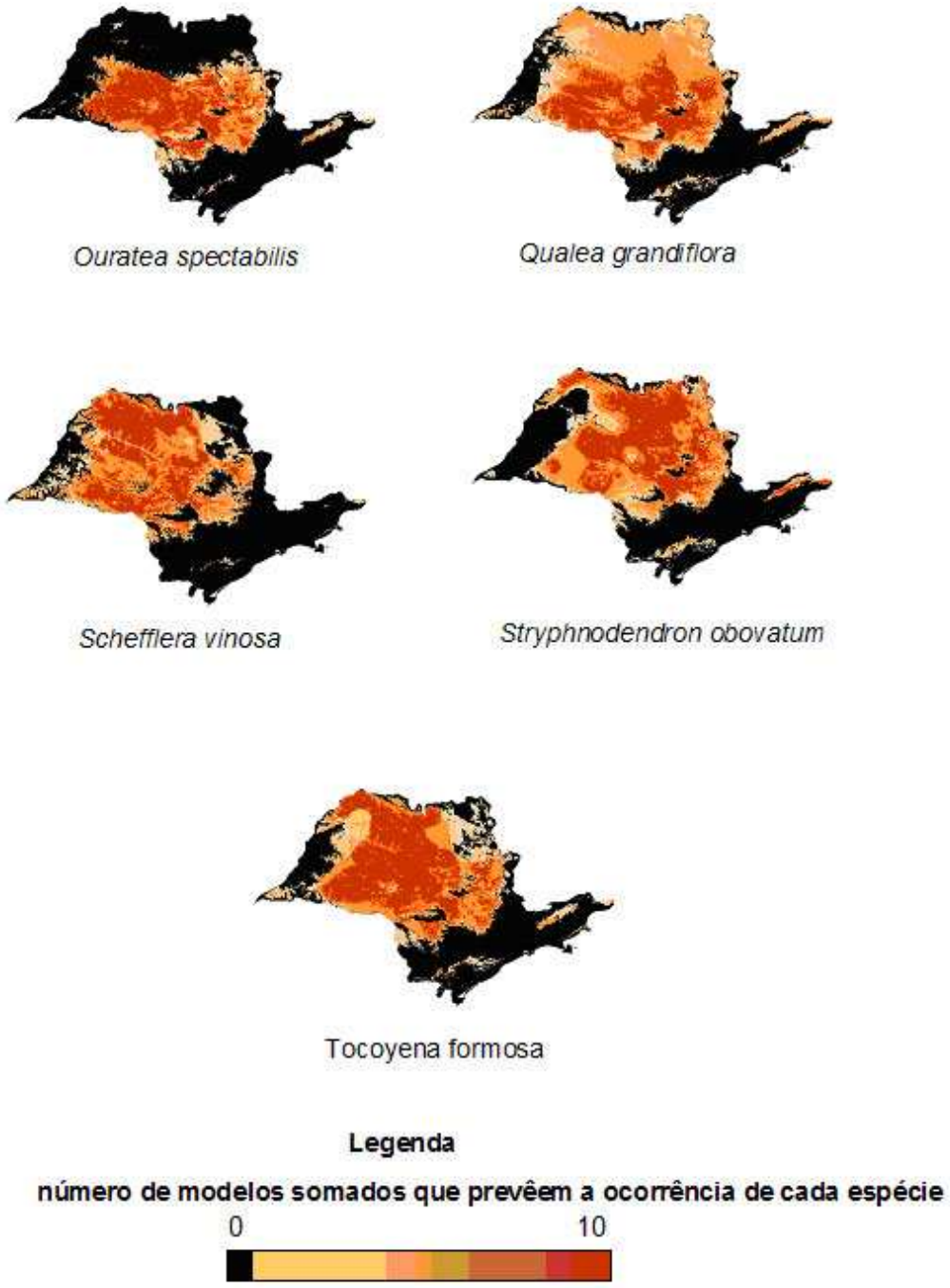


Figura 17C: Distribuição potencial de espécies de cerrado no Estado de São Paulo, obtida a partir da soma dos 10 melhores modelos entre 100 gerados pelo GARP.

Aplicações dos modelos resultantes

O mapa potencial da distribuição de cerrado no Estado de São Paulo (Figura 14) pode servir de base para o desenvolvimento de projetos conservacionistas,

como, por exemplo, para a formação de corredores de vegetação ou áreas de conservação *ex-situ*.

Para as espécies analisadas, os mapas de distribuição potencial de espécies (Figura 17) podem servir de base para definir quais as espécies mais indicadas para recuperação de áreas degradadas dentro do domínio dos Cerrados no Estado. O uso de mapas de modelagem em recuperação de áreas degradadas será abordado mais profundamente no Capítulo 2.

Conclusões

- A metodologia empregada e os resultados obtidos em relação à porcentagem de acerto foram considerados satisfatórios para a geração de modelos de distribuição geográfica de espécies vegetais, baseados em dados abióticos, para a região de estudo.
- A eficácia do modelo em predizer a ocorrência de espécies do cerrado é maior se forem utilizados apenas pontos de amostragem com fisionomias de cerrado, excluindo-se áreas de transição.
- Para minimizar problemas decorrentes da falta de convergência do GARP, considera-se que devem ser gerados no mínimo 100 modelos para cada espécie modelada.
- O uso de modelagem pode auxiliar no entendimento dos padrões de distribuição de um bioma ou ecossistema em uma análise regional, como é o caso da distribuição de cerrado no Estado de São Paulo. Esta área foi sistematicamente subestimada no passado. Segundo a modelagem, as áreas potenciais de ocorrência de espécies de cerrado são mais amplas do que as áreas apontadas pelos mapas existentes.
- Os resultados esperados de modelagem de espécies, em uma escala estadual, deverão ser bastante melhorados com o acréscimo de dados contínuos de solo, como por exemplo, a composição química e a granulometria.
- Diferentes espécies apresentam diferentes padrões de distribuição no Estado, evidenciando as diferentes necessidades ambientais e diferentes fatores limitantes para a distribuição dessas espécies.

Capítulo 2: Modelagem de distribuição geográfica para espécies lenhosas da Bacia do Médio Paranapanema

Introdução

Desenvolvendo estudos na região do Médio Paranapanema, a Fundação de Apoio à Pesquisa Agrícola (FUNDAG) e o Consórcio Intermunicipal do Escritório da Região de Governo de Assis (CIERGA) produziram um mapeamento de dados básicos da região da Bacia compreendendo o diagnóstico completo do meio físico, o potencial de uso sustentável das terras, suas limitações e possibilidades, o uso atual e sua adequação à capacidade de uso, e ainda a caracterização climática da região (FUNDAG 1998). Os dados resultantes deste projeto estão disponíveis no site da CGH-MP[12] (CPTI 2000). Além destes dados, a existência de um conjunto de dados de registros de ocorrências de espécies, tanto das áreas de Cerrado como das áreas de Floresta Estacional e Mata Ciliar, originados de uma parceria com a Estação Ecológica de Assis, constituíram-se na base biótica e abiótica para a geração dos modelos e para as análises pós-modelagem deste estudo. A existência e a disponibilidade de uso desses dados motivaram a escolha desta região para o desenvolvimento de modelagem em escala de Bacia Hidrográfica, visando ao desenvovimento de uma metodologia para auxiliar na recuperação de áreas degradadas.

A integração de dados através de Sistemas de Informações Geográficas (SIG´s) tem contribuído de forma significativa para a coleta de dados, aplicação de modelos, monitoramento e avaliação dos processos de degradação ambiental face à ocupação antrópica e às potencialidades do meio físico (FUNDAG 1998).

Existem poucas informações claras e organizadas, com recomendações sobre a distribuição geográfica e as características ecológicas das espécies arbóreas (DURIGAN et al. 2004b). Considerando-se a importância da preservação e

[12] http://www.sigrh.sp.gov.br/

recuperação da Bacia do Médio Paranapanema, torna-se essencial uma ferramenta que possa dar subsídios para pesquisadores, prefeituras e também para Comitês de Bacias Hidrográficas em questões conservacionistas. Ou seja, a metodologia testada pode servir de base para orientar trabalhos envolvendo a escolha de espécies arbustivo-arbóreas mais adequadas para a recuperação ambiental das diferentes situações ambientais apresentadas na região.

Objetivos

Esse capítulo trata de modelos de distribuição potencial de espécies lenhosas e seu uso em processos de recuperação ambiental na escala de uma bacia hidrográfica, com os seguintes objetivos específicos.

1. Testar os diferentes tipos de bases de dados para distribuição potencial de espécies na escala de uma bacia hidrográfica.
2. Avaliar o efeito do número de amostras na porcentagem de acerto dos modelos.
3. Usar a modelagem de distribuição potencial de espécies vegetais para uma escala local.
4. Aplicar a modelagem de distribuição potencial de espécies e Sistemas de Informação Geográfica no desenvolvimento de uma metodologia de auxílio na recuperação de áreas degradadas.

Material e Métodos

Neste trabalho foram desenvolvidas análises direcionadas para a Bacia do Médio Paranapanema. Foram utilizadas 28 espécies arbustivo-arbóreas, de um total de 482 espécies registradas em 63 diferentes pontos da Bacia (Figura 18), no desenvolvimento de modelagem de distribuição geográfica, através da utilização de mapas temáticos abrangendo dados climáticos e do meio físico.

Área de estudo

A Bacia do Médio Paranapanema localiza-se no sudoeste do Estado de São Paulo (Figura 18), tem uma área de abrangência de 16.763km^2 delimitada

pelas coordenadas 22º01’59” e 23º23’43”S e 50º11’49”e 53º00’01”W, com altitudes que variam de 334 a 1007m.

As 63 áreas de amostragem abrangendo 19 municípios da Bacia do Médio Paranapanema são apresentadas no Anexo 2.

Esta região apresenta uma amplitude térmica que vai de picos mínimos e máximos de temperaturas (médias das temperaturas mínimas dos três meses mais frios e máximas dos três meses mais quentes) que variam de 9.9^{o}C a 31,5^{o}C respectivamente, sendo a temperatura média anual de 20,7^{o}C. A curva climática da Bacia (Figura 19) evidencia a ocorrência de invernos secos e verões chuvosos. As geadas são freqüentes em alguns pontos da Bacia (Figura 20).

Os fatores determinantes da ocorrência de geadas estão condicionados a baixas temperaturas do ar, à existência ou não de vento, condições de cobertura de céu e umidade relativa do ar, cobertura do terreno e tipo de relevo local (FUNDAG 1998). Sendo assim, as áreas mais propensas à ocorrência de geadas são as de meia encosta em direção ao topo e áreas de baixadas. No estudo realizado pela FUNDAG e CIERGA foram consideradas a altitude e a latitude para gerar um mapa de “probabilidade de temperaturas mínimas ≤ 2 ^{o}C”. As áreas mais sujeitas a esse tipo de evento climático são as áreas mais escuras do mapa de geada apresentado na Figura 20.

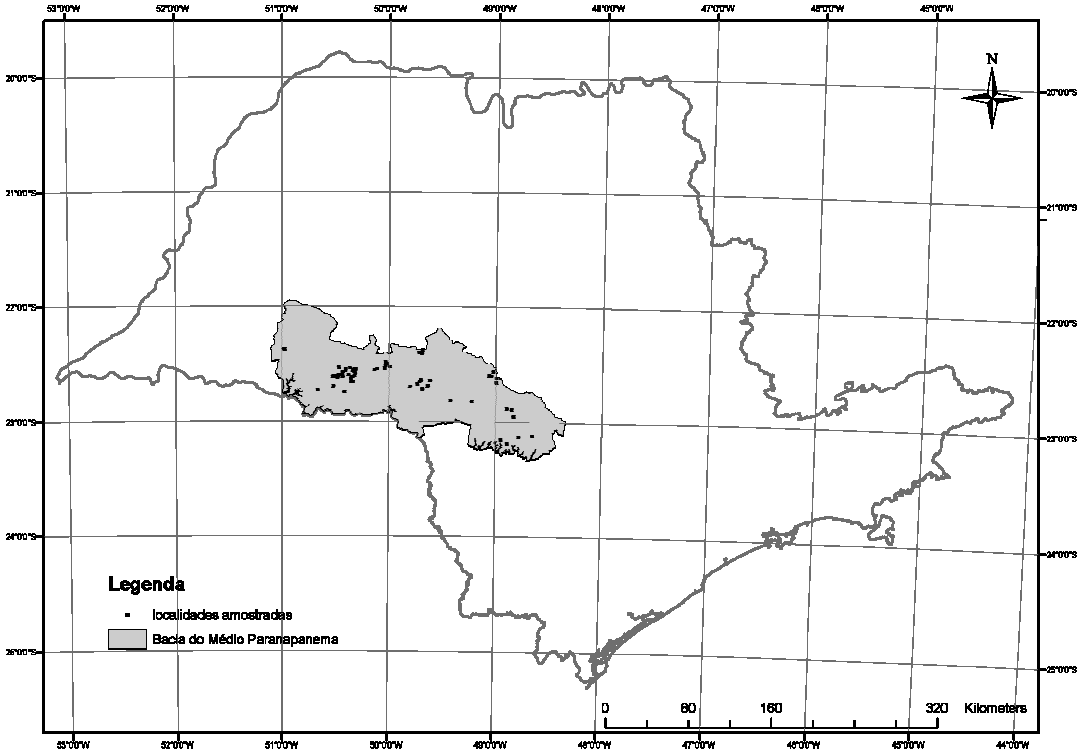


Figura 18: Localização geográfica da Bacia do Médio Paranapanema no Estado de São Paulo com os 63 pontos de amostragem da vegetação utilizados neste estudo.

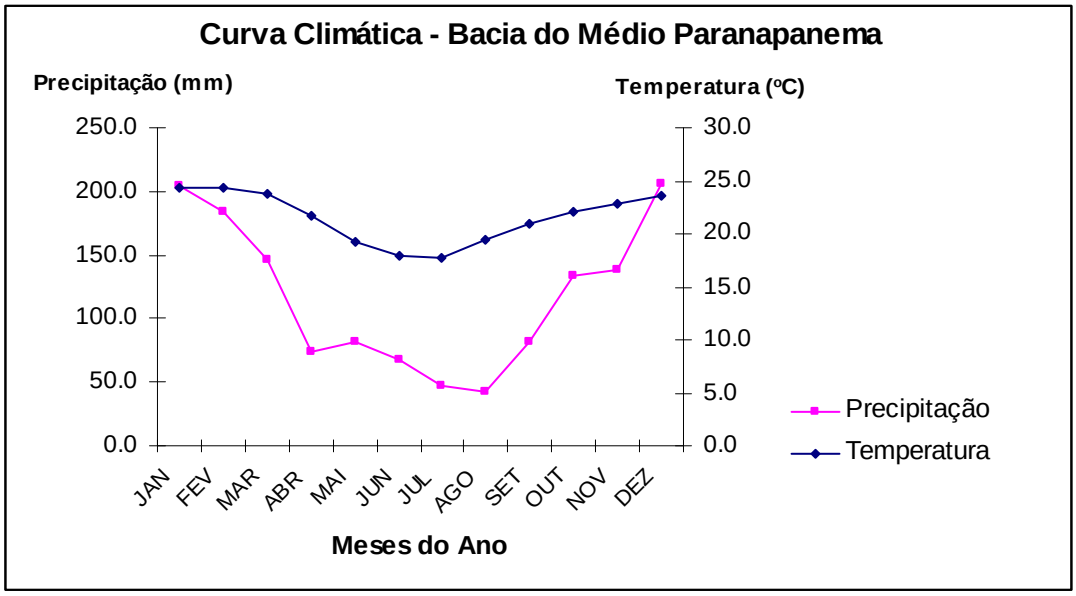


Figura 19: Curva climática da Bacia do Médio Paranapanema (FUNDAG 1998).

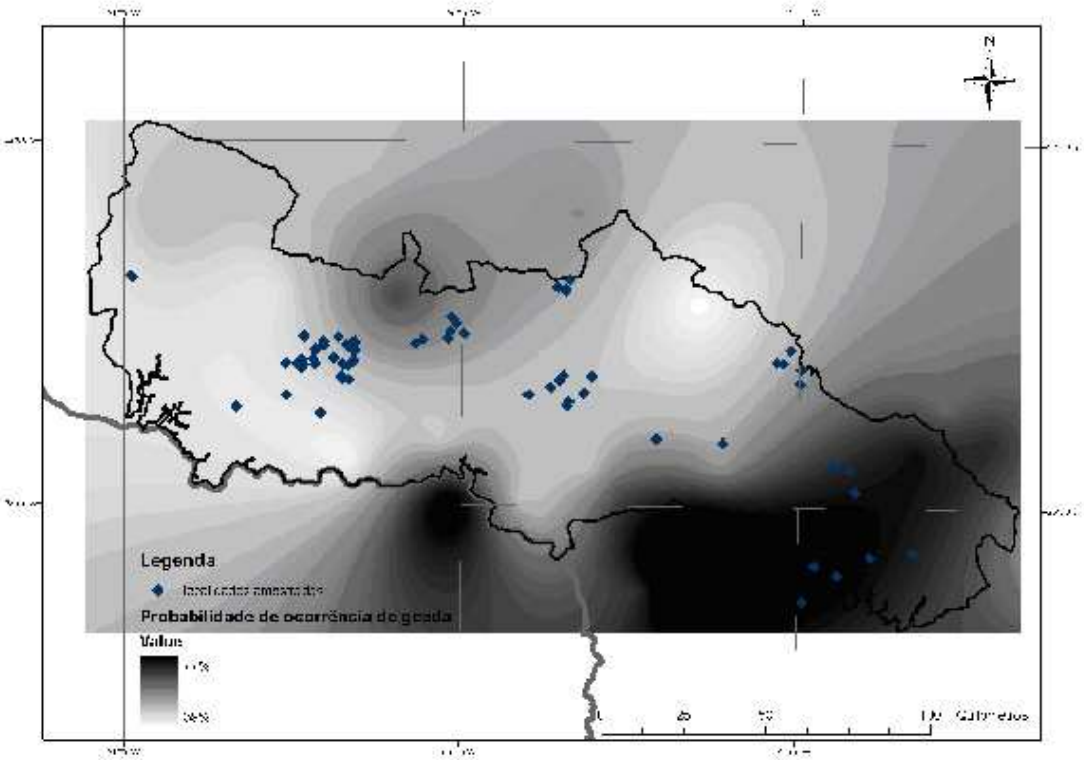


Figura 20: Mapa da probabilidade de ocorrência de geada na Bacia do Médio Paranapanema e a distribuição das localidades amostradas na região.

