

LARISSA MAIARA FRAGA

Orientador: Frederico Gadelha Guimarães

**MÓDULO DE CALIBRAÇÃO AUTOMÁTICA DE
MODELOS ESPACIALMENTE-EXPLÍCITOS PARA A
MODELAGEM AMBIENTAL**

Ouro Preto
Novembro de 2010

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**MÓDULO DE CALIBRAÇÃO AUTOMÁTICA DE
MODELOS ESPACIALMENTE-EXPLÍCITOS PARA A
MODELAGEM AMBIENTAL**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

LARISSA MAIARA FRAGA

Ouro Preto
Novembro de 2010



UNIVERSIDADE FEDERAL DE OURO PRETO

FOLHA DE APROVAÇÃO

Módulo de Calibração Automática de Modelos Espacialmente-Explícitos
para a Modelagem Ambiental

LARISSA MAIARA FRAGA

Monografia defendida e aprovada pela banca examinadora constituída por:

Dr. FREDERICO GADELHA GUIMARÃES – Orientador
Universidade Federal de Minas Gerais

Dr. TIAGO GARCIA DE SENA CARNEIRO
Universidade Federal de Ouro Preto

Dr. HAROLDO GAMBINI SANTOS
Universidade Federal de Ouro Preto

Mestrando ANDRÉ SIQUEIRA RUELA
Universidade Federal de Ouro Preto

Ouro Preto, Novembro de 2010

Resumo

Os modelos desenvolvidos em modelagem ambiental podem apresentar muitos parâmetros, podem ser não lineares, complexos e estocásticos, tornando o problema de calibração neste contexto um problema relativamente complexo do ponto de vista da otimização. Este trabalho descreve a implementação de um módulo de calibração de parâmetros de modelos desenvolvidos no ambiente TerraME utilizando o paradigma da Computação Evolutiva. De forma a superar os desafios identificados acima, é importante incorporar ao framework TerraME um módulo de calibração automática e flexível de modelos complexos, como os tipicamente encontrados em modelagem ambiental. Este artigo apresenta o módulo de calibração baseado em algoritmos genéticos e ilustra sua aplicação em um estudo de caso, especificamente, um modelo para dinâmica populacional do *Aedes aegypti* baseado em autômatos celulares. Os resultados ilustram a utilização do módulo de calibração e o potencial da ferramenta em modelagem ambiental de fenômenos complexos.

Abstract

Models developed in environmental modeling usually present many parameters, are nonlinear, complex and stochastic models, making the calibration of their parameters a relatively complex optimization problem. This paper describes the implementation of a calibration module based on evolutionary computation for calibrating models developed using the TerraME modeling environment. In order to overcome challenges mentioned before, it is important to incorporate a flexible and automatic calibration module to the TerraME framework for complex models, such as the ones typically found in environmental modeling. This paper presents the calibration module based on genetic algorithms and illustrates its application in a case study, specifically, the epidemiologic model based on cellular automata for Dengue disease at the state of Rio de Janeiro. The results illustrate the utilization of the calibration module and the potential of the proposed tool in modeling complex phenomena.

Dedico este trabalho à minha família e meus amigos, únicos motivos que me fazem buscar mais realizações.

Agradecimentos

Agradeço à minha família pelo apoio e suporte em todos os momentos difíceis, em especial a minha mãe Beth, vó Anita, “irmãos” Italo e Bruna, primas e primos, tias e tios.

Aos meus amigos da computação, pelas companhias nas noites não dormidas, pelo incentivo e cooperação nas “ferrações”, principalmente a Ciça, Gegel, Tales, Rodrigo, Kokinha, Marcão, Titi e Rodolfo. Sem vocês seria impossível.

À família que Ouro Preto me trouxe pelo bem enorme que me fazem, Maria, Gabi, Ed, Vini, Gabriel e Ju, obrigada.

Aos mestres, responsáveis pela minha formação, Lucília, Zé Maria, Rabelo, David e especialmente à meus orientadores, Fred e Tiago.

Sumário

1	Introdução	1
1.1	Motivação e Relevância	3
1.2	Objetivos Gerais e Específicos	4
1.3	Organização da monografia	4
2	Modelagem Ambiental e a Plataforma TerraME	6
2.1	Modelagem Ambiental	6
2.2	O Ambiente TerraME	7
2.3	Conclusão	9
3	Algoritmos Genéticos e Calibração de Modelos	11
3.1	Algoritmos Genéticos	11
3.1.1	Representação	12
3.1.2	Cruzamento	13
3.1.3	Mutação	13
3.1.4	Seleção	13
3.2	Calibração de Modelos	14
3.2.1	Modelos Estocásticos	16
3.3	Conclusão	18
4	Resultados Obtidos	19
4.1	Estudo de Caso: Aedes Aegypti	19
4.1.1	Modelo Determinístico	23
4.1.2	Modelo Estocástico	24
5	Discussão e Conclusão	27
	Referências Bibliográficas	28

Lista de Figuras

2.1	Ambiente de desenvolvimento e suporte a modelagem ambiental TerraME.	8
2.2	Arquitetura do Ambiente TerraME.	9
3.1	Fluxograma AG	13
3.2	Variação do erro	18
4.1	Modelo Integrado	20
4.2	Bairro de Higienópolis - Rio de Janeiro	20
4.3	Armadilhas espalhadas no bairro de Higienópolis no Rio de Janeiro, RJ.	22
4.4	Monte Carlo x Algoritmo Genético	24
4.5	Modelo Estocástico	25
4.6	Taxa de oviposição ao longo das 53 semana, usando o modelo com múltiplos parâmetros.	26

Lista de Tabelas

3.1	CalibracaoAG	15
4.1	Resultados AG	23
4.2	Resultados AG	23
4.3	Múltiplos parâmetros de calibração	25

Capítulo 1

Introdução

A todo tempo estamos submetidos a mudanças cada vez mais intensas que vêm alterando o funcionamento dos Sistemas Terrestres. Essas mudanças podem ser naturais, antropogênicas ou uma combinação dos dois e têm impacto direto na integridade do meio ambiente e na qualidade de vida das pessoas. Estudos recentes apontam as ações humanas como a principal força direcionadora das alterações sofridas por esses sistemas e indicam que elas estarão associadas a alterações drásticas dos biomas (Salazar et al., 2007). O aumento da temperatura global, por exemplo, poderá implicar em explosões populacionais de vetores de doenças tropicais e resultar em epidemias globais, a mudança na vegetação tem multiplicado desertos e cada vez mais espécies de animais são ameaçadas de extinção.

Planejadores, decisores e todos aqueles que se ocupam em projetar e estabelecer políticas públicas precisam de ferramentas de modelagem que sejam confiáveis, e capazes de capturar a dinâmica e os resultados das ações humanas (Lambin, 1994)(Kaimowitz e Angelsen, 1998)(Turner II et al., 1995). A simulação de processos naturais ou a simulação das interações humano-ambiente são instrumentos de pesquisa de impactos e previsão (Batistella e Moran, 2005) que têm potencial como meio de detectar, identificar, mapear e monitorar alterações nos ecossistemas, independentemente dos seus agentes causais (Coppin et al., 2004). Essa simulação pode ser feita em vários tipos de sistemas, como os sistemas ecológicos, os sistemas climáticos, os sistemas públicos de saúde, os sistemas hidrológicos, os sistemas produtivos ou os sistemas de uso do solo. Modelos computacionais que reproduzem de uma forma satisfatória o fenômeno geográfico sob estudo contribuem para o ganho do conhecimento científico no que diz respeito ao seu funcionamento e este conhecimento pode servir de alicerce para o planejamento e definição de políticas públicas.

O *framework* TerraME é uma plataforma de domínio público para o desenvolvimento de modelos ambientais espacialmente explícitos integrados a um Sistema de Informação Geográfica (SIG), desenvolvido pela parceria TerraLAB-UFOP¹ (Universidade Federal de Ouro Preto) e INPE (Instituto Nacional de Pesquisas Espaciais). Esta plataforma fornece uma linguagem

¹<http://www.terralab.ufop.br/>

de modelagem de alto nível que é utilizada para a descrição dos modelos/algoritmos e sua posterior integração aos bancos de dados geográficos. Atualmente o laboratório dispõe de séries temporais de dados espaciais devidamente armazenadas em bancos de dados geográficos tais como (i) Uso do solo: como no estudo das mudanças de uso e cobertura do solo (LUCC - *Land-Use/Cover Change*, ver (Lambin e Geist, 2006)) na região da Amazônia Brasileira; e (ii) Saúde pública: como no caso do estudo do modelo epidemiológico para a Dengue na cidade do Rio de Janeiro, RJ (SAUDAVEL) (Lana, 2009). Analisando essas massas de dados é possível entender o impacto do fenômeno sob estudo em relação ao modelo desenvolvido, o que lhe dá confiança e credibilidade. O TerraME também dispõe de algoritmos para calibração e validação de modelos espacialmente explícitos que serão utilizados para avaliar o desempenho dos modelos/algoritmos.

