



MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

**Epidemiologia da Malária através de Autômatos Celulares e Modelagem
baseada em Indivíduos**

Jones Albuquerque

Instituições participantes

Departamento de Estatística e Informática - DEINFO/UFRPE

Centro de Pesquisas Aggeu Magalhães - CPqAM/FIOCRUZ

Departamento de Física e Engenharia Nuclear - Universidade Politécnica da Catalúnia

Recife

Maior de 2011

Sumário

1	Qualificação do principal problema a ser abordado	1
1.1	A Malária como doença endêmica Mundial	1
1.2	Por que o Uso de Autômatos Celulares para Modelagem de Epidemiologia?	2
2	Objetivos gerais	2
3	Metas	3
4	Metodologia científica e atividades de execução	3
4.1	Modelos matemático-computacionais	3
4.2	Atividades e processo de execução	6
5	Principais contribuições científicas e tecnológicas esperadas	7
5.1	Prêmios	7
5.2	Eventos Científicos	8
5.3	Software e negócios	8
5.4	Disciplinas ministradas	8
5.5	Considerações sobre as Contribuições para a UFRPE e para o Brasil	9
6	Por que os estudos na Universidade Politécnica da Catalúnia?	10
7	Interesse de empresas e parceiras de colaboração já estabelecidas	10
8	Cronograma de Atividades	11
9	Aspectos éticos	12

1 Qualificação do principal problema a ser abordado

1.1 A Malária como doença endêmica Mundial

A malária é reconhecida como grave problema de saúde pública no mundo, ocorrendo em quase 50% da população, em mais de 109 países e territórios. Sua estimativa é de 300 milhões de novos casos e 1 milhão de mortes por ano, principalmente em crianças menores de 5 anos e mulheres grávidas do continente africano. No mundo, a doença vem sendo estudada por vários órgãos mundiais, como o CDC (*Centers for Disease Control and Prevention*)¹, cujo mapa da Figura 1 ilustra países da África e Ásia, além dos Países das Américas cobertos pela Floresta Amazônica, como áreas endêmicas da Malária.

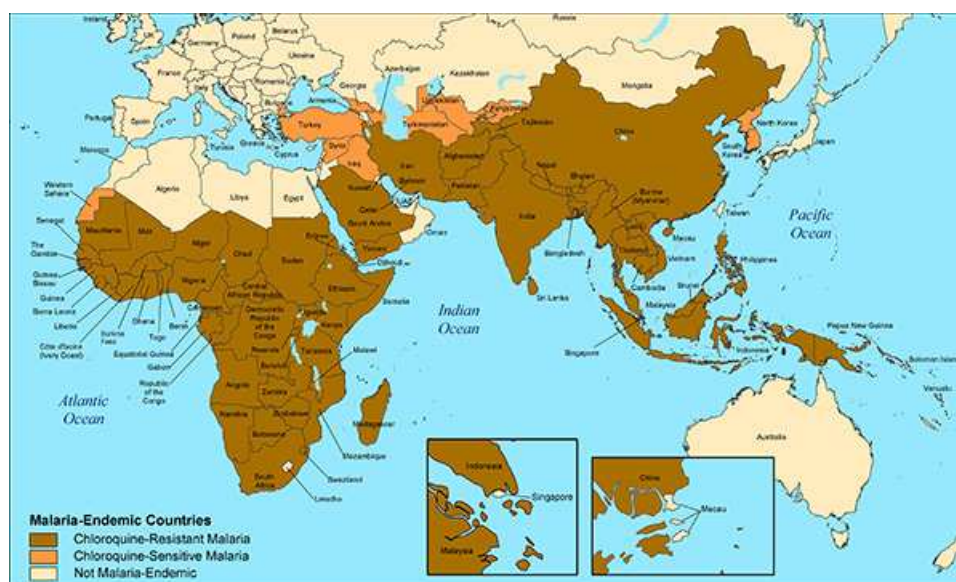


Figura 1: Mapa com áreas endêmicas da Malária na África e Ásia.

Fonte: [CDC - Centers for Disease Control and Prevention - www.cdc.org].