Em relação ao meio físico, a Bacia do Médio Paranapanema apresenta 38 diferentes categorias de solo (Figura 21). Na Tabela 4 é apresentada uma legenda detalhada dos tipos de solos da Bacia.

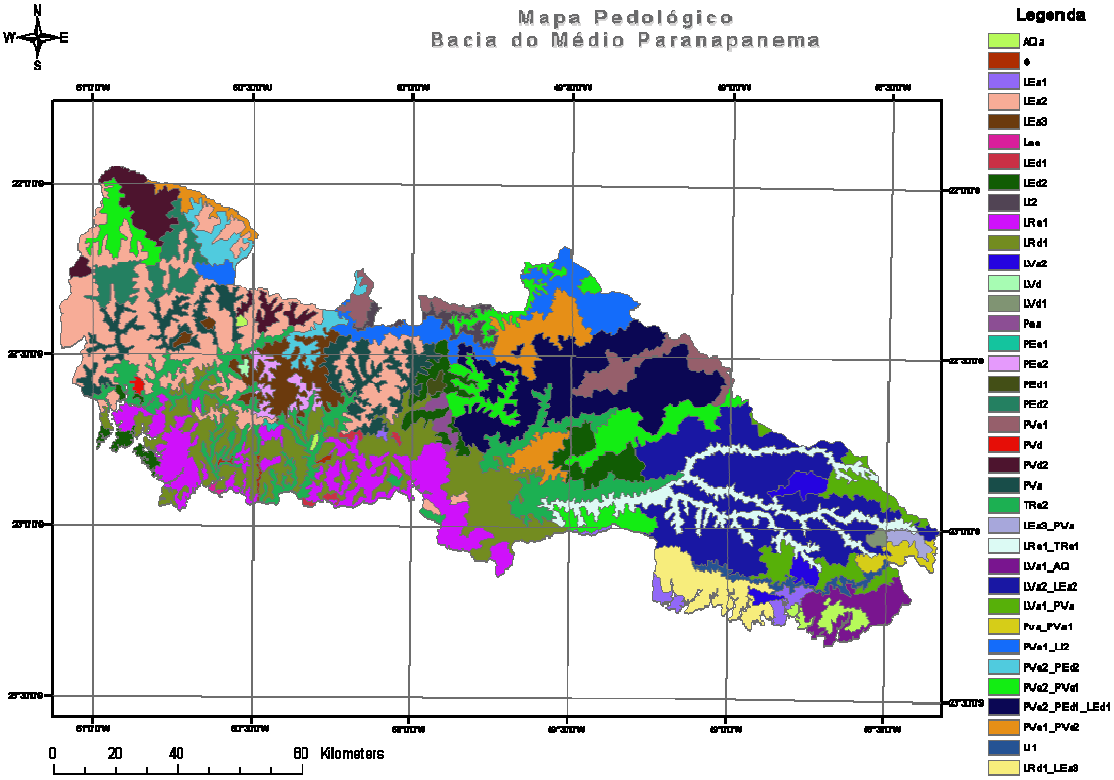


Figura 21: Mapa Pedológico da Bacia do Médio Paranapanema. Fonte: Centro de Desenvolvimento Agropecuário do Médio Vale do Paranapanema – CDVale.

Tabela 4: Detalhamento da legenda para os 38 tipos de solos da Bacia do Médio Paranapanema.

Categorias de Solos	Legenda	Área (ha)	Porcentagem Relativa
LEa2	Latossolo Vermelho Escuro álico ou distrófico, A moderado, textura média	215728.63	10.99
LVa2_LEa2	LATOSSOLO VERMELHO AMARELO Álico ou distrófico, A moderado, textura média. - Latossolo Vermelho Escuro álico ou distrófico, A moderado, textura média.	168129.67	8.57
PVe2_PEd1 _LEd1	Podzólico Vermelho Amarelo eutrófico, não abrupto, A moderado, textura arenosa/média. - PODZÓLICO VERMELHO ESCURO distrófico ou álico, Tb, A moderado, textura média ou arenosa/média. - LATOSSOLO VERMELHO ESCURO DISTRÓFICO A moderado, textura média.	161191.29	8.21
TRe2	Terra Roxa Estruturada Eutrófica Latossólica, teor muito argiloso	141193.19	7.20
LEd2	LATOSSOLO VERMELHO ESCURO DISTRÓFICO A moderado, textura argilosa ou muito argilosa	123954.97	6.32

LRd1	LATOSSSOLO ROXO DISTRÓFICO, A moderado, textura argilosa ou muito argilosa	121249.72	6.18
LRe1	LATOSSOLO ROXO EUTRÓFICO A moderado, textura muito argilosa	116433.43	5.93
PVe2_PVe1	Podzólico Vermelho Amarelo eutrófico, não abrupto, A moderado, textura arenosa/média. - Podzólico Vermelho Amarelo eutrófico abrupto, A moderado, textura arenosa/média.	95073.52	4.85
LRe1_TRe1	LATOSSOLO ROXO EUTRÓFICO A moderado, textura muito argilosa. - TERRA ROXA ESTRUTURADA, eutrófica, A moderado ou chernozênico, textura argilosa ou muito argilosa.	82225.48	4.19
PVa	PODZÓLICO VERMELHO AMARELO álico ou distrófico, Tb, A moderado, textura média	79156.40	4.03
PVd2	PODZÓLICO VERMELHO AMARELO distrófico, não abrupto, A moderado, textura média ou arenosa/média.	76024.26	3.87
PVe1_Li2	Podzólico Vermelho Amarelo eutrófico abrupto, A moderado, textura arenosa/média. -	73898.49	3.77
PVe1_Pee2	Podzólico Vermelho Amarelo eutrófico abrupto, A moderado, textura arenosa/média + PODZÓLICO VERMELHO ESCURO eutrófico, A moderado , textura média/argilosa	67129.43	3.42
LVa1_AQ	LATOSSOLO VERMELHO AMARELO álico, A moderado, textura média. - Areia Quartzoza.	57097.45	2.91
LVa1_PVa	LATOSSOLO VERMELHO AMARELO álico, A moderado, textura média. - PODZÓLICO VERMELHO AMARELO álico ou distrófico, Tb, A moderado, textura média.	47445.04	2.42
LEa3	Grupo Indiferenciado de LATOSSOLO VERMELHO ESCURO, textura média e AREIA QUARTZOSA ambos álicos, A moderado	46930.51	2.39
PVe1	Podzólico Vermelho Amarelo eutrófico abrupto, A moderado, textura arenosa/média.	44370.94	2.26
LRd1_LEa3	LATOSSSOLO ROXO DISTRÓFICO, A moderado, textura argilosa ou muito argilosa. - Grupo Indiferenciado de LATOSSOLO VERMELHO ESCURO, textura média e AREIA QUARTZOSA ambos álicos, A moderado.	42367.25	2.16
PEd2	PODZÓLICO VERMELHO ESCURO distrófico ou álico, Tb, A moderado, textura média /argilosa.	40258.40	2.05
PVe2_PEd2	Podzólico Vermelho Amarelo eutrófico, não abrupto, A moderado, textura arenosa/média. - PODZÓLICO VERMELHO ESCURO distrófico ou álico, Tb, A moderado, textura média /argilosa.	31679.40	1.61
LVa2	LATOSSOLO VERMELHO AMARELO Álico ou distrófico, A moderado, textura média	19046.05	0.97
LEa1	Latossolo Vermelho Escuro álico, A moderado, textura média	14263.61	0.73
Li1	Solos LITÓLICOS eutróficos, fase substrato basalto+TERRA ROXA estruturada eutrófica, A chernozênico, textura muito argilosa.	14307.15	0.73
Li2	Substrato basalto ou diabásio Serra Geral Hi - Solos Hidromórficos	13801.35	0.70
PEe2	PODZÓLICO VERMELHO ESCURO eutrófico, A moderado , textura média/argilosa	12860.59	0.66

Pva_LVa1	PODZÓLICO VERMELHO AMARELO álico ou distrófico, Tb, A moderado, textura média - LATOSSOLO VERMELHO AMARELO álico, A moderado, textura média	12602.70	0.64
AQa	Areia Quartzosa álica	12165.85	0.62
Pea	Podzólico Vermelho Escuro Álico Tb, A moderado, textura média a argilosa	7363.37	0.38
LEa3_PVa	Grupo Indiferenciado de LATOSSOLO VERMELHO ESCURO, textura média e AREIA QUARTZOSA ambos álicos, A moderado. - PODZÓLICO VERMELHO AMARELO álico ou distrófico, Tb, A moderado, textura média.	6270.92	0.32
LEd1	LATOSSOLO VERMELHO ESCURO DISTRÓFICO A moderado, textura média.	4073.22	0.21
LVd1	LATOSSOLO VERMELHO AMARELO distrófico, A moderado, textura média	2948.21	0.15
PEd1	PODZÓLICO VERMELHO ESCURO distrófico ou álico, Tb, A moderado, textura média ou arenosa/média.	2847.08	0.15
G	GLEISSOLOS A moderado	2599.59	0.13
PVd1	PODZÓLICO VERMELHO AMARELO distrófico abrupto, A moderado textura arenosa/média	1712.79	0.09
LVd	Latossolo Vermelho – Amarelo distrófico	1232.53	0.06
PEe1	PODZÓLICO VERMELHO ESCURO eutrófico, abrupto, A moderado, textura arenosa ou médiaso	934.98	0.05
Lee	Grupo Indiferenciado de LATOSSOLO VERMELHO ESCURO E LATOSSOLO VERMELHO ESCURO PODZÓLICO ambos eutróficos A moderado, textura média ou argilosa	797.81	0.04
LRd2	Grupamento indiferenciado de LATOSSOLO ROXO distrófico e LATOSSOLO ROXO ÁLICO, ambos A moderado, textura argilosa ou muito argilosa, inclusões de LATOSSOLO ROXO ÁCRICO, A moderado textura argilosa ou muito argilosa	826.41	0.04
TOTAL		1962191.64	100

Origem dos dados de ocorrência das espécies

As 63 localidades utilizadas neste estudo pertencem aos seguintes grupos de dados:

- 31 remanescentes de áreas de cerrado cujos dados são provenientes do Projeto "Viabilidade da Conservação dos remanescentes de Cerrado do Estado de São Paulo", obtidos a partir de inventários expeditos e disponíveis no site do Projeto SinBiota[13].
- 29 fragmentos de vegetação natural (cerrado, floresta estacional semidecidual e ecótono), dados provenientes do "Plano para o Desenvolvimento Sustentável do Entorno da Estação Ecológica de Assis"

[13] http://sinbiota.cria.org.br/

(dados secundários), obtidos a partir de inventários expeditos e dados de Franco (2002) e Meira-Neto (1991), obtidos através de levantamento florístico e fitossociológico tradicional (dados secundários).

- Três remanescentes de mata ciliar, levantamento florístico convencional, dados secundários provenientes de Durigan e Leitão-Filho (1995) e Silveira (2001).

Escolha das espécies

Para a modelagem referente à distribuição de espécies arbóreas na bacia do Médio Paranapanema foram utilizadas todas as espécies registradas em 30 ou mais localidades diferentes dentro da região de estudo, apresentadas na Tabela 5. A escolha deste número seguiu as recomendações para o algoritmo utilizado GARP (*Genetic Algorithm for Rule Set Production*) utilizado para gerar os modelos (STOCKWELL e NOBLE 1992) no presente estudo.

Também considerou-se que, para se obter o melhor resultado na aplicação de técnicas de modelagem de nicho fundamental a recomendação é de que sejam escolhidas espécies sem problemas de identificação no campo. Desse modo podem ser minimizados os problemas advindos desse tipo de erro, já que os dados utilizados na modelagem, na maior parte das vezes, são provenientes de diferentes fontes, como é o caso deste estudo. Dentro deste subconjunto, foram escolhidas 28 espécies para utilização no processo de modelagem para a Bacia (Tabela 5).

Tabela 5: Espécies arbustivo-arbóreas utilizadas na modelagem para a Bacia do Médio Paranapanema, com as fisionomias em que ocorrem. Dados ecológicos, silviculturais e de fisionomia obtidos em Durigan et al. (2004ª) e Durigan et al. (2004b). **Cce** = campo cerrado; **Ce** = cerrado *s.s.* ; **CE** = cerradão; **FESxC** = região de transição entre a Floresta Estacional Semidecidual e o Cerrado; **MC** = Mata Ciliar; **FES** = Floresta Estadual Semidecidual; **A** = árvore; **a** = arbusto; **Pre** = preferencial de locais úmidos; **Ind** = Indiferente à umidade do solo; **Int** = intolerante a solos úmidos; **M** = moderado; **L** = lento; **R** = rápido; **SG** = suscetível à geada; **TG** = tolerante à geada e **T** = tolerante; **I** = intolerante. * *Xylopia aromatica* não compõe o conjunto dase 28 espécies utilizada na modelagem da Bacia porque não contem o número mínimo de pontos estabelecidos (30 pontos) para esse experimento, mas foi utilizada no experimento referente à análise dos fatores ambientais envolvidos na modelagem.

Espécies	**Famílias**	**CCe**	**Ce**	**CE**	**FESxC**	**MC**	**FES**	**Porte**	**Tolerância à umidade do solo**	**Ritmo de crescimento**	**Suscetibilidade à geada**	**Tolerância à sombra**
Anadenanthera falcata (Benth.) Speg.	Mimosaceae	x	x	x				A	Ind	R	TG	I
Bauhinia rufa (Bong.) Steud.	Caesalpiniaceae	x	x	x				a	Ind	M	SG	I
Casearia sylvestris Sw.	Flacourtiaceae	x	x	x	x	x	x	A	Ind	M	SG	T
Cedrela fissilis Vell.	Meliaceae	x	x	x	x	x	x	A	Ind	M	TG	T
Copaifera langsdorffii Desf.	Caesalpiniaceae	x	x	x	x	x	x	A	Ind	M	TG	T
Croton floribundus Spreng.	Euphorbiaceae			x	x		x	A	-	R	SG	I
Cupania vernalis Cambess.	Sapindaceae				x	x	x	A	Ind	M	TG	T
Gochnatia polymorpha (Less.) Cabrera	Asteraceae		x	x	x	x	x	A	Pre	R	TG	I
Luehea grandiflora Mart. & Zucc.	Tiliaceae		x	x	x		x	A	Ind	M	TG	T
Machaerium acutifolium Mart. ex Benth.	Fabaceae	x	x	x	x		x	A	Ind	M	SG	-
Machaerium brasiliense Vogel	Fabaceae			x	x	x		A	Ind	M	TG	I
Matayba elaeagnoides Radlk.	Sapindaceae			x	x	x	x	A	Pre	M	TG	T
Myrcia multiflora (Lam.) DC.	Myrtaceae			x	x	x		A	Ind	L	TG	T
Ocotea corymbosa (Meisn.) Mez	Lauraceae			x	x	x	x	A	Ind	M	SG	T
Ocotea velutina (Nees) Rohwer	Lauraceae				x	x	x	A	Ind	M	-	T
Peltophorum dubium (Spreng.) Taub.	Caesalpiniaceae				x		x	A	Ind	R	TG	I
Pera obovata Bail.	Euphorbiaceae		x	x	x	x		A	Ind	M	TG	T
Platypodium elegans Vogel	Fabaceae		x	x	x		x	A	Int	M	TG	I

Rapanea umbellata (Mart.) Mez	Myrsinaceae			x	x	x	x	A	Pre	M	TG	T
Siparuna guianensis Aubl.	Monimiaceae			x	x	x	x	A	Ind	M	SG	T
Stryphnodendron obovatum Benth.	Mimosaceae		x	x	x			A	Int	L	SG	I
Syagrus romanzoffiana (Cham.) Glassman	Arecaceae			x	x	x	x	A	Ind	L	TG	T
Tabebuia ochracea (Cham.) Standl.	Bignoniaceae	x	x	x	x	x	x	A	Ind	L	TG	I
Tabernaemontana hystrix Steud.	Apocynaceae			x	x	x	x	A	Ind	M	-	T
Tapirira guianensis Aubl.	Anacardiaceae			x	x	x	x	A	Pre	M	SG	T
Terminalia glabrescens Mart.	Combretaceae			x	x		x	A	Ind	M	SG	I
Vochysia tucanorum Mart.	Vochysiaceae		x	x	x	x		A	Ind	L	SG	I
* *Xylopia aromatica* (Lam.) Mart.	Annonaceae	x	x	x					Int	M	SG	I
Zanthoxylum rhoifolium Lam.	Rutaceae		x	x	x		x	A	Ind	M	TG	T

Pontos de amostragem

Para testar a eficácia dos modelos gerados e garantir a independência dos dados utilizados, o conjunto de pontos de amostragem foi dividido, *a priori,* em dois conjuntos de dados: 50% dos pontos foram utilizados para construir os modelos (conjunto de treino) e 50% foram utilizados para avaliar a eficácia dos modelos gerados (conjunto de teste).

Método de modelagem

Para gerar os modelos de distribuição geográfica para as 28 espécies foi utilizado o algoritmo genético GARP (*Genetic Algorithm for Rule Set Production*) (STOCKWELL e NOBLE 1992). Este algoritmo permite prever a distribuição geográfica potencial a partir de modelos de nicho fundamental de espécies (ver a Introdução Geral e o Anexo 1).