A sustentabilidade de ecossistemas tornou-se um objetivo primordial na atualidade trazendo como uma de suas consequências, a contínua necessidade de precisão e atualização de banco de dados (Coppin et al., 2004). Dentre os principais problemas em modelagem ambiental, podemos destacar a calibração dos modelos desenvolvidos. O desenvolvimento de modelos de simulação confiáveis não pode prescindir das etapas de calibração e validação do modelo. Dependendo da complexidade do modelo e do método escolhido, o tempo despendido na execução dessa calibração pode ser bem elevado. Esse retardo influencia diretamente na evolução do modelo, pois qualquer alteração visando melhora em seu desenvolvimento terá uma reflexão mais demorada. Em particular, ao desenvolver modelos espacialmente explícitos baseados em autômatos celulares, o número de parâmetros a serem ajustados pode ser muito elevado, e a função de ajuste do modelo pode apresentar relação não linear em relação aos seus parâmetros que também podem ser complexos ou multimodais. Por fim, é comum que modelos em modelagem ambiental possuam parâmetros estocásticos ou mistos, fazendo com que a função de avaliação da qualidade do modelo seja estocástica, ao invés de determinística. Todas essas características podem tornar o problema de calibração em modelagem ambiental um problema relativamente complexo do ponto de vista da otimização, em que técnicas heurísticas podem ser indicadas.

Alguns trabalhos na literatura têm abordado o uso de técnicas de inteligência computacional e algoritmos evolutivos em modelagem ambiental. Por exemplo, D'Ambrosio et al. (D'Ambrosio et al., 2007) utilizam algoritmos genéticos paralelos para a calibração de um modelo de deslizamento de terras baseado em autômatos celulares. Algoritmos genéticos paralelos também são usados para calibração de modelos de processos geológicos tais como fluxo de detritos e lava em (D'Ambrosio e Spataro, 2007). Uma revisão da aplicação de técnicas de inteligência artificial e computacional em sistemas de modelagem ambiental é apresentada em (Chen et al., 2008), cobrindo raciocínio baseado em casos, lógica fuzzy, redes neurais artificiais, algoritmos genéticos, sistemas multiagentes, autômatos celulares e inteligência de enxame (*swarm intelligence*). Métodos de computação bio-inspirada e de aprendizagem de máquina

aplicados no âmbito da ecologia computacional são discutidos em (Recknagel, 2006). Algoritmos genéticos também têm sido amplamente estudados para evoluir regras e comportamentos em autômatos celulares (Mitchell et al., 1994) (Breukelaar e Bäck, 2005).

Nesse contexto, uma ferramenta de calibração que atenda a todas as peculiaridades de um fenômeno modelado se faz necessária. Essa ferramenta deve ser, principalmente, eficiente no quesito de tempo de retorno de resultados. A proposta a seguir descreve a implementação de um módulo de calibração de parâmetros de modelos desenvolvidos no ambiente TerraME utilizando o paradigma da Computação Evolutiva. De forma a superar os desafios identificados acima, é importante incorporar ao *framework* TerraME um módulo de calibração automática e flexível de modelos complexos, como os tipicamente encontrados em modelagem ambiental. Este trabalho apresenta o módulo de calibração baseado em algoritmos genéticos e ilustra sua aplicação em um estudo de caso, especificamente, o modelo epidemiológico para a Dengue no estado do Rio de Janeiro. Os resultados ilustram a utilização do módulo de calibração fazendo um paralelo entre a calibração proposta com a anteriormente adotada - utilizando Monte Carlo - e potencial da ferramenta em modelagem ambiental de fenômenos complexos.

1.1 Motivação e Relevância

O AG é uma opção interessante como ferramenta de uso geral de calibração em modelagem ambiental pelas razões listadas a seguir:

- Modelos baseados em autômatos celulares, pela própria natureza de modelagem espacialmente distribuída, em geral apresentam muitos parâmetros desconhecidos a serem ajustados;
- Modelos desenvolvidos em modelagem ambiental em geral apresentam função de erro de calibração não linear em relação aos parâmetros de calibração, gerando funções de erro não convexas e multimodais;
- Os parâmetros de calibração podem ser mistos, ou seja, alguns parâmetros podem assumir valores contínuos e outros parâmetros podem assumir valores discretos, tornando necessário um método de calibração flexível e capaz de lidar com variáveis mistas;
- É comum que modelos desenvolvidos em modelagem ambiental possuam parâmetros estocásticos no modelo, de forma que cada simulação do modelo produz um resultado diferente para um mesmo conjunto de valores para os parâmetros de calibração. Portanto, o problema de calibração torna-se um problema de otimização em que a função objetivo apresenta ruído. Algoritmos evolutivos em geral têm se mostrado eficientes no contexto de otimização de funções com ruído;

- A avaliação da qualidade (erro de ajuste) de um dado conjunto de valores para os parâmetros de calibração requer a simulação de um modelo complexo e em geral computacionalmente caro de simular. Dessa forma, a calibração por métodos puramente estocásticos como o Monte Carlo demanda elevado esforço computacional. Essa melhora em relação ao tempo, se deve ao fato do direcionamento da busca no AG ser condicionado a uma heurística caracterizada pela pressão seletiva, tornando o AG um método de busca “estocástico informado”, isto é, direcionado pela pressão seletiva e acumulativa da operação de seleção.

1.2 Objetivos Gerais e Específicos

O objetivo desse trabalho é apresentar uma alternativa de calibração que seja mais vantajosa que a anterior, no sentido de propiciar economia principalmente de tempo. A alternativa proposta foi implementar um módulo de calibração genérico para o ambiente TerraME baseado em algoritmos genéticos (Mitchell, 1998; Eiben e Smith, 2003). Um algoritmo genético (AG) é uma técnica de procura utilizada para encontrar soluções aproximadas em problemas de otimização e busca. Eles consistem numa classe particular de algoritmos evolutivos que usam técnicas inspiradas na biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação.

A ferramenta a ser desenvolvida tem por objetivo ser parametrizada para a calibração de qualquer modelo implementado, independente de suas peculiaridades. O modelador deve apenas chamar a função de calibração e passar como parâmetro as configurações desejadas para o seu problema específico. A linguagem Lua permite que funções sejam passadas como parâmetro a outras funções, o que torna essa chamada ainda mais fácil. Ela também deve oferecer de forma generalizada suporte para modelos deterministas e estocásticos.

Após a conclusão da etapa de implementação, objetiva-se que testes em modelos práticos sejam feitos para validar a ferramenta, testando sua confiança.

1.3 Organização da monografia

O Capítulo 1 é uma introdução objetivando, principalmente, apresentar as motivações e objetivos dessa monografia.

O Capítulo 2 conceitua modelagem ambiental e TerraME. Conhecendo as principais características do fundamento da modelagem podemos entender melhor o seu funcionamento e assim contribuir para a construção. O TerraME é apresentado como ambiente de desenvolvimento e suporte a modelagem espacial dinâmica. Características de sua estrutura e funcionamento são descritas e vantagens da utilização do TerraME por modeladores são apresentadas.

O Capítulo 3 apresenta Algoritmos Genéticos e calibração de modelos. A justificativa e funcionamento do AG são especificados e os detalhes dos métodos específicos usados no módulo

são explicados. Na sequência, conceitua-se calibração de modelos e o módulo desenvolvido com esse intuito é apresentado, descrevendo passo a passo suas características.

O Capítulo 4 descreve os testes efetuados utilizando o módulo de calibração desenvolvido. Um estudo de caso é feito utilizando um modelo de integração de modelos dinâmicos espaciais para ecologia do vetor *Aedes aegypti*, que é resultado de uma dissertação de mestrado em Ecologia de Biomas Tropicais, na Universidade Federal de Ouro Preto. Algumas abordagens acerca desse modelo são ilustradas.

O Capítulo 5 apresenta conclusões e discussões sobre os resultados obtidos.

Capítulo 2

Modelagem Ambiental e a Plataforma TerraME

Neste capítulo são apresentadas noções de modelagem ambiental. A partir dessas definições podemos identificar seus objetivos e características principais, especificando seu funcionamento.

A plataforma TerraME também será apresentada como o ambiente de desenvolvimento e suporte a modelagem. Características de sua estrutura e funcionamento serão detalhadas.

Ao final do capítulo, são apresentadas as vantagens que um modelador encontra utilizando o ambiente TerraME.

2.1 Modelagem Ambiental

Computadores vêm sendo utilizados como ferramentas de apoio a procedimentos de estudos, de análises e de simulações em vários campos do conhecimento humano. Sistemas complexos para análises e para modelagens foram desenvolvidos para se trabalhar dados relacionados com áreas específicas como finanças, transportes, geologia, solos, etc.. Nessa mesma tendência, sistemas de armazenamento, manipulação e apresentação de dados espaciais, conhecidos como SIG, foram criados e estão sendo utilizados no campo das ciências ambientais (Felgueiras, 1999).