A região amazônica é considerada a área endêmica do País para malária. Em 2008 no Brasil, aproximadamente 97% dos casos de malária se concentraram em seis estados da região amazônica: Acre, Amapá, Amazonas, Pará, Rondônia e Roraima. Os outros três estados da região amazônica Maranhão, Mato Grosso e Tocantins foram responsáveis por menos de 3% dos casos de malária no país. A maioria dos casos ocorre em áreas rurais, mas há registro da doença também em áreas urbanas (cerca de 15%). Mesmo na área endêmica, o risco de contrair a doença não é uniforme. Este risco é medido pela incidência parasitária anual (IPA), que serve para classificar as áreas de transmissão em alto, médio e baixo risco, de acordo com o número de casos por mil habitantes (vide

¹Vide <http://wwwnc.cdc.gov/travel/yellowbook/2010/chapter-2/malaria.htm>

(portal.saude.gov.br)).

1.2 Por que o Uso de Autômatos Celulares para Modelagem de Epidemiologia?

1. Porque, aqui no Brasil, temos tido sucesso com a modelagem de Esquistossomose, onde no *XXI International Symposium on Schistosomiasis* (Congresso Internacional de Esquistossomose), de frequência bi-anual, e que ocorreu em Salvador-BA nos dias 20 a 22 de agosto de 2008, o projeto Xiscanoé (www.xiscanoe.org), que modelou o processo de expansão da Esquistossomose, foi escolhido como o **melhor trabalho científico do Congresso**;
2. Porque o sucesso do estudo de Autômatos Celulares Determinísticos quando aplicados à Epidemiologia da Esquistossomose sugere sua aplicabilidade em outras endemias;
3. Porque vários modelos matemático-computacionais têm sido propostos com o objetivo de capturar as particularidades e características de comportamento dos vetores de propagação de doenças na área de epidemiologia. Nesta direção está-se trabalhando com Redes Neurais artificiais, Algoritmos Genéticos, Redes Complexas (vide Pro(www.xiscanoe.org), link "Projects").

Por estes fatores, especificamente, **esta proposta objetiva projetar e desenvolver modelos matemático-computacionais mais generalistas (Probabilísticos) que capture variáveis mais relevantes de todo do processo de comportamento e expansão da Malária como objeto de estudo, estendendo o que já vem sendo realizado com o Projeto Xiscanoé em Carne de Vaca para o caso de Autômatos Determinísticos**. E, a partir de então, fornecer ferramentas computacionais com os modelos calibrados para que se tornem efetivas ferramentas de auxílio e trabalho dos pesquisadores deste projeto e em seus futuros estudos epidemiológicos.

2 Objetivos gerais

Este projeto pretende prover a sociedade, o Estado e a comunidade científica de modelos matemático-computacionais para auxiliar o controle da Malária. A sociedade, pois estará-se melhorando o nível de controle e, conseqüentemente, do serviço prestado pelos órgãos públicos à população afetada pela doença. O Estado, pois possuirá cenários para poder otimizar a distribuição de seus recursos financeiros e humanos no monitoramento, controle

e combate da doença. A comunidade científica, pois espera-se que os modelos propostos para a previsibilidade de comportamentos em Malária contribuam para o estado-da-arte em epidemiologia e matemática computacional, uma vez que são áreas carentes de contribuições computacionais que as viabilizem como parceiras. Além de prover um estudo de comportamento de máquinas computacionais abstratas como os Autômatos Celulares, as quais, no Brasil, não há grupo completamente consolidado que as estude em seus aspectos teórico-computacionais caracterizando assim o que a literatura mundial chama de **Epidemiologia Computacional**.

3 Metas

1. Modelar a epidemiologia da Malária em Regiões Endêmicas e com baixa endemicidade. Criar instrumentos de comparação entre as epidemias causadas por *P. falciparum* e *P. Vivax*;
2. Comparar a abordagem de modelagem da Epidemiologia da Malária com Autômatos Celulares com outras abordagens;
3. Avaliação dos modelos desenvolvidos como ferramentas efetivas de combate e análise da expansão da doença;
4. Avaliar a implementação de modelos em máquinas paralelas de modo a melhorar as performance das simulações;
5. Divulgar nos meios de divulgação científica os resultados obtidos.