Para gerar uma metodologia aplicável em vários casos, como na identificação de áreas prioritárias para recuperação ambiental, é preciso associar técnicas de geoprocessamento através do uso de SIGs (neste trabalho utilizou-se o *software* de Sistema de Informação Geográfica - ArcView 8.2 da ESRI) aos métodos de modelagem. Essas técnicas de geoprocessamento foram utilizadas para a visualização e cruzamento de dados entre os modelos resultantes e as tabelas de dados ambientais e biológicos, objetivando gerar mapas mais completos para várias finalidades. Estas ferramentas são complementares e o seu uso neste capítulo tem por objetivo principal fornecer ferramentas de análise para orientar trabalhos envolvendo a tomada de decisão sobre conservação da biodiversidade e do ambiente no qual as espécies estão inseridas.

Para a Bacia, todos os modelos foram gerados a partir dos dados de ocorrência das espécies e de dados ambientais referentes à topografia, clima (temperatura e precipitação), solos, geologia e imagens de satélite.

Parâmetros ambientais utilizados para a modelagem

- Mapa de Solos da Bacia do Médio Paranapanema - Instituto Agronômico de Campinas - FUNDAG. Escala 1:250.000 (1 camada).

- Mapa Geológico da Bacia do Médio Paranapanema - Instituto Agronômico de Campinas - FUNDAG. Escala 1:250.000 (1 camada).
- Multitemporal NDVI (*Normalized Difference Vegetation Index*) (greenness) médias mensais. Fonte: AVHRR (*Advanced Very High Resolution Radiometer*)[14] imagem de satélite. Resolução de 1km (12 camadas).
- Dados climáticos*: Worldclim* version 1.1 *Global Cimate Surface* [15]. Resolução 0.83km.
  - ⇒ precipitação mensal (12 camadas)
  - ⇒ temperaturas máximas mensais (12 camadas)
  - ⇒ temperaturas médias mensais (12 camadas)
  - ⇒ temperatura mínimas mensais (12 camadas).
- Dados topográficos: U.S. *Geological Surveys* [16]. Resolução de 1km.
  - ⇒ elevação (1 camada)
  - ⇒ relevo (1 camada)
  - ⇒ aspecto do terreno (1 camada)
  - ⇒ capacidade de retenção de umidade (1 camada)

Resultados e discussão

A Tabela 5 traz a listagem das espécies arbóreas da Bacia que foram selecionadas por representarem o conjunto de espécies com mais de 30 registros de ocorrência na região de estudo.

Os resultados apresentados são provenientes de vários experimentos com modelagem de nicho fundamental de espécies arbustivo-arbóreas realizados para a Bacia do Médio Paranapanema (Figura 18).

Teste dos diferentes tipos de bases de dados abióticos para distribuição potencial em uma escala local

Na Figura 22 são apresentados os modelos resultantes da utilização de quatro conjuntos diferentes de variáveis ambientais. Em todos os modelos foram utilizados dados de topografia (quatro camadas), solos e geologia. A diferença

---

[14] http://www.fs.fed.us/r4/rsgis_fire/
[15] http://biogeo.berkeley.edu/worldclim/worldclim.htm

entre eles está nos dados de clima utilizados, sendo **A**: imagem de satélite – NDVI – médias mensais (12 camadas). **B**: médias mensais dos dados climáticos de temperatura (máxima, média e mínima) e precipitação, fonte – Worldclim (48 camadas). **C**: médias anuais dos dados climáticos contendo os picos de máximas e mínimas de temperatura (máxima, média e mínima) e precipitação, fonte – Worldclim (12 camadas). **D**: apenas as médias anuais de temperatura (máxima, média e mínima) e precipitação, fonte – Worldclim (quatro camadas).

Na Figura 23 é apresentada uma comparação entre as porcentagens de acerto de presença (quando o modelo diz que a espécie ocorre e ela realmente ocorre, baseado no conjunto de dados de teste) utilizando-se os pontos de treino para cada espécie (Tabela 6) para os quatro conjuntos de dados abióticos diferentes.

---

[16] http://edcdaac.usgs.gov/gtopo30/hydro/

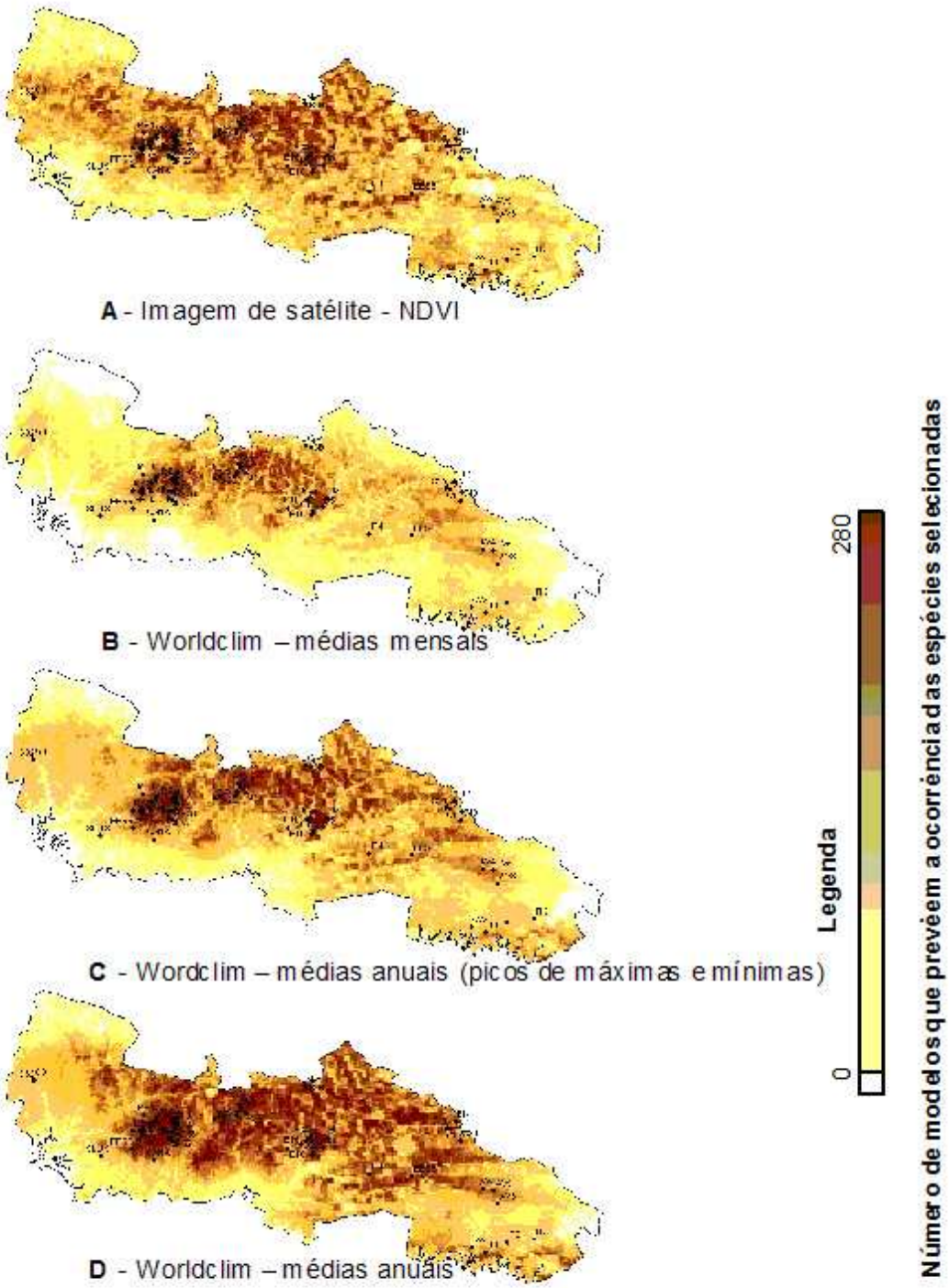


Figura 22: Resultado da modelagem para a probabilidade de ocorrência de 28 espécies arbóreas, utilizando diferentes conjuntos de dados ambientais na Bacia do Médio Paranapanema, Estado de São Paulo.

Tabela 6: Comparação entre as porcentagens de acerto total para os diferentes conjuntos de dados ambientais utilizados para modelar as espécies da Bacia do Médio Paranapanema. **A** e **A1**: Imagem de Satélite NDVI, duas repetições. **B** e **B1**: Worldclim médias mensais, duas repetições. **C**: Worldclim média anual com valores máximos, médios e mínimos. **D**: Worldclim médias anuais.

	Porcentagem de acerto								
especies	**A %**	**A1 %**	**B %**	**B1 %**	**C %**	**D %**	**média %**	**Pontos de treino**	**N total de pontos**
Casearia sylvestris	79	80	79	79	79	79	79.17	25	50
Gochnatia polymorpha	76	77	76	73	73	74	74.83	25	50
Syagrus romanzoffiana	73	74	74	73	73	76	73.83	25	50
Copaifera langsdorffii	77	77	68	69	73	71	72.50	24	48
Ocotea corymbosa	77	77	69	79	74	77	75.50	24	48
Croton floribundus	66	71	73	71	69	69	69.83	22	44
Pera obovata	79	77	76	71	76	76	75.83	22	44
Matayba elaeagnoides	55	60	61	57	65	66	60.67	21	42
Tabernaemontana hystrix	66	63	65	65	66	71	66.00	21	42
Tapirira guianensis	71	68	68	73	71	68	69.83	21	42
Machaerium acutifolium	68	69	68	66	60	61	65.33	20	40
Siparuna guianensis	65	66	66	68	68	65	66.33	20	40
Vochysia tucanorum	71	65	63	63	65	65	65.33	20	40
Machaerium brasiliense	63	61	63	60	60	58	60.83	19	38
Platypodium elegans	71	71	68	68	69	66	68.83	19	38
Stryphnodendron obovatum	65	60	65	65	61	61	62.83	19	38
Luehea grandiflora	63	73	69	71	66	71	68.83	18	36
Tabebuia ochracea	60	61	61	61	61	63	61.17	18	36
Terminalia glabrescens	57	61	71	68	69	61	64.50	18	36
Anadenanthera falcata	60	61	60	61	61	61	60.67	17	34
Cupania vernalis	55	55	57	58	55	53	55.50	17	34
Ocotea velutina	58	55	58	58	57	55	56.83	17	34
Rapanea umbellata	60	58	53	57	57	53	56.33	17	34
Zanthoxylum rhoifolium	55	58	55	58	55	57	56.33	17	34
Myrcia multiflora	57	58	65	60	60	58	59.67	16	32
Peltophorum dubium	52	52	57	55	63	49	54.67	16	32
Bauhinia rufa	53	50	74	69	61	58	60.83	15	30
Cedrela fissilis	50	53	55	53	50	47	51.33	15	30
média total spp	**64**	**65**	**66**	**65**	**65**	**64**			

Verifica-se, pela análise da Tabela 6 e Figura 23, que a porcentagem de acerto total está diretamente relacionada com o número de registros de ocorrência utilizados no modelo.

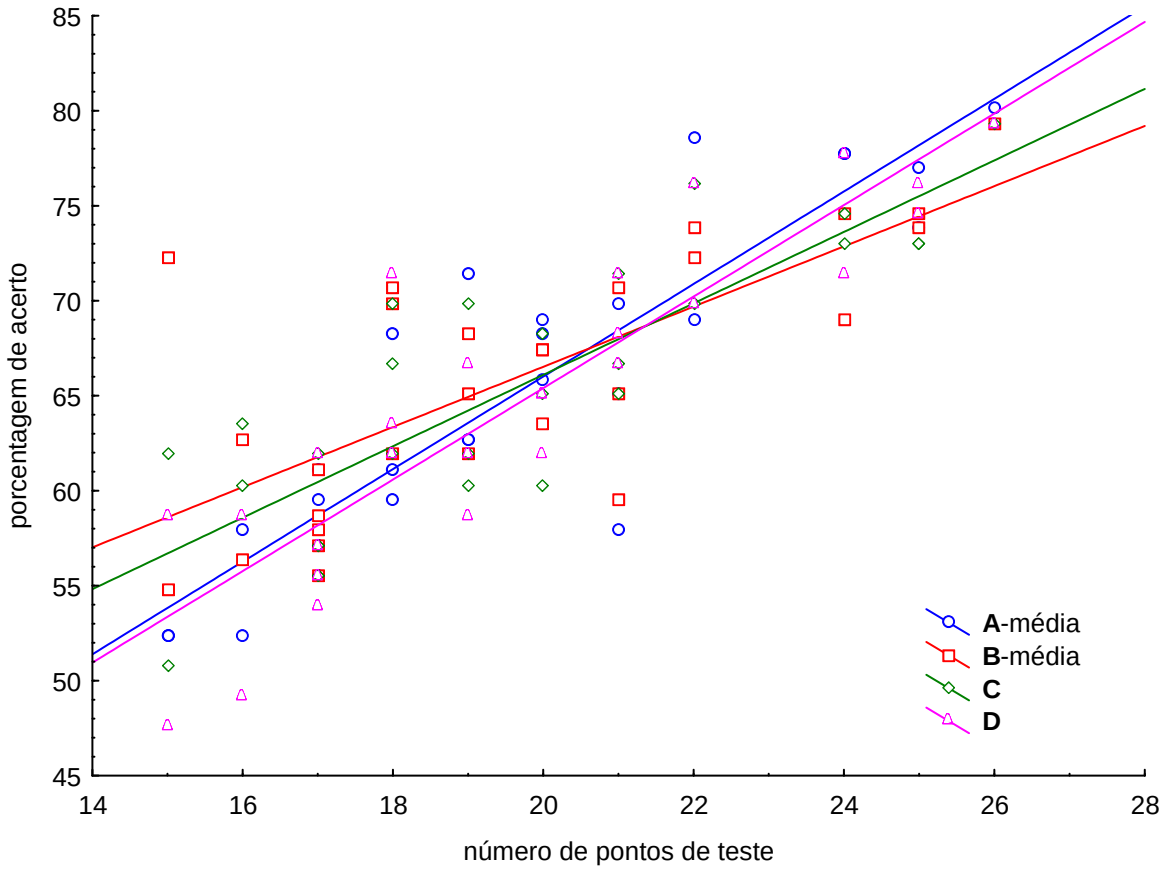


Figura 23: Número de pontos de ocorrência de espécies e respectivas porcentagens de acerto entre os diferentes conjuntos de dados ambientais. **A-média**: média dos valores da porcentagem de acerto para os experimentos **A** e **A1**. **B-média**: média dos valores da porcentagem de acerto para os experimentos **B** e **B1**.

A porcentagem de acerto foi afetada pelo tipo de base de dados e pelo número de pontos de teste, mas houve uma interação significativa entre esses fatores (Tabela 7), representando que os efeitos do tipo de base de dados foram diferentes para diferentes números de pontos de teste. Os modelos baseados na imagem de Satélite (A) e no Worldclim médias anuais (D) apresentaram menor porcentagem de acerto com baixos tamanhos amostrais, mas foram os melhores em eficiência, com tamanhos amostrais acima de 20 pontos.

As comparações *a posteriori* (Figuras 24 e 25) corroboram a afirmativa anterior, mostrando que **A** e **D** têm menores valores de intercepto e maiores valores para inclinação. Nessa situação não é possível determinar um conjunto único de dados abióticos que represente uma situação homogeneamente superior em termos de probabilidade de acerto. Com isso, outros critérios podem ser

utilizados para determinar qual conjunto deve ser utilizado nas próximas análises. Sendo assim, optou-se pelo conjunto de dados **D** - Worldclim – calculando-se as médias anuais para cada camada climática (temperaturas máximas, médias, mínimas e precipitação), porque este conjunto apresenta uma menor quantidade de layers ambientais, otimizando assim o tempo envolvido no processo de modelagem.

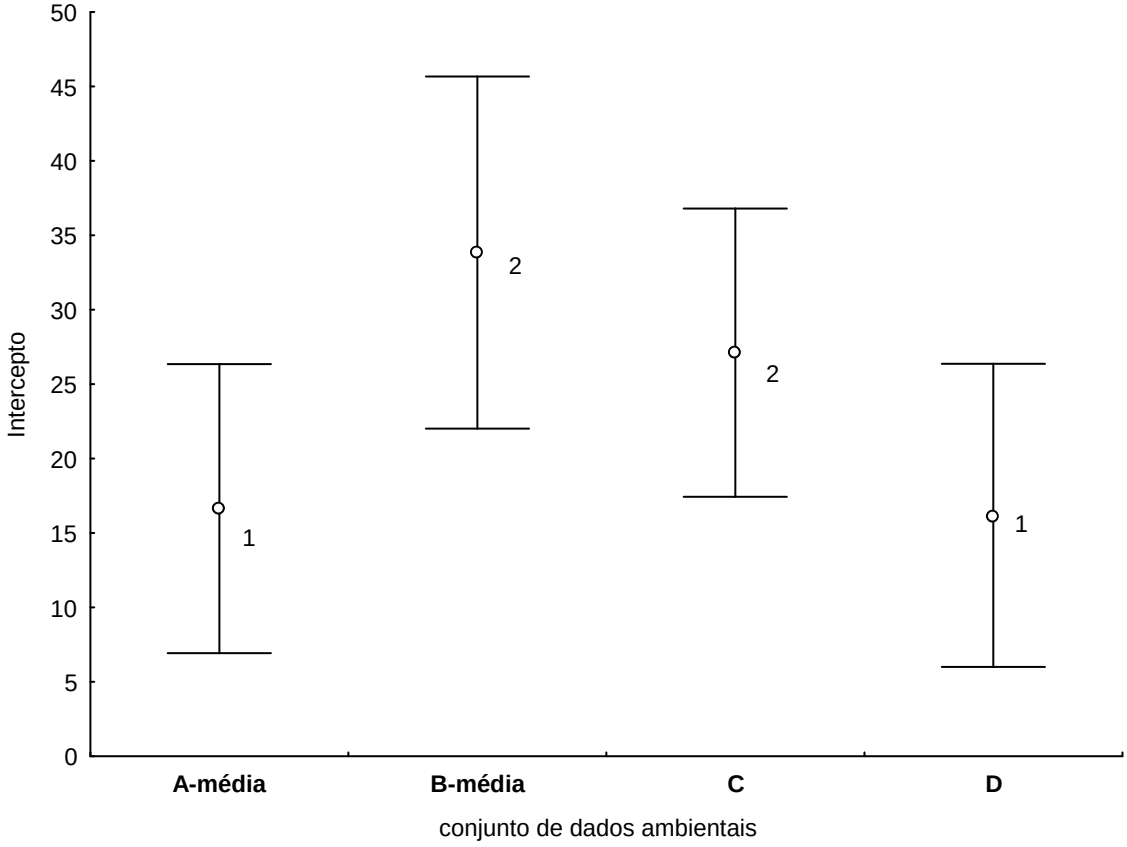


Figura 24: Comparação entre os valores de intercepto da regressão linear entre as porcentagens de acerto e número de pontos de teste para as diferentes bases de dados abióticos. Números iguais representam valores estatisticamente iguais a partir de um teste *a posteriori* de Tukey em um nível de 5% de significância.

