

ERINALDO LEITE SIQUEIRA JÚNIOR

Algoritmos genéticos como mecanismo de obtenção de regras de evolução de Autômatos Celulares - Estudo de caso com infecção por esquistossomose

RECIFE-PE - FEV/2014

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Algoritmos genéticos como mecanismo de obtenção de regras de evolução de Autômatos Celulares - Estudo de caso com infecção por esquistossomose

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada para defesa pública como exigência parcial para obtenção do título de Doutor.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Prof. Dr. Tiago Alessandro Espínola Ferreira

Co-orientador: Prof. Dr. Jones Oliveira de Albuquerque

RECIFE-PE - FEV/2014.

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**Algoritmos genéticos como mecanismo de obtenção de regras de evolução de
Autômatos Celulares - Estudo de caso com infecção por esquistossomose**

Erinaldo Leite Siqueira Júnior

Tese julgada adequada para obtenção do título de doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 25/02/2014 pela Comissão Examinadora.

Orientador:

Prof. Dr. Tiago Alessandro E. Ferreira
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Jones Oliveira de Albuquerque
Universidade Federal Rural de Pernambuco
DEInfo - UFRPE

Prof. Dr. Wilson Rosa de Oliveira Junior
Universidade Federal Rural de Pernambuco
PGBIOM - UFRPE

Prof. Dr. Cláudio Tadeu Cristino
Universidade Federal Rural de Pernambuco
PGBIOM - UFRPE

Prof. Dr. Sílvio de Barros Melo
Universidade Federal de Pernambuco
CIn - UFPE

Dedicado aos meus pais, por toda cobrança
e compreensão.

"The chess-board is the world; the pieces are the phenomena of the universe; the rules of the game are what we call the laws of Nature."

T. H. Huxley

Resumo

A esquistossomose ou bilharzÍase é uma doença crônica causada por platelmintos parasitas e multicelulares do gênero *Schistosoma*. É a mais grave forma de parasitose por organismo multicelular, matando milhares de pessoas por ano.

O presente trabalho tem objetivo de modelar a dinâmica de infecção do hemoparasita *S. mansoni*, através do uso de modelo baseado autômatos celulares. Para tanto, foi necessário o conhecimento prévio das particularidades do ciclo do trematodo, bem como sua forma de contato e o meio em que tal contato ocorre. Diante dessas informações foi modelada a dinâmica da população de um sítio artificial com características semelhantes às encontradas em nosso litoral. Foram realizadas simulações com três níveis para cada parâmetros: baixo, médio e alto; de acordo com a amplitude que cada parâmetro é compreendido na literatura corrente e analisado o nível de infecção da região para 5 passos temporais bem definidos. Foram obtidos os mais diversos cenários, desde quadros de diminuição do nível de infecção do sítio, até cenários em que o nível de infecção subiu na ordem de 334 %.

Posterior a essa análise, foi realizado um aprofundamento quanto aos parâmetros que tinham maior significância na determinação do avanço da infecção. Para isso, utilizamos a árvore de decisão para determinarmos tais parâmetros. Através do quadro apresentado, o pesquisador terá total conhecimento sobre o cenário que apresenta maior probabilidade de ocorrer comparando os parâmetros de campo. Além disso, de posse da estimativa dos parâmetros de campo, poderá confrontar o quadro de infecção e através dessa informação identificar os principais parâmetros que merecem atenção para eliminar ou atenuar o quadro atual.

Para os parâmetros, através da coleta dos moluscos, questionários aplicados a população do sítio em estudo e exames laboratoriais, é possível aproximar o nível de infecção da localidade. A partir dessa informação, e realizada nova consulta a população sob um intervalo de tempo definido e estimado o novo nível de infecção, foi possível estimar a coleção de parâmetros (ou uma coleção de parâmetros equivalentes que culminam num mesmo quadro) através de uma estratégia de otimização baseada em algoritmo genético.

Palavras-chave: Esquistossomose; Algoritmo genético; Autômatos celulares; Simulação.

Abstract

Schistosomiasis or bilharzia is a chronic disease caused by parasites and multicellular flatworms of the genre *Schistosoma*. It is the most severe form of parasitism by multicellular organism, killing thousands of people a year.

The present study aimed to model the dynamics of infection hemoparasite *S. mansoni*, through the use of model-based cellular automata. Therefore, it was necessary prior knowledge of the particularities of the trematode cycle as well as their contact form and the environment in which such contact occurs. Given this information was modeled population dynamics of an artificial site with similar characteristics to those found on our coast. Simulations with three levels for each parameter were performed: low, medium and high; according to the amplitude of each parameter is comprised in the current literature, and considering the level of infection of region to 5 time steps well-defined. The most varied scenarios were obtained from pictures of decreased level of infection of the region, until scenarios in which the level of infection increased in the order of 334 %.

Subsequent to this analysis, was performed a deepening for parameters that were most significant in determining the progress of the infection. For this, we use the decision tree to determine such parameters. Through the framework presented, the researcher will have full knowledge about the scenario that is most likely to occur comparing field parameters. Furthermore, in possession of the estimate parameters of the field, you can confront the scenario of infection through this information and identify key parameters that need attention to remove or mitigate the current frame.

For the parameters through the collection of molluscs, questionnaires population of the site in study and laboratory tests, it is possible to approximate the level of infection of the locality. From that information, and held new query the population under a defined time interval and estimated the new level of infection was possible to estimate the parameters collection (or collection of equivalent parameters that culminate in the same framework) through a strategy optimization based on genetic algorithm.

Key words: Schistosomiasis; Genetic Algorithm; Cellular automaton; Simulation.

Lista de Figuras

1	Ilustração do deslocamento da cercária (Carvalho, Coelho e Lenzi, 2008).	p. 4
2	Ilustração da penetração pelas cercárias (Carvalho, Coelho e Lenzi, 2008).	p. 5
3	Ilustração da penetração pelas cercárias (Carvalho, Coelho e Lenzi, 2008).	p. 7
4	Ilustração de um ovo de <i>Schistosoma mansoni</i> (Lima, 2013).	p. 8
5	Ilustração produzida por microscopia a laser confocal, a imagem exibe larvas do <i>S. mansoni</i> (miracídio) (Jurberg, 2013).	p. 9
6	Esporocistos e cercárias em várias fases de desenvolvimento. Em A e B os esporocistos aparecem indiferenciados, compactamente arrumados e limitados por uma fina membrana externa. Em C já se pode ver uma cercária em diferenciação (seta) no meio de esporocistos; D mostra cercárias em várias etapas de diferenciação (Carvalho, Coelho e Lenzi, 2008).	p. 11
7	Ilustração na natureza de um processo baseado em fita codificada (DNA) que é submetida a uma caixa de leitura (RNA polimerase) para a criação de uma outra fita codificada (RNA-m). A caixa de leitura, por si só não é uma fita e não atende a definição de auto reprodução.	p. 16
8	Ilustração das vizinhanças de Von Neumann e Moore para os casos de $r = 1$ e $r = 2$.	p. 20
9	Ilustração da geometria da malha para o caso bidimensional, na ordem, geometrias: Toro, cilindro, cone e esférica.	p. 21
10	Ilustração do comparativo entre computação evolutiva e a evolução natural em suas abordagens para solucionar seus problemas. (Eiben, 2008)	p. 24
11	Ilustração da característica monótona e não negativa da função de avaliação (Eiben, 2008).	p. 32
12	Ilustração do ciclo realizado pelo processo de busca; inicializando a coleção, passado por todos os operadores e avaliando se atende ou não o critério de parada.	p. 34

13	Ilustração de um genótipo em sua codificação binária. Note que, neste exemplo, a cada 4 bits formamos um alelo.	p. 34
14	Ilustração do processo de mutação na codificação binária, através do processo bit flip. Em (a) o genótipo inicial e em (b) o mesmo genótipo após a mutação.	p. 35
15	Ilustração do operador de recombinação atuando sobre dois genótipos pais através de uma recombinação de um único ponto.	p. 36
16	Ilustração do operador de recombinação atuando sobre dois genótipos pais através de uma recombinação de três pontos. Note que as partes 2 e 4 de cada fenótipo foram permutadas.	p. 36
17	Ilustração da motivação da configuração da malha	p. 43
18	Boxplot para o agrupamento 1.	p. 51
19	Boxplot para o agrupamento 2.	p. 51
20	Boxplot para o agrupamento 3.	p. 52
21	Boxplot para o agrupamento 4.	p. 52
22	Boxplot para o agrupamento 5.	p. 53
23	Boxplot para o agrupamento 6.	p. 53
24	Histograma para as médias de infecção do agrupamento 1.	p. 56
25	Histograma para as médias de infecção do agrupamento 2.	p. 56
26	Histograma para as médias de infecção do agrupamento 3.	p. 57
27	Histograma para as médias de infecção do agrupamento 4.	p. 57
28	Histograma para as médias de infecção do agrupamento 5.	p. 58
29	Histograma para as médias de infecção do agrupamento 6.	p. 58
30	Q-Q plot para o agrupamento 1 na simulação para 6 meses.	p. 59
31	Árvore de decisão para os agrupamentos.	p. 64
32	Ilustração da malha com as suas taxas de infecção convertidas em cores.	p. 67