4 Metodologia científica e atividades de execução

4.1 Modelos matemático-computacionais

Um modelo pode ser visto como a representação aproximada de algum problema real utilizando uma determinada linguagem (matemática, lógica, geográfica, física, etc.) e respeitando uma ou mais teorias. Modelos que utilizam linguagem matemática são denominados modelos matemáticos e podem ser representados por um conjunto de equações e/ou expressões. Construir um modelo permite colocar a complexidade de um problema real dentro de uma estrutura lógica passível de ser analisada. Com isso, é possível evidenciar as alternativas de decisão e seus efeitos previstos, indicando dados que são relevantes e levando a conclusões informativas [8].

O processo de construção de modelos é chamado modelagem. Existem diversas técnicas para modelagem matemática e muitas delas têm sido usadas para representar fenômenos epidemiológicos. A seguir, serão descritas algumas das técnicas a serem ensaiadas na modelagem matemático-computacional da Epidemiologia da Malária a exemplo do que já vem sendo realizado pelo Projeto Xiscanoé com a Modelagem dea Epidemiologia da Esquistossomose em Pernambuco.

Autômatos Celulares

Autômatos Celulares [13] que representam sistemas dinâmicos, onde o tempo e o espaço são discretos, vêm sendo utilizados na literatura como modelos matemático-computacionais para simulação de objetos incluindo em epidemiologia [4, 12, 6]. Autômatos Celulares são definidos como a evolução dos estados das células que o compõe. O estado de uma célula $\sigma_i^t \in \{0, 1\}$ indica que na posição i no tempo t a célula assume um dos estados definidos, neste caso 0 ou 1. Assumindo uma rede N-dimensional de células, tem-se um Autômato N-dimensional. A evolução dos estados das células é dada por uma função, assim a regra de evolução é definida como:

$$\sigma_i^{t+1} = f(\sigma_{i-k}^t, \dots, \sigma_i^t, \dots, \sigma_{i+k}^t),$$

onde k é o índice de iterações. A regra de evolução é aplicada simultaneamente em todas as células. O estado de uma célula no tempo $t + 1$ depende do estado das $2k + 1$ células no tempo t , o que constitui sua vizinhança, como ilustrado na Figura 2. Tais sistemas

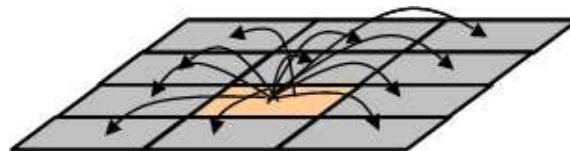


Figura 2: Ilustração de uma célula de Autômato Celular e a influência de sua vizinhança em suas gerações.

conseguem gerar espaços de solução os mais variados possíveis configurando cenários de previsibilidade. Assim, é possível, com auxílio de especialistas filtrar tais cenários para garantir determinado grau de confiança nas respostas do modelo. Mesmo assim, quando o conjunto de variáveis é grande, o grau de previsibilidade pode não colaborar para uma aplicação prática na qual se deseja obter planejamento estratégico a partir das respostas dos modelos. Além do tempo computacional previsto para simulação de tais modelos ser um aspecto restritivo quando o conjunto de cenários se torna complexo [10]. Entretanto, para o caso de localidades pequenas, como publicado pelo Projeto Xiscanoé em Carne de Vaca, esta análise se torna viável computacionalmente.

Autômatos Celulares Probabilísticos

Autômatos Celulares onde as regras e equações de definição de gerações obedecem um processo estocástico são definidos como probabilísticos. Assim, incorporar processos markovianos às regras de evolução dos modelos podem contribuir para melhorar o grau de certeza sobre os cenários gerados pelos modelos, como já foi exercitado preliminarmente no projeto Xiscanoé em um de seus estudos [9].

Redes Bayseanas

Matematicamente, uma Rede Bayesiana é uma representação compacta de uma tabela de conjunção de probabilidades do universo do problema. Por outro lado, do ponto de vista de um especialista, Redes Bayesianas constituem um modelo gráfico que representa de forma simples as relações de causalidade das variáveis de um sistema. Uma Rede Bayesiana consiste do seguinte:

- Um conjunto de variáveis e um conjunto de arcos ligando as variáveis;
- Cada variável possui um conjunto limitado de estados mutuamente exclusivos;
- As variáveis e arcos formam um grafo dirigido sem ciclos (DAG);
- Para cada variável A que possui como pais B_1, \dots, B_n , existe uma tabela $P(A|B_1, \dots, B_n)$.