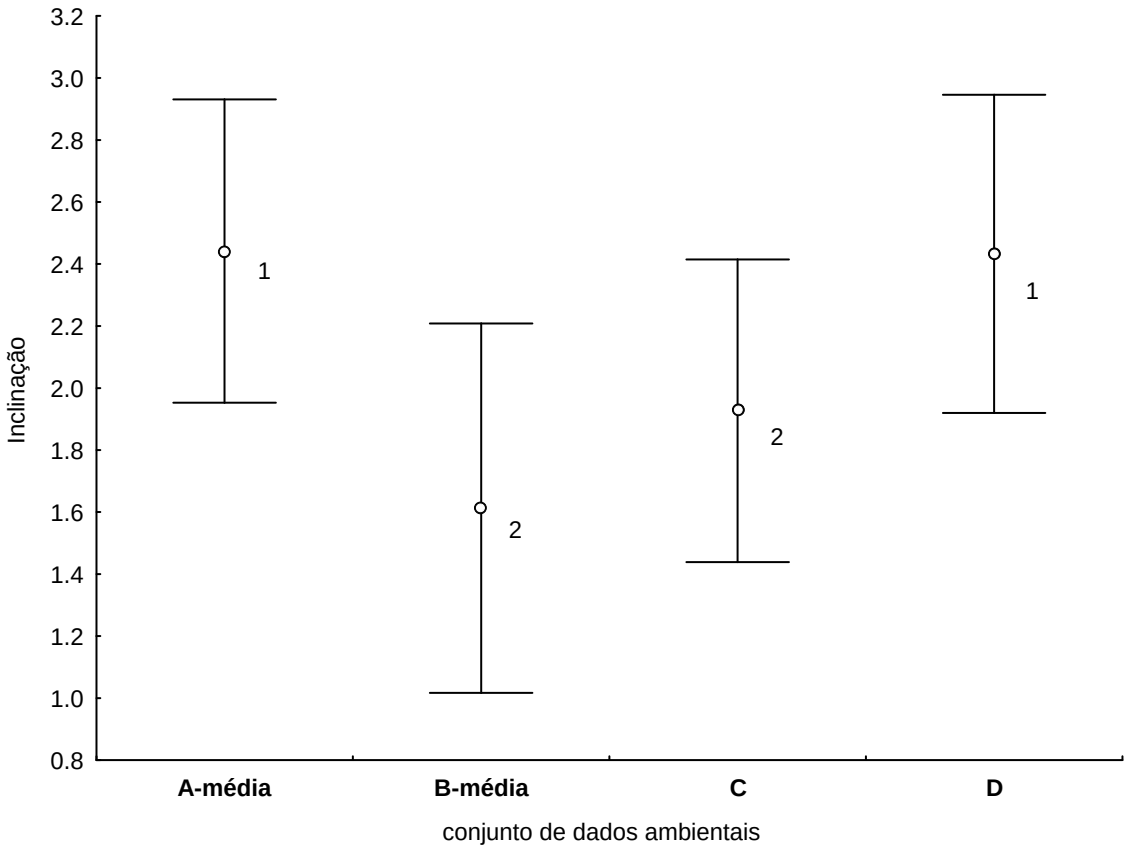


Figura 25: Comparação entre os valores de inclinação da regressão linear entre as porcentagens de acerto e número de pontos de teste para as diferentes bases de dados abióticos. Números iguais representam valores estatisticamente iguais a partir de um teste *a posteriori* de Tukey em um nível de 5% de significância.

Tabela 7: Análise de covariância para o efeito do número de registros de ocorrência e diferentes conjuntos de dados ambientais sobre a porcentagem de acerto no modelo GARP para espécies arbóreas na Bacia do Médio Paranapanema. **gl trat** = graus de liberdade do tratamento; **gl erro =** graus de liberdade do erro; **P** = probabilidade do resultado ser dada ao acaso.

Comparação	F	gl trat	Gl erro	P
Diferença de inclinação	2.676	3	104	**0.051**
Efeito do fator (dados_amb)	2.9724	3	107	**0.035**
Efeito do número de pontos no teste	266.3876	1	107	**<0.001**

Avaliação do efeito do número de amostras na porcentagem de acerto dos modelos

Para determinar o número ideal de pontos para este conjunto de dados foi realizado um experimento comparando-se a porcentagem de acerto total para o mesmo conjunto de dados, ou seja, para os mesmos dados ambientais e para uma mesma espécie (*Casearia sylvestris*), variando-se o número de pontos de treino utilizados no modelo. O resultado é apresentado na Figura 26. Esta espécie foi escolhida por representar o melhor conjunto de dados disponíveis, tanto em relação ao número de pontos como na porcentagem de acerto (Tabela 6). Para essa espécie, a partir de 23 pontos de treino ocorre uma estabilização da porcentagem de acerto. Esses resultados corroboram aqueles encontrados por Stockwell e Peterson (2002), que recomendam o uso de, no mínimo, 20 pontos de amostragem para cada espécie modelada.

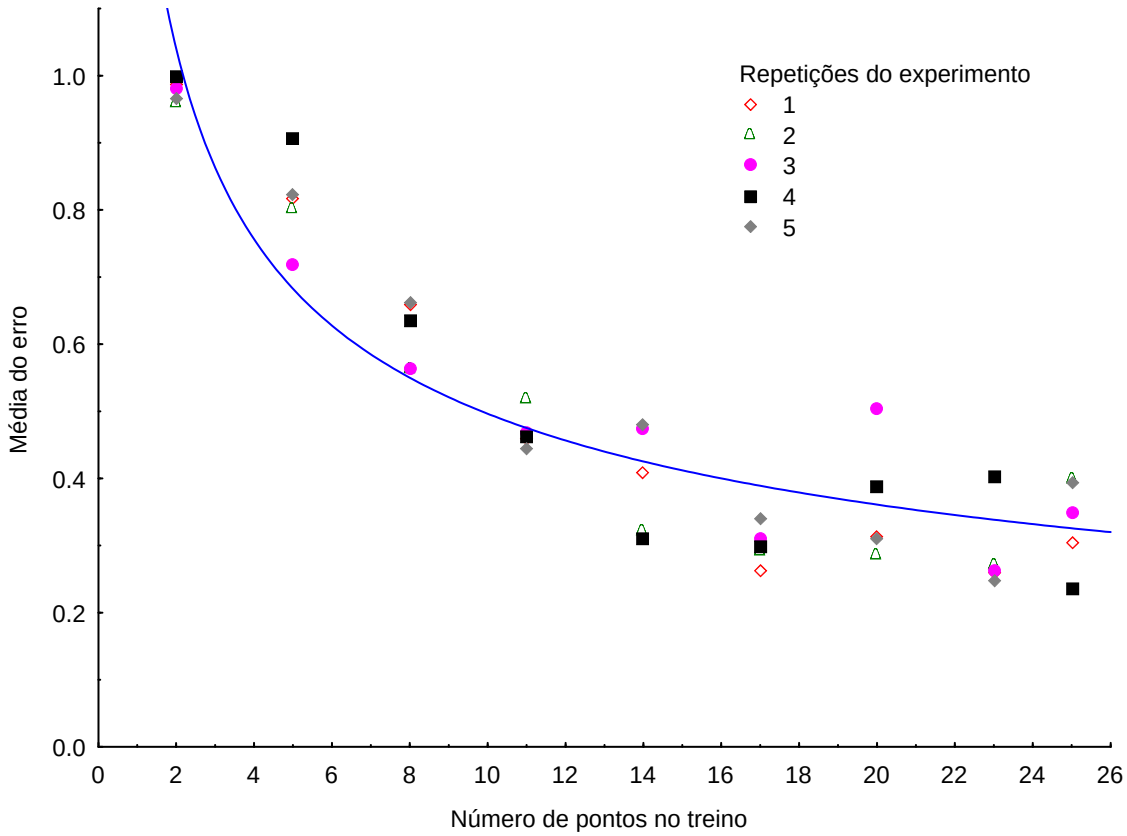


Figura 26: Efeito do número de pontos de treino sobre a média do erro e a variação entre cinco repetições do modelo de distribuição potencial de *Casearia sylvestris,* utilizando dados ambientais **D** (Worldclim – médias anuais).

O modelo não-linear explica 87,7% da variação nos dados, seguindo a equação $Y = 1{,}40315X^{-0{,}45956}$

Com isso, a partir da função gerada é possível estimar qual seria a porcentagem de acerto aumentando-se o número de pontos para esta espécie.

Sendo assim, como o conjunto de dados utilizados para gerar o modelo final para todas as 28 espécies variou de 30 a 50 pontos de amostragem, uma vez que foram utilizados todos os pontos disponíveis (tantos os de treino como os de teste) para gerar o modelo final (Tabela 6), é de se esperar que a média do erro fique entre 0,294 e 0,232. Ou seja, para o conjunto de dados utilizado, espera-se uma porcentagem de acerto de presença entre 71% e 78%, aproximadamente, para as 28 espécies modeladas.

Modelagem de distribuição potencial de espécies vegetais para uma escala local

Com base nos resultados anteriores, o mapa resultante (Figura 27) foi gerado para as 28 espécies, utilizando-se todos os pontos de amostragem disponíveis para cada espécie (Tabela 6).

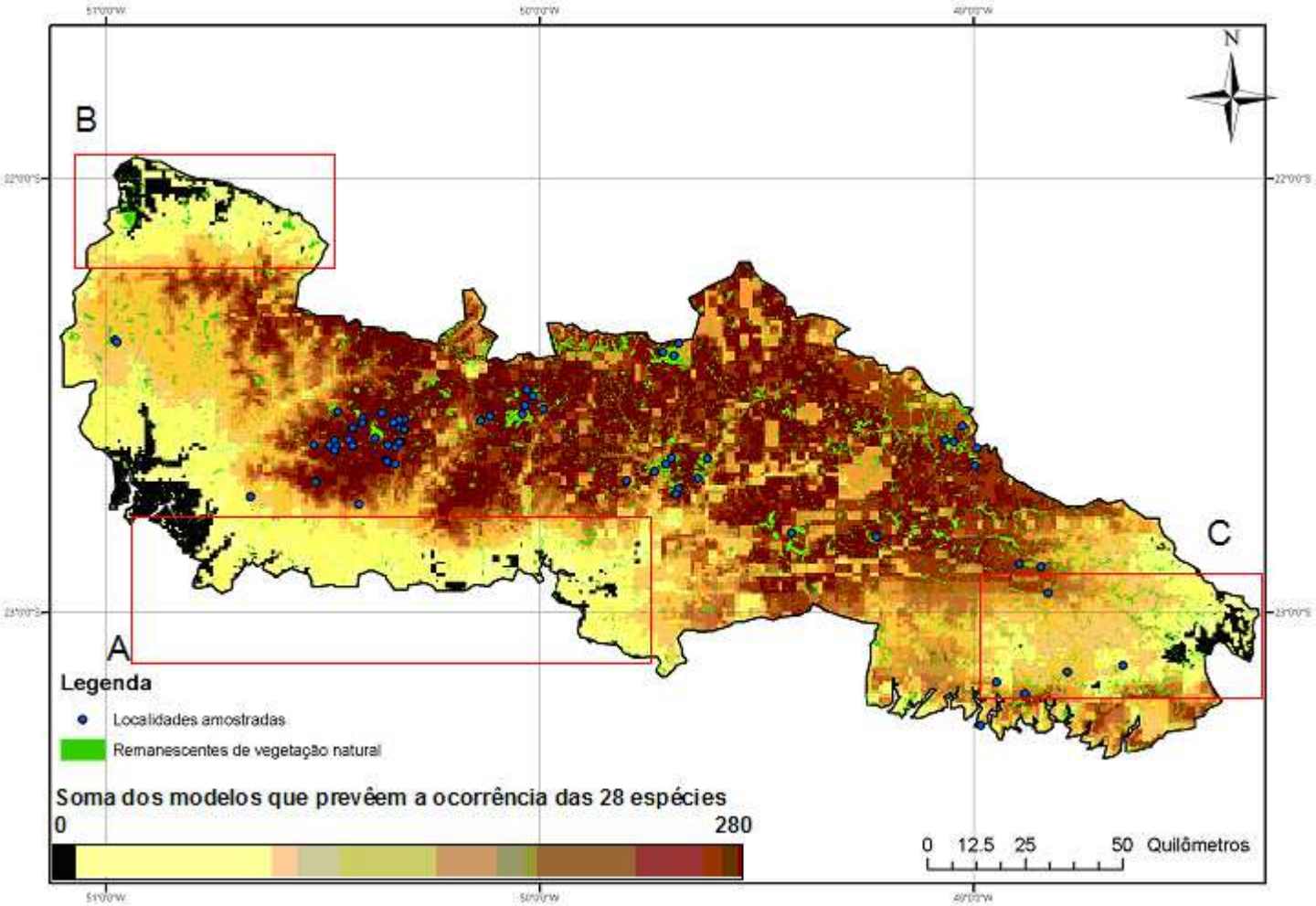


Figura 27: Mapa da riqueza potencial para 28 espécies arbóreas da Bacia do Médio Paranapanema, produzido a partir da soma dos 10 melhores modelos (entre os 100 gerados para cada espécie) utilizando todos os registros (pontos de treino e pontos de teste) disponíveis na área de estudo.

Influência de fatores ambientais na distribuição de espécies arbóreas

O modelo de riqueza potencial apresenta uma baixa probabilidade de ocorrência das espécies nas áreas A, B e C (Figura 27). A área **A** apresenta solos predominantemente do tipo LRe1; LRd1 e TRe2 (Latossolos Roxos e Terra Roxa). As espécies modeladas, em sua maioria plantas de cerrado, estão distribuídas, em sua maior parte, por Latossolos e Podzólicos Vermelho Escuro (Tabela 4), o que resulta uma baixa probabilidade de ocorrência dessas espécies nessa área. Na área **B**, tanto o tipo de solo quanto o clima parecem ser adequados à ocorrência das espécies (baixa probabilidade de ocorrência de geada e solos do tipo Podzólico Vermelho Amarelo), mas o modelo apresenta baixa probabilidade de ocorrência das espécies nessa região. A baixa capacidade do modelo em prever a distribuição pode estar associado aos poucos pontos de amostragem disponíveis na área. A área **C** apresenta uma baixa probabilidade de ocorrência de espécies, não devido ao tipo de solo, que é composto, nessa região, na maior parte, por Latossolos Vermelho Amarelos e Latossolos Vermelho Escuros, mas talvez pela alta probabilidade de ocorrência de geada na região (Figura 20) uma vez que cerca de 35% das espécies utilizadas são suscetíveis à geada (Tabela 5).

Na tentativa de compreender quais parâmetros ambientais melhor explicam a distribuição das espécies, ou ainda, qual o fator mais limitante da sua distribuição, foram gerados alguns modelos com o objetivo específico de observar a influência da ocorrência de geadas em cinco espécies (*Xylopia aromatica, Bauhina rufa, Terminalia glabrescens, Pera obovata* e *Copaifera langsdorffii)*. Essas espécies foram escolhidas porque tanto *Xylopia aromatica* quanto *Bauhina rufa* e *Terminalia glabrescens* são suscetíveis à geada, enquanto *Pera obovata* e *Copaifera langsdorffii* são tolerantes.

Os resultados da modelagem para *Xylopia aromatica* são apresentados na Figura 28, na qual os pontos em azul mostram os registros de ocorrência e os registros em verde mostram a provável **não** ocorrência dessa espécie.

Sobrepondo-se o mapa de probabilidade de geada aos modelos de distribuição potencial dessas espécies, observa-se que a não ocorrência da espécie

coincide com as áreas de maior incidência de geada (áreas mais escuras da Figura 28).

Cabe ressaltar que o mapa de geada não foi utilizado na modelagem. Apenas os mapas de temperaturas máximas, médias e mínimas, assim como a altitude, foram suficientes para que o algoritmo pudesse prever esse padrão de distribuição.