Lista de Tabelas

1	Índice pluviométrico da praia de Carne de Vaca.	p. 44
2	Estatísticas descritivas para o agrupamento 1.	p. 54
3	Estatísticas descritivas para o agrupamento 2.	p. 54
4	Estatísticas descritivas para o agrupamento 3.	p. 54
5	Estatísticas descritivas para o agrupamento 4.	p. 55
6	Estatísticas descritivas para o agrupamento 5.	p. 55
7	Estatísticas descritivas para o agrupamento 6.	p. 55
8	Ajuste dos parâmetros para distribuição normal.	p. 59
9	Ajuste dos parâmetros para distribuição normal.	p. 60
10	Ajuste dos parâmetros para o autômato celular	p. 69

Sumário

1	Introdução	p. 1
1.1	Histórico	p. 1
1.2	Migração e desenvolvimento de <i>Schistosoma mansoni</i> no hospedeiro definitivo	p. 3
1.3	Evolução de <i>Schistosoma mansoni</i> no hospedeiro intermediário	p. 8
1.4	Contribuições relevantes	p. 12
1.5	Organização do trabalho	p. 13
2	Autômatos Celulares	p. 15
2.1	Histórico	p. 15
2.2	Definição formal	p. 18
2.3	Variações de um autômato celular	p. 19
3	Algoritmo genético	p. 23
3.1	Computação evolutiva	p. 23
3.2	Histórico computação evolutiva	p. 24
3.3	Contexto biológico	p. 25
3.4	Componentes de um algoritmo evolutivo	p. 27
3.4.1	Representação	p. 27
3.4.2	Função de avaliação (Função de fitness)	p. 28
3.4.3	População	p. 28
3.4.4	Mecanismo de seleção dos reprodutores	p. 29
3.4.5	Operadores de variação: mutação e recombinação	p. 30

3.4.5.1	Mutação	p.31
3.4.5.2	Recombinação	p.31
3.4.6	Mecanismo de seleção dos sobreviventes (Substituição)	p.32
3.5	Representação de um problema através da codificação de um algoritmo genético	p.33
3.5.1	Representação do genótipo - AG	p.34
3.5.2	Operador de Mutação - AG	p.35
3.5.3	Operador de recombinação - AG	p.35
3.5.4	Modelo de população	p.36
3.5.5	Mecanismo de seleção dos reprodutores	p.37
3.5.5.1	Seleção proporcional ao fitness	p.37
3.5.5.2	Seleção por ranking	p.38
3.5.5.3	Seleção pelo algoritmo roleta	p.38
3.5.5.4	Seleção por torneio	p.39
3.5.6	Mecanismo de seleção dos sobreviventes (substituição)	p.39
3.5.6.1	Sobrevivência dos genótipos mais novos	p.39
3.5.6.2	Sobrevivência baseada no valor da função de fitness	p.40
3.6	Algoritmo genético utilizado no trabalho	p.40
3.7	Unificando os conhecimentos	p.40
4	Modelagem matemática	p.41
4.1	O problema	p.41
4.1.1	Mapa	p.41
4.1.2	Camada de indivíduos sadios	p.43
4.1.3	Camada de indivíduos infectados	p.43
4.1.4	Camada de indivíduos recuperado	p.43
4.1.5	Camada de moluscos sadios	p.44

4.1.6	Camada de moluscos infectados	p. 44
4.1.7	Alagamentos	p. 44
4.1.8	Mobilidade dos indivíduos	p. 45
4.1.9	Mobilidade dos moluscos	p. 45
4.1.10	Dinâmica de infecção	p. 46
4.1.11	Modelagem para o saneamento	p. 46
4.1.12	Parâmetros para a modelagem	p. 47
4.1.13	Quantificando a infecção	p. 48
5	Resultados	p. 49
5.1	Simulações dos parâmetros	p. 49
5.2	Árvore de decisão	p. 57
5.3	Ajuste dos parâmetros	p. 65
6	Conclusão	p. 71
	Referências	p. 73

1 Introdução

1.1 Histórico

Segundo o trabalho de Davis (1980, 1992), o gênero *Schistosoma* e os moluscos a ele relacionados evoluíram na Gondwana antes da desagregação da Pangéia e chegaram à Ásia pela placa Indiana dentro dos últimos 50 milhões de anos. Daí, seus ancestrais e dos moluscos já conviveriam nos territórios que hoje constituem a América do Sul e a África.

Estudando a filogenia molecular em sete espécies de esquistossomos, Deprès et al. (1992) sugerem que eram parasitos de roedores e ruminantes na África até há 1-10 milhões de anos, quando os primeiros hominídeos invadiram as áreas de savana, ambientes favoráveis para transmissão de parasitos.

O tráfico de escravos, Files (1951), introduziu da África para o hemisfério ocidental o *Schistosoma mansoni* (*S. mansoni*) em época relativamente recente. Já por parte dos moluscos, investigações de sistemática molecular (Campbell et al., 2000) tem indicado a origem americana do gênero *Biomphalaria*, a partir de um ancestral '*glabrata-like*' que produziu as doze espécies africanas há cerca de cinco milhões de anos, ou mais recentemente, há 1,1 - 1,8 milhões.

A análise filogenética de 23 espécies de *Biomphalaria* (16 neotropicais e 7 africanas) também sugere a origem americana desse gênero (Dejong et al., 2001).

O caso mais antigo de esquistossomose humana de que se tem notícia foi identificado por imunodiagnose (*Enzyme-Liked Immunosorbent Assay - Elisa*) na múmia de um adolescente egípcio infectado pelo *S. haematobium*, que viveu há mais de cinco mil anos (Deelder et al., 1990).

As primeiras observações sobre o agente etiológico da esquistossomose foram feitas em 1851, no Egito, pelo patologista alemão Theodor Bilharz. Nascido em 1825, foi estudante de von Siebold, a quem devia seu interesse em parasitologia. Ao necropsiar um jovem, encontrou na veia porta um helminto o qual compartilhou com seu mestre von Si-

ebold, este por sua vez, apresentou a descoberta em um congresso de naturalistas em Gotha (von Siebold, 1852). Até então era desconhecida a existência de um trematódeo sanguícola.

Em 1902, Manson refere-se a um paciente inglês que tinha residido por 15 anos em várias ilhas do Caribe. O exame microscópico revelou: "Fezes, ovos de bilharzia não numerosos; distribuídos por toda a massa fecal; com espículo lateral. Urina, sem ovos nem células do sangue". No ano seguinte, Manson (1903) escreveu: "Possivelmente há duas espécies de Bilharzia, uma com ovos de espículo lateral depositando-os somente no reto, a outra infestando indiretamente o reto e a bexiga".

Os primeiros autores a investigar a evolução do *Schistosoma* supunham que, como outros trematódeos, devia passar por uma metamorfose em um hospedeiro intermediário, possivelmente molusco, antes de serem capazes de infectar outras pessoas. Segundo Cobbold (1864), as formas larvais deveriam ser encontradas em moluscos gastrópodos próprios às localidades onde existissem os vermes adultos.

Na China, foram encontrados ovos de *Schistosoma japonicum* no fígado e no reto de um cadáver bem conservado de mulher por volta de 2100 a.C. (Liu, 1978). Segundo uma antiga lenda, a esquistossomose japônica originou-se quando um navio naufragou, derramando laca próximo a Katayama, passando os habitantes locais, tanto quanto os animais, a sofrer irritação cutânea ao entrarem em contato com a água, donde o nome 'doença de Katayama', reconhecida depois em outras partes do Japão. A descrição original da doença foi publicada por Fujii em 1847 e republicada em 1909.

Confirmado o ciclo vital do *S. japonicum*, Leiper foi enviado pelo War Office ao Egito, por indicação de Manson, para "investigar a doença bilharziana, recomendar medidas preventivas em relação às tropas". Instalou-se em uma área de alta endemicidade nos arredores do Cairo, ocupando um laboratório da Escola de Medicina e passando a coletar todos os moluscos de água doce da região. Identificou 15 espécies de sete gêneros, inclusive quatro *Bulinus* (*contortus*, *innesi*, *dybowski* e *alexandrinus*) e dois *Planorbis* (*boissyi* e *mareoticus*) (Leiper e Atkinson, 1915).

Já no Brasil, a história da esquistossomose foi iniciada com referência a uma carta de Griesinger enviada de Berlim em 1866 a Otto Wucherer, sugerindo-lhe verificar se as urinas de pacientes com hemalúria continham ovos de *Schistosoma haematobium*. O mesmo não encontrou ovos de *Schistosoma*, mas teve a surpresa (Wucherer, 1868) de achar microfilárias em vários pacientes com hematúria e quilúria, adotando as denominações *Filaria wuchereri* e *Wuchereria bancrofti* (Cobbold, 1859).

Em 1908, Pirajá da Silva, examinando sangue fresco encontrou "hematozoários em forma de crescente e um ovo muito semelhante ao de *Schistosoma mansoni*". Descrevendo brevemente o ciclo do parasito: "ovos expelidos do organismo humano caem na água e um miracídio é libertado; o hospedeiro intermediário ainda é desconhecido, assim como as migrações e transformações do trematódeo não são conhecidas". Ovos do *Schistosoma* também foram vistos nas fezes (Pirajá da Silva, 1912).

Os estudos de Pirajá da Silva constituíram uma das contribuições mais importantes para a demonstração da especificidade do *S. mansoni*.

Independentemente dos pesquisadores japoneses e de Leiper, cujos trabalhos só veio a conhecer quando já estavam adiantadas suas investigações e já havia reproduzido todo o ciclo vital do parasito, iniciou Lutz, em 1916, uma série de estudos sobre a esquistossomose no Brasil. Os resultados foram publicados em várias notas preliminares (Lutz 1916, 1917a, 1917b) e depois reunidas em monografia (Lutz, 1919). Todos os resultados a que chegou Leiper no Egito foram confirmados por Lutz em relação à esquistossomose mansoni, no Brasil.