Observa-se que, caso A não possua um pai, a tabela de probabilidades é reduzida para uma probabilidade incondicional $P(A)$. Uma vez definida a topologia da rede, basta especificar as probabilidades dos nós que participam em dependências diretas, e utilizar estas para computar as demais probabilidades que se deseje. Redes Bayesianas têm sido utilizadas no diagnóstico e prevenção de doenças em epidemiologia [11]. Em [7], uma rede bayesiana é utilizada para diagnosticar grupos de risco para contração da doença tuberculose em populações endêmicas. A Figura 3 ilustra a Rede Bayesiana utilizada.

Programação Linear Modelos de programação linear têm como objetivo determinar uma solução que minimiza uma função, sujeita a um conjunto de equações e/ou inequações lineares. Estão presentes em muitas áreas de aplicações, com ênfase em problemas de economia e engenharia.

Um modelo de programação linear pode ser descrito como,

$$\begin{aligned} \min \quad & c^T x \\ \text{s.a.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{1}$$

sendo $A \in \mathbb{R}^{m \times n}$ a matriz de restrições, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ e $x \in \mathbb{R}^n$ os vetores de termos independentes, custos e variáveis de decisão, respectivamente.

zar o conhecimento dos pesquisadores colaboradores deste projeto que já estão imersos neste estudo em suas atividades individuais para especificar e gerenciar a implementação dos sistemas computacionais envolvendo os respectivos estudantes associados aos programas de pós-graduação aos quais estão vinculados. Os pesquisadores colaboradores deste projeto possuem experiência na implementação de modelos computacionais como ocorreu com o Projeto SIMULARE - An Open Tool for Simulation, disponível em <http://xsimulare.tigris.org/>, no ANKOS - A New Kind Of Simulator, disponível em <http://ankos.sourceforge.net>. e no próprio Epischisto;

2. Selecionar os melhores modelos e viabilizá-los como ferramenta de trabalho para os pesquisadores em parasitologia;

A execução desta atividade se dará por meio da experiência de membros do grupo em atividades como esta: transformar um protótipo em um produto com conceitos de engenharia de software utilizando o conceito de Fábricas de Software, vide <http://200.17.137.110:8080/in953>. Para o caso particular do projeto Xiscanoé, já há uma fábrica de software em definição e operação inicial. As atividades e projetos atuais da fábrica podem ser acompanhadas pelo site: <http://www.geemap.com.br/>.

3. Divulgar nos meios de divulgação científica os resultados obtidos.

5 Principais contribuições científicas e tecnológicas esperadas

Até então, de **novembro de 2006 a dezembro de 2009**, o projeto Xiscanoé, chamado Epischisto desde **janeiro de 2010**, produziu, além de vários Resumos, Artigos em Conferências e Periódicos todos disponíveis em (<http://www.epischisto.org>) os seguintes produtos:

5.1 Prêmios

1. **Melhor Trabalho Científico** do *XXI International Symposium on Schistosomiasis* (Congresso Internacional de Esquistossomose), de frequência bi-anual, e que ocorreu em Rio de Janeiro-RJ nos dias 05 a 08 de outubro de 2010. Organizado pelo PIDE - Programa Integrado De Esquistossomose da Fundação Oswaldo Cruz.
2. **Melhor Trabalho Científico** do *XXI International Symposium on Schistosomiasis* (Congresso Internacional de Esquistossomose), de frequência bi-anual, e que ocorreu

em Salvador-BA nos dias 20 a 22 de agosto de 2008. Organizado pelo PIDE - Programa Integrado De Esquistossomose da Fundação Oswaldo Cruz.

5.2 Eventos Científicos

1. **I WMMC - I Workshop de Modelagem Matemático-Computacional Aplicada à Saúde Pública** tem por objetivo reunir pesquisadores das áreas envolvidas na modelagem de fenômenos epidemiológicos relacionados à Saúde Pública. Financiado pela FACEPE/FINEP/CNPq sob processo número: ARC-0080-1.03/07. O site do evento pode ser encontrado em <http://200.17.137.110:8080/wmmc>;
2. Organizou juntamente com o Programa de Pós-Graduação em Biologia Molecular da UFPE o **Curso de BioInformática Aplicada a Genética**, cujo material, entre outros, encontra-se disponível em <http://200.17.137.109:8081/xiscanoe/courses-1/>;
3. Participou das IV e V Semanas Nacionais de Ciência e Tecnologia com palestras sobre Epidemiologia Computacional;
4. Ministrou, em Semanas acadêmicas na UFRPE, JED -Jornada de Educação Digital e Semana de Biologia, cursos de modelagem computacional e epidemiologia computacional.