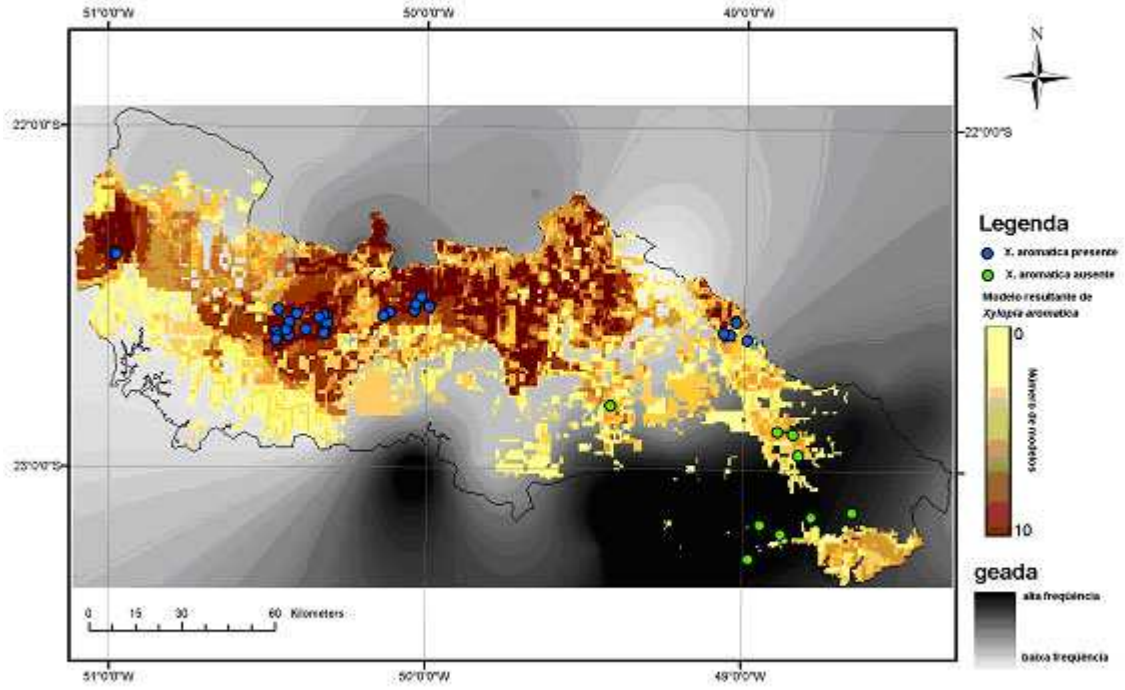


Figura 28: Distribuição potencial de *Xylopia aromatica* para a Bacia do Médio Paranapanema e a probabilidade de ocorrência de geada na região.

A distribuição potencial das demais espécies é apresentada na Figura 29, na qual observa-se que para as espécies tolerantes à geada (TG) a região de geada parece não influenciar na distribuição da espécie enquanto o oposto é observado para as espécies suscetíveis à geada (SG).

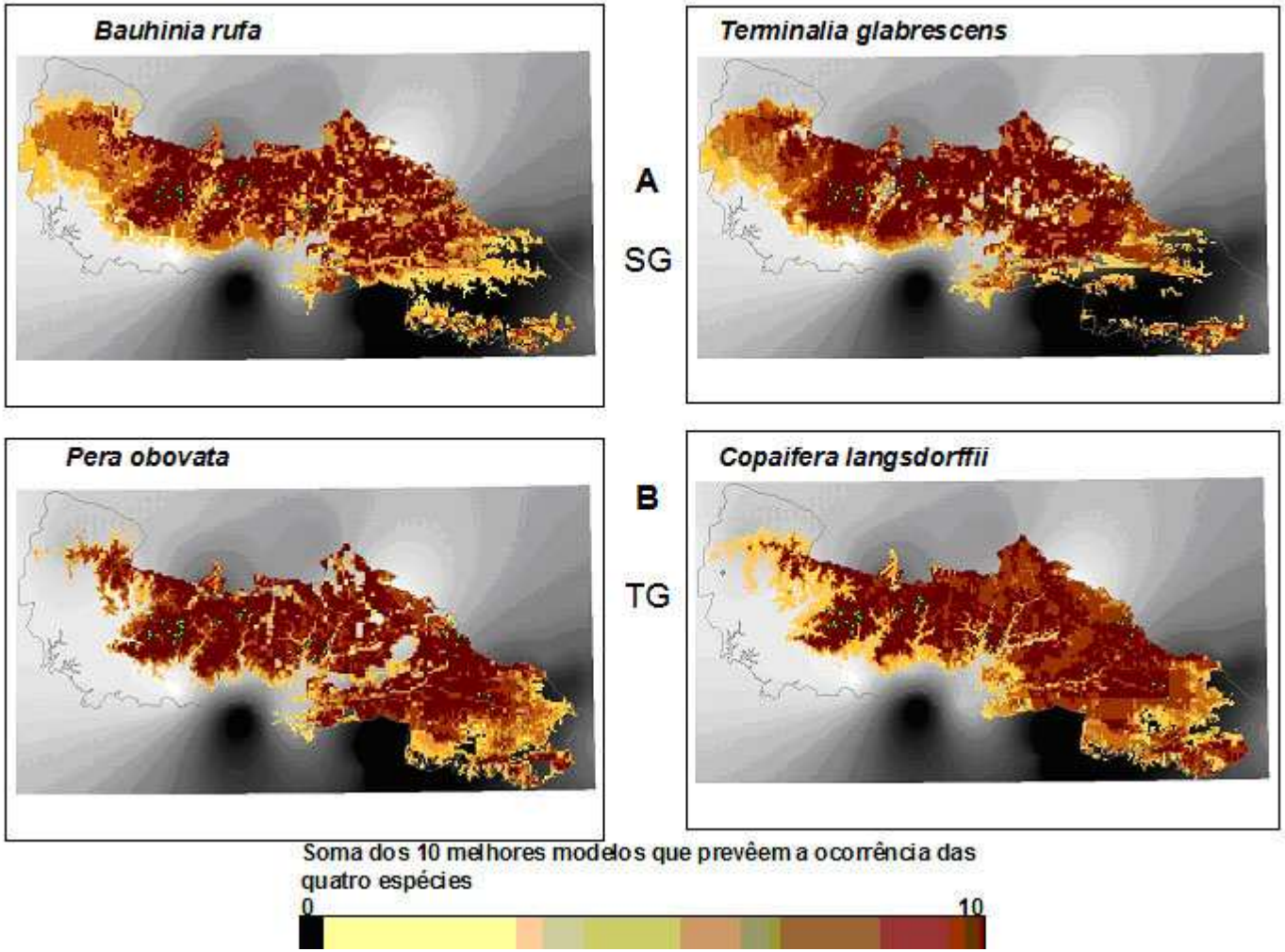


Figura 29: Mapas de distribuição potencial para quatro espécies arbóreas ocorrentes na Bacia do Médio Paranapanema. SG: espécies suscetíveis à geada. TG: espécies tolerantes à geada. Os pontos verdes representam os registros de ocorrência para cada espécie.

Aplicações de modelagem para auxiliar trabalhos de recuperação ambiental

Uma vez obtido o melhor conjunto de espécies para modelar a riqueza potencial da Bacia, é possível cruzar essa informação com os dados do meio físico, (mapas de risco de erosão, uso do solo, capacidade de uso do terreno etc) para indicar quais seriam as espécies (e as áreas) mais apropriadas para projetos de recuperação de áreas degradadas. Para isso, é necessário usar recursos disponíveis em Sistemas de Informação Geográfica, como os apresentados nas Figuras 30, 31 e 32.

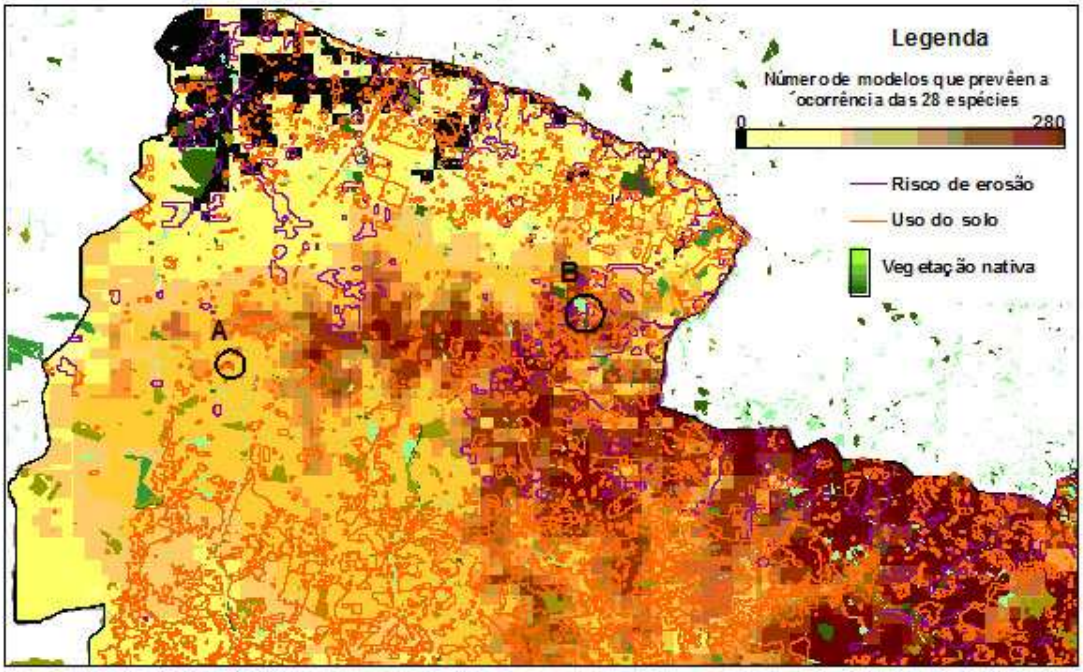
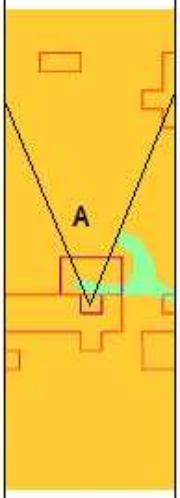


Figura 30: Cruzamento de dados entre o modelo de riqueza de espécies e as áreas de risco de erosão, uso do solo e remanescentes de vegetação nativa para a Bacia do Médio Paranapanema. As áreas em destaque A e B representam 2 pontos quaisquer da Bacia.

Ponto A	Descrição da categoria
Uso da terra	Solo descoberto: categoria representadas pelos solos preparados, glebas recem colhidas com palha seca sobre o terreno, glebas em pousio e parte de pastagens muito degradadas, pastoreadas em excesso com grande porcentagem de solos parcialmente expostos.
Risco de erosão	Alto
Adequação do uso do solo	Pouco adequado para uso da terra em práticas agrícolas
Capacidade de uso do solo	Reflorestamento/refúgio fauna e flora/lazer
Geologia	Ka4: Arenitos finos a muito finos, quartzosos, com freqüentes intercalações de argilitos e siltitos, formando bancos espessos. Localmente arenitos com pelotas de argila. Presença moderada de cimentação carbonática. Formação Adamantina.
Solo	PEd2: Podzólico vermelho amarelo distrófico, não abrupto, A moderado, textura média ou arenosa/média.



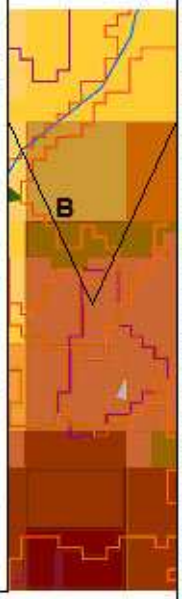
Espécies ponto A – 22 spp	Número de modelos
Rapanea umbellata	9
Zanthoxylum rhoifolium	9
Peltophorum dubium	8
Stryphnodendron obovatum	8
Bauhinia rufa	7
Matayba elaeagnoides	7
Terminalia glabrescens	7
Casearia sylvestris	6
Cedrela fissilis	6
Siparuna guianensis	6
Tabebuia ochracea	6
Gochnatia polymorpha	4
Anadenanthera falcata	3
Luehea grandiflora	3
Syagrus romanzoffiana	3
Vochysia tucanorum	3
Copaifera langsdorffii	2
Machaerium acutifolium	2
Machaerium brasiliense	2
Ocotea corymbosa	2
Ocotea velutina	2
Platypodium elegans	2

Legenda

Risco de erosão — Uso do solo — 220 modelos

Figura 31: Lista das espécies com potencial de ocorrência no ponto A da Bacia do Médio Paranapanema, classificadas de acordo com o número decrescente de modelos que previram a ocorrência de cada espécie entre os 10 melhores modelos (dentre os 100 gerados) e a informação ambiental disponível para o ponto A.

Ponto B	Descrição das categorias
Uso da terra	Solo descoberto: categoria representada pelos solos preparados, glebas recem colhidas com palha seca sobre o terreno, glebas em pousio e parte de pastagens muito degradadas, pastoreadas em excesso com grande porcentagem de solos parcialmente expostos.
Risco de erosão	alto
Adequação do uso do solo	Não adequado para uso agrícola
Capacidade de uso	Reflorestamento / refúgio fauna e flora / lazer
Geologia	Km: arenitos com granulometria fina a grossa, formando, geralmente, bancos maciços. Incluem lentes e intercalações subordinadas de siltitos, argilitos e arenitos muito finos e níveis rudáceos. Formação Marília.
Solo	PVe2_PEd2: Podzólico Vermelho Amarelo eutrófico, não abrupto, A moderado, textura arenosa/média. Podzólico Vermelho Escuro distrófico ou álico, Tb, A moderado, textura média /argilosa.



Espécies ponto B – 28 spp	Número de modelos
Matayba elaeagnoides	10
Tabernamontana hystrix	10
Tabebuia ochracea	10
Zanthoxylum rhoifolium	10
Casearia sylvestris	9
Peltophorum dubium	9
Syagrus romanzoffiana	9
Croton floribundus	8
Luehea grandiflora	8
Machaerium acutifolium	8
Machaerium brasiliense	8
Rapanea umbellata	8
Stryphnodendron obovatum	8
Vochysia tucanorum	8
Bauhinia rufa	7
Pera obovata	7
Terminalia glabrescens	7
Cupania vernalis	6
Ocotea velutina	6
Siparuna guianensis	6
Gochnatia polymorpha	5
Ocotea corymbosa	5
Anadenanthera falcata	4
Cedrela fissilis	4
Platypodium elegans	4
Copaifera langsdorffii	3
Myrcia multiflora	2
Tapirira guianensis	1

Legenda

Risco de erosão — Uso do solo

número de modelos somados 1 – 280

Figura 32: Lista das espécies com potencial de ocorrência no ponto B da Bacia, classificadas de acordo com o número decrescente de modelos que previram a ocorrência de cada espécie entre os 10 melhores modelos (dentre os 100 gerados) e a informação ambiental disponível para o ponto B.

É importante ressaltar que estas informações podem ser levantadas para qualquer ponto da bacia, através do uso de um SIG, configurando assim um sistema de auxílio na tomada de decisão sobre quais espécies devem ser utilizadas em um determinado local e quais são as características e necessidades ambientais dessa região. Esse sistema pode tanto ajudar na obtenção de sucesso no plantio dessas espécies como também ser usado para facilitar o processo de conscientização das instituições gerenciadoras dos recursos naturais da região.

Foram utilizadas nesta análise apenas 28 espécies, para as quais existem registros suficientes de ocorrência para modelar a sua distribuição. Naturalmente, essa análise preditiva poderia ser aplicada à maioria das 482 espécies arbustivo-arbóreas do Médio Paranapanema (DURIGAN et al. 2004b),

mediante um esforço para ampliar o número de locais de amostragem de áreas naturais remanescentes.

Conclusões

- Para uma escala local, o emprego de técnicas de modelagem e de análise espacial por SIGs pode auxiliar no processo de recuperação de áreas degradadas, fornecendo informações sobre onde existem as condições abióticas mais propícias para o desenvolvimento de um conjunto determinado de espécies.
- A probabilidade de acerto do modelo está diretamente relacionada com o número de registros de ocorrência das espécies. Isso significa que um maior esforço em levantamentos de biodiversidade pode fortalecer os modelos resultantes para a predição da ocorrência das espécies.
- A modelagem preditiva pode ser utilizada para predizer a composição florística de áreas desmatadas dentro de uma bacia hidrográfica, desde que existam dados ambientais de clima e do meio físico em escala adequada e, principalmente, registros suficientes de ocorrência das espécies, georeferenciados e bem distribuídos por toda a bacia. Quanto maior o número de levantamentos, mais extensa pode ser a lista de espécies recomendadas e maior a precisão do modelo gerado.

Capítulo 3: Modelagem para espécies com distribuição restrita ou com poucos pontos de amostragem

Introdução

A vegetação de cerrado apresenta alto índice de endemismo, sendo que cerca de 30% das espécies arbóreas são conhecidas por poucos de amostragem (RATTER et al. 2000). Para as espécies vegetais como um todo, considerando as espécies lenhosas e não lenhosas, este índice sobe para 44%, segundo Mittermeier et al. (1999).

Os poucos pontos de amostragem podem representar casos de espécies com distribuição restrita, como é o caso de endemismos, que pode ser definido como espécie nativa e restrita a uma determinada área geográfica (ACIESP 1997), ou ainda pode ser o caso de espécies raras, espécies que apresentam menos de um indivíduo por hectare (KAGEYAMA e GANDARA 2000) ou organismo ou espécie que ocorre em apenas 5% ou menos de amostras realizadas (ACIESP 1997). Por outro lado, os poucos pontos de amostragem podem representar uma situação de deficiência de coleta.

O enquadramento em um ou outro caso é de fundamental importância para a adoção de estratégias eficazes de conservação. No caso de deficiência de coletas é preciso investir em trabalho de campo. No caso de distribuição restrita a situação é mais complicada, necessitando de estratégias conservacionistas mais elaboradas. Para separar os dois casos, são necessários esforços de campo para verificar se essas espécies com registros únicos ocorrem em outras áreas sob condições ambientais semelhantes ou se são, realmente, ocorrências únicas. O desenvolvimento de metodologias de análise e direcionamento de pesquisa de campo através de modelagem pode diminuir significativamente os recursos necessários e reduzir o tempo associado ao trabalho de campo. Em uma situação crítica de devastação ambiental, a redução de tempo é especialmente importante.