Ao mesmo tempo em que Lutz estudava a esquistossomose mansoni no Brasil, Iturbe e González (1916) obtinham, na Venezuela, a infecção experimental do *Planorbis guadeloupensis* (= *Biomphalaria glabrata*). Seguiram o desenvolvimento completo do parasito no molusco, infectaram animais com as cercárias e obtiveram os vermes adultos no sistema porta. Encontraram o mesmo planorbídeo altamente infectado em condições naturais, demonstrando assim seu papel de hospedeiro intermediário naquele país. Observações adicionais sobre a epidemiologia e a profilaxia da esquistossomose e a morfologia da cercária foram também publicadas por Iturbe (1917a, 1917b).

1.2 Migração e desenvolvimento de *Schistosoma mansoni* no hospedeiro definitivo

O agente penetrante em mucosa ou pele, a cercária, inicia sua movimentação quando exposto a uma alteração na intensidade luminosa (aumento ou decréscimo) (Saladin, 1982). Sua movimentação é dada através de movimentos natatórios originados por sua cauda e se desloca para a superfície da água, e seguido de momentos de passividade, torna a afundar. A sua aproximação em direção ao hospedeiro é influenciada por perturbações na água (Haas, 1992), à formação de um gradiente termal tão fraco como 0,15° C/mm (Haas et al., 1990), estímulos químicos emitidos pela pele (Wilsonm, 1987) e a estímulos fóticos

(Saladin, 1982).

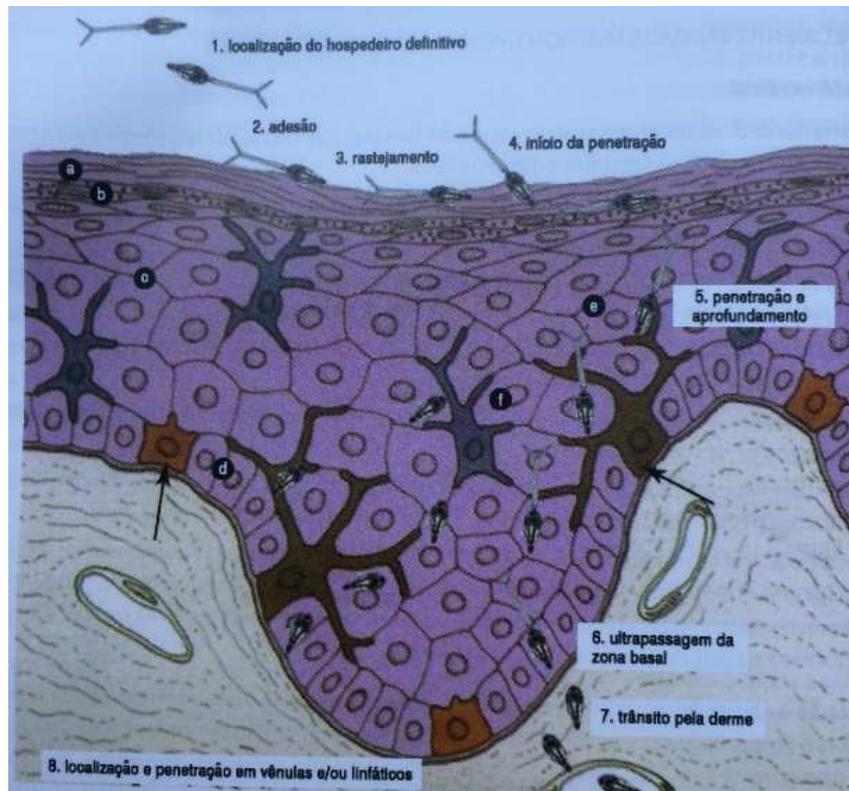


Figura 1: Ilustração do deslocamento da cercária (Carvalho, Coelho e Lenzi, 2008).

A Figura 1 apresenta as diferentes etapas do ingresso da cercária através da pele do indivíduo. Começando pela sua localização, adesão, rastejamento e penetração; até, sua ultrapassagem da membrana basal epidérmica, assim, tendo acesso livre pela derme até encontrar com vênulas e/ou linfáticos.

Ao entrar em contato com o hospedeiro a cercária se adere através de sua ventosa oral, em seguida, modifica seu padrão de deslocamento abandonando a movimentação pela cauda (Wilson, 1987). Através de seu rastejamento sobre a pele do hospedeiro, inicia-se a busca por sítios que permitam a penetração na pele da vítima; uma vez realizada a penetração, novas cercárias que entrem em contato com a pele do mamífero seguirão pelo mesmo caminho fazendo uso dos túneis deixados pelos primeiros penetrantes (Griffiths, 1953). Essa postura visa economia de energia e conteúdo grandular, além de contornar mais facilmente o sistema imunológico do hospedeiro (Stirewalt, 1971).

A duração do ato de penetração das cercárias na pele varia de hospedeiro para hospedeiro, mas pode variar de 3,1 a 17 minutos na pele humana (Stirewalt e Hackey, 1956).

Dessa forma, a penetração em si ocorre de forma rápida e é antecedida por uma breve fase de adesão e rastejamento pela superfície da pele.

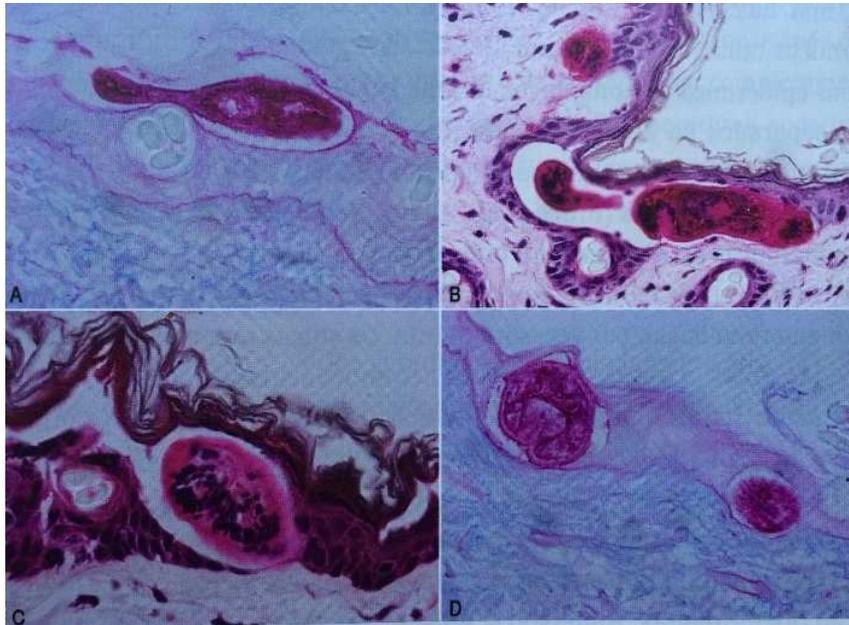


Figura 2: Ilustração da penetração pelas cercárias (Carvalho, Coelho e Lenzi, 2008).

A Figura 2 apresenta quatro imagens, a primeira, mostrando a penetração horizontal sob o estrato córneo (camada mais externa da pele). A segunda, com duas cercárias, uma atingindo a membrana basal da epiderme e a outra localizada parcialmente na derme superficial. A terceira imagem, cercária atingindo a membrana basal epidérmica, e a última, duas cercárias intra-epidérmicas com parte da face inferior tocando na membrana basal epidérmica.

Entre uma a três horas após a penetração das cercárias na pele do hospedeiro, é dado início a uma série de adaptações bioquímicas e morfológicas na tentativa de assegurar a sobrevivência dos parasitos neste ambiente. O parasito passa a não tolerar mais água fresca (Stirewalt, 1974), ocorre também o desvio do metabolismo oxidativo para anaerobiose facultativa, com excreção de ácido láctico (Coles, 1972). Através dessas mudanças, após 40 horas de infecção, as glândulas acetabulares e seus ductos desaparecem caracterizando a completa transformação das cercárias em esquistossômulos.

A saída dos esquistossômulos da derme pode ocorrer pelo sistema venoso (vênulas), mais raramente por capilares venosos dilatados ou por linfáticos ou por ambos os sistemas vasculares (Abboud, Hanna e Regab, 1971).

A via linfática é utilizada entre 1,25%-15% dos esquistossômulos, em murinos (Gui et al., 1995), e por pelo menos 1%, em babuínos (*Papio anubis*) (Pearce e Maclaren, 1986). Parte dessa pouca procura pelos esquistossômulos se deve ao pequeno número desse tipo de vasos em comparação com o grande número de vasos sanguíneos. Por essa via, os esquistossômulos atingem os linfonodos regionais da pele já a partir do segundo dia de infecção pelos próximos cinco ou seis dias (Gui, 1995).

Acessando o sistema circulatório, seja venoso ou linfático, são levados através do fluxo sanguíneo venoso para o lado direito do coração, e daí, são distribuídos para os pulmões, via artéria pulmonar. O número de esquistossômulos dentro dos capilares pulmonares aumenta significativamente entre o segundo e o sétimo dias após a infecção (Pinto e Almeida, 1948). Após atingir seu pico de permanência, os parasitos apresentam decaimento em sua concentração nos capilares pulmonares seguindo um decaimento quase exponencial (Miller e Wilson, 1980). Em sua permanência nos capilares pulmonares, os esquistossômulos perdem a camada intersticial fibrosa subtegumentar e parte da musculatura anterior, ficando assim com diâmetro menor e alongado facilitando sua migração através dos capilares sanguíneos.