Modelagem ambiental pode ser basicamente definida como qualquer abstração do mundo natural com algum poder de prever eventos com base em princípios gerais. A todo tempo estamos submetidos a mudanças cada vez mais intensas que vêm alterando o funcionamento dos Sistemas Terrestres. Essas mudanças podem ser naturais, antropogênicas ou uma combinação dos dois e têm impacto direto na integridade do meio ambiente e na qualidade de vida das pessoas. Estudos recentes apontam as ações humanas como a principal força direcionadora das alterações sofridas por esses sistemas e indicam que elas estarão associadas a alterações drásticas dos biomas (Salazar et al., 2007). Por exemplo, a floresta Amazônica dará lugar a

uma vegetação pobre em biodiversidade cuja fisionomia se assemelhará a de um cerrado pobre (D., 2007). O aumento da temperatura global poderá implicar em explosões populacionais de vetores de doença tropicais e resultar em epidemias globais (S. et al., 2002) (W. et al., 2000) (A. et al., 1998) (M. et al., 1995). Essas simulações podem ser feitas em vários tipos de sistemas, como os sistemas ecológicos, os sistemas climáticos, os sistemas públicos de saúde, os sistemas hidrológicos, os sistemas produtivos ou os sistemas de uso do solo. Modelos matemáticos, aritméticos e lógicos, buscando representar propriedades e processos do meio físico natural, têm sido implementados, nos SIG, com o objetivo de facilitar o seu estudo e compreensão para que se possa atuar sobre o meio ambiente de forma responsável e cooperativa (Felgueiras, 1999).

As ferramentas de modelagem devem ser capazes de simular os fenômenos escolhidos para estudo levando em consideração a dinâmica e os resultados das ações humanas. De acordo com os resultados obtidos na modelagem aqueles que são responsáveis por projetar e estabelecer políticas públicas tem condições de analisar os impactos das suas decisões (Lambin, 1994) (Kaimowitz e Angelsen, 1998) (Turner II et al., 1995). A simulação de processos naturais ou a simulação das interações humano-ambiente são instrumentos de pesquisa de impactos e previsão (Batistella e Moran, 2005) que têm potencial como meio de detectar, identificar, mapear e monitorar alterações nos ecossistemas, independentemente dos seus agentes causais (Coppin et al., 2004). A sustentabilidade de ecossistemas tornou-se um objetivo primordial na atualidade trazendo como uma de suas consequências, a contínua necessidade de precisão e atualização de banco de dados (Coppin et al., 2004). Desta forma, é clara a necessidade do desenvolvimento de técnicas e tecnologias que permitam o melhor entendimento do funcionamento dos sistemas Terrestres e a simulação de seu comportamento. Tais artefatos deverão ser utilizados para a análise de cenários alternativos gerados por simulações computacionais que servirão de base para o planejamento e definição de políticas públicas (Lana, 2009).

2.2 O Ambiente TerraME

TerraME é um ambiente de desenvolvimento e suporte a modelagem ambiental espacial dinâmica que suporta modelos de computação baseados em autômatos celulares (Wolfram, 2002) e conceitos de autômatos celulares aninhados (Nested-CA) (Carneiro, 2006). Um modelo espacial dinâmico é uma representação abstrata de um fenômeno que evolui no tempo e no espaço, baseado em descrições de entidades, processos e relações entre eles. Dessa forma, o TerraME está associado a um Sistema de Informação Geográfica (SIG) que fornece a localização espacial dos dados. Os resultados destes modelos são mapas que mostram a distribuição espacial de um padrão ou de uma variável contínua. Maiores detalhes sobre o TerraME podem ser encontrados em (Carneiro e Câmara, 2007, 2009).

O TerraME permite simulação em duas dimensões de espaços celulares regulares e irregulares. Entre as aplicações típicas de TerraME podem ser citadas a modelagem e simulação de

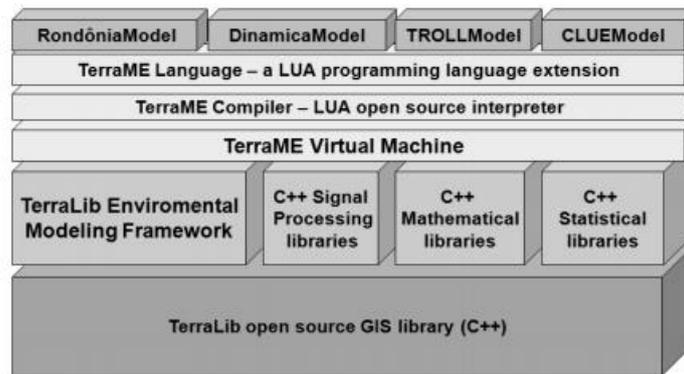


Figura 2.2: Arquitetura do Ambiente TerraME.

A biblioteca TerraLib é usada para leitura das entradas do modelo e para a gravação dos resultados da simulação além de fornecer mecanismos para o gerenciamento da base espacial de dados. Por fim, o aplicativo TerraView, um Sistema de Informação Geográfica (SIG) desenvolvido sobre a biblioteca C++ TerraLib, é utilizado para a visualização, a análise e o gerenciamento dos dados.

A Figura 2.2 descreve o arquitetura da plataforma de modelagem TerraME. Na primeira camada, a TerraLib oferece sistemas de informação geográfica (SIG), gestão de dados espaciais, serviços de análise e funções extras para a manipulação de dados temporais. Na segunda, o *framework* TerraME provê a simulação e os serviços de calibração e validação. Foi desenvolvido para ser independente da plataforma. Este *framework* pode ser utilizado diretamente para desenvolvimento de modelo, porém o desenvolvimento de modelos na linguagem C++ pode ser um desafio para modeladores, por isso, o TerraME proporciona uma linguagem de modelagem de alto nível. A terceira camada da arquitetura implementa a linguagem de modelagem TerraME e o ambiente de execução. A interface TerraME/LUA estende a linguagem de programação LUA com os novos tipos de dados espaciais de modelagem dinâmica e serviços para simulação e avaliação do modelo. Para tornar possível utilizar o *framework* através do interpretador LUA, foi necessário exportar a API do TerraME para a API de LUA, assim o TerraME reconhece os tipos descritos no modelo. Através dessa exportação outras aplicações escritas em C ou C++ podem ter suas APIs exportadas para a linguagem LUA. A última camada é chamada camada de aplicação e inclui os modelos de usuário final.

2.3 Conclusão

Como os modeladores, em geral, não são programadores e sim especialistas na área de conhecimento a ser modelada, o TerraME toma o cuidado de escolher uma linguagem de fácil implementação onde a preocupação principal seja o modelo e não a ferramenta. Uma linguagem de programação de extensão projetada para dar suporte à programação procedimental em geral

e que oferece facilidades para a descrição de dados é a linguagem LUA. Ela foi planejada para ser utilizada por qualquer aplicação que necessite de uma linguagem de *script* leve e poderosa e é facilmente acoplada à linguagem C.

O TerraME procura ser de fácil utilização para que a dificuldade encontrada na modelagem procure ser o modelo a ser desenvolvido e não a programação relacionada a simulação favorecendo assim a interdisciplinalidade, tão importante nesta área, pois torna ampla a possibilidade de múltiplos fenômenos serem modelados.

Capítulo 3

Algoritmos Genéticos e Calibração de Modelos

Este capítulo inicia-se com a definição e apresentação das características dos Algoritmos Genéticos. Após explicar a motivação, detalha a estrutura de seu funcionamento e os métodos específicos adotados nesse trabalho.

A calibração de modelos é definida e a implementação dessa calibração desenvolvida aqui é apresentada.

Alguns comentários e conclusões sobre as vantagens da utilização de algoritmos genéticos na calibração de modelos ambientais fecham o presente capítulo.

3.1 Algoritmos Genéticos

Algoritmos genéticos são uma classe particular de algoritmos evolutivos que utilizam do processo de evolução natural como um paradigma de solução de problemas, como os de otimização. Eles são inspirados no princípio Darwiniano da evolução das espécies e na reprodução genética (Goldberg, 1989). De acordo com esse princípio, a seleção de indivíduos é feita de modo a privilegiar aqueles mais aptos à sobrevivência, que, conseqüentemente, têm a maior probabilidade de reprodução. O indivíduo que se reproduz, transmite seu código genético a seus descendentes e com isso suas características são mantidas na população. A manipulação dessas características define o perfil da população.

A vantagem mais notória da computação evolutiva está na generalização da ferramenta de resolução de problemas, possibilitando a sua solução a partir da simples descrição matemática da solução desejada, pois se houvesse a necessidade de especificar os passos que levam ao resultado, eles certamente seriam específicos de cada caso. Isso proporciona ao algoritmo robustez e flexibilidade o que faz com que seus procedimentos genéricos e adaptáveis sejam aplicados na solução de problemas complexos.

Os Algoritmos Genéticos são probabilísticos e constituem uma técnica de busca heurística, altamente paralela e adaptativa. Os princípios biológicos como hereditariedade, mutação, seleção natural e recombinação (ou crossing over) são codificados em algoritmos computacionais que buscam solucionar da melhor forma um determinado problema evoluindo populações direcionadas por uma função objetivo.

A implementação se dá através de uma simulação onde uma população de representações abstratas sofre as mudanças resultantes da aplicação dos processos evolutivos, direcionadas por um objetivo pré estabelecido. A evolução se inicia com a geração de uma população de indivíduos inicial formada pela representação de várias soluções aleatórias. Cada indivíduo representa um candidato à solução do problema e pode ser representado através de uma estrutura de dados. Dentre os principais tipos de representação estão binária e real.