Porém, para utilizar algoritmos de modelagem geralmente é necessário um número mínimo de registros de ocorrência para cada espécie. No caso do

GARP (*Genetic Algorithm for Rule Set Production*), é aconselhável o uso de pelo menos 20 pontos (STOCKWELL e PETERSON 2002). Para os casos em que não existem pontos suficientes, há a possibilidade de se utilizarem modelos mais simples, baseados em distância ambiental.

Objetivos

A pesquisa de que trata o presente capítulo teve como objetivos:

- Desenvolver um algoritmo para modelar a distribuição geográfica de espécies visando a orientar trabalhos de localização, no campo, de espécies consideradas endêmicas, raras ou dificilmente coletadas.
- Verificar se *Byrsonima subterranea* é uma espécie realmente rara ou se é um caso de deficiência de coletas.

Materiais e Métodos

Regra geral, as espécies que apresentam distribuição restrita são mais suscetíveis à ameaça de extinção. Este parece ser o caso da *Byrsonima subterranea* Brad. & Mark. - Malpighiaceae (Figuras 33 e 34), espécie de fisionomias abertas de cerrado e de porte sub-arbustivo, que foi enquadrada na categoria “presumivelmente extinta” na lista oficial das espécies ameaçadas de extinção no Estado de São Paulo (D.O.E. de 10/03/1998).

Figura 33: *Byrsonima subterranea* Brad. & Mark. (Malpighiaceae). Registro da planta no município de Assis, SP.

Figura 34: Fisionomia de Cerrado aberto no qual foi encontrada a espécie Byrsonima subterranea no município de Avaré, SP.

O critério de inclusão dessa espécie na lista de espécies ameaçadas foi: plantas sem coletas e/ou registros datados dos últimos 30 anos. No final dos

anos 90, porém, esta espécie foi registrada no município de Assis, Bacia do Médio Paranapanema, SP (DURIGAN et al. 1999), e este ponto foi posteriormente georeferenciado.

Área de estudo

O ponto original de ocorrência de *Byrsonima subterranea* foi registrado no município de Assis, Bacia do Médio Paranapanema, sudoeste de São Paulo (Figura 35).

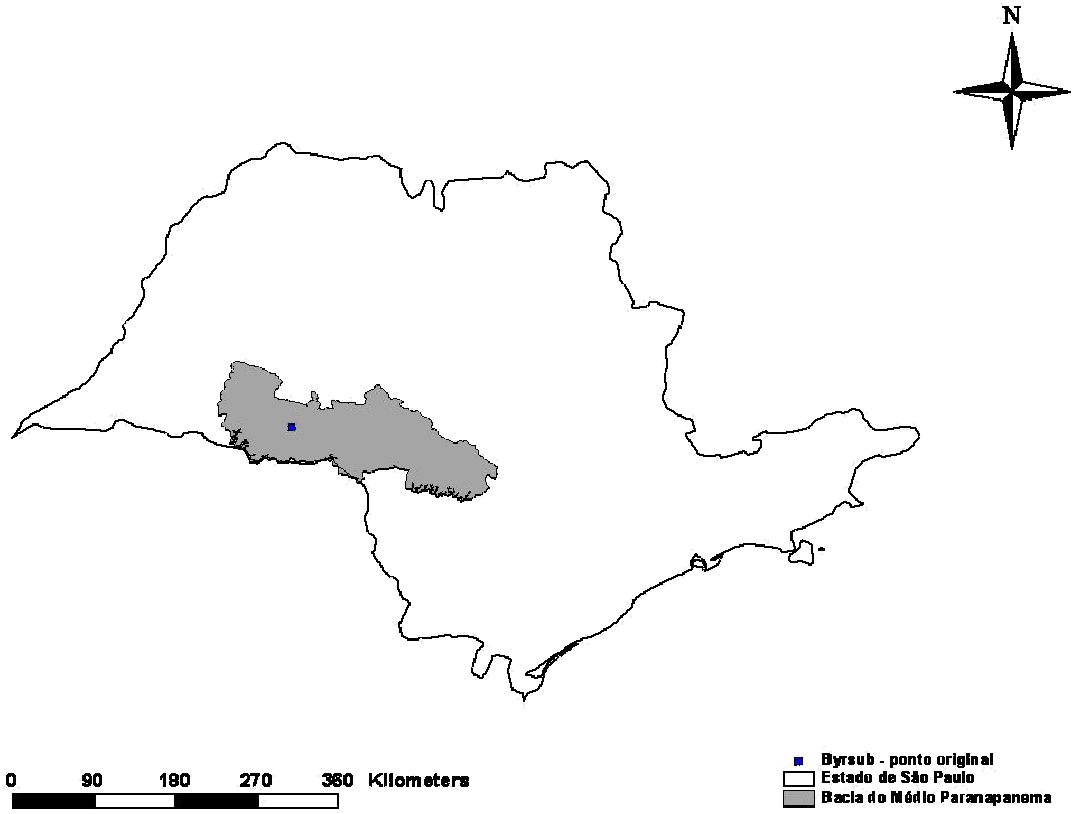


Figura 35: Localização da Bacia do Médio Paranapanema no Estado de São Paulo e do ponto original de ocorrência de *Byrsonima subterranea*.

Desenvolvimento do algoritmo

Para tratar os casos de espécies com ocorrência única, foi desenvolvido um modelo baseado em distância ambiental, que mede a distância entre o valor dos parâmetros ambientais no ponto geográfico de ocorrência da espécie e os valores desses mesmos parâmetros na área de estudo. Quanto menor a distância, maior é a similaridade ambiental entre os pontos.

A distância entre os valores (previamente normalizados) dos parâmetros ambientais foi calculada através da fórmula de distância euclidiana.

$$D = \sqrt{\sum (Xi - Yi)^2}$$

O processo consiste em caracterizar a região adjacente ao ponto de origem para saber o quanto essas áreas são similares ou não em relação às variáveis ambientais (Figura 36) utilizadas. O modelo final é resultado da projeção desses valores de volta ao espaço geográfico.

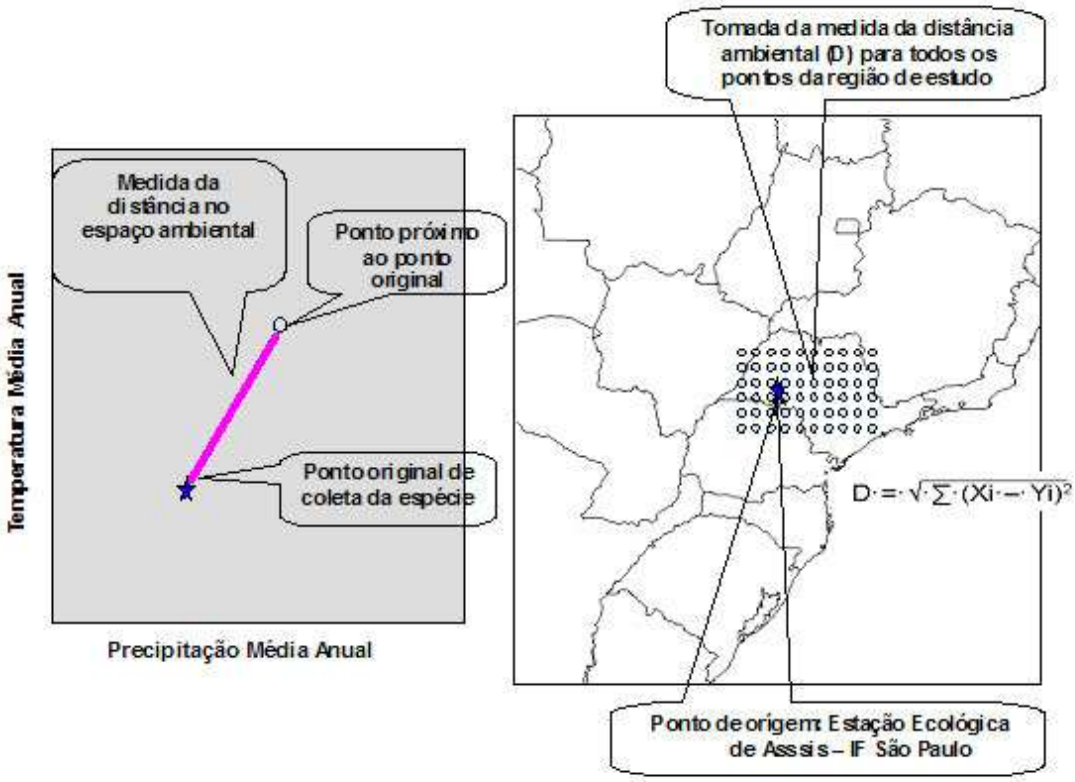


Figura 36: Método de modelagem de nicho fundamental, baseado em distância ambiental (distância euclidiana) para espécies com um único ponto de amostragem.

Parâmetros ambientais utilizados na modelagem

- Clima (precipitação e temperatura) – IPCC (Intergovernmental Panel on Climate Change) dados de 1961-1990. Resolução de 50Km http://www.ipcc.ch
- Dados topográficos do U.S. Geological Surveys (grupo de dados do Hydro-1k, com resolução de 1km) que incluem 17 camadas de dados de

topografia (elevação, relevo, aspecto, e capacidade de retenção de umidade, etc.). http://edcdaac.usgs.gov/gtopo30/hydro/

Busca da espécie no campo

A partir do mapa gerado para a Bacia do Médio Paranapanema, indicando as áreas de maior probabilidade de ocorrência de *Byrsonima subterranea* (Figuras 37 e 38), foram realizadas algumas excursões de coleta visando a verificar a ocorrência da espécie nas áreas indicadas. Com o apoio de um GPS e do mapa dos fragmentos remanescentes de cerrado na região (KRONKA et al. 1998), foram visitadas seis áreas, dentre as mais prováveis, especialmente aquelas que continham fisionomias abertas de cerrado.

A busca em campo dentro de cada fragmento foi direcionada para a fisionomia em que a espécie foi observada no ponto original de coleta, ou seja: cerrado com estrato arbóreo baixo e descontínuo, com grande quantidade de luz direta sobre o solo (Figura 34).

Método de modelagem

As técnicas de modelagem que envolvem análise ambiental utilizam algoritmos que procuram por condições ambientais semelhantes àquelas onde as espécies foram encontradas.

O uso de técnicas de modelagem de distribuição geográfica de espécies é particularmente indicado em situações nas quais é preciso tomar decisões mas não há um conjunto grande de informações disponível. Isto pode acontecer tanto para regiões como para táxons específicos, para os quais a coleta de dados ainda é bastante precária para o tamanho, riqueza biológica ou importância da região ou táxon estudado.

Esses resultados podem auxiliar na indicação de novas áreas de coleta, onde existam as condições ambientais ideais para a sobrevivência das espécies analisadas.

Resultados e discussão

O modelo gerado (Figura 37) de distribuição geográfica potencial para *Byrsonima subterranea* no Estado de São Paulo e Estados e países vizinhos

mostra uma área potencial de ocorrência bastante grande. Aproximando-se para a escala da Bacia do Médio Paranapanema (Figura 38), verifica-se que as duas primeiras classes de cor delimitam as áreas mais propícias de se encontrar a espécie no campo.

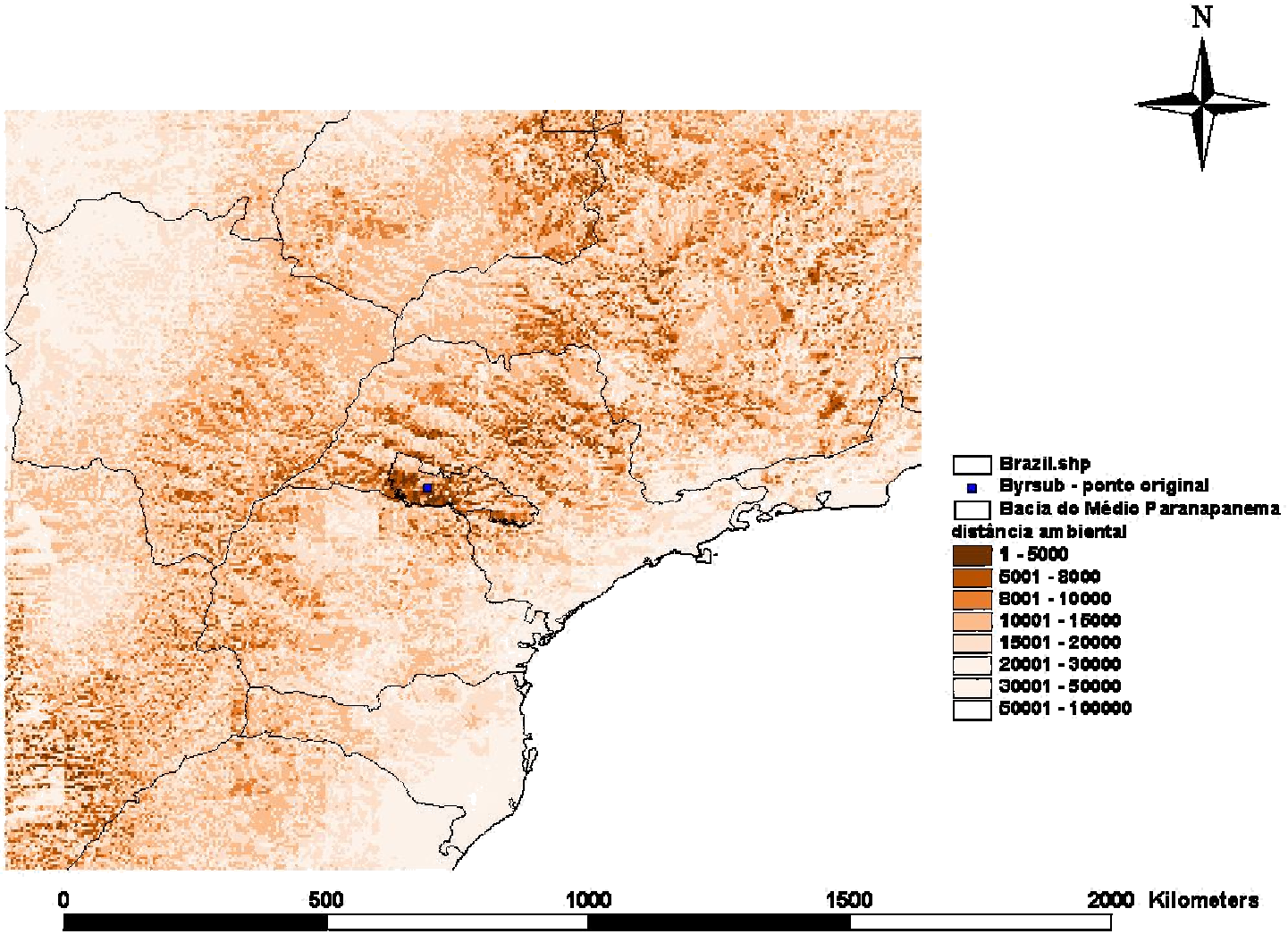


Figura 37: Modelo de distribuição potencial de *Byrsonima subterranea* baseado em um único registro de ocorrência (Estação Ecológica de Assis, Município de Assis, SP) utilizando dados climáticos e topográficos. O detalhe representa a Bacia do Médio Paranapanema e é apresentado na Figura 38.

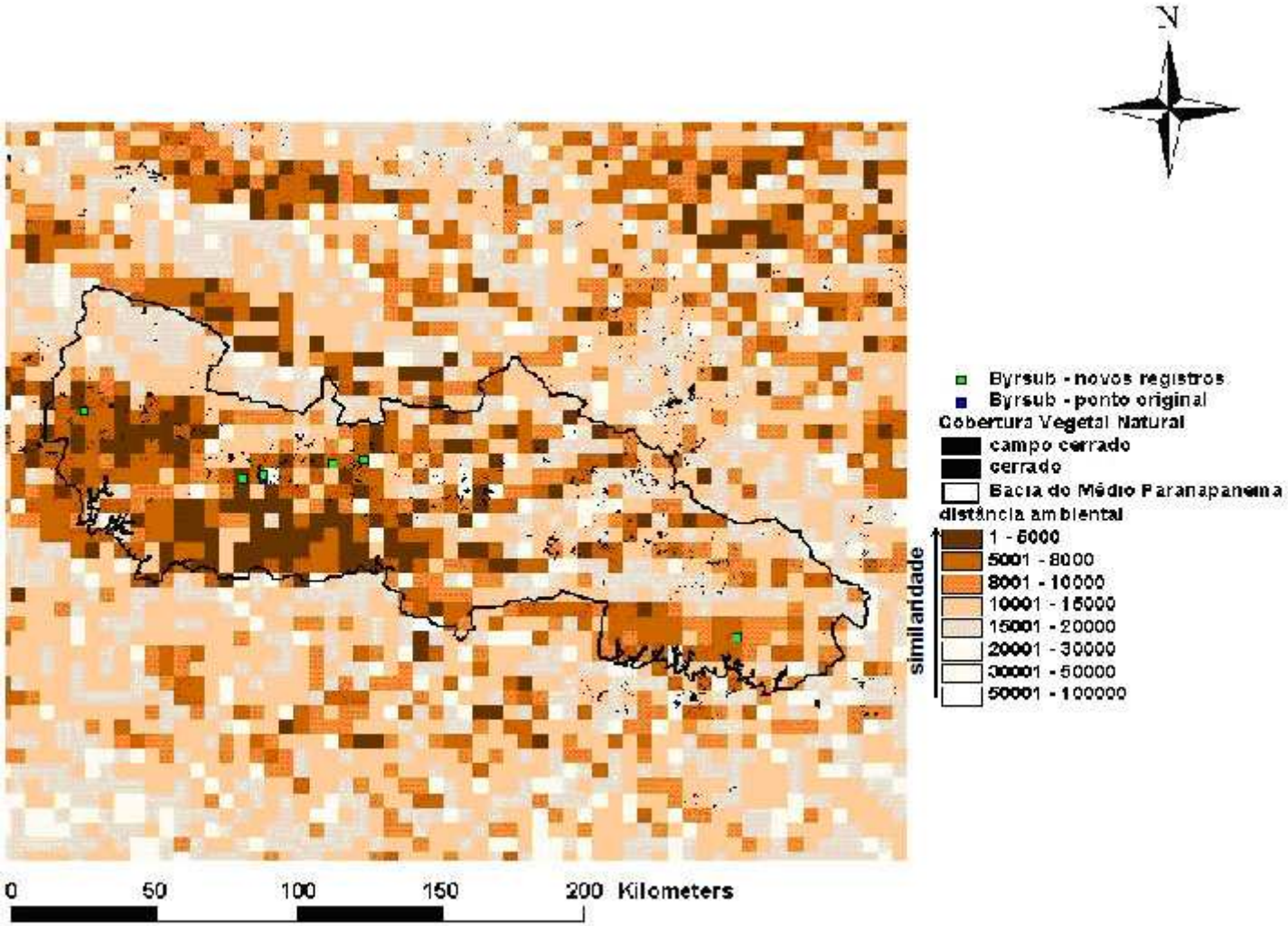


Figura 38: Resultado do trabalho de campo com a inclusão de mais seis registros de ocorrência para a espécie *Byrsonima subterranea* coletadas nas áreas de maior probabilidade de ocorrência da espécie associadas ao tipo de vegetação na região da Bacia do Médio Paranapanema.

