Calibração em Modelagem Ambiental na Plataforma TerraME usando Algoritmos Genéticos

Larissa M. Fraga[1], **Raquel M. Lana**[1], **Tiago G. S. Carneiro**[1]
Frederico G. Guimarães[2]

[1] Departamento de Computação
Universidade Federal de Ouro Preto (UFOP)
35400-000 – Ouro Preto – MG – Brasil

[2]Departamento de Engenharia Elétrica
Universidade Federal de Minas Gerais (UFMG)
31270-010 – Belo Horizonte – MG – Brasil

{laryfraga, raquelmlana, tiagogsc, frederico.g.guimaraes}@gmail.com

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.*

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.*

1 Introdução

A todo tempo estamos submetidos a mudanças cada vez mais intensas que vêm alterando o funcionamento dos Sistemas Terrestres. Essas mudanças 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 (2007)]. O aumento da temperatura global, por exemplo, poderá implicar em explosões populacionais de vetores de doenças tropicais e resultar em epidemias globais.

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 (1998), Turner II (1995)]. A simulação de processos naturais ou a simulação das interações humano-ambiente são instrumentos de pesquisa de impactos e predições [Batistella (2005)]. Essa simulação pode ser feita em vários tipos de sistemas, como os sistemas ecológicos, os sistemas climáticos, os sistemas públicos de saúdes, 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-UFOP[1] (Universidade Federal de Ouro Preto) e INPE (Instituto Nacional de Pesquisas Espaciais). Esta plataforma fornece uma linguagem 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 (2006)]) na região da Amazônia Brasileira; e (ii) Saúde pública: como no caso do estudo do modelo de dinâmica populacional para o *Aedes aegypti* na cidade do Rio de Janeiro, RJ (SAUDAVEL) [Lana (2009)].

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. Por fim, é comum que modelos em modelagem ambiental possuam parâmetros estocásticos, fazendo com que a função de avaliação da qualidade do modelo seja estocástica, ao invés de determinística. Todas essas características

[1] http://www.terralab.ufop.br/

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 (2007)a] 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 (2007)b]. Uma revisão da aplicação de técnicas de inteligência artificial e computacional em sistemas de modelagem ambiental é apresentada em [Chen (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 (1994), Breukelaar (2005)].

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, o modelo de dinâmica populacional para o *Aedes aegypti* no estado do Rio de Janeiro. Os resultados ilustram a utilização do módulo de calibração e o potencial da ferramenta em modelagem ambiental de fenômenos complexos.

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 (2007), Carneiro (2009)].

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 processos de mudança de uso e cobertura do solo para toda a região Amazônica no âmbito do projeto GEOMA[2], desenvolvimento de modelos de drenagem da água das chuvas, modelos de propagação de incêndios em florestas, e modelos epidemiológicos junto à FIOCRUZ, como no caso do controle da Dengue nas cidades do Rio de Janeiro (RJ) e Recife (PE).