Essa população sofre modificações de acordo com a taxa de cruzamento e mutação estabelecidas na proposta da resolução do problema. O cruzamento é responsável pela criação de novos indivíduos através da combinação de dois outros. Essa combinação pode ser feita através da troca de fragmentos da sequência de representações de soluções, garantindo a coerência da solução para o caso de representações reais. Os indivíduos que passarão por esse processo são selecionados na população de acordo com algum direcionamento pré-estabelecido. A mutação é feita modificando-se o conteúdo de um atributo específico do indivíduo, como uma posição aleatória da representação de um elemento da população, elemento esse também escolhido aleatoriamente.

Os filhos resultantes dessas modificações podem substituir automaticamente os pais ou substituir somente se forem avaliados como soluções melhores que eles, dependendo da implementação adotada. Essa nova população formada pela mesclagem da população anterior com os elementos agora criados forma uma nova geração que é avaliada. Cada solução avaliada possui uma medida de adaptação ou fitness. O processo recomeça da mesma forma com a nova população e, geracionalmente, a tendência é evoluir as soluções com indivíduos mais adaptados até alcançar a condição de parada, o que geralmente indica a existência de um indivíduo que represente uma solução que possua o resultado esperado. Outro possível critério de parada poderia ser um número definido de gerações sem melhora na população, ou o valor máximo de gerações ser atingido.

A estrutura básica de um Algoritmo Genético pode ser ilustrada na figura 3.1:

As opções de operações adotadas nesse trabalho serão detalhadas a seguir:

3.1.1 Representação

A representação que será utilizada é a binária. Nela, as soluções são codificadas em cadeias binárias de acordo com uma precisão pré estabelecida. Essas cadeias podem ser decodificadas para o valor original a qualquer momento, através de cálculos matemáticos. Todas as operações seguintes são feitas nos indivíduos representados nesse formato.

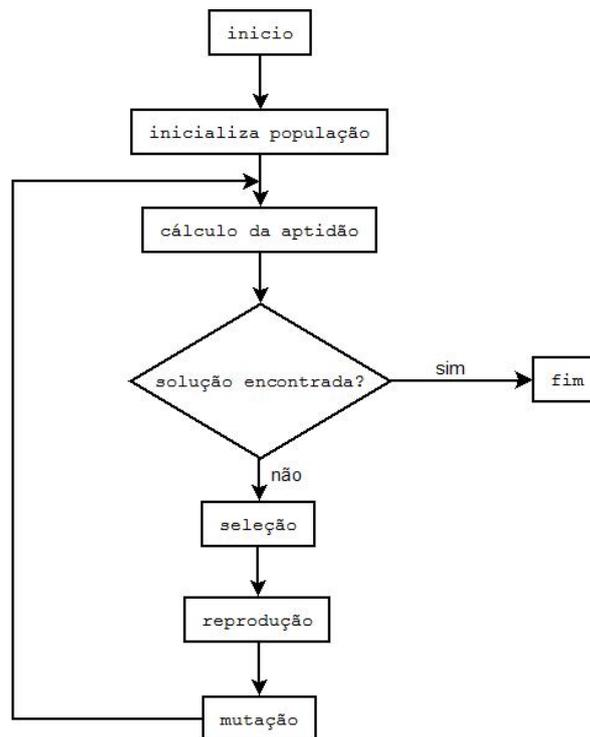


Figura 3.1: Fluxograma AG

3.1.2 Cruzamento

O cruzamento adotado foi o que possui dois pontos de corte. Nele, os indivíduos pais são selecionados e dois pontos da cadeia sorteados. A sequência binária do trecho compreendido entre esses dois pontos é trocada entre os elementos pais, formando dois elementos filhos geralmente diferentes dos originais.

3.1.3 Mutação

A mutação adotada é Bit flip. Uma posição aleatória da cadeia binária é sorteda e o bit correspondente a essa posição tem seu valor invertido. Essa operação sempre faz com que um indivíduo pai gere um filho diferente dele.

3.1.4 Seleção

A seleção utilizada é a seleção por torneio dois a dois. Nela dois indivíduos são selecionados aleatoriamente na população e avaliados. Aquele que apresenta melhor resultado é selecionado para a operação desejada. No caso do cruzamento, a seleção é chamada uma vez para cada pai. Com a seleção por torneio, diminuímos a probabilidade de indivíduos menos adaptados serem selecionados para as operações.

3.2 Calibração de Modelos

Modelos ambientais desenvolvidos são validados, geralmente, através de testes que comparam os resultados obtidos em simulações com aqueles realmente coletados acerca do fenômeno estudado. Quanto mais parecido o simulado é do real, melhor representado está o problema e com isso maior segurança ele tem para prever um comportamento futuro desse fenômeno. Para que um modelo tenha bons resultados, ele deve levar em consideração as peculiaridades do problema, todos os agentes que contribuem no comportamento do fenômeno em questão devem estar presentes, muitos deles como parâmetros do modelo criado. A calibração de parâmetros desse modelo consiste em encontrar a combinação de valores que melhor contribui para que o resultado encontrado na simulação seja satisfatório.

Para a calibração dos parâmetros dos modelos desenvolvidos no ambiente TerraME, o modelador dispunha até então somente da técnica de simulação de Monte Carlo (Rubinstein e Kroese, 2007), que é uma das técnicas bastante utilizadas no âmbito da modelagem ambiental. A técnica de Monte Carlo consiste num processo de amostragem estocástica cujo objetivo é permitir a observação do desempenho de uma variável de interesse em razão do comportamento de parâmetros do modelo que encerram elementos de incerteza. A base para o processo de amostragem realizado nas simulações de Monte Carlo é a geração de números aleatórios. Contudo, o processo de calibração por Monte Carlo pode ser muito custoso em diversas aplicações de modelagem ambiental, uma vez que a avaliação da função de erro do modelo requer a simulação do mesmo, o que pode levar alguns segundos ou até mesmo minutos, dependendo da complexidade do modelo. Dessa forma, o processo de calibração de apenas alguns poucos parâmetros pode consumir dias de simulação.

O objetivo desse trabalho é apresentar uma alternativa de calibração que seja mais vantajosa que a anterior, no sentido de propiciar economia principalmente de tempo. A alternativa proposta foi implementar um módulo de calibração genérico para o ambiente TerraME baseado em algoritmos genéticos (Mitchell, 1998) (Eiben e Smith, 2003). A ferramenta desenvolvida tem por objetivo ser parametrizada para a calibração de qualquer modelo implementado, independente de suas peculiaridades. O modelador deve apenas chamar a função de calibração e passar como parâmetro as configurações desejadas para o seu problema específico. A linguagem Lua permite que funções sejam passadas como parâmetro a outras funções, o que torna essa chamada ainda mais fácil.

A função de calibração tem a seguinte chamada:

```
CalibracaoAG(boolEstocastico, mModelo, fAvaliacao, boolFlag, precisao,  
             nTamPop, matrizLimites, pc, pm, nMaxGeracoes)
```

em que cada parâmetro é detalhado na Tabela 3.1. A estrutura geral do algoritmo genético implementado é mostrada no Algoritmo 3.2.

Tabela 3.1: CalibracaoAG

<code>boolEstocastico</code>	Variável booleana que indica se o modelo a ser calibrado é estocástico ou não.
<code>mModelo</code>	Contém o modelo a ser calibrado, descrito em linguagem LUA. O modelo recebe os parâmetros de calibração como argumento.
<code>fAvaliacao</code>	É a função de avaliação da qualidade do modelo, em geral o erro de ajuste a dados previamente fornecidos. Representa a função objetivo a ser minimizada ou maximizada pelo AG.
<code>boolFlag</code>	Variável que indica se a função objetivo deve ser minimizada ou maximizada.
<code>precisao</code>	Indica a precisão desejada para cada variável, internamente, reflete o número a ser usado na representação binária do indivíduo do AG.
<code>nTamPop</code>	Tamanho da população.
<code>matrizLimites</code>	Matriz ($n \times 2$) contendo os limites máximos e mínimos para cada um dos parâmetros de calibração do modelo.
<code>pc</code>	Probabilidade de cruzamento.
<code>pm</code>	Probabilidade de mutação. A probabilidade de mutação usada é por indivíduo, isto é, indica a probabilidade de um indivíduo sofrer mutação ou não. Caso o indivíduo sofra mutação, um de seus bits é escolhido aleatoriamente e invertido.
<code>nMaxGeracoes</code>	Número máximo de gerações sem melhora, usado como critério de parada do algoritmo.

`boolEstocastico, mModelo, fAvaliacao, boolFlag, precisao, nTamPop, matrizLimites, pc, pm, nMaxGeracoes`. Valores estimados dos parâmetros ótimos de calibração. $t \leftarrow 0$ {contador de gerações} $P_t = \{p_t^{(1)}, \dots, p_t^{(\mu)}\} \leftarrow$ Inicializar População(`nTamPop`) $t < \text{nMaxGeracoes}$ $F_t \leftarrow$ Avaliar População(P_t , `mModelo`, `fAvaliacao`) $t^*, p^* \leftarrow$ Armazenar Melhor Indivíduo(P_t, F_t) Até que P_{t+1} esteja completa $p_t^{(i)}, p_t^{(j)} \leftarrow$ Selecionar pais por torneio(P_t, F_t) $q_t^{(1)}, q_t^{(2)} \leftarrow$ Aplicar cruzamento(`pc`, $p_t^{(i)}, p_t^{(j)}$) $q_t^{(1)}, q_t^{(2)} \leftarrow$ Aplicar mutação(`pm`, $q_t^{(1)}, q_t^{(2)}$) Efetuar Competição($p_t^{(i)}, p_t^{(j)}, q_t^{(1)}, q_t^{(2)}$) Atualizar(P_{t+1}) $t \leftarrow t + 1$ Algoritmo Genético para a calibração de modelos no ambiente TerraME.