5.3 Software e negócios

1. RIBEIRO, R. A., Mariz, L., ALBUQUERQUE, J. O., Souza, Marco A. A., BARBOSA, C. C. G. S. ANKOS - A New Kind of Simulator, 2008. <http://ankos.sourceforge.net>;
2. Brandão, E. Araújo, M. Pereira, P. Wagner, R. Geemap - geographical Environment for Epidemic Mapping, 2008. <http://www.geemap.com.br/>;
3. Encontra-se em fechamento de negociação com a WRI - Wolfram Research Inc. (<http://www.wolfram.com>) a disponibilização de vinte (20) licenças do Mathematica 6.0 para uso e treinamento de capital humano para o projeto e divulgação do ambiente de desenvolvimento produzido pela WRI.

5.4 Disciplinas ministradas

As disciplinas responsáveis pela formação do Capital Humano para o projeto foram criadas e ofertadas aos estudantes de cursos de graduação (Licenciatura em Matemática, Licen-

ciatura em Física, Licenciatura em Computação, Bacharelado em Biologia e Bacharelado em Sistemas de Informação) na UFRPE. As disciplinas são:

1. Elementos de Epidemiologia Computacional;
2. Modelagem Computacional Aplicada a Epidemiologia;
3. Computação Evolutiva;
4. Introdução a Inteligência Artificial;
5. Tópicos em Otimização;
6. Modelos em Redes Complexas.

5.5 Considerações sobre as Contribuições para a UFRPE e para o Brasil

Observa-se que nos últimos 4 anos, o projeto Xiscanoé, agora Epischisto, tem sido bastante produtivo. Neste período, também foram formados vários estudantes na área de Epidemiologia Computacional totalizando 18 estudantes-pesquisadores, destes, seis (8) atualmente cursam programas de Pós-Graduação.

Também, com a aglutinação de pesquisadores da UFRPE em torno do tema de modelagem, foi constituído um programa de Mestrado em Informática Aplicada que tem como uma área de pesquisa Computação Inteligente e Modelagem e que tem entre suas linhas de pesquisa: (1) Modelagem Matemático Computacional e Simulação e (2) Epidemiologia Computacional Aplicada a Saúde Pública. Além disso, o Projeto Epischisto é um dos principais projetos em andamento do recém-criado Mestrado, vide: <http://www.ppgia.ufrpe.br>

Então, para esta proposta pretende-se investigar o uso de Autômatos Celulares na Epidemiologia da Malária e, pelo menos, repetir os mesmos índices do projeto Xiscanoé/Epischisto em formação de Capital Humano, Produção Científica e Tecnológica que vem se mantendo desde novembro de 2006. Ainda, pretende-se com a incorporação de empresas interessadas (vide Seção *Interesse de empresas e parceiras já estabelecidas*) consolidar a inserção das tecnologias computacionais desenvolvidas pelo Epischisto na sociedade de pesquisadores de outras endemias.

6 Por que os estudos na Universidade Politécnica da Catalúnia?

Um dos principais grupos de pesquisa em Modelagem de Epidemiologia, parceiros da FIOCRUZ, está no Departamento de Física e Engenharia Nuclear da Universidade Politécnica da Catalunya(UPC). O *Discrete Modelling and Simulation of Biological Systems* (MOSIMBIO) é um grupo consolidado, com vários projetos, associado a um programa de Pós-Graduação com várias Teses e Dissertações defendidas no tema, vide: <http://mosimbio.upc.edu/>

Além disso, na UFRPE, o Departamento de Estatística e Informática, teve um Programa de Pós-Graduação, nível Mestrado, aprovado e em suas Linhas de Pesquisa estão contempladas Epidemiologia Computacional e Modelagem Computacional. Assim esta proposta está em completa sintonia científica com a UFRPE.