Após sua passagem nos pulmões o desenvolvimento dos esquistossômulos completa-se nos vasos intra-hepáticos, onde se acumulam lentamente (Coelho, 1970). O trajeto que os esquistossômulos seguem do pulmão até as veias do sistema porta-hepático tem sido motivo de discussões. Há três vias postuladas:

- Intravascular direta, sem recirculação (modelo vascular passivo) (Miyagawa, 1912): Os esquistossômulos, após saírem dos capilares arteriais e atingirem o compartimento venoso do pulmão, seriam transportados passivamente através do coração para a aorta, e daí para o sistema porta-hepático através das artérias hepáticas ou mesentéricas;
- Intravascular reversa (modelo vascular ativo) (Kruger et al., 1969): Propõe que os esquistossômulos após atingirem o pulmão, mudariam de direção e migrariam, contra a corrente sanguínea, passando pela artéria pulmonar, ventrículo e átrio direitos, veia cava posterior e penetrariam no fígado através das veias hepáticas;
- Transdiafragmática (modelo transdiafragmático) (Bruce et al. 1974): Conjectura a escapada dos capilares pulmonares para os alvéolos, migrando, através do tecido pulmonar, para caírem na cavidade pleural e atravessariam o diafragma e a cápsula hepática, atingindo os vasos portais intra-hepáticos.

É no sistema porta-hepático que se dá o amadurecimento sexual do parasito e a junção da fêmea com o macho através de seu canal ginecóforo para auxiliar na digestão do sangue e no transporte para os sítios de postura de ovos.

Os ovos são postos na corrente sanguínea e precisam sair do território vascular e atravessar a parede do intestino para serem liberados junto com as fezes, para assim, dar continuidade ao ciclo biológico do parasito. Porém, cerca de 50% desses ovos liberados pelas fêmeas são carregados pela circulação portal e ficam retidos nos tecidos do hospedeiro (Warren, 1978), induzindo a formação de granulomas hepáticos e intestinais, hepatoesplenomegalia e fibrose periportal. Em média cada fêmea libera de 150 a 300 ovos por dia (Watson, 1960; Brumpt, 1930).

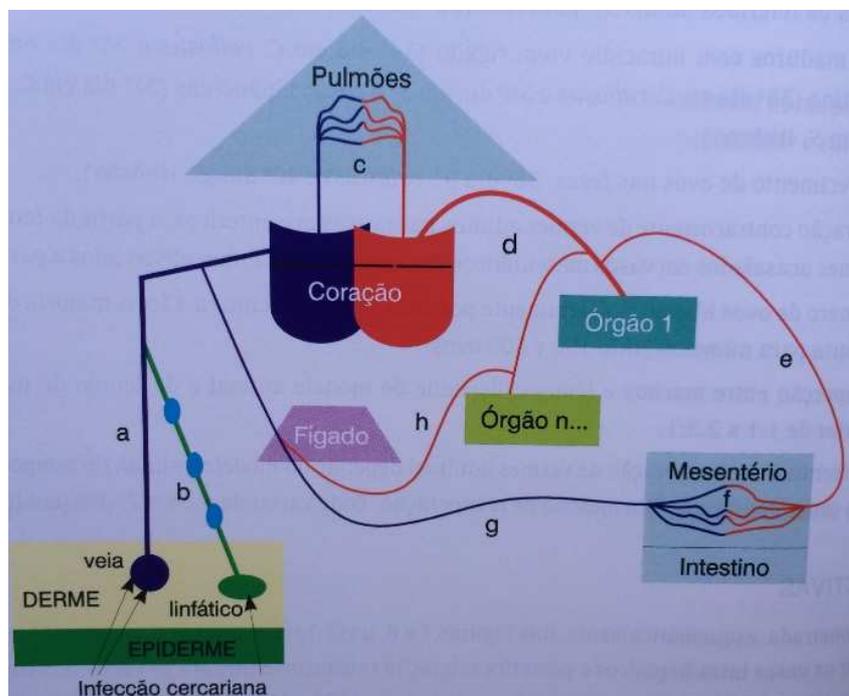


Figura 3: Ilustração da penetração pelas cercárias (Carvalho, Coelho e Lenzi, 2008).

A figura 3 apresenta o esquema de migração do parasito através do hospedeiro desde o momento do contato com a pele até a liberação dos ovos através do intestino.

1.3 Evolução de *Schistosoma mansoni* no hospedeiro intermediário

No seu hospedeiro intermediário, hospedeiro invertebrado do gênero *Biomphalaria*, o *Schistosoma mansoni* desenvolve sua fase assexuada através de complexos processos de multiplicação celular e mecanismos adaptativos. Parte dos ovos de esquistossômulos adultos são depositados nas fezes a partir do 35º dia (*Calomys callosus*) e 40º dia (*Swiss Webster*) em quantidade entre 150 e 300 ovos, Figura 4. Estes por sua vez, ao entrarem em contato com o ambiente aquático sofrem a influência da baixa osmolaridade do meio, de estímulos luminosos e da temperatura do ambiente, o que leva a intensos movimentos da larva e provoca a ruptura transversal da casca do ovo, e daí, permitindo a eclosão do miracídio.



Figura 4: Ilustração de um ovo de *Schistosoma mansoni* (Lima, 2013).

A eclosão seria praticamente inibida na ausência de luminosidade e não ocorreria em temperaturas abaixo de 10°C e superiores a 37°C (Standen, 1951).

O miracídio, Figura 5, é uma larva que apresenta uma intensa atividade natatória, tem forma cilíndrica e dimensões médias de 180 micrômetros de comprimento por 64 micrômetros de diâmetro. Apresenta células epidérmicas, onde se implantam os cílios, os quais permitem movimentos no meio aquático, além de várias estruturas contidas no meio líquido no interior da larva. A extremidade apresenta uma papila apical, ou terebratorium em que são encontradas as terminações das glândulas adesivas e as terminações da glândula de penetração. Também são encontrados no terebratorium um conjunto de cílios maiores e

espículos anteriores, provavelmente importantes no processo de penetração nos moluscos, e por fim, terminações nervosas que estão associadas as funções tácteis e sensoriais (Ottolina, 1957).

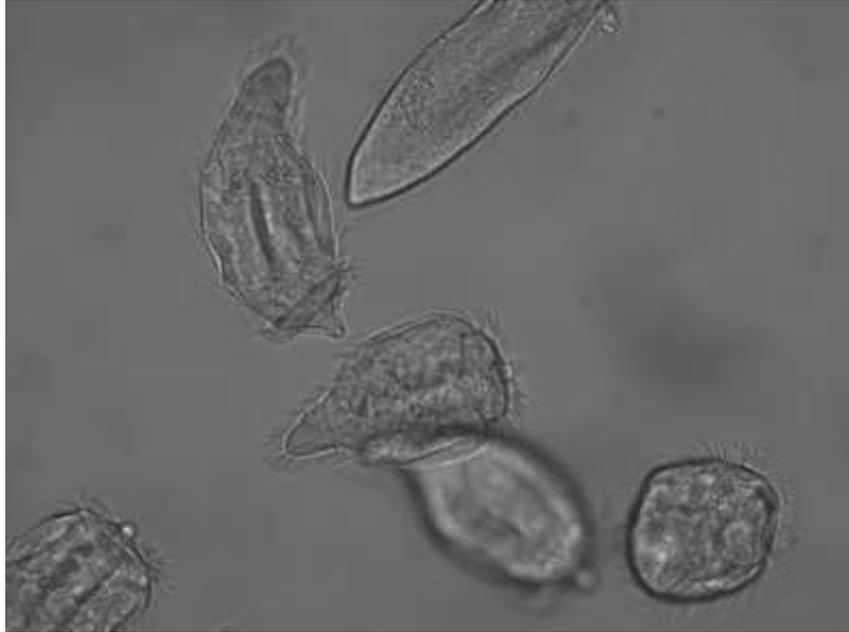


Figura 5: Ilustração produzida por microscopia a laser confocal, a imagem exhibe larvas do *S. mansoni* (miracídio) (Jurberg, 2013).

As células germinativas, em número de 50 a 100, darão continuidade ao ciclo do caramujo. Ottolina em 1957, sugere que elas estariam dispostas em fileiras chamadas de cordões germinativos inseridos ao longo da parede interna do miracídio, entretanto, o trabalho de Pan (1965) mostrou ser sua localização na parte posterior do miracídio.

Após sua eclosão, há um tempo de vida de 8 horas para localizar e penetrar em algum molusco, entretanto, após 5 horas poucos conseguem penetrar nos tecidos do hospedeiro. O comportamento que favorece a essa localização é a fototaxia positiva presente em ambos, já que há atração do miracídio a fontes luminosas (Chaia, 1956) e o molusco *B. glabrata* também mostraram ocorrência de fototaxia positiva (Williams e Coelho, 1975). Outra característica marcante é a exposição da massa cefalopodal do molusco no meio aquático que é estimulada pela luminosidade (Pimentel-Souza, 1976), o que favoreceria o processo de infecção, já que são estas as partes preferidas para a penetração miracidiana (Coelho, 1957).

O direcionamento do miracídio em relação ao seu hospedeiro intermediário depende da espécie e linhagem de ambos os envolvidos. Conforme trabalho realizado entre cepa

de *S. mansoni* do Egito e duas do Brasil, foi demonstrado que somente a linhagem egípcia apresentava comportamento de movimento direcional em face do alvo com substâncias obtidas da água de incubação de seus hospedeiros intermediários específicos, o que não se verificou nas duas linhagens brasileiras (Kalbe et al., 1996).