O algoritmo começa com a inicialização aleatória de uma população com `nTamPop` indivíduos, que codificam configurações candidatas para o problema. O número de bits para cada variável depende da precisão desejada para aquela variável, um parâmetro que é fornecido pelo usuário. Caso o usuário não forneça uma precisão, o valor *default* é igual a 10^{-3} vezes a faixa de variação da variável (limite superior menos o limite inferior).

Cada geração t consiste na execução iterativa dos operadores genéticos que caracterizam o algoritmo. Na avaliação da população, linha 5, o modelo desenvolvido na linguagem LUA e indicado por `mModelo` é simulado usando os parâmetros de calibração codificados pelo genótipo do indivíduo. O resultado da simulação do modelo é então analisado para se estimar a qualidade do modelo, isto é, seu ajuste aos dados previamente fornecidos. A função de aptidão dos indivíduos é portanto uma função da qualidade do modelo em modelar os dados do fenômeno sob análise.

Nas linhas 7 a 13, uma nova população de soluções candidatas é criada em função da população atual e de seus valores de aptidão. A etapa de seleção dos indivíduos “pais” para a reprodução é realizada de forma estocástica usando torneio binário. Nesta forma de seleção,

dois indivíduos são selecionados aleatoriamente entre a população e competem entre si de forma determinística, isto é, o melhor indivíduo entre os dois vence e é selecionado para a reprodução. Os melhores indivíduos possuem portanto maior probabilidade de serem selecionados para a reprodução. Dois pais $p_t^{(i)}, p_t^{(j)}$ são escolhidos dessa forma, ver linha 8. Nas linhas 9 e 10, duas novas soluções são produzidas a partir dos operadores de cruzamento e mutação. Esses operadores são constituídos por heurísticas específicas e em geral com algum componente de aleatoriedade. Convém destacar que embora os operadores genéticos apresentem algum grau de aleatoriedade, a busca realizada pelo algoritmo genético está longe de ser uma busca aleatória, pois o operador de seleção possui uma componente de determinismo que orienta o algoritmo na direção das melhores regiões do espaço de busca. Assim como na natureza a seleção natural é força motriz na evolução e na criação de complexidade, nos algoritmos genéticos o operador de seleção é o mecanismo responsável pelo progresso acumulativo e direcionado do algoritmo. O módulo de calibração conta atualmente com operador de cruzamento com dois pontos de corte por variável e mutação por inversão de bit.

Finalmente, na linha 11, as duas novas soluções geradas (indivíduos “filhos”) competem de forma determinística com seus pais, isto é, os descendentes substituem os pais caso sejam melhores do que eles. A substituição pode ser vista como um operador de sobrevivência. Os indivíduos escolhidos na competição são armazenados em P_{t+1} até que a nova população esteja completa com $n\text{TamPop}$ indivíduos.

3.2.1 Modelos Estocásticos

Quando o modelo a ser calibrado é estocástico, o resultado de sua simulação varia para um mesmo conjunto de valores nos parâmetros de calibração. Algoritmos evolutivos têm se mostrado eficientes em otimização com ruído em diversos contextos. Na computação evolucionária, as incertezas são divididas em quatro categorias distintas?:

- Função de fitness ruidosa;
- Parâmetros mudam depois da otimização;
- Função de fitness aproximada;
- O valor ótimo do problema muda.

Para cada uma dessas categorias, encontramos propostas de soluções na literatura. Nesse trabalho, será adicionada à ferramenta de calibração de modelos estocásticos para a categoria de funções de fitness ruidosas ?. Nesse caso, podemos utilizar três possíveis abordagens:

- Média Explícita;
- Média Implícita;

- Seleção Modificável.

A abordagem escolhida para implementação foi a seleção modificável. Nela, há competição entre pais e filhos e permanece na população aquele que apresenta melhor aptidão. Essa competição pode ser baseada em porcentagem ou limiar (*threshold*).

Na calibração de modelos estocásticos levando em conta um limiar (*threshold*), define-se esse valor baseado na variância estimada do modelo e aplica-se no momento da competição entre os pais e seus descendentes. Essa escolha torna o AG implementado bastante versátil caso o modelo a ser calibrado seja estocástico ou não. Se o modelo for determinístico, o limiar de comparação para a substituição é zero, caso contrário, o limiar é definido em função da variância do modelo.

A variância do modelo é estimada de forma automática pelo módulo de calibração, executando-se o modelo inúmeras vezes para uma combinação fixa dos parâmetros e calculando o desvio padrão dos resultados dessas execuções. Esse procedimento é repetido para algumas combinações fixas dos parâmetros, fazendo uma média dos valores estimados para o desvio padrão. O resultado encontrado é utilizado como um limiar na competição dos filhos em relação aos pais. Após o cruzamento, um filho substitui o pai se seu valor de aptidão for melhor que o dele em um valor superior a esse limiar, caso contrário, o pai permanece na população. Essa abordagem é interessante porque assegura de certa forma que indivíduos melhores permaneçam na população, pois, apesar da execução de um modelo retornar valores diferentes para um mesmo valor de parâmetro, flutuações próximas ao desvio padrão da função não são considerados.

Neste trabalho, a abordagem utilizada na calibração de modelos estocásticos leva em conta a variância estimada do modelo para definir um limiar (*threshold*) na competição entre os pais e seus descendentes, conforme linha 11. Essa escolha torna o AG implementado bastante versátil caso o modelo a ser calibrado seja estocástico ou não. Se o modelo for determinístico, o limiar de comparação para a substituição é zero, caso contrário, o limiar é definido em função da variância do modelo.

Analisando o comportamento dessa estratégia, conseguimos manter a variação do erro dentro de um limite e com isso assegurar de certa forma que indivíduos melhores permaneçam na população, já que esse valor a ser superado dá margem a desconsideração de valores de flutuação próximos ao desvio padrão.

O resultado da aplicação da estratégia do limiar na variação do erro em uma sequência de execuções pode ser observado a seguir:

A variância do modelo é estimada de forma automática pelo módulo de calibração, executando-se o modelo inúmeras vezes para uma combinação fixa dos parâmetros e calculando o desvio padrão dos resultados dessas execuções. Esse procedimento é repetido para algumas combinações fixas dos parâmetros, fazendo uma média dos valores estimados para o desvio padrão. O resultado encontrado é utilizado como um limiar na competição dos filhos

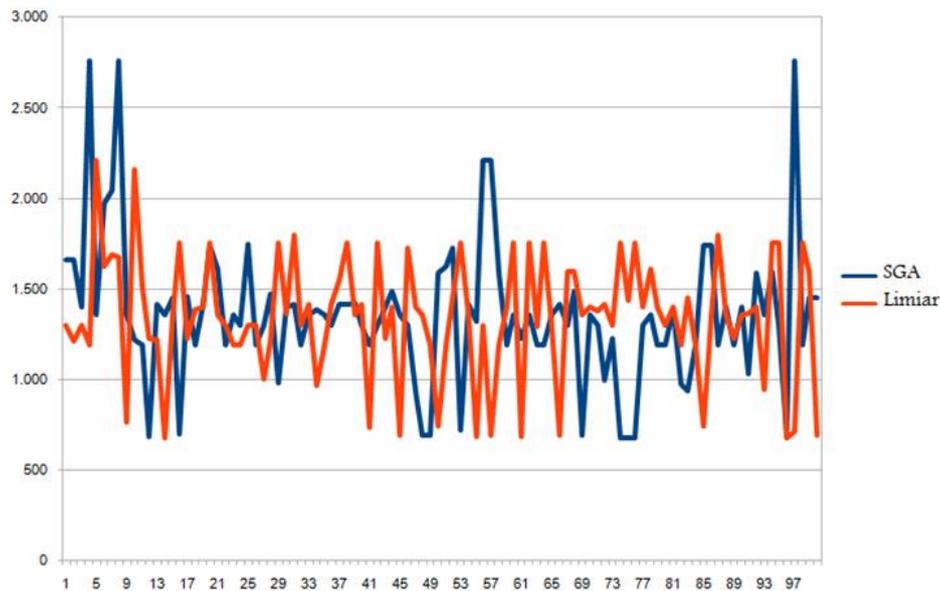


Figura 3.2: Variação do erro

em relação aos pais. Após o cruzamento, um filho substitui o pai se seu valor de aptidão for melhor que o dele em um valor superior a esse limiar, caso contrário, o pai permanece na população. Essa abordagem é interessante porque assegura de certa forma que indivíduos melhores permaneçam na população, pois, apesar da execução de um modelo retornar valores diferentes para um mesmo valor de parâmetro, flutuações próximas ao desvio padrão da função não são considerados.