Por outro lado, o projeto Episichisto vem formando diversos estudantes de Iniciação Científica e e estudantes de graduação com a oferta de várias disciplinas optativas de caráter formativo e científico contribuindo para a complementação acadêmica destes discentes como pode ser visto em vide: <http://200.17.137.109:8081/xiscanoe/courses-1>.

Assim, observa-se uma completa inserção da pesquisa proposta e da cooperação internacional a vir a se estabelecer entre a UFRPE e a UPC.

7 Interesse de empresas e parceiras de colaboração já estabelecidas

Atualmente, o projeto Episichisto está em negociação com o WRI (<http://www.wolfram.com>), empresa que produz o Mathematica, ferramenta utilizada nas simulações por Autômatos Celulares, para que este financie os pesquisadores em seus estudos de Pós-doc em seus laboratórios em Illinois - EUA.

Também, a Geemap (<http://www.geemap.com.br/>), empresa constituída pelo Porto Digital/CESAR e está interessada em produtizar os protótipos do Projeto Episichisto para serem comercializados pela FIOCRUZ e OMS. A exemplo do software de catalogação e descrição de pontos de coleta desenvolvido em parceria com a Geemap e disponível em seu site e reproduzido na Figura 4.

Ainda, a SOFISTIC (<http://sofistic.tecnologia.ws/site/>), empresa em constituição pelo Porto Digital/CESAR.edu está desenvolvendo uma versão mobile para dispositivos móveis do ANKOS. Assim, agentes de saúde e pesquisadores poderão sincronizar

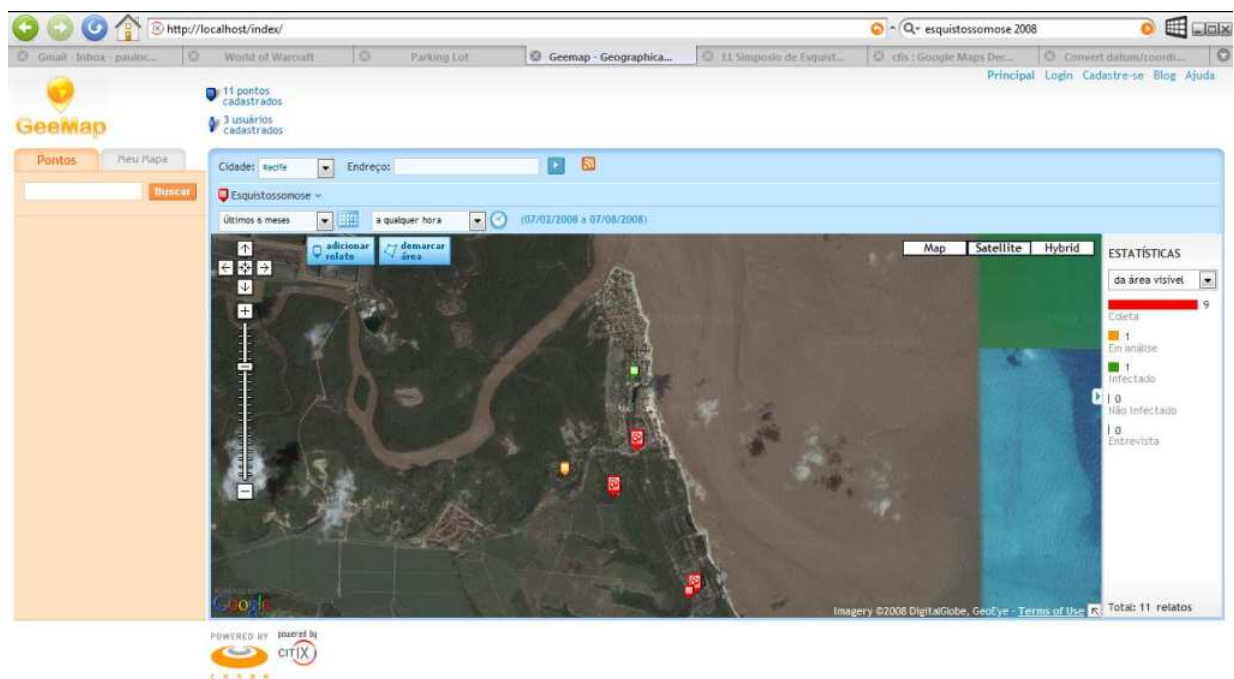


Figura 4: Tela do Geemap com sistema integrado a fotos de satélite. Neste caso, foto de Carne de Vaca - Goiana com as marcações dos pontos de coleta.

informações de coleta diretamente de seus dispositivos celulares.