A temperatura é fator determinante na penetração do miracídio, já que redução de 1°C na temperatura (de 26°C para 25°C) levaria a uma redução na taxa de infecção dos moluscos de 80% para 40% (Maldonado e Acosta-Maltienzo, 1947).

Na infecção, o miracídio ao se encontrar com o molusco faz com que o terebratorium assumira a forma de ventosa, e daí, descarregando o conteúdo de sua glândula de adesão enquanto o invasor agita-se intensamente com movimentos contráteis e giratórios comandados pelas vibrações ciliares e pela ação da musculatura subepitelial. O miracídio, após ter descarregado o conteúdo de sua glândula, dá início ao processo de digestão dos tecidos através das enzimas proteolíticas. A ação das enzimas e dos movimentos permite a ultrapassagem do epitélio, e assim, a larva se estabelece no tecido subcutâneo. O processo de penetração dura entre 10 e 15 minutos, e nesta fase, apenas 30% dos miracídios são capazes de penetrar e evoluir em *B. glabrata*, 30% penetram e não evoluem; e 40% são incapazes até mesmo de penetrar no molusco (Chernin e Antolics, 1975).

Após a inserção no molusco, as larvas perdem as glândulas de adesão/penetração e continuam a perder outras estruturas durante todo o processo de penetração. Sendo a perda do epitélio ciliado junto com a degeneração do terebratorium o passo seguinte, a perda da musculatura subepitelial, e por último, a perda do sistema nervoso que pode resistir por mais alguns dias. Com exceção do desaparecimento do sistema nervoso, todos os demais ocorrem num período de 48 horas após a penetração.

Após 72 horas, as células germinativas entram em processo de multiplicação, e agora passando a se chamar: esporocisto primário, esporocisto mãe ou esporocisto I. A formação de esporocisto secundário é iniciada com um aglomerado de células germinativas nas paredes do esporocisto primário e uma formação de vacúolo na parte central da larva. Assim, um esporocisto dá origem aos septos, ficando o esporocisto primário dividido em 150 a 200 camadas, em que cada septo ou camada pode ser considerado um esporocisto secundário, Figura 6.

A migração dos esporocistos inicia-se em torno do 18º dia, a migração leva de 2 a 3 dias e o destino final será nos espaços intertubulares da glândula digestiva, devido à

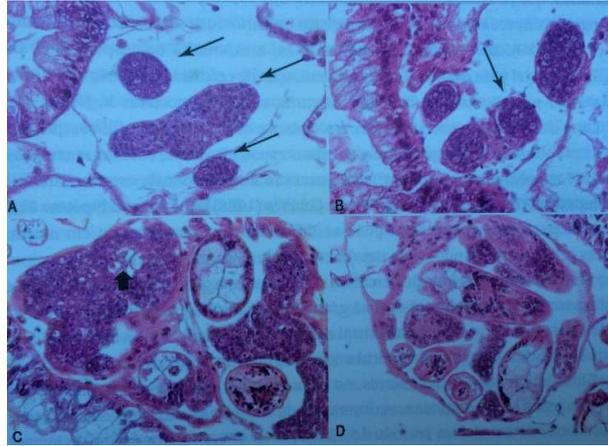


Figura 6: Esporocistos e cercárias em várias fases de desenvolvimento. Em A e B os esporocistos aparecem indiferenciados, compactamente arrumados e limitados por uma fina membrana externa. Em C já se pode ver uma cercária em diferenciação (seta) no meio de esporocistos; D mostra cercárias em várias etapas de diferenciação (Carvalho, Coelho e Lenzi, 2008).

sua riqueza nutritiva. No destino final, os esporocistos secundários apresentam três áreas estruturais bem definidas: A primeira, é o poro de nascimento; a segunda, representadas pelas cercárias desenvolvidas ou em desenvolvimento; e a terceira, apresenta formas embrionárias. Essa configuração explica a prolongada eliminação de cercárias pelo molusco, já que os esporocistos secundários dão origem a esporocistos terciários além das cercárias que serão eliminadas (Jourdan et al, 1980).

Pan (1965) verificou que a formação cercariana inicia-se com a disposição das células germinativas e que até a emergência das cercárias para o meio aquático é necessário um período de 27 a 30 dias, em condições ideais de temperatura (cerca de 28°C). Um único miracídio pode gerar cerca de 300 mil cercárias (Faust e Hoffman, 1934), sendo que cada miracídio já leva definido o sexo das cercárias que serão produzidas.

Quanto à morfologia das cercárias: Comprimento total de 500 micrômetros, cauda bifurcada medindo 230 por 50 micrômetros e corpo cercariano com 190 por 70 micrômetros. Duas ventosas estão presentes: A ventosa oral apresenta as terminações das chamadas glândulas de penetração e a ventosa ventral com musculatura mais desenvolvida.

A emergência das cercárias ocorre com a saída dos organismos dos esporocistos filhos. A migração das cercárias é feita tanto pelo espaço intercelulares cheios de hemolinfa, como através sistema venoso do caramujo. A passagem para o meio exterior processa-se pela formação de vesículas, no epitélio do manto, e de pseudobrânquia (Pan, 1965).

Uma característica marcante é a grande infectabilidade em caramujos juvenis (7 a 9 mm). Théron et al. (1998) sugere vários fatores para justificar esta preferência, e especulam sobre possíveis fatores de atração derivados do processo de amadurecimento sexual dos caramujos subadultos (hormônios), maior disponibilidade de matéria-prima para desenvolvimento dos esporocistos e uma menor mortalidade em relação aos juvenis (3 a 5 mm). Acrescente-se ainda que o adulto teria menor expectativa de vida e seu sistema inato de defesa seria mais eficiente.

Ainda, associado aos fatores ligados ao mecanismo de defesa do hospedeiro intermediário, o ciclo evolutivo do parasito sofre influência de fatores ambientais como temperatura e luminosidade naturais. Assim, a temperatura e luminosidade exercem influência na eclosão miracidiana, no encontro entre hospedeiro/parasito e na emergência das cercárias. Já intramolusco, a temperatura ambiental é fundamental para o desenvolvimento dos esporocistos e a ausência de contato com a água pelo molusco também compromete a evolução dos esporocistos.

Diante disso, a transmissão ocorre nos períodos de maior luminosidade e com temperaturas elevadas, estes por si só já justificaria a razão pelo qual esquistossomose mansoni ser transmitida nas regiões tropicais do planeta.

1.4 Contribuições relevantes

Em 2006, PRIXIA NIETO publicou seu trabalho sobre modelagem de nicho ecológico para leishmaniose visceral utilizando algoritmo genético para previsão do conjunto de regras. Esse trabalho confirma a eficácia da utilização de algoritmos de busca na tentativa de obter o conjunto de regras que explicam a dinâmica de infecção e valida a utilização de metodologia análoga para obter o conjunto de regras para o caso da expansão da esquistossomose (NIETO *et al.*, 2006). Neste mesmo ano, STENSGAARD publicou seu trabalho utilizando algoritmo genético na obtenção do conjunto de regras para modelar habitats para caramujos e áreas potenciais de transmissão de doenças entre gados e caramujos em Uganda, assegurando a aplicabilidade da metodologia e ratificando as limitações da mesma devido a não utilização/pouca informação de dados parasitológicos dificultando identificar a magnitude atual da transmissibilidade da enfermidade (STENSGAARD *et al.*, 2006).

Em 2010, CHEN Z. propôs uma abordagem para modelar a dinâmica de contágio na província de Hubei na China. Sua modelagem é baseada em seis equações diferenciais que descrevem a suscetibilidade e infectabilidade humana, do gado e do caramujo. O

autor verificou que um controle direto entre a dinâmica gado-caramujo causa muito mais impacto do que a dinâmica humano-caramujo para a região estudada, porém, o mesmo adverte a necessidade de maiores compreensões entre os fatores ambientais na tentativa de quebrar o ciclo gado-caramujo presente naquela área (Chen *et al.*, 2010)

Estudos realizados no lago Poyang utilizando modelo matemático baseado em equações diferenciais para avaliar a transmissão do *Schistosoma japonicum* através das flutuações sazonais no nível do lago mostraram que é possível reduzir consideravelmente o número de infecções associando um controle através de políticas de conscientização e também um controle focado no tratamentos dos reservatórios de animais que entram em contato com essas águas (XIANG *et al.*, 2013).

C. Wan publicou um artigo apresentando a modelagem por autômatos celulares para infecção por *Schistosoma japonicum* garantindo a sua eficiência em simular o processo de transmissão, e o autor ainda ratifica que o uso de modelos baseados em autômato celular é uma ferramenta útil para entender a dinâmica de infecção e para construir programas efetivos de combate a infecção (Wan *et al.*, 2013).

Já J. Albuquerque apresentou um estudo preditivo detalhado baseado em dados reais com relação a processo expansão da esquistossomose pela costa brasileira. Neste trabalho foi utilizado modelo baseado em autômatos celulares que caracterizam fatores ambientais e sociais para a previsão da expansão da enfermidade, foi possível constatar a instalação de áreas endêmicas o que associado a um controle prévio pelos agentes sanitários seria crucial para assegurar a qualidade de vida dos moradores do litoral brasileiro (Albuquerque, 2013).