3.3 Conclusão

Algoritmos Genéticos tem se apresentado como boa opção para resolver problemas complexos em vários contextos. Na modelagem ambiental encontramos muitas dificuldades que se transformam em barreiras para uma boa realização do trabalho. Um elevado tempo de simulação de um modelo, por exemplo, dificulta a visualização de resultados e consequentemente um aprimoramento de seu conteúdo. Algoritmos Genéticos usado na otimização dessas dificuldades tem resultados satisfatórios. A calibração de modelos, geralmente, é uma etapa relativamente demorada da modelagem onde a otimização desse processo tem uma contribuição mais significativa.

Capítulo 4

Resultados Obtidos

Este capítulo descreve a implementação de cada etapa dos testes realizados num modelo de estudo de caso. O modelo descreve a integração de modelos dinâmicos espaciais para ecologia do vetor *Aedes aegypti*, que é resultado de uma dissertação de mestrado em Ecologia de Biomas Tropicais, na Universidade Federal de Ouro Preto. Este estudo de caso foi apresentado em (?).

4.1 Estudo de Caso: *Aedes Aegypti*

Sistemas de Informação Geográfica (SIG) vêm sendo utilizados para armazenar, recuperar, visualizar e analisar informações a respeito de fenômenos que ocorrem no espaço geográfico, como por exemplo, mudanças de uso e cobertura do solo, evolução de epidemias, etc. O uso de recursos computacionais para tais estudos permitem que volumes de dados, que não poderiam ser processados pela mente humana, possam ser considerados simultaneamente para a produção de prognósticos a respeito do estado futuro de um determinado fenômeno que se desenvolve no espaço e no tempo (Lana, 2009).

Na Saúde Pública, SIGs são utilizados principalmente como uma ferramenta para espacialização de dados epidemiológicos. O principal objetivo do projeto desenvolvido na dissertação (Lana, 2009) é analisar de forma integrada submodelos para dengue que englobem tanto variáveis abióticas quanto bióticas e determinar como a adição de complexidade aos modelos existentes afetam o desempenho dos mesmos e do modelo integrado. Assim, este trabalho exercita o ambiente EpiTerra, que é fruto da integração dos *frameworks* TerraME e EpiGrass, para o desenvolvimento de modelos integrados e de larga escala que simulem o comportamento da dengue e a posterior aplicação desse modelo à cidade do Rio de Janeiro, RJ. Uma ilustração desse integração pode ser entendida da seguinte forma:

O modelo foi validado com dados obtidos na cidade do Rio de Janeiro. Observa-se destacado na figura abaixo o bairro de Higienópolis, local escolhido para o estudo pela disponibil-

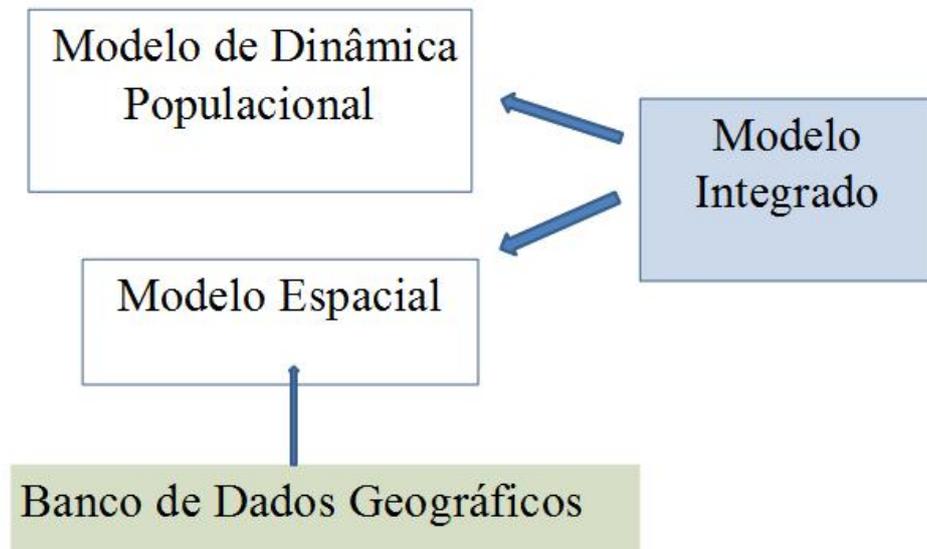


Figura 4.1: Modelo Integrado

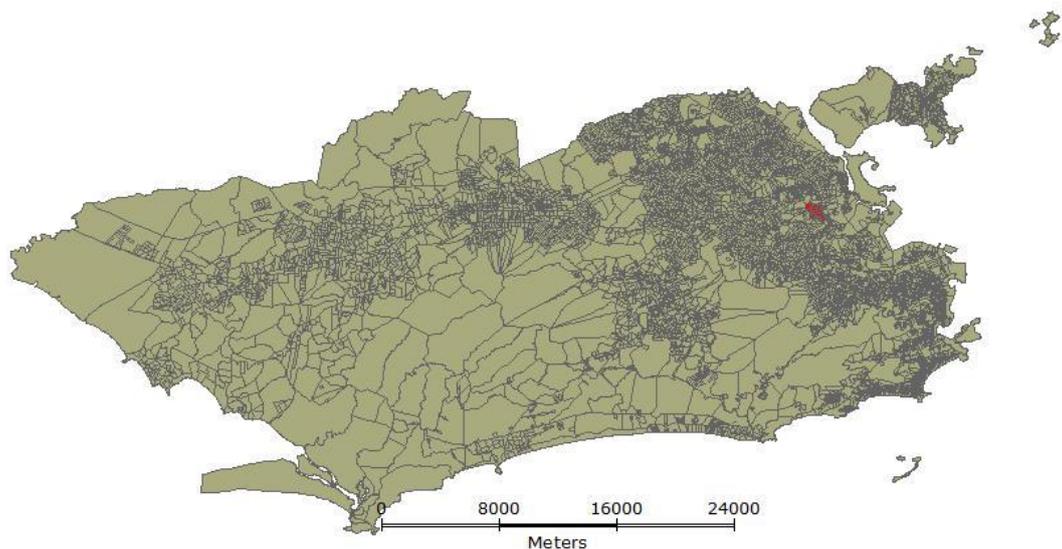


Figura 4.2: Bairro de Higienópolis - Rio de Janeiro

idade de dados relevantes para a modelagem do problema e pela localização em uma região reconhecida por ser palco de várias epidemias da dengue, o Rio de Janeiro.

Foi necessária a aquisição de dados ambientais sobre o local, como por exemplo, modelo digital de elevação (relevo), mapas de cobertura e uso do solo, mapa de segregação social, arruamento, mapas de domicílios. Além disso, foram necessários dados que descrevessem estatisticamente a população de hospedeiros da cidade e seu padrão de mobilidade. Também foram necessários dados estatísticos sobre epidemias de dengue ocorridas nessa cidade. Entre as principais fontes de dados pode-se citar:

- Fiocruz: no esforço de modelar epidemias de dengue na cidade do Rio de Janeiro, RJ, o PROCC da Fiocruz avaliou pesquisas de campo que permitiram estabelecer estatisticamente o perfil da população dessa cidade assim como o padrão de mobilidade. Estes dados se tornaram disponíveis para a equipe executora deste trabalho. Os mapas de arreamento, domicílios, de segregação social e cobertura do solo também foram fornecidos pela equipe do PROCC. A equipe da Fiocruz também disponibilizou dados estatísticos sobre epidemias passadas de dengue na cidade do Rio de Janeiro, RJ.
- SRTM: o modelo digital de elevação para a cidade do Rio de Janeiro foi obtido na resolução de 90x90 metros por meio do projeto *Shuttle Radar Topographic Mission* (SRTM) da NASA.
- Imagens de sensores remotos: imagens dos satélites CBERS-II e LandSat TM para a cidade do Rio de Janeiro, RJ, foram disponibilizadas gratuitamente pelo INPE.

Com os dados em mãos a modelagem foi feita levando em consideração todos os fatores ambientais que colaboram para a dispersão do mosquito. Uma espécie de armadilha, as ovitrampas, foram espalhadas em lugares estratégicos no bairro de Higienópolis no Rio de Janeiro, RJ. Os espécimes coletados, em todos os estágios do *Aedes aegypti*, ovos, larvas, pupas e adultos, foram amostrados por 54 semanas. As amostras foram separadas aleatoriamente em dois grupos, um para validação e outro para calibração do modelo.