Como resultado do trabalho de campo foram registradas mais seis ocorrências dessa espécie, em quatro municípios da Bacia: Assis, Echaporã, Martinópolis e Campos Novos Paulista (Figura 38).

Com base nos resultados obtidos neste estudo, considera-se que o caso de *Byrsonima subterranea* parece ser mais um problema de coleta, associado à dificuldade de visualização da espécie no campo, do que o caso de uma espécie com distribuição restrita e realmente ameaçada de extinção.

O modelo resultante se mostrou bastante eficaz no direcionamento do trabalho de campo. Esta metodologia pode ser reaplicada para outros grupos taxonômicos que se apresentem sob ameaça de extinção, com distribuição restrita ou que sejam lacunas de conhecimento e de coleta. É de se esperar que com a inclusão de dados pedológicos o resultado seja ainda melhor. No caso específico do algoritmo utilizado, os dados pedológicos a serem utilizados na modelagem devem ser obrigatoriamente contínuos (não categorizados), uma vez que não é possível comparar, através de distância, diferentes classes de informação, no caso, diferentes classes de solos.

Conclusões

- Métodos de modelagem podem auxiliar na busca de espécies no campo, técnica que pode ser utilizada tanto para espécies com distribuição restrita, ameaçadas de extinção, ou espécies dificilmente coletadas.
- A espécie *Byrsonima subterranea* não deve ser considerada em risco de extinção uma vez que trata-se de um caso típico de coletas deficientes.
- Estes métodos podem tornar mais eficiente o trabalho de campo, diminuindo o custo e o tempo empregados em viagens de coleta.
- Em escalas locais, os dados ambientais precisam incluir fatores que delimitam a ocorrência de determinado tipo de vegetação. No caso deste tipo de trabalho, a inclusão de dados pedológicos contínuos (não categorizados) é de fundamental importância para melhorar a qualidade dos modelos gerados.

Comentários gerais

Aplicação de modelagem para estudo de distribuição geográfica de espécies vegetais

Para se conseguir bons resultados com modelagem é preciso observar alguns pontos e realizar algumas análises pré e pós modelagem. A seguir são comentados alguns procedimentos para se alcançar este objetivo.

Qualidade dos dados bióticos e abióticos

A qualidade dos resultados da aplicação de técnicas de modelagem é altamente dependente da qualidade das informações biológicas (registros de ocorrência geográfica do objeto de estudo) e dos dados ambientais (mapas temáticos que caracterizam as condições ambientais de uma determinada região). Quanto melhor for a resolução, ou seja, quanto mais detalhada for a escala dos mapas e quanto mais precisos forem os registros dos pontos de ocorrência das espécies, melhor será o resultado final da modelagem.

Compatibilidade de escalas

Outro fator importante a ser observado em modelagem é a compatibilidade entre os dados bióticos e os dados abióticos disponíveis. Por exemplo, dados ambientais com resolução de 1km não são adequados para modelar a partir de registros de ocorrência de espécies com precisão muito menor. Exemplo comum dessa situação acontece com boa parte do material científico depositado em coleções, em que localidades ou mesmo municípios são utilizados para georeferenciar a ocorrência de uma uma determinada espécie. Neste caso é melhor usar dados ambientais com precisão menor para adequar a questão a ser analisada pela modelagem com a escala dos dados utilizados. O emprego desses dados deve ser adequado para a abrangência geográfica ou à questão a ser avaliada pela modelagem, e é importante observar que

nesses casos o modelo resultante terá sempre a precisão da pior resolução utilizada.

Então, é de fundamental importância que, ao se iniciar o trabalho com modelagem, alguns aspectos sejam previamente analisados:

- Quanto melhor for a qualidade dos dados e mais detalhada for a informação ambiental e biológica disponível, melhor será o modelo resultante.
- Deve-se sempre buscar a compatibilidade máxima possível de resolução entre os dados bióticos e abióticos utilizados.
- A resolução dos dados precisa ser adequada para a abrangência geográfica (tamanho) da área de estudo ou para a pergunta que se deseja responder.
- A escolha do algoritmo para a modelagem deve adequar-se tanto à questão a ser respondida como ao número de pontos georeferenciados disponíveis.

Referências Bibliográficas

ACIESP. (1997). *Glossário de Ecologia*. 352p.

ANDERSON, R. P.; LAVERDE, M.; PETERSON, A. T. (2002). *Geographical distributions of spiny pocket mice in South America: Insights from predictive models*. Global Ecology and Biogeography. 11: 131-141.

ANDERSON, R. P.; LEW, D.; PETERSON, A. T. (2003). *Using intermodel variation in error components to select best subsets of ecological niche models.* Ecological Modelling. 162: 211-232.

AUSTIN, M. P.; NICHOLLS, A. O.; DOHERTY, M. D.; MEYERS, J. A. (1994). *Determining species response functions to an environmental gradient by means of a beta-function*. Journal of Vegetation Science. 5: 215-228.

BEHLING, H. (1998). *Late Quaternary vegetational and climatic changes in Brazil*. Review of Palaeobotany and Palynology. 99: 143-156.

---------. (2003). *Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland*. Vegetation History and Archaeobotany. 12: (4): 263-270.

BOJORQUEZ-TAPIA, L. A.; AZUARA, I.; EZCURRA, E.; FLORES V., O. A. (1995). *Identifying conservation priorities in Mexico through geographic information systems and modeling*. Ecological Applications. 5: 215-231.

BORGONOVI, M.; CHIARINI, J. V. (1965). *Cobertura vegetal do Estado de São Paulo. I. Levantamento por fotointerpretação das áreas cobertas com cerrado, cerradão e campo em 1962*. Bragantia. 24: 159-172.

BOSTON, T.; STOCKWELL, D. R. B. (1994). *Interactive species distribution reporting, mapping and modeling using the World Wide Web*. http://kaos.erin.gov.au/database/WWW-Fall94/species_paper_new.html

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. (1984). *Classification and regretion trees*. Chapman and Hall. New York.

CHEN, G.; PETERSON, A. T. (2002). *Prioritization of areas in China for biodiversity conservation based on the distribution of endangered bird species.* Bird Conservation International. 12: 197-209.

CPTI. (2000). *Relatório da Situação dos Recursos Hídricos da UGRHI do Médio Paranapanema*. Assis, SP. CRH-MP.

DARWIN, C. R. (1859). *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in te Struggle for Life*. John Murray, Albermarle Street. London.

DAVIS, L. (1996). *Handbook of Genetic Algorithms*. Thomson Computer Press. Boston, Massachusetts.

DURIGAN, G.; BACIC, M. C.; FRANCO, G. A. D. C.; SIQUEIRA, M. F. (1999). *Inventário florístico do cerrado na Estação Ecológica de Assis, SP*. Hoehnea. 26: (2): 149-172.

DURIGAN, G.; BAITELO, J. B.; FRANCO, G. A. D. C.; SIQUEIRA, M. F. D. (2004a). *Plantas do Cerrado Paulista. Imagens de uma paisagem ameaçada*. Instituto Florestal, Secretaria do Meio Ambiente do Estado de São Paulo. Japan International Cooperation Agency - JICA. Páginas e Letras Editora e Gráfica. São Paulo, SP.

DURIGAN, G.; LEITÃO-FILHO, H. D. F. (1995). *Florística e fitossociologia de matas ciliares do oeste paulista*. Revista do Instituto Florestal. 7: (2): 197-239.

DURIGAN, G.; RATTER, J. A.; BRIDGEWATER, S.; SIQUEIRA, M. F. D.; FRANCO, G. A. D. C. (2003a). *Padrões fitogeográficos do cerrado paulista sob uma perspectiva regional*. Hoehnea. 30: (1): 39-51.

DURIGAN, G.; SIQUEIRA, M. F. D.; FRANCO, G. A. D. C.; BRIDGEWATER, S.; RATTER, J. A. (2003b). *The vegetation of priority areas for cerrado conservation in São Paulo State, Brazil*. Edinburgh Journal of Botany. 60: (2): 217 - 241.

DURIGAN, G.; SIQUEIRA, M. F. D.; FRANCO, G. A. D. C.; CONTIERI, W. A. (2004b). *A Flora Arbustivo-Arbórea do Médio Paranapanema: Base para a Restauração de Ecossistemas Naturais*. Pesquisas em Conservação e Recuperação Ambiental no Oeste Paulista. Resultados da Cooperação Brasil/Japão. Bôas, O. V. & Durigan, G. nstituto Florestal - Secretaria do Meio Ambiente do Estado de São Paulo e Japan International Cooperation Agency - JICA. Páginas e Letras Editora e Gráfica. São Paulo, SP.

EGBERT, S. L.; PETERSON, A. T.; SANCHEZ-CORDERO, V.; PRICE, K. P. (1998). *Modeling conservation priorities in Veracruz, Mexico*. GIS in natural resource management: Balancing the technical-political equation. Morain, S. High Mountain Press. Santa Fe, New Mexico. 141-150.

EITEN, G. (1970). *A vegetação do estado de São Paulo*. Boletim do Instituto de Botânica. 7: 1-77.

---------. (1972). *The cerrado vegetation of Brazil*. The Botanical Review. 38: (2): 201-341.

FLORIANO, P. R. (2004). *Gestão do Conhecimento em Comunidades de Prática*. http://www.terraforum.com.br/lib/pages/download.php?l_intDocCod=147&l_intView=1

FOGEL, L. J.; OWENS, A. J.; WALSH, M. J. (1966). *Artificial Intelligence through Simulated Evolution*. Wiley.

FRANCO, G. A. D. C. (2002). *Florística e fitossociologia de duas unidades do mosaico florestal da Estação Ecológica dos Caetetus - Floresta Estacional Semidecidual, Gália - SP.* ESALQ-Universidade de São Paulo. Piracicaba.

FUNDAG. (1998). *Diagnóstico do Meio Físico da Região do Médio Paranapanema*. Assis. FUNDAÇÃO DE APOIO A PESQUISA AGRÍCOLA.

GOLDBERG, D. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Inc., Reading. Massachusetts.

GUISAN, A.; ZIMMERMANN, N. E. (2000). *Predictive habitat distribution models in ecology*. Ecological Modelling. 135: 147-186.

HOLLAND, J. (1975). *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor. Michigan.

HUNTLEY, B.; BERRY, P. M.; CRAMER, W.; MCDONALD, A. P. (1995). *Modelling present and potential future ranges of some European higher plants using climate response surfaces*. Journal of Biogeography. 22: 967-1001.

HUTCHINSON, G.E. (1957). Concluding remarks. *Cold Spring Harbor Symp Quantitative Biol.*, **22,** 415–427

IBGE. 1988. Mapa de Vegetação do Brasil. Brasília. Ministério da Agricultura. Instituto Brasileiro de Desenvolvimento Florestal. Secretaria de Planejamento e Coordenação da Presidência da República. Fundação Instituto Brasileiro de Geografia e Estatística - IBGE.

KAGEYAMA, P.; GANDARA, F. B. (2000). *Recuperação de áreas ciliares*. Matas Ciliares: Conservação e Recuperação. RODRIGUES, R. R. & LEITÃO FILHO, H. F. Editora da Universidade de São Paulo/FAESP. São Paulo.

KRONKA, F. J. D. N.; NALON, M. A.; BAITELLO, J. B.; MATSUKUMA, C. K.; PAVÃO, M.; YWANE, M. S. S.-I.; LIMA, L. M. P. R.; KANASHIRO, M. M.; BARRADAS, A. M. F.; BORGO, S. C. (2003). *Levantamento da Vegetação Natural e Caracterização de Uso do Solo no Estado de São Paulo*. XI SBSR - Simpósio Brasileiro de Sensoriamento Remoto. INPE. Belo Horizonte. 2779-2785.

KRONKA, F. J. N.; NALON, M. A.; MATSUKUMA, C. K.; PAVÃO, M.; GUILLAUMON, J. R.; CAVALLI, A. C.; GIANNOTTI, E.; IWANE, M. S. S.; LIMA, L. M. P. R.; MONTES, J.; DEL-CALI, I. H.; HAACK, P. G. (1998). *Áreas de domínio de cerrado no Estado de São Paulo*. Secretaria de Estado do Meio Ambiente, Instituto Florestal. São Paulo. 84.

LEDRU, M.-P. (2002). *Late Quaternary History and Evolution of the Cerrados as Revealed by Palynological Records*. The Cerrados of Brazil - Ecology and Natural History of a Neotropical Savanna. Oliveira, P. E. & Marquis, R. J. Columbia University Press. New York. 33-50.

LEDRU, M.-P.; SALGADO-LABOURIAU, M. L.; LORSCHEITTER, M. L. (1998). *Vegetation dynamics in southern and central Brazil during the last 10,000 yr B.P.* Review of Palaeobotany and Palynology. 99: 131-142.

MAGANA, V.; CONDE, C.; SANCHEZ, O.; GAY, C. (1997). *Assessment of current and future regional climate scenarios for Mexico*. Climate Research. 9: 107-114.

MEIRA-NETO, J. A. A. (1991). *Composicao floristica e fitossociologia de fisionomias de cerrado da Estacao Ecologica de Santa Barbara, Municipio de Aguas de Santa Barbara - SP.* Universidade Estadual de Campinas. Campinas.

MITCHELL, M. (1996). *An Introduction To Genetic Algorithms*. MIT Press.

MITTERMEIER, R. A.; MYERS, N.; MITTERMEIER, C. G.; GIL, P. R. (1999). *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions.* CEMEX Conservation International. Mexico. 430.

MMA. (1998). *Primeiro Relatório Nacional para a Convenção sobre Diversidade Biológica*. Ministério do Meio Ambiente, dos Recursos Hidricos e da Amazônia Legal. Brasilia, DF.

MOTTA, P. E. F.; CURI, N.; FRANZMEIER, D. P. (2002). *Relation of Soils and Geomorphic Surfaces in the Brazilian Cerrado*. The Cerrados of Brazil - Ecology and Natural History of Neotropical Savanna. Oliveira, P. S. & Marquis, R. J. Columbia University Press. New York.

MYERS, N. (1988). *Threatened biotas: "Hot spots" in tropical forests*. Environmentalist. 8: 187-208.

NIX, H. A. (1986). *A biogeographic analysis of Australian elapid snakes*. Atlas of Australian Elapid Snakes. Longmore, R. Australian Government Publishing Service. Canberra. 4-15.

OBERHAUSER, K.; PETERSON, A. T. (2003). *Modelling current and future potencial wintering distributions of eastern North American monarch butterflies*. PNAS (Proceedings of the National Academy of Sciences of the United States of America). 100: (24): 14063-14068.

OBITKO, M. (1998). *Genetic Algorithms*. http://cs.felk.cvut.cz/~xobitko/ga/

OLDEN, J. D.; JACKSON, D. A. (2002). *Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks*. Ecological Modelling. 154: 135-150.

OLIVEIRA, J. B.; CAMARGO, M. N.; ROSSI, M.; CALDERANO FILHO, B. 1999. Mapa pedológico do Estado de São Paulo: legenda expandida. Campinas. Instituto Agronômico/EMBRAPASolos.

OLIVEIRA-FILHO, A. T. D.; RATTER, J. A. (2002). *Vegetation Physiognomies and Woddy Flora of the Cerrado Biome*. The Cerrados of Brazil - Ecology and Natural History of a Neotropical Savanna. Oliveira, P. E. & Marquis, R. J. Columbia University Press. New York. 91-120.

ORTEGA-HUERTA, M. A.; PETERSON, A. T. (2004). *Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico*. Diversity and Distributions. 10: 39-54.

PEARSON, R. G.; DAWSON, T. P.; BERRY, P. M.; HARRISON, P. A. (2002). *SPECIES: A spatial evaluation of climate impact on the envelope of species*. Ecological Modelling. 154: 289-300.

PETERSEN, L. R.; ROEHRIG, J. T. (2001). *West Nile virus: A reemerging global pathogen*. Emerging Infectious Diseases. 7: 1-10.

PETERSON, A. T. (2001). *Predicting Species' Geographic Distributions Based On Ecological Niche Modeling*. Condor. 103: (3): 599–605.

PETERSON, A. T.; ORTEGA-HUERTA, M. A.; BARTLEY, J.; SANCHEZ-CORDERO, V.; SOBERON, J.; BUDDEMEIER, R. H.; STOCKWELL, D. R. B. (2002a). *Future projections for Mexican faunas under global climate change scenarios*. Nature. 416: 626-629.