1.5 Organização do trabalho

Ao longo dos anos foram apresentados artigos enfatizando a importância da modelagem matemática para previsão e acompanhamento da evolução das doenças em estudo. Característica marcante é a presença de algoritmo genético na tentativa de obter a coleção de regras que melhor explica a dinâmica de contágio e deslocamento/evolução da moléstia. Contudo, a utilização única do algoritmo genético vem apresentando resultados satisfatórios porém com limitações baseadas em suposições parasitológicas e sociais da região trabalhada, essa é a principal fraqueza do modelo. Por outro lado, os estudos baseados em equações diferenciais/autômatos celulares visam agregar a maior quantidade de informação possível baseada em aspectos sociais, ambientais e parasitológicos inerentes

a região e aos agentes infecciosos em foco. Contudo, tais modelos são limitados a escolha dos parâmetros que serão utilizados em sua modelagem, e dessa forma, dependentes do bom senso do pesquisador. Essa característica, a subjetividade, limita os resultados a diversos cenários para cada coleção de parâmetros que venham a ser utilizados.

Dessa forma, é natural a busca por modelos híbridos que associem essas características de autômatos celulares quanto a utilização de informações do ambiente em estudo e a otimização para a coleção de regras obtida pela utilização dos algoritmos genéticos.

Diante do exposto, a estratégia é utilizar as informações particulares dos ciclos evolutivos para criar um modelo matemático para mimetizar a dinâmica de infecção dos indivíduos da região em estudo, e a partir disso, poder identificar os parâmetros que levam ao quadro endêmico.

O presente trabalho se encontra organizado da seguinte forma: Capítulo 2, trata dos autômatos celulares e sua utilidade para modelar cenários dinâmicos; Capítulo 3, aborda os algoritmos genéticos apresentando suas **particulares** e seu poder de otimização; Capítulo 4, versa sobre o modelo matemático adotado para simular a evolução da dinâmica de infecções; e Capítulo 5, apresentando os resultados conquistados através do modelo matemático baseado em autômato celular e os resultados do ajuste dos parâmetros do modelo através da abordagem por algoritmo genético.

2 Autômatos Celulares

The chess-board is the world; the pieces are the phenomena of the universe; the rules of the game are what we call the laws of Nature.

T. H. Huxley

2.1 Histórico

O estudo de autômatos celulares, denotado neste trabalho por AC, teve início nos trabalhos de John Von Neumann em seus estudos sobre auto reprodução (Neumann, 1966). Em seu trabalho, *Theory of self-reproducing automata*, o autor apresenta um paralelo entre o corpo humano bem como suas capacidades de processamento e os computadores disponíveis na época. Chegando inicialmente à conclusão de que o suporte tecnológico existente não se comparava à máquina humana, e nesse contexto, apresenta a fragilidade dos componentes tecnológicos que ao longo do tempo poderiam sofrer melhoras.

So our present machinery is about 200 thousand times less efficient than the nervous system is. Computing machines will be improved in the next few years, perhaps by replacing vacuum tubes with amplifying crystals, but even then they will be of the order of 10 thousand times less efficient than the nervous system. (NEUMANN, 1966, p. 67)

No trabalho de Neumann é possível notar as primeiras preocupações em mimetizar os organismos vivos e sua complexidade. A tentativa de construir uma máquina com o mesmo poder de processamento, e além disso, a capacidade de se auto reproduzir (o próximo ponto abordado por Neumann) com objetivo de obter um comportamento livre da interferência seu construtor (McMullin, 2000).

Na tentativa de justificar um autômato celular, sua definição, listagem dessas partes, síntese de autômatos por outros autômatos, e por fim, o problema da auto reprodução, Neumann explana sobre Alan Turing e sua máquina, um dispositivo hipotético que manipula uma fita sob o controle de um conjunto de regras, máquina esta que pode ser adaptada para simular qualquer algoritmo computacional.

It is evident in the case of a Turing's automaton, which is a box with a finite number of states. Its outputs are modifications of another entity, which, for a sake of convenience, i call of the punching tape. This tape is not itself an object which has states between which it can move of its own accord. Furthermore, it is not finite, but is assumed infinite in both direction. Thus, this tape is qualitatively completely different from an automaton which does the punching, and so the automaton is working into a qualitatively different medium. (NEUMANN, 1966, p. 74-75)

O procedimento apresentado por Turing é conveniente, porém, não atende à necessidade do problema da auto-reprodução de Neumann, já que a fita não é parte do meio. Ou seja, o meio é alimentado pelo autômato e o que é produzido é completamente diferente deste. Neumann imagina um autômato como uma peça inerente ao meio em que vive e que sua atividade é apenas de pegar partes e agrupá-las ou se agrupando a partes que encontrar para então desmontá-las (pensamento de processo enzimático), por fim, em essência o produto final dessa ação seria novamente outro autômato com as mesmas funcionalidades (Neumann, 1966). A figura 7 ilustra um sistema que não atende a condição de auto reprodução como o caso do autômato de Turing, porém, terá particular interesse em nossa motivação para modelagem associada ao conceito de algoritmo genético.

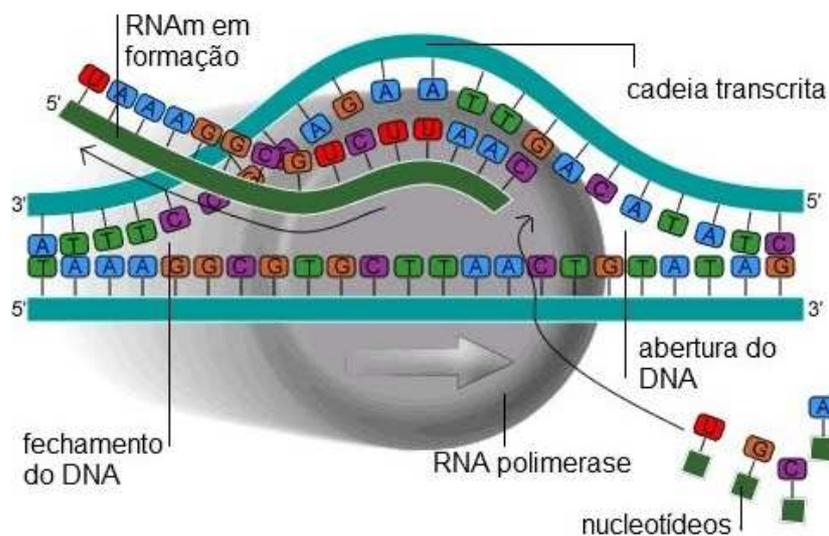


Figura 7: Ilustração na natureza de um processo baseado em fita codificada (DNA) que é submetida a uma caixa de leitura (RNA polimerase) para a criação de uma outra fita codificada (RNA-m). A caixa de leitura, por si só não é uma fita e não atende a definição de auto reprodução.

Embora a máquina de Turing não atenda às necessidades de auto reprodução de Neumann, sua estrutura é interessante para o entendimento de um autômato. Alan Turin, em

1936, apresenta sua *a-machine* (máquina automática) no artigo, *On Computable Numbers, with an Application to the Entscheidungs problem* (Turing, 1937), que deu origem ao teorema de Church-Turing, neste texto, é possível obter definições tais como:

- Fita - Dividida em unidades em sequência, denominadas células, tal fita pode ser indefinidamente extensível para ambos os lados ou indefinidamente extensível somente para a direita;
- Alfabeto - Conjunto de símbolos que podem ser escritos e lidos na fita. Em especial, existe o símbolo para representar quando a célula da fita está em branco, isto é, não há nenhum símbolo escrito nela;
- Caixa de leitura - Tem a capacidade de ler e escrever na fita, podendo movê-la para direita ou esquerda um passo por vez;
- Registrador de estados - Armazena o estado atual do sistema, contém uma quantidade finita de estados distintos e há um estado inicializador que põe em funcionamento o registrador;
- Tabela de ações - É o conjunto de instruções possíveis para a caixa de leitura. São todos os comandos que a caixa de leitura poderá executar na fita, tal conjunto é a tentativa de simplificar as opções de escolha da caixa.

essa configuração é muito geral, e pode ser aplicada a qualquer situação, inclusive para reproduzir o comportamento de um sistema mais complexo.

It is possible to invent a single machine which can be used to compute any computable sequence. If this machine Λ is supplied with a tape on the beginning of which is written the S.D. (Standard Description) of some computing machine Π , then Λ will compute the same sequence as Π . (TURING, 1936, p. 241-242)

A máquina auto replicadora de Neumann, sem uso de computador foi projetado na década de 40 e implementada em 1995, no artigo, *An implementation of von Neumann's self-reproducing machine* (Pesavento, 1995).

Uma máquina computacional é dita ser computacionalmente universal se for capaz de simular qualquer outra máquina computacional. O construtor universal de Neumann é uma máquina capaz de simular o que puder ser incorporado em seu autômato celular e descrita em uma fita longa e finita, porém arbitrária (Pesavento, 1995).

