

Universidade Federal de Ouro Preto - UFOP
Instituto de Ciências Exatas e Biológicas - ICEB
Departamento de Computação - DECOM

Calibração em Modelagem Ambiental na Plataforma
TerraME usando Algoritmos Genéticos

Aluno: Larissa Maiara Fraga
Matricula: 07.1.4152

Orientador: Frederico Gadelha Guimarães

Ouro Preto
1 de outubro de 2010

Universidade Federal de Ouro Preto - UFOP
Instituto de Ciências Exatas e Biológicas - ICEB
Departamento de Computação - DECOM

Calibração em Modelagem Ambiental na Plataforma TerraME usando Algoritmos Genéticos

Proposta de monografia apresentada ao curso de Bacharelado em Ciência da Computação, Universidade Federal de Ouro Preto, como requisito parcial para a conclusão da disciplina Monografia II (BCC391).

Aluno: Larissa Maiara Fraga
Matricula: 07.1.4152

Orientador: Frederico Gadelha Guimarães

Ouro Preto
1 de outubro 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 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 trabalho propoe um módulo de calibração baseado em algoritmos genéticos.

Palavras-chave: Modelagem Ambiental. Algoritmo Genético. Calibração.

Sumário

1	Introdução	1
2	Justificativa	3
3	Objetivos Gerais e Específicos	4
4	Metodologia	5
5	Cronograma de atividades	6

Lista de Figuras

Lista de Tabelas

1	CalibraçãoAG	5
2	Cronograma de Atividades.	6

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 [13]. O aumento da temperatura global, por exemplo, poderá implicar em explosões populacionais de vetores de doenças tropicais e resultar em epidemias globais [16].

Planejadores, decisores e todos aqueles que se ocupam em projetar e estabelecer políticas públicas precisam de ferramentas de modelagem que sejam conáveis, e capazes de capturar a dinâmica e os resultados das ações humanas [9] [10] [1]. A simulação de processos naturais ou a simulação das interações humano-ambiente são instrumentos de pesquisa de impactos e previsões [2]. 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 através da sua otimização e obtenção de melhores resultados, essas são chamadas políticas saudáveis [4].

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 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[11]) 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 (SAUDEL) [12].

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 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. [6] 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 [5]. Uma revisão da aplicação de técnicas de inteligência artificial e computacional em sistemas de modelagem ambiental é apresentada em [18], 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 [8]. Algoritmos genéticos também têm sido amplamente estudados para evoluir regras e comportamentos em autômatos celulares [14][3].

Esta proposta de trabalho objetiva 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.

2 Justificativa

É clara e reconhecida a importância do desenvolvimento de modelos ambientais baseados no impacto que eles podem causar no cotidiano da sociedade. A simulação de processos ambientais ou a simulação das interações humano-ambiente são instrumentos de pesquisa de impacto e predição. 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 [9] [10] [1].

Para a calibração dos parâmetros dos modelos desenvolvidos no ambiente TerraME, o modelador dispunha até então somente da técnica de simulação de Monte Carlo [17], 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.

3 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 [15][7]. 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 e 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.

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 vários modelos práticos sejam feitos para validar a ferramenta, testando sua confiança.

4 Metodologia

A etapa mais custosa do desenvolvimento da ferramenta será a etapa de implementação. A função de calibração terá 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

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	matrizLimites Matriz (n x 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.

Tabela 1: CalibraçãoAG

5 Cronograma de atividades

Na Tabela 2, está detalhado o cronograma com as seguintes atividades:

1. Revisão bibliográfica sobre Algoritmos Genéticos.
2. Implementação do AG aplicado à ferramenta.
3. Implementação da generalização da ferramenta.
4. Revisão bibliográfica de Algoritmos Evolutivos adequados para calibração de modelos estocásticos.
5. Implementação do Algoritmo Evolutivo escolhido para a aplicação.
6. Testes da ferramenta desenvolvida em modelos práticos.

Atividades	Ago	Set	Out	Nov	Dez
1	X				
2	X	X			
3		X	X		
4			X		
5				X	
6				X	X
Redigir a Monografia			X	X	X
Apresentação do Trabalho					X

Tabela 2: Cronograma de Atividades.

Referências

- [1] Steven Sanderson Gunther Fischer Louise Fresco B. L. Turner II, David Skole and Rik Leemans. 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, 1995.
- [2] Mateus Batistella and Emilio F. Moran. *Dimensões humanas do uso e cobertura das terras na amazônia: uma contribuição do LBA*. Acta Amazonical, 2 edition, April/June 2005.
- [3] R. Breukelaar and Th. Bäck. Using a genetic algorithm to evolve behavior in multi dimensional cellular automata. *Proc. of the Genetic and Evolutionary Computation Conference (GECCO'05)*, page 107114, 2005.
- [4] Ministério da Saúde. Abordagens espaciais na saúde pública. *Fundação Oswaldo Cruz, Brasília*, 1:1186–1194, 2006.
- [5] Donato D'Ambrosio and William Spataro. Parallel evolutionary modelling of geological processes. *Parallel Computing*, pages 186–212, 2007.
- [6] William Spataro Donato D'Ambrosio, Giulio Iovine and Hideaki Miyamoto. A macroscopic collisional model for debris-flows simulation. *Environmental Modelling Software*, pages 1417–1436, 2007.
- [7] Agoston E. Eiben and James E. Smith. Multi-histogram equalization methods for contrast enhancement and brightness preserving. *Introduction to Evolutionary Computing.*, 2003.
- [8] editor F. Recknagel. Ecological informatics: Understanding ecology by biologically inspired computation. *Springer-Verlag*, 1996.
- [9] David Kaimowitz and Arild Angelsen. Economic models of tropical deforestation: A review. *Technical report, Center for International Forestry Research, Bogor, Indonesia*, 1998.
- [10] E. F. Lambin. Modeling deforestation processes, a review. trees, tropical ecosystem environment observation by satellites. *Research Report 1, European Commission Joint Research Center / European Space Agency*, 1994.
- [11] E. F. Lambin and editors Helmut J. Geist. Land-use and land-cover change: Local processes and global impacts. *Global Change - The IGBP Series*, 2006.
- [12] Raquel Martins Lana. Modelos dinâmicos acoplados para simulação da ecologia do vetor aedes aegypti. *Master's thesis*, 2009.
- [13] Carlos A. Nobre Luis F. Salazar and Marcos D. Oyama. Climate change consequences on the biome distribution in tropical south america. *Geophysical Research Letters*, pages 1–6, 2007.

- [14] J. P. Crutchfield M. Mitchell and P. T. Hraber. Evolving cellular automata to perform computations: Mechanisms and impediments. *Physica D*, pages 361–391, 1994.
- [15] Melanie Mitchell. An introduction to genetic algorithms. complex adaptive systems. *MIT Press*, 1998.
- [16] Steven Riley. Large-scale spatial-transmission models of infectious disease. 2007.
- [17] R. Y. Rubinstein and D. P. Kroese. Simulation and the monte carlo method. *John Wiley Sons*, 2007.
- [18] Anthony J. Jakeman Serena H. Chen and John P. Norton. Artificial intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, pages 379–400, 2008.