[2]Ver `http://www.geoma.lncc.br`

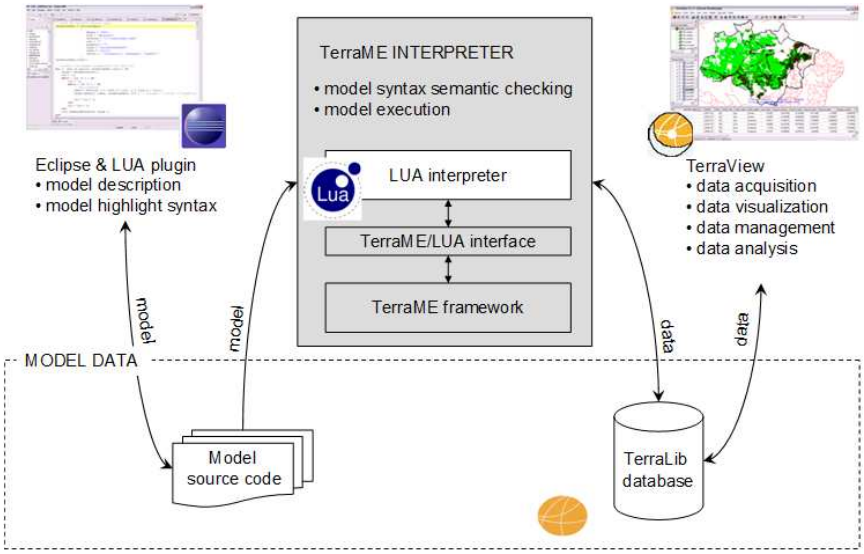

Figura 1. Ambiente de desenvolvimento e suporte a modelagem ambiental TerraME.

A parte fundamental do desenvolvimento do ambiente TerraME é o *interpretador TerraME*, conforme ilustrado na Figura 1. Ele lê um programa escrito na linguagem de modelagem TerraME, que por sua vez é uma linguagem estendida da linguagem LUA, interpreta o código fonte, e chama as funções do *framework* TerraME. Este *framework* é um conjunto de módulos escritos em C++ que oferecem funções e classes para modelagem espacial dinâmica e acesso ao banco de dados espacial TerraLib. Os resultados da simulação dos modelos desenvolvidos podem ser vistos com a aplicação TerraView.

Para a elaboração do código descritivo do modelo na linguagem usada pelo TerraME, o pesquisador pode utilizar qualquer editor de texto puro, tal como Crimson[3] ou Notepad++[4], porém um ambiente de desenvolvimento integrado (IDE - *Integrated Development Environment*), como por exemplo o Eclipse[5], provê recursos que facilitam essa elaboração, tais como o destaque da sintaxe e a facilidade de chamar o interpretador do TerraME dentro deste mesmo ambiente. O interpretador faz a verificação da sintaxe e da semântica do modelo, sendo também o responsável pela execução do modelo.

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 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

[3] http://www.crimsoneditor.com
[4] http://notepad-plus.sourceforge.net
[5] http://www.eclipse.org

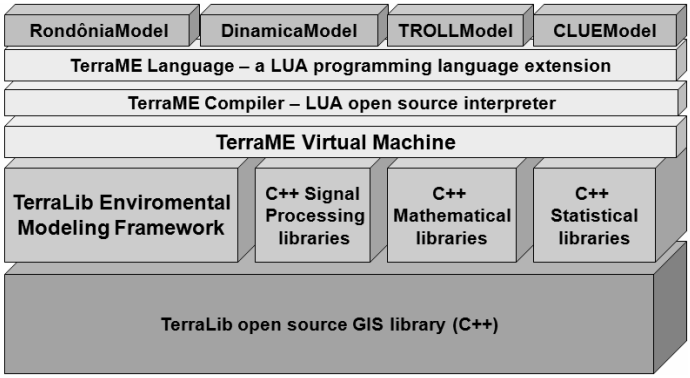

Figura 2. Arquitetura do Ambiente TerraME.

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.

3 Módulo de Calibração de Modelos Ambientais

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 (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 (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. 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 asumir 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.

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 1. A estrutura geral do algoritmo genético implementado é mostrada no Algoritmo 1.

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.

Tabela 1. CalibracaoAG

`boolEstocastico`	Variável booleana que indica se o modelo a ser calibrado é estocástico ou não.
`mModelo`	Contém o modelo a ser calibrado, descrito em linguagem LUA. O modelo recebe os parâmetros de calibração como argumento.
`fAvaliacao`	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.
`boolFlag`	Variável que indica se a função objetivo deve ser minimizada ou maximizada.
`precisao`	Indica a precisão desejada para cada variável, internamente, reflete o número a ser usado na representação binária do indivíduo do AG.
`nTamPop`	Tamanho da população.
`matrizLimites`	Matriz $(n \times 2)$ contendo os limites máximos e mínimos para cada um dos parâmetros de calibração do modelo.
`pc`	Probabilidade de cruzamento.
`pm`	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.
`nMaxGeracoes`	Número máximo de gerações sem melhora, usado como critério de parada do algoritmo.

Algorithm 1: Algoritmo Genético para a calibração de modelos no ambiente TerraME.

Entrada: `boolEstocastico, mModelo, fAvaliacao, boolFlag, precisao, nTamPop, matrizLimites, pc, pm, nMaxGeracoes`.
Saída: Valores estimados dos parâmetros ótimos de calibração.

1 **início**
2 $t \leftarrow 0$ {*contador de gerações*};
3 $P_t = \{p_t^{(1)}, \ldots, p_t^{(\mu)}\} \leftarrow$ `Inicializar População(nTamPop)`;
4 **enquanto** $t <$ *nMaxGeracoes* **faça**
5 $F_t \leftarrow$ `Avaliar População(`P_t`, mModelo, fAvaliacao)`;
6 $t^*, p^* \leftarrow$ `Armazenar Melhor Indivíduo(`P_t, F_t`)`;
7 **repita**
8 $p_t^{(i)}, p_t^{(j)} \leftarrow$ `Selecionar pais por torneio(`P_t, F_t`)`;
9 $q_t^{(1)}, q_t^{(2)} \leftarrow$ `Aplicar cruzamento(`pc, $p_t^{(i)}, p_t^{(j)}$`)`;
10 $q_t^{(1)}, q_t^{(2)} \leftarrow$ `Aplicar mutação(`pm, $q_t^{(1)}, q_t^{(2)}$`)`;
11 `Efetuar Competição(`$p_t^{(i)}, p_t^{(j)}, q_t^{(1)}, q_t^{(2)}$`)`;
12 `Atualizar(`P_{t+1}`)`;
13 **até** *Até que* P_{t+1} *esteja completa*;
14 $t \leftarrow t + 1$;
15 **fim**
16 **fim**

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. O módulo de calibração conta atual-

mente 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 `nTamPop` indivíduos.

Quando o modelo a ser calibrado é estocástico, o resultado de sua simulação varia para um mesmo conjunto de valores para os parâmetros de calibração. Dessa forma, a função objetivo do problema de calibração torna-se uma função com ruído. Algoritmos evolutivos têm se mostrado eficientes em otimização com ruído em diversos contextos [Eiben (2003)].

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.

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.

4 Estudo de Caso: *Aedes aegypti*

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 construir modelos computacionais que ajudem a entender o comportamento dinâmico espacial de populações de *Aedes aegypti* em áreas urbanas. Esses modelos são de especial importância para a avaliação computacional de diferentes estratégias de controle da Dengue, além de permitirem a análise dos efeitos que diferentes variáveis bióticas e abióticas exercem sobre o comportamento dessas populações. Assim, este trabalho exercita o desenvolvimento de um modelo espacialmente-explícito que permita apontar áreas de risco de transmissão da doença e a posterior aplicação desse modelo à cidade do Rio de Janeiro, RJ.

O estudo foi feito para o bairro de Higienópolis (22º 52"25' S, 43º 15"41' W) na cidade do Rio de Janeiro, RJ, e foi necessária a aquisição de dados ambientais, como por exemplo, mapas de divisão de setores censitários, arruamento e domicílios. Além disso, foram necessários dados climáticos (Estação Meteorológica do Galeão) e estatística

de oviposição do mosquito para o bairro estudado [Honório (2009)]. 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 arruamento, domicílios, de segregação social e cobertura do solo também foram fornecidos pela equipe do PROCC. A equipe da Fiocruz também disponibilizou dados estatísticos sobre estatísticas de oviposição da mosquito no bairro de Higienópolis na cidade do Rio de Janeiro, RJ.
- 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 1,5 ano, mas o estudo em questão utiliza dados de 53 semanas. As amostras foram separadas aleatoriamente em dois grupos, um para validação e outro para calibração do modelo.

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

4.1 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 simuação é fornecida pela Estação Meteorológica do Galeão[6]. 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.

Com as 30 execuções do AG, a Capacidade de Suporte encontrada foi de $464,5 \pm 26,9$, com erro de calibração médio de $1879,60 \pm 35,5$. 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.

4.2 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

[6] `http//www.rio.rj.gov.br/georio/alerta/tempo`

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 foi considerado satisfatório para o modelo.

4.2.1 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 respeito às taxas de desenvolvimento do mosquito, na qual ovos se transformam em larvas, larvas se transformam em pupas e pupas que conseguem 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 2. 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 2 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 3 ilustra o desempenho do modelo em relação aos dados amostrados. Nesta figura o modelo foi simulado com a melhor solução encontrada pelo AG.

Tabela 2. 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` $0,515 \pm 0,22$

5 Conclusões

Este artigo apresentou o 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

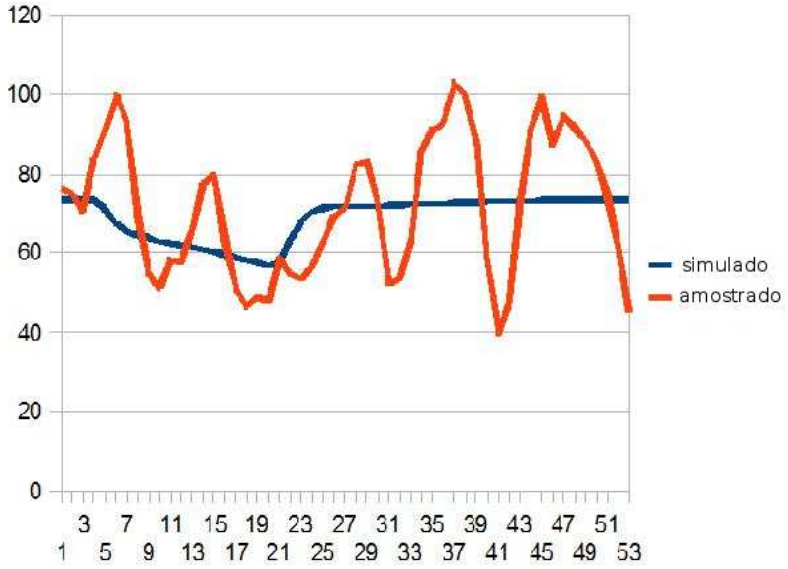

Figura 3. Taxa de oviposição ao longo das 53 semanas, usando o modelo com múltiplos parâmetros.

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 bliblioteca 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

Batistella, M. and Moran, E. F. (2005). Dimensões humanas do uso e cobertura das terras na amazônia: uma contribuição do LBA. *Acta Amazonica*, 35(2).

Breukelaar, R. and 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)*, pages 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. and Câmara, G. (2007). A gentle introduction to TerraME. Technical report, Instituto Nacional de Pesquisa Espacial (INPE), São José dos Campos, Brasil.

Carneiro, T. and 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., and Boaventura, M., editors, *Notas em Matemática Aplicada*, volume 40. Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC).

Chen, S. H., Jakeman, A. J., and Norton, J. P. (2008). Artificial intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, 78:379–400.

D'Ambrosio, D., Iovine, G., Spataro, W., and Miyamoto, H. (2007a). A macroscopic collisional model for debris-flows simulation. *Environmental Modelling & Software*, 22:1417–1436.

D'Ambrosio, D. and Spataro, W. (2007b). Parallel evolutionary modelling of geological processes. *Parallel Computing*, 33:186–212.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing*. Natural Computing Series. Springer.

Honório, N. A., Codeço, C. T., Alves, F. C., Magalhães, M. A. F. M., and de Oliveira, R. L. (2009). Temporal distribution of Aedes aegypti in different districts of Rio de Janeiro, Brazil, measured by two types of traps. *Journal of Medical Entomology*, 46(5):1001–1014.

Kaimowitz, D. and 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. and Geist, H. J., editors (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.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Complex Adaptive Systems. MIT Press, Massachusetts.

Mitchell, M., Crutchfield, J. P., and 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. and Kroese, D. P. (2007). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, 2nd edition.

Salazar, L. F., Nobre, C. A., and 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., and 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.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media Inc., Champaign, Illinois, USA.