To obtain complete self-reproduction, the tape is interpreted in two different ways. In the first interpretation, the tape is simply copied, while in the second the tape is decoded to

produce excitations corresponding to the cell assembly to be built. To implement these interpretations, our universal constructor possesses an organ for the representation of its internal state. This organ governs the changing interconnections between the other organs of the constructor. (PESAVENTO, 1995, p. 338)

2.2 Definição formal

Cosma R. Shalizi (2001) em sua tese de doutorado intitulada *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata* apresenta uma visão bem lúdica para ilustrar o entendimento do que se trata um autômato celular:

Take a board, and divide it up into squares, like a chess-board or checker-board. These are the cells. Each cell has one of a finite number of distinct colors - red and black, say, or (to be patriotic) red, white and blue. (We don't allow continuous shading, and every cell has just one color.) Now we come to the "automaton" part. Sitting somewhere to one side of the board is a clock, and every time the clock ticks the colors of the cells change. Each cell looks at the colors of the nearby cells, and its own color, and then applies a definite rule, the transition rule, specified in advance, to decide its color in the next clock-tick; and all the cells change at the same time. (The rule can sometimes tell the cell to stay the same.) Each cell is a sort of very stupid computer - in the jargon, a finite-state automaton - and so the whole board is called a cellular automaton, or CA. To run it, you color the cells in your favorite pattern, start the clock, and stand back. (SHALIZI, 2001, p. 66)

Seja Δ uma malha definida pelo produto \mathbb{Z}^k , isto é, uma malha de dimensão k constituída por células (hipercubos do \mathbb{Z}^k). Cada célula δ tem uma vizinhança $v(\delta)$ de outras células, entende-se vizinhança $v(\delta)$ como os componentes da malha que tem relação direta com a célula δ ou que segue alguma relação lógica que o conecta com tal célula. Todas as células tem o mesmo tamanho e forma vizinhança igualmente a qualquer outra célula.

A vizinhança-padrão se define por todas as células que distam até r unidades, assim, existe uma hiperesfera de dimensão k com raio r que tem seu centro na célula δ que envolve um conjunto de células.

O conjunto A , de cardinalidade $\#(A) = p$, contém os valores que as células podem assumir em qualquer configuração. Denota-se o valor assumido na célula δ por s^δ e a configuração na vizinhança da célula δ por $s^{v(\delta)}$, o tempo t para cada evolução é definido discretamente, com isso, cada passo s_t^δ é dado de forma regular e sob um mesmo intervalo e a configuração global associada ao autômato é denotado por s_t .

A regra de transição, é definida como um mapeamento f que tem como variáveis independentes a configuração $s^{v(\delta)}$ e como variável dependente o valor atualizado da célula

δ . Nesse passo, se o mapeamento f tiver saída fixa para cada entrada, então tal mapeamento é determinístico. Caso contrário, isto é, se houver um conjunto de saídas possíveis em que cada uma tem uma probabilidade de ocorrer, o mapeamento será estocástico.

A equação de movimento do AC se dá aplicando a regra para cada célula separadamente:

$$s_{t+1}^\delta = f(s_t^{v(\delta)})$$

é possível perceber que na função acima, aplica-se f na configuração da vizinhança no momento t e como resposta obtém-se o valor assumido por δ no instante $t + 1$.

A aplicação simultânea do mapeamento f em todas as células da malha define a *regra de evolução global* Φ , uma aplicação que leva o conjunto $(\#A)^{\mathbb{Z}^k}$ em $(\#A)^{\mathbb{Z}^k}$.

Daí, temos a aplicação:

$$\Omega_{t+1} = \Phi\Omega_t$$

em que $\Omega_t = \{s_t\}$ é um conjunto de configurações e o operador Φ atua sobre conjuntos. (SHALIZI, 2001)

2.3 Variações de um autômato celular

Um AC tem sua caracterização definida por quatro partes:

- A geometria do meio que contém as células;
- A regra de transição, isto é, a regra que atualiza o estado atual da célula;
- Os estados das células;
- A vizinhança que se comunica com a célula em estudo.

Partindo dessa caracterização, qualquer modificação em uma dessas quatro características produz uma classe distinta de AC.

Geometria - Diz respeito à configuração da grade, podendo ser k -dimensional infinita. Sendo infinita a condição de borda é irrelevante. Por outro lado, se for finita, a borda apresenta comportamento singular, já que a vizinhança não estará completa em muitas definições (por exemplo, a vizinhança de Neumann ou vizinhança de Moore conforme Figura 8), daí, é necessário que sejam admitidas condições para a borda. Ao assumir-se condições distintas, será modificado completamente a dinâmica da evolução do autômato.

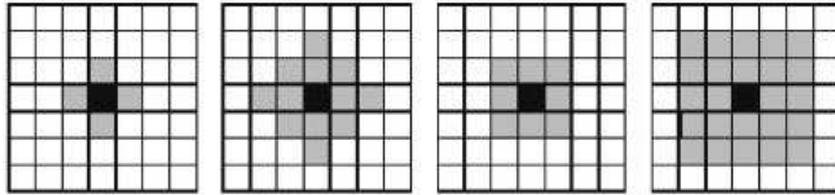


Figura 8: Ilustração das vizinhanças de Von Neumann e Moore para os casos de $r = 1$ e $r = 2$.

Ao considerar $k = 2$ e que as células extremas são vizinhas das células extremas opostas, estamos dando uma geometria de toro; se tomar que somente as células extremas horizontais ou verticais sejam vizinhas das células extremas opostas horizontais ou verticais, respectivamente, e as extremas demais estão em estado de repouso, esta sendo assumida uma geometria cilíndrica; se definir que somente as células extremas horizontais ou verticais sejam vizinhas das células extremas opostas horizontais ou verticais, respectivamente, e as células extremas superiores (ou de um dos lados no caso vertical) são vizinhas de um único ponto e as células extremas inferiores (ou do outro lado no caso vertical) estão em estado de repouso, configura-se uma geometria de cone; já se for definido que os extremos laterais tendo vizinhança com os extremos laterais opostos, e os extremos superiores e inferiores cada um são vizinhos de um único ponto, está assegurada uma geometria de uma esfera para nossa malha. A Figura 9 ilustra a geometria de cada um desses exemplos.

Esses casos ilustrados, fazem parte de uma classe de autômatos em que a regra de borda é fixa e imutável ao longo da evolução do autômato, sendo chamado de AC estático. Porém, há a classe de autômatos em que a regra de borda pode ser modificada segundo uma lógica ou estocasticidade controlada, tal classe ainda assim é considerada estática e tem recebido pouca atenção na literatura (Varshavsky, 1969). Diante disso, a classe de AC dinâmico é definida pelo grupo de autômatos que modificam sua regra de borda, bem como sua regra de vizinhança. Este é amplamente estudado devido a sua capacidade de modelar sistemas biológicos (Lindenmayer, 1968).

Regra de transição - Tradicionalmente, a regra de transição é definida de forma determinística. Ou seja, para cada estado da célula em estudo e de sua vizinhança há apenas um estado seguinte na linha de evolução do autômato para tal célula, porém, é possível definir um mapeamento não determinístico para controlar a dinâmica de evolução de cada célula do AC. Se o AC tem uma regra de transição para cada célula, então este AC é dito híbrido (Sarkar, 1998).

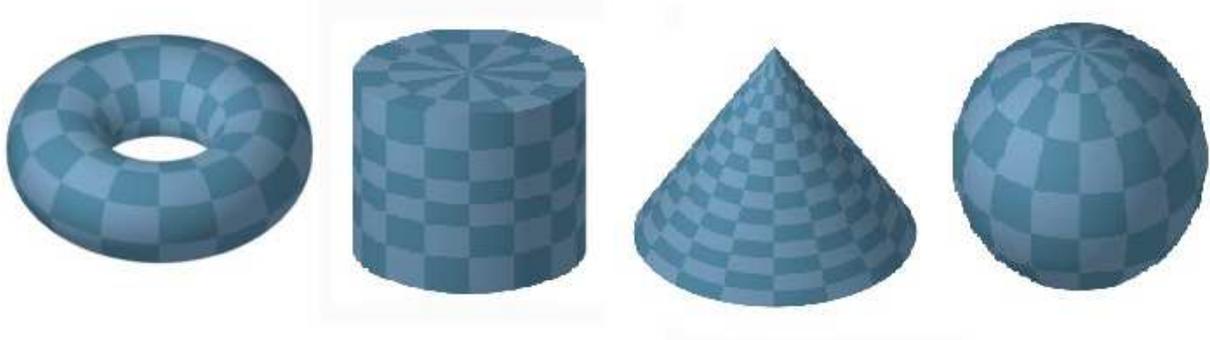


Figura 9: Ilustração da geometria da malha para o caso bidimensional, na ordem, geometrias: Toro, cilindro, cone e esférica.

Estados das células - Normalmente, um AC pode assumir uma quantidade finita e fixa de possíveis estados. Na maioria dos casos, há também um estado especial, chamado estado de repouso, em que a célula entrará em estado de repouso se todos os seus vizinhos estiverem neste estado. Note que ao ser definida uma quantidade fixa e finita de estados, assume-se que essas são todas as possibilidades de mudança que as células vão assumir durante toda a evolução do autômato. Porém, uma modificação nessa característica, isto é, definir um AC em que as células podem assumir diferentes conjuntos de estados (Não confundir com diferentes estados, mas, diferentes conjuntos de estados, inclusive, dois conjuntos completamente distintos em que nenhum estado seja coincidente. Em outras palavras, é assumir que há sítios na malha do autômato que assumem uma coleção de estados, enquanto outros locais da malha podem assumir coleções diferente de estados) recebem o nome AC's polimorfos (Burks, 1970).

Para se abordar um AC polimorfo deve-se encarar o próprio AC como uma coleção finita de autômatos, em que cada célula representa um AC usual. Dessa forma, é possível que cada célula tenha seu conjunto de estados e o AC polimorfo como um todo (O conjunto de células, isto é, a grade) terá uma característica de diferentes conjuntos de estado que dependerá de sua localização na malha.