PETERSON, A. T.; PAPES, M.; KLUZA, D. A. (2003a). *Predicting the potential invasive distributions of four alien plant species in North America*. Weed Science. 51: (6): 863–868.

PETERSON, A. T.; SANCHEZ-CORDERO, V.; BEARD, C. B.; RAMSEY, J. M. (2002b). *Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico*. Emerging Infectious Diseases. 8: 662-667.

PETERSON, A. T.; SCACHETTI-PEREIRA, R.; KLUZA, D. A. (2003b). *Assessment of Invasive Invasive Potential of Homalodisca coagulata in Western North America and South America*. Biota Neotropica. 3: (1):

PETERSON, A. T.; VIEGLAIS, D. A. (2001). *Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem*. BioScience. 51: (5): 363-371.

RATTER, J. A.; BRIDGEWATER, S.; RIBEIRO, J. F.; DIAS, T. A. B.; SILVA, M. R. (2000). *Distribuição das espécies lenhosas da fitofisionomia cerrado sentido restrito nos estados compreendidos no bioma cerrado*. Boletim do Herbário Ezechias Paulo Heringer. 5: 5-43.

RATTER, J. A.; DARGIE, T. C. D. (1992). *An Analysis of the Floristic Composition of 26 Cerrado Areas in Brazil*. Edinburg Journal of Botany. 49: 235
250.

RECHENBERG, I. (1965). *Cybernetic Solution Path of An Experimental Problem*. Ministry of Aviation, Royal Aircraft Establishment. United Kingdom.

SALA, O. E.; CHAPIN-III, F. S.; ARMESTO, J. J.; BERLOW, E.; BLOOMFIELD, J.; DIRZO, R.; HUBER-SANWALD, E.; HUENNEKE, L. F.; JACKSON, R. B.; KINZIG, A.; LEEMANS, R.; LODGE, D. M.; MOONEY, H. A.; OESTERHELD, M.; POFF, N. L.; SYKES, M. T.; WALKER, B. H.; WALKER, M.; WALL, D. H. (2000). *Global biodiversity scenarios for the year 2100*. Science. 287: (5459): 1770-1774.

SALGADO-LABOURIAU, M. L. (1998). *Late Quaternary palaeoclimate in the savannas of South America*. Journal of Quaternary Science. 12: (5): 371-379.

SALGADO-LABOURIAU, M. L.; BARBIERI, M.; K.R.FERRAZ-VICENTINI; M.G.PARIZII. (1998). *A dry climatic event during the late Quaternary of tropical Brazil*. Review of Palaeobotany and Palynology. 99: 115-129.

SALGADO-LABOURIAU, M. L.; CASSETI, V.; FERRAZ-VICENTINI, K. R.; MARTIN, L.; SOUBIÈS, F.; SUGUIO, K.; TURCQ, B. (1997). *Late Quaternary vegetational and climatic changes in cerrado and palm swamp from Central*

Brazil. Palaeogeography, Palaeoclimatology, Palaeoecology. 128: (1-4): 215-226.

SERRA FILHO, R.; CAVALLI, A. C.; GUILLAUMON, J. R.; CHIARINI, J. V.; NOGUEIRA, F. P.; IVANCKO, C. M. A. M.; BARBIERI, J. L.; DONIZELI, P. L.; COELHO, A. G. S.; BITTENCOURT, I. (1974). *Levantamento da cobertura natural e do reflorestamento no Estado de São Paulo*. Boletim Técnico do Instituto Florestal. 11: 1-56.

SILVEIRA, É. R. D. (2001). *Recuperação da mata ciliar do Córrego Tarumã (Tarumã, SP): aspectos estruturais, florísticos e ambientais de quatro diferentes modelos florestais, dez anos após o plantio.* Universidade de São Paulo. São Carlos.

SIQUEIRA, M. F. D.; PETERSON, A. T. (2003). *Consequences of Global Climate Change for Geographic Distributions of Cerrado Tree Species*. Biota Neotropica. 3: (2): http://www.biotaneotropica.org.br/v3n2/pt/abstract?article+BN00803022003

SMA-SP. (1997). *Cerrado: bases para a conservação e uso sustentável das áreas de cerrado do Estado de São Paulo*. Secretaria de Estado do Meio Ambiente de São Paulo. São Paulo, SP. 113.

SOBERON, J.; PETERSON, A. T. (2005). *Interpretation of models of fundamental ecological niches and species' distributional areas*. Biodiversity Informatics. 2: 1-10.

STOCKWELL, D. R. B. (1999). *Genetic algorithms II*. Machine learning methods for ecological applications. Fielding, A. H. Kluwer Academic Publishers. Boston. 123-144.

STOCKWELL, D. R. B.; NOBLE, I. R. (1992). *Induction of sets of rules from animal distribution data: A robust and informative method of analysis*. Mathematics and Computers in Simulation. 33: 385-390.

STOCKWELL, D. R. B.; PETERS, D. P. (1999). *The GARP modelling system: Problems and solutions to automated spatial prediction*. International Journal of Geographic Information Systems. 13: 143-158.

STOCKWELL, D. R. B.; PETERSON, A. T. (2002). *Effects of sample size on accuracy of species distribution models*. Ecological Modelling. 148: 1-13.

THOMAS, C. D.; CAMERON, A.; GREEN, R. E.; BAKKENES, M.; BEAUMONT, L. J.; COLLINGHAM, Y. C.; ERASMUS, B. F. N.; SIQUEIRA, M. F. D.; GRAINGER, A.; HANNAH, L.; HUGHES, L.; HUNTLEY, B.; JAARSVELD, A. S. V.; MIDGLEY, G. F.; MILES, L.; ORTEGA-HUERTA, M. A.; PETERSON, A. T.; PHILLIPS, O. L.; WILLIAMS, S. E. (2004). *Extinction risk from climate change*. Nature. 427: 145-148.

VARGAS, M. A. T.; HUNGRIA, M. (1997). *Biologia dos solos dos Cerrados*. EMBRAPA-CPAC. Planaltina. 524.

VICTOR, M. A. D. M. (1975). *A Devastação Florestal*. Sociedade Brasileira de Silvicultura. 48p.

YEE, T. W.; MITCHELL, N. D. (1991). *Generalized additive models in plant ecology*. Journal of Vegetation Science. 2: 587-602.

Anexo 1 - Algoritmos Genéticos

Segundo Mitchell (1996), entre as décadas de 1950 e 60 vários grupos de cientistas da computação começaram a estudar os sistemas evolucionários que se baseavam na idéia da teoria da evolução das espécies (DARWIN 1859). Esta idéia 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, por exemplo em problemas de engenharia aeronáutica (RECHENBERG 1965) e em problemas de ciência da computação (FOGEL et al. 1966).

No início da década de 70 Holland e seu grupo de pesquisa começaram a estudar o fenômeno da adaptação das espécies como ocorre na natureza visando 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 tinha 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

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 (DAVIS 1996; MITCHELL 1996).

Para ilustrar como estes conceitos são empregados na resolução de um problema através de um algoritmo genético, será utilizado 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, 4) e q = (16, 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 = 1001110010010100
Q = 1000010001100100
```

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 16 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 39 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*.

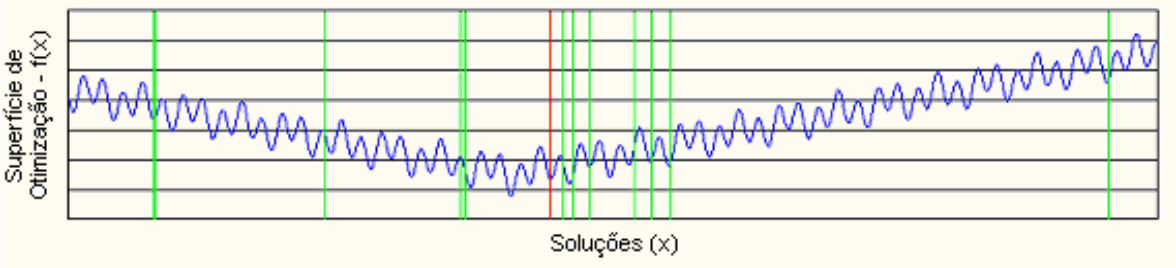


Figura 39: Exemplo de superfície de otimização. Fonte: (OBITKO 1998). As barras verticais representam exemplos de soluções presentes no espaço de soluções para um problema arbitrário. A curva em azul representa a função de otimização.

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 da 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, aplica-se um conjunto de operadores heurísticos para modificar soluções existentes, obtendo novas possíveis 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 pais. Um exemplo de permutação entre os cromossomos P e Q definidos anteriormente na posição 7 pode ser assim representado:

Antes da permutação tinha-se P e Q:

P = 10011100-10010100 p = (156, 18, 4); f(p) = 140

Q = 10000100-01100100 q = (16, 8, 100); f(q) = 500

Após a permutação tem-se P' e Q':

P' = 10011100-01100100 p' = (156, 8, 100); f(p') = 646

Q' = 10000100-10010100 q' = (16,18, 4); f(q') = 0

Como resultado tem-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. Como exemplo de uma mutação no cromossomo P, na posição 17, temos:

Antes da mutação:

P = 10011100<1>0010100 p = (156, 18, 4); f(p) = 140

Após a mutação P':

P = 10011100<0>0010100 p' = (156, 2, 4); f(p') = 172

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

Estrutura do 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 apresentada.

Tipos de Regras utilizadas

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. No algoritmo genético a pré-condição é uma expressão lógica que define uma região do espaço ambiental 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. Essas regras usam apenas um único valor de cada uma das variáveis ambientais na sua pré-condição. Um exemplo de regra atômica seria:

*Se **Temp = 23° C e Elevação = 2000m***

*então **táxon está presente***

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 28° C 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). Este algoritmo procura determinar uma região com características climáticas similares às de onde a espécie é encontrada, buscando uma região única que identifique todas as áreas com clima similar ao dos pontos de origem (BOSTON e STOCKWELL 1994). Já o GARP considera todas as formas de dados ambientais (dados contínuos ou categorizados) e tipos que forem incluídos pelo pesquisador, não somente o clima, e trabalha com regras menos restritivas que o algoritmo anterior.

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 28° C 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.

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

$$y = c_0 + c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

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 40).

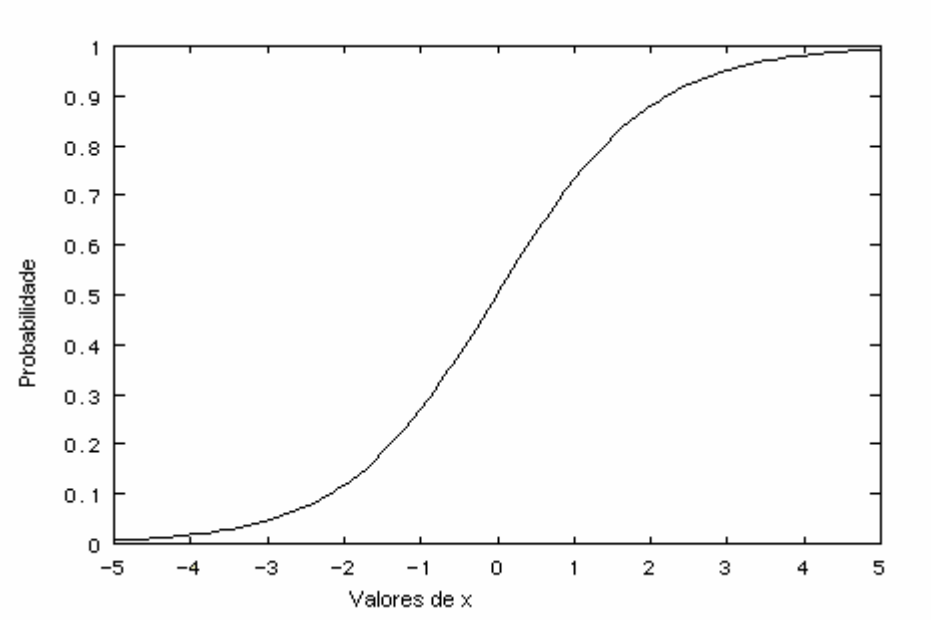


Figura 40: Exemplo de função de probabilidade na regressão logística.

GARP

O algoritmo GARP define o modelo de nicho fundamental das espécies através de um conjunto de regras como as apresentadas anteriormente. As regras presentes num modelo de nicho 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 10° C e Elevação está entre 1000 e 2000m**

então **táxon está presente** (regra de faixa)

Regra 2: Se **Temp é de 18° C e Elevação é 1500m**

então **táxon está ausente** (regra atômica)

Regra 3: Se **$p = 1 / [1 + e^{-y}] > 0.75$**,

onde **$y = 0.2 + Temp \times 0.45 + Elevação \times -0.12$**

então **táxon está presente** (regra tipo logit)

<u>*Codificação das Regras no Algoritmo Genético*</u>

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 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 ou Ausência") também é codificada como um gene, podendo sofrer alterações durante a execução do algoritmo.

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.

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%

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%

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, baseado nos pontos de teste.

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. Nesta operação aquelas regras que têm valor da função de adaptação abaixo de um certo limiar pré-definido são eliminadas da população (conjunto de regras).

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

Quando um número pré-determinado 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 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.

Anexo 2: Lista dos fragmentos de vegetação natural utilizados na modelagem das espécies da Bacia do Médio Paranapanema.

F.E.S = Floresta Estadual Semidecidual

M.C. = Mata Ciliar

Código Localidade	Longitude	Latitude	Município	Bioma/Ecossistema
A1	-50.04	-22.5428	Campos Novos Paulista	Cerrado
A16	-50.0339	-22.5231	Campos Novos Paulista	Cerrado
A17	-49.9922	-22.5303	Campos Novos Paulista	Cerrado
A20	-50.1151	-22.547	Campos Novos Paulista	Cerrado
A29	-50.1365	-22.558	Echaporã	Cerrado
A6	-50.029	-22.4863	Ocauçu	Cerrado
A8	-50.0164	-22.5012	Ocauçu	Cerrado
AV1	-48.984	-23.2607	Avaré	Cerrado
AV2	-48.9478	-23.1608	Avaré	Cerrado
AV3	-48.8286	-22.9559	Ávaré	Cerrado
AV4	-48.8952	-22.8881	Avaré	Cerrado
AV5	-48.8454	-22.8954	Avaré	Cerrado
B1	-49.7093	-22.6571	São Pedro do Turvo	Cerrado
B10	-49.8001	-22.6968	São Pedro do Turvo	Cerrado
B16	-49.7358	-22.6757	São Pedro do Turvo	Cerrado
B2	-49.6135	-22.6452	Ubirajara	Cerrado
B34	-49.697	-22.6439	São Pedro do Turvo	Cerrado
B5	-49.6359	-22.6915	São Pedro do Turvo	Cerrado
B6	-49.6808	-22.714	São Pedro do Turvo	Cerrado
B9	-49.6865	-22.7271	São Pedro do Turvo	Cerrado
BERR	-50.5167	-22.7	Tarumã	M.C.
C13	-49.0459	-22.6065	Agudos	Cerrado
C14	-49.0262	-22.5693	Agudos	Cerrado
C15	-49.0671	-22.6029	Agudos	Cerrado
C22	-48.9921	-22.6214	Borebi	Cerrado
C30	-48.9977	-22.6616	Borebi	Cerrado
CANA	-50.4167	-22.75	Cândido Mota	M.C.
EEA2	-50.3784	-22.5977	Assis	F.E.S.
EESB	-49.2238	-22.8257	Águas de Santa Bárbara	F.E.S.
G0	-50.9788	-22.3714	Rancharia	Cerrado
G32	-50.9739	-22.3772	Rancharia	Cerrado
GA	-49.68	-22.38	Gália	F.E.S.
GB	-49.6917	-22.4083	Gália	F.E.S.
GG	-49.7167	-22.4	Alvinlândia	F.E.S.
IT1	-48.8823	-23.1879	Avaré	Cerrado
IT2	-48.7841	-23.1377	Avaré	Cerrado
IT3	-48.6561	-23.1229	Itatinga	Cerrado
IT4	-49.4203	-22.8163	Santa Cruz do Rio Pardo	Cerrado
MP1	-50.4653	-22.5372	Paraguaçu Paulista	Ecótono
MP10	-50.4314	-22.5758	Assis	F.E.S.

MP14	-50.4314	-22.6158	Assis	Cerrado
MP15	-50.3503	-22.615	Assis	Ecótono
MP16	-50.4078	-22.5517	Assis	Ecótono
MP19	-50.4719	-22.6036	Assis	Ecótono
MP2	-50.3531	-22.6478	Assis	F.E.S.
MP20	-50.3239	-22.5747	Assis	Ecótono
MP21	-50.3192	-22.6061	Assis	Ecótono
MP22	-50.3525	-22.6528	Assis	Ecótono
MP24	-50.3131	-22.5544	Platina	Cerrado
MP25	-50.3131	-22.5775	Platina	Ecótono
MP27	-50.4836	-22.6144	Assis	F.E.S.
MP28	-50.4406	-22.6003	Assis	Cerrado
MP29	-50.4694	-22.6175	Assis	F.E.S.
MP37	-50.4731	-22.6269	Assis	Ecótono
MP4	-50.3333	-22.6181	Assis	F.E.S.
MP49	-50.3233	-22.6083	Assis	Ecótono
MP50	-50.3639	-22.5397	Assis	F.E.S.
MP51	-50.4081	-22.5631	Assis	F.E.S.
MP52	-50.52	-22.6131	Assis	F.E.S.
MP53	-50.3325	-22.6581	Assis	F.E.S.
MP7	-50.3236	-22.5544	Platina	Cerrado
MP8	-50.3392	-22.5619	Assis	Ecótono
SLUIS	-50.6667	-22.7333	Tarumã	M.C.