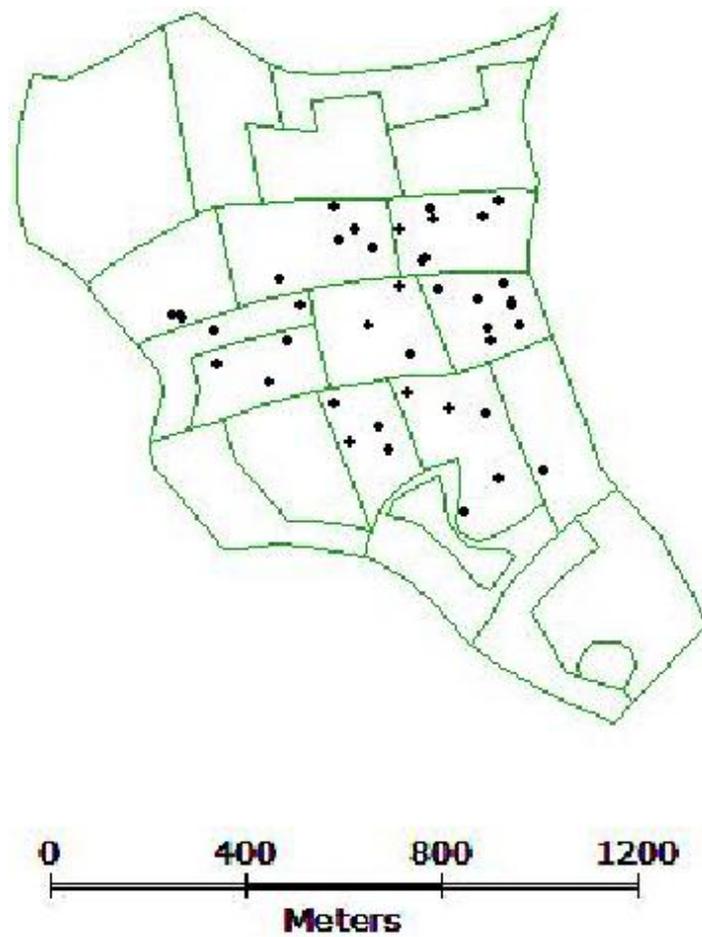


Figura 4.3: Armadilhas espalhadas no bairro de Higienópolis no Rio de Janeiro, RJ.

A seguir são detalhados três modelagens diferentes realizadas e os resultados obtidos com a calibração no ambiente TerraME.

4.1.1 Modelo Determinístico

Neste primeiro caso, o modelo possui um parâmetro único, definido como “Capacidade de Suporte”. Esse parâmetro é um número real compreendido no intervalo [100, 1000]. A temperatura usada nessa simulação é fornecida pela Estação Meteorológica do Galeão¹. O algoritmo genético implementado foi utilizado para a calibração desse modelo, realizando-se 30 execuções do algoritmo e validando o resultado obtido com o método de Monte Carlo, único disponível até então na plataforma TerraME.

Uma única simulação do modelo no TerraME consome aproximadamente 3 segundos de computação. O método de Monte Carlo foi usado com 10,000 avaliações, levando aproximadamente 8 horas. O AG foi executado com 10 indivíduos ao longo de 20 gerações, levando apenas 7 minutos para calibrar o modelo.

Na tabela a seguir são apresentados os resultados obtidos em uma média de 40 execuções do modulo de calibração.

População	Gerações	Capacidade de Suporte	Erro
10	10	464,5 ±26,87	1879,60 ±35,5
10	20	462,9 ±17,43	1868,64 ±24,07
10	30	457,7 ±0,823	1860,99 ±0,0241

Tabela 4.1: Resultados AG

O gráfico a seguir apresenta o resultado gerado através da solução obtida a partir do menor número de gerações apresentadas(10) coincidentemente também aquela que possui o maior erro comparando o resultado encontrado com a solução obtida a partir da simulação de Monte Carlo e AG:

Comparando-se numericamente as soluções:

	Monte Carlo	Algoritmo Genético
Capacidade de Suporte	457,3	464,5
Erro	1860,98	1861,04
Número de experimentos	10000	140
tempo estimado(minutos)	500	7

Tabela 4.2: Resultados AG

Comparando-se os resultados obtidos a partir da aplicação das duas técnicas podemos notar que os valores da “Capacidade de suporte” e do erro possuem diferenças aceitáveis. O número

¹<http://www.rio.rj.gov.br/georio/alerta/tempo>

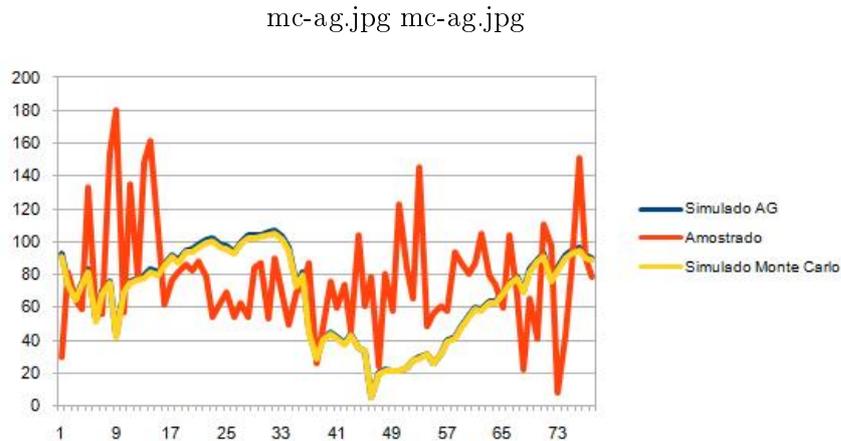


Figura 4.4: Monte Carlo x Algoritmo Genético

de experimentos executados é consideravelmente menor utilizando AG e conseqüentemente o tempo também é. Com AG podemos chegar a um resultado aceitavelmente parecido utilizando 1,4% do tempo gasto pela simulação de Monte Carlo, logo, seu uso se torna vantajoso. Porém, embora a calibração tenha sido correta, uma vez que foi validada pelo método de Monte Carlo, o modelo ainda não apresenta desempenho satisfatório para os dados coletados.

4.1.2 Modelo Estocástico

Os resultados obtidos anteriormente ainda não se mostraram satisfatórios para o modelador. Em seguida, foi introduzido na modelagem mais um parâmetro referente a uma porcentagem de probabilidade de oviposição. Neste caso, a temperatura é simulada dentro de um intervalo específico semelhante ao amostrado onde ela é maior nas 16 semanas referentes ao verão. Dessa forma, o problema de calibração agora consiste em um problema de otimização com dois parâmetros contínuos, sendo que a função de erro é estocástica. Novamente o algoritmo genético foi utilizado, considerando função objetivo com ruído, e usando 20 indivíduos por 100 gerações.

Dessa vez, o valor encontrado para a Capacidade de Suporte foi de $271,5 \pm 13,55$, com erro de $973,4 \pm 166,3$. Este resultado ilustra a capacidade do módulo implementado em tratar a calibração de modelos com parâmetros estocásticos. Entretanto, este resultado ainda não representa bem o fenômeno a ser modelado. Analisando o gráfico podemos observar que o modelo sofre grande influência da temperatura e ainda não é considerado satisfatório.

4.1.2.1 Modelo determinístico com mais parâmetros

Uma nova versão da estruturação do modelo leva em conta novos parâmetros que eram usados como constantes baseados em estudos encontrados na literatura. Essas constantes dizem

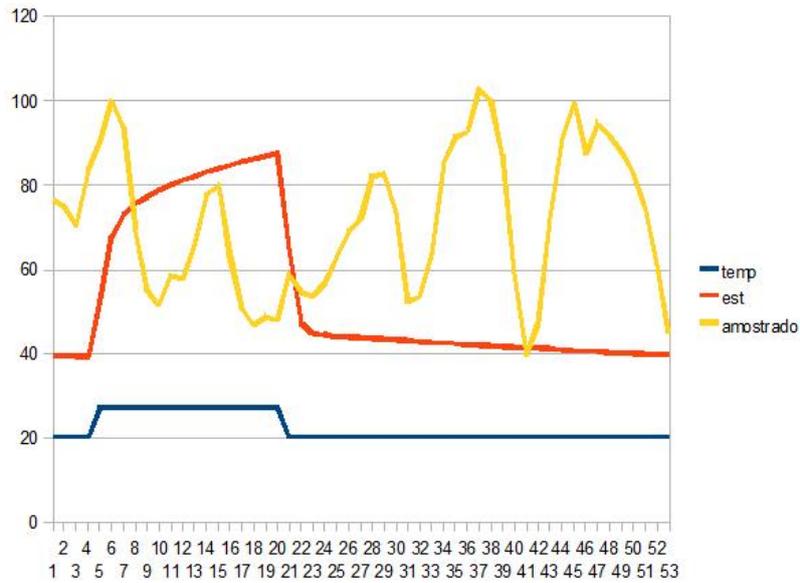


Figura 4.5: Modelo Estocástico

respeito à taxa de ovos que conseguem se transformar em larvas, a taxa de larvas que conseguem se transformar em pupas e a taxa de pupas que consegue evoluir para o estágio adulto, isso tanto no verão quanto no inverno, totalizando mais 6 parâmetros no intervalo $[0, 1]$. O parâmetro estocástico no modelo anterior foi retirado, tornando o modelo determinístico novamente. O problema de calibração possui agora 7 variáveis, a Capacidade de Suporte C mais as 6 taxas mencionadas.

O algoritmo genético foi utilizado com 50 indivíduos por 100 gerações. A calibração consumiu desta vez em média 2 horas de simulação. Os valores obtidos para os parâmetros de calibração são relacionados na Tabela 4.3. Vale ressaltar que a melhor solução obtida não corresponde à média obtida para os valores dos parâmetros. Se o modelo for executado com a configuração média na Tabela 4.3 referente a 40 execuções, o resultado obtido para o erro do modelo é superior. O erro médio obtido com a calibração pelo AG foi de $365,09 \pm 76,38$, bem inferior ao encontrado nos casos anteriores. A Figura 4.6 ilustra o desempenho do modelo em relação aos dados amostrados. Nesta figura o modelo foi simulado com a melhor solução encontrada pelo AG.

Tabela 4.3: Múltiplos parâmetros de calibração

CapacidadeSuporte $275,19 \pm 178,79$	txVeraoOvos $0,367 \pm 0,24$	txVeraoLarv $0,326 \pm 0,32$	txVeraoPulp $0,396 \pm 0,34$
	txInvernoOvos $0,465 \pm 0,31$	txInvernoLarv $0,484 \pm 0,28$	txInvernoPulp $0,515 \pm 0,22$

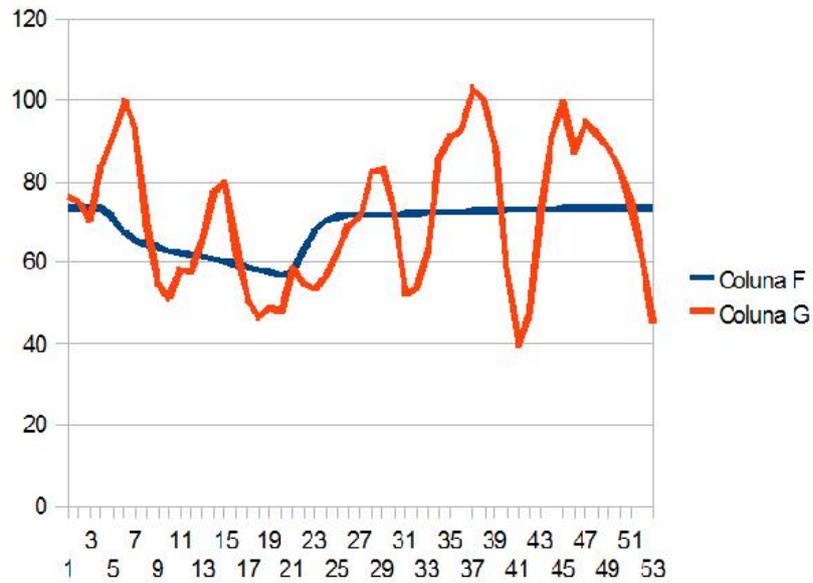


Figura 4.6: Taxa de oviposição ao longo das 53 semana, usando o modelo com múltiplos parâmetros.

Analisando o gráfico podemos observar que a flexibilização que considera valores constantes como potenciais parâmetros a serem calibrados observa uma melhor tendência do fenômeno, porém, ainda não foi obtido um ajuste satisfatório entre a oviposição simulada e aquela observada. Isso pode ser referente a uma escala espacial inadequada para o estudo, a um período de amostragem muito curto. A resolução temporal também precisa ser refinada através de técnicas de suavização e interpolação.

Capítulo 5

Discussão e Conclusão

Este trabalho apresentou um módulo de calibração baseado em algoritmos genéticos para o ambiente de modelagem ambiental TerraME. O módulo de calibração baseado em algoritmos genéticos introduz flexibilidade para o modelador no desenvolvimento e calibração de modelos, sendo capaz de calibrar modelos não lineares complexos com múltiplos parâmetros, e modelos determinísticos e estocásticos.

O desenvolvimento de modelos de simulação confiáveis não pode prescindir das etapas de modelagem, calibração e validação do modelo. Dependendo da complexidade do modelo e do método escolhido, o tempo despendido na execução dessa calibração pode ser bem elevado, retardando a evolução do modelo. O estudo de caso apresentado ilustra o processo de modelagem da dengue no TerraME usando modelos espaciais baseados em autômatos celulares. O estudo de caso ilustra o processo de modelagem, desde um modelo mais simples até um modelo mais complexo, envolvendo parâmetros estocásticos ou não. A calibração usando algoritmos genéticos fica bastante facilitada para o modelador e mais rápida, facilitando e beneficiando o processo de modelagem como um todo.

Como trabalhos futuros, pretende-se incorporar o módulo implementado na linguagem LUA à biblioteca de funções TerraLib C++, camada 2 na arquitetura do ambiente TerraME. Com relação ao modelo da dengue, pretende-se investigar modelos mais complexos, em particular envolvendo parâmetros espacialmente distribuídos. Nesse caso, a calibração pode envolver centenas de parâmetros, dependendo da resolução espacial do modelo.

Referências Bibliográficas

- A., P. J.; M., M. W. J.; A., F. D. e H., J. T. (1998). Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environ Health Perspective*, 106(3):147153.
- Batistella, M. e Moran, E. F. (2005). Dimensões humanas do uso e cobertura das terras na amazônia: uma contribuição do LBA. *Acta Amazonica*, 35(2).
- Breukelaar, R. e Bäck, T. (2005). Using a genetic algorithm to evolve behavior in multi dimensional cellular automata. In *Proc. of the Genetic and Evolutionary Computation Conference (GECCO'05)*, pp. 107–114. ACM Press.
- Carneiro, T. (2006). *Nested-CA: A Foundation for Multiscale Modeling of Land Use and Land Change*. PhD thesis, Instituto Nacional de Pesquisa Espacial (INPE), São José dos Campos, Brasil.
- Carneiro, T. e Câmara, G. (2007). A gentle introduction to TerraME. Technical report, Instituto Nacional de Pesquisa Espacial (INPE), São José dos Campos, Brasil.
- Carneiro, T. e Câmara, G. (2009). TerraME: Suporte a modelagem ambiental multi-escalas integrada a bancos de dados geográficos. In Barcelos, C. A. Z.; de Andrade, E. X. L. e Boaventura, M., editores, *Notas em Matemática Aplicada*, volume 40. Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC).
- Chen, S. H.; Jakeman, A. J. e Norton, J. P. (2008). Artificial intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, 78:379–400.
- Coppin, P.; Jonckheere, I.; Nackaerts, K.; Muys, B. e Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25(9):1565–1596.
- D., V. L. F. S. . N. C. A. . O. M. (2007). Climate change consequences on the biome distribution in tropical south america. *Geophysical Research Letters*, 34:1 – 6.

- D'Ambrosio, D.; Iovine, G.; Spataro, W. e Miyamoto, H. (2007). A macroscopic collisional model for debris-flows simulation. *Environmental Modelling & Software*, 22:1417–1436.
- D'Ambrosio, D. e Spataro, W. (2007). Parallel evolutionary modelling of geological processes. *Parallel Computing*, 33:186–212.
- Eiben, A. E. e Smith, J. E. (2003). *Introduction to Evolutionary Computing*. Natural Computing Series. Springer.
- Felgueiras, C. A. (1999). *Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação*. PhD thesis, INPE - Instituto Nacional de Pesquisas Espaciais.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.
- Kaimowitz, D. e Angelsen, A. (1998). Economic models of tropical deforestation: A review. Technical report, Center for International Forestry Research, Bogor, Indonesia.
- Lambin, E. F. (1994). Modeling deforestation processes, a review. trees, tropical ecosystem environment observation by satellites. Research Report 1, European Commission Joint Research Center / European Space Agency, Brussels.
- Lambin, E. F. e Geist, H. J., editores (2006). *Land-Use and Land-Cover Change: Local Processes and Global Impacts*. Global Change - The IGBP Series. Springer.
- Lana, R. M. (2009). Modelos dinâmicos acoplados para simulação da ecologia do vetor aedes aegypti. Master's thesis, Universidade Federal de Ouro Preto, Programa de Pós-Graduação em Ecologia de Biomas Tropicais, Ouro Preto, Brasil.
- M., M. W. J.; W., N. L.; J., R. e H., J. T. (1995). Potential impact of global climate change on malaria risk. *Environ Health Perspective*, p. 103(5).
- Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Complex Adaptive Systems. MIT Press, Massachusetts.
- Mitchell, M.; Crutchfield, J. P. e Hraber, P. T. (1994). Evolving cellular automata to perform computations: Mechanisms and impediments. *Physica D*, 75:361–391.
- Recknagel, F., editor (2006). *Ecological Informatics: Understanding Ecology by Biologically Inspired Computation*. Springer-Verlag, Berlin.
- Rubinstein, R. Y. e Kroese, D. P. (2007). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, 2nd edição.

- S., H.; N., W.; J., M. e A., W. (2002). Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet*, pp. 1 – 5.
- Salazar, L. F.; Nobre, C. A. e Oyama, M. D. (2007). Climate change consequences on the biome distribution in tropical South America. *Geophysical Research Letters*, 34:1–6.
- Turner II, B. L.; Skole, D.; Sanderson, S.; Fischer, G.; Fresco, L. e Leemans, R. (1995). Land-use and land-cover change science/research plan. IHDP Report 07, The International Geosphere-Biosphere Programme (IGBP) and The Human Dimensions of Global Environmental Change Programme (HDP), Stockholm and Geneva.
- W., G. A. K. L. S.; E., C. U. e A., P. J. (2000). Climate change and vector- borne diseases: a regional analysis. *Bulletin- World Health Organization*, 78/79:1136–1147.
- Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media Inc., Champaign, Illinois, USA.