Outra abordagens, mais abstrata, é assumir um conjunto de estados tão grande quanto se queira, e daí, obter o seu conjunto das partes (coleção que contém todas as possibilidades de subconjuntos do conjunto de estados) e assumir uma probabilidade não nula para cada subconjunto para se tornar o conjunto de estados da célula em estudo no momento. Note que para isso é necessário que a escolha seja feita de forma aleatória em função da probabilidade associada a cada conjunto.

Vizinhança - Assumindo uma grade k -dimensional é possível definir diversos formatos de vizinhança, como a vizinhança de Neumann (que consiste de uma hipersfera de dimensão k de raio r) e Moore (que consiste de um hipercubo de dimensão k cujo lado tem comprimento $2r + 1$) já falados anteriormente. Nesses termos, é possível definir a entrada

e a saída da vizinhança de uma célula, isto é, uma célula obtém as informações de seu estado e do estado dos seus vizinhos e através da regra de transição modifica seu estado (ou de seus vizinhos, inclusive). Se a quantidade de células requisitadas na entrada como parâmetro de transição é igual à quantidade de células modificadas na saída, então o AC é balanceado. Uma variante para AC em que a regra de transição é decidida pela soma dos estados dos vizinhos, chamada de autômato totalístico, foi introduzindo por Wolfram e a universalidade desse tipo de AC foi provada por Culik (Culik, 1990).

Algumas variantes tem atenção especial, são elas:

Tessellation Automata - Consiste de um AC em que há um conjunto finito de regras de transição e uma ação (por exemplo, o lance de uma moeda) é usada para decidir que regra de transição será aplicada naquele passo. Um problema interessante com relação *tessellation* teórica é o problema da completude e sua relação com o problema do Jardim do Eden para AC. Tal problema é definido como:

Partindo de uma configuração inicial com somente um estado fora de repouso, é possível aplicar regras de transição que levem o autômato a qualquer configuração finita?

Se a resposta for sim para alguma subclasse de AC, então esta subclasse é dita completa. Já houve solução parcial para esse problema (Yamada, 1970; Marouka, 1974; 1977)

One-Way - Consiste de autômatos celulares que utilizam uma única forma de informação, em uma lista unidimensional cada célula depende de si mesma ou da vizinha a sua esquerda (Ou da vizinha a sua direita, porém, somente um lado oferece informação). Alguém poderia achar que tal restrição poderia diminuir o poder do autômato, entretanto, Morita em seu trabalho intitulado *Computation-universality of one-dimensional one-way reversible cellular automata* demonstrou a universalidade para autômato unidimensional reversível. Há aplicações desse tipo AC em reconhecimento de linguagem (Chang, 1988) e em séries temporais (Morita, 1992).

3 Algoritmo genético

Não é o mais forte que sobrevive, nem o mais inteligente, mas o que melhor se adapta às mudanças.

Charles Darwin

3.1 Computação evolutiva

É possível definir a computação evolutiva como uma área de pesquisa da modelagem computacional em que tem como principal ponto de inspiração o processo de evolução natural. Notadamente, é possível perceber como colônias de formigas, a própria evolução genética, enxame de abelhas entre outros, apresentam formas características de eliminar suas tentativas contraproducentes e perpetuar na população os casos mais adaptáveis (Karaboga, 2007; Al Salami, 2009).

Tal filosofia é usada como metodologia para resolução de problemas que podem ser transcritos como um dilema evolucionário, com base nisso e se apropriando da observação de como a natureza soluciona tais dilemas, a computação evolucionária surge como uma forma de resolução de problemas baseado em tentativa-e-erro (direcionado) (Eiben, 2008).

Mas como definir a evolução natural e sua interpretação em computação evolutiva? Tome um cenário ambiental em que neste vive uma população que tem como objetivo sobreviver e se reproduzir. A adaptabilidade desses indivíduos será governada pelas características do cenário em que a população está inserida, essa adaptabilidade indica o quanto eles conseguiram seu objetivo (sobreviver e se multiplicar). Sob o ponto de vista de resolução de problema por tentativa-e-erro, a população servirá como uma coleção de possíveis soluções para o problema, como são soluções, é possível medir o quanto essas possíveis soluções respondem ao problema (e quão boa é a solução, e tal critério será utilizado para definir as chances dessa possível solução permanecer na coleção e servir de parâmetro para soluções futuras).

Evolução	Resolução do Problema
Ambiente	Problema
Indivíduo	Possível solução
Ajuste	Qualidade

Figura 10: Ilustração do comparativo entre computação evolutiva e a evolução natural em suas abordagens para solucionar seus problemas. (Eiben, 2008)

3.2 Histórico computação evolutiva

A utilização dos princípios apresentados por Charles Darwin na tentativa de resolução de problemas de forma autônoma teve início na década de 40, bem antes da disseminação de computadores; David B. Fogel em seu livro *Evolutionary Computation: The Fossil Record* (Fogel, 1998) apresenta os artigos que fundamentam a computação evolutiva. Em 1948, Turing propôs o processo de busca intitulado “Busca genética” (Turing, 1948) e em 1962, Bremermann (Bremermann, 1962) executou experimentos em otimização através de evolução e recombinação.

Variações da metodologia ocorreram na década de 60 em três locais distintos:

- Nos Estados Unidos, nos anos de 1965 e 1966, os primeiros artigos sobre programação evolutiva foram publicados por Fogel, Owens e Walsh. A programação evolucionária foi inventada por Lawrence J. Fogel (1928-2007) enquanto fazia parte da Fundação da ciência nacional em 1960, em sua variação, Fogel trabalhava em uma série de experimentos em que as máquinas de estados finitos representam organismos individuais em uma população de solucionadores de problemas. Estes modelos gráficos são usados para descrever o comportamento, é por isso que chamou sua abordagem de programação evolutiva (Fogel, 1965; 1966).
- Na Alemanha, Rechenberg e Schwefel na Universidade Técnica de Berlin apresentou a estratégia evolutiva para resolução de problemas de otimização contínua de parâmetros de controle e são descritas de forma completa no trabalho *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien des Biologischen Evolution* de 1973; a grande mudança apresentada nessa técnica é dada pela utilização de um operador mutação baseado em uma distribuição Normal (Rechenberg, 1973; Schwefel, 1981).
- Novamente nos Estados Unidos, John Henry Holland denominou seu método por Algoritmo genético, método este que é usado para achar soluções aproximadas em problemas de otimização e busca. Os algoritmos genéticos diferem dos algoritmos tradicionais de otimização em quatro aspectos (Holland, 1973, 1992; De Jong, 1975):

- Baseiam-se em uma codificação do conjunto das soluções possíveis, e não nos pa-

râmetros da otimização em sí;

- Os resultados são apresentados como uma população de soluções e não como uma solução única;
- Usam transições probabilísticas e não determinísticas;
- Não necessitam de nenhum conhecimento inerente ao problema, apenas de uma forma de avaliação dos resultados (particularmente, o aspecto mais interessante do método. Pois, uma vez que é retirada a responsabilidade de se conhecer a singularidade da resolução de um problema, o algoritmo está apto a utilizar o procedimento em problemas de natureza aleatória ou que se tenha pouco conhecimento acerca de sua dinâmica)(Goldberg, 1989).

Por cerca de 15 anos essas áreas foram desenvolvidas separadamente, porém, na década de 90 passaram a ser vistos como representações diferentes de uma mesma tecnologia que ficou conhecida como computação evolutiva. Uma quarta corrente seguindo as idéias gerais surgiu, programação genética, defendida por Koza (Koza, 1992; 1994). Na programação genética, consiste numa técnica automática de programação que propicia a evolução de programas de computadores para resolverem (ou aproximar a resolução) de problema.

A terminologia contemporânea denota todos os campos da computação evolutiva, os algoritmos contidos nessa terminologia são chamados de algoritmos evolutivos, e esta classificação considera a programação evolutiva, estratégia evolutiva, algoritmo genético e programação genética como sub-áreas pertencentes a variantes deste.

Uma vertente mais recente é a computação de enxames, o algoritmo conhecido como Otimização por Enxame de Partículas (PSO, particle swarm optimization). O algoritmo PSO foi proposto inicialmente por J. Kennedy e R. Eberhart (Kennedy, 1995) e é um técnica inspirada no comportamento social de bandos de pássaros. A busca por alimento e a interação entre os pássaros ao longo do vôo são modeladas como um mecanismo de otimização. No caso, a área sobrevoada é equivalente ao espaço de busca e encontrar o local com maior quantidade de comida corresponde a encontrar a solução ótima.

3.3 Contexto biológico

Pode-se dizer que a teoria da evolução tem participação fundamental na compreensão deste capítulo (e com isso, compreensão dos algoritmos genéticos), daí, é necessário conhecer algumas definições:

- Um *gene* é uma sequência de bases de DNA que codificam para uma característica, por exemplo, cor de cabelo ou da pele;
- Um *alelo* é o valor de uma característica. O gene da cor do cabelo pode ter um alelo preto ou um alelo vermelho em diferentes pessoas.

Com isso, é possível definir a evolução:

Evolução é a variação da frequência alélica na população ao longo do tempo (Telles, 2001).

Nessa definição, conceitua-se sobre a população. Ou seja, toda vez que nasce uma criança de cabelos pretos, então a frequência da população quanto ao gene cor de cabelo tem a frequência do alelo preto aumentada (analogamente, para o caso de morte de um homem de cabelos pretos), daí, é possível notar que a frequência alélica muda a todo tempo já que nascem e morrem pessoas a todo instante. Enfim, nota-se que a evolução acontece a todo tempo (podemos definir o