8 Cronograma de Atividades

O período proposto para desenvolvimento do projeto será de 12 meses, com início previsto para o 2o semestre de 2011. A seguir encontra-se um cronograma para as atividades a serem desenvolvidas:

1. Modelar a epidemiologia da Malária em Regiões Endêmicas e com baixa endemicidade. Criar instrumentos de comparação entre as epidemias causadas por *P. falciparum* e *P. Vivax* - 6 meses;
2. Comparar a abordagem de modelagem da Epidemiologia da Malária com Autômatos Celulares com outras abordagens - 2 meses;
3. Avaliação dos modelos desenvolvidos como ferramentas efetivas de combate e análise da expansão da doença - 3 meses;
4. Avaliar a implementação de modelos em máquinas paralelas de modo a melhorar as performance das simulações - 1 mês;
5. Divulgar nos meios de divulgação científica os resultados obtidos - Durante todo o período.

9 Aspectos éticos

O desenvolvimento deste projeto não representa risco para a população humana das localidades estudadas. Todo o procedimento de coleta de material físico-químico e biológico não representa risco para a população local, que não está envolvida nas atividades. Os técnicos e pesquisadores envolvidos no estudo são profissionais já treinados e capacitados nas referidas condutas. Todo o material biológico e físico-químico será analisado no laboratório de Esquistossomose do Centro de Pesquisa Aggeu Magalhães / Fiocruz, Serviço de Referência em Esquistossomose da região Nordeste, e, portanto, já certificado e apto a estas atividades.

Referências

- [1] J. O. Albuquerque and C Coelho JR. *A Stochastic Approach for Conceptual Level Codesign*. PhD thesis, Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, 2001.
- [2] J. O. Albuquerque, C. Coelho JR, D. Cecílio JR, A. O. Fernandes, and G. R. Mateus. Solving a stochastic formulation for hardware/software codesign. In *9th International Conference on Stochastic Programming*, 2001.
- [3] S. Bocanegra, F. F. Campos, and A. R. Oliveira. Using a hybrid preconditioner for solving large-scale linear systems arising from interior point methods. *Computational Optimization and Applications*, 36:149–164, 2007.
- [4] N. Boccara, K. Cheong, and M. Oram. A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behaviour. *Journal of Physics A: Mathematical and General*, 27:1585–1597, 1994.
- [5] Universidade de São Paulo. Modelos matemáticos em epidemiologia. Disciplina do Programa Pós-Graduação em Matemática, 2005.
- [6] Shih Ching and Fu. *Modelling Epidemic Spread using cellular automata*. Master thesis, Department of Computer Science and Software Engineering, The University of Western Australia, 2002.
- [7] L. Getoor, J. T. Rhee, and D. Koller. Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine*, 30:233–256, 2004.
- [8] M. C. Goldbarg and H. L. L. Pacca. *Otimização Combinatória e Programação Linear*. Editora Campus, 2a Edição, 2005.

- [9] B. Miranda, C. G. Melo, J. Albuquerque, S. Bocanegra, H. Paredes, R. Souza-Santos, M. A. Souza, and C. S. Barbosa. Autômatos probabilísticos para simulação do processo de expansão da esquistossomose entre humanos. In *XI International Symposium on Schistosomiasis*, Salvador-BA, 2008. FIOCRUZ.
- [10] W. R. Oliveira, M. C. P. Souto, and T. B. Ludermir. Turing’s analysis of computation and artificial neural networks. *Journal Of Intelligent And Fuzzy Systems*, pages 63–91, 2003.
- [11] A. Rodin, T.H. Mosley, A. G. Clark, C. F. Sing, and E. Boerwinklei. Mining genetic epidemiology data with bayesian networks application to apoe gene variation and plasma lipid levels. *J Comput Biol.*, 12(1):1–11, 2005.
- [12] G. Rousseau, B. Giorgini, R. Livi, and H. Chate. Dynamical phases in a cellular automaton model for epidemic propagation. *Physica D*, 103:554–563, 1997.
- [13] S. Wolfram. *A New Kind of Science*. Published by Wolfram Media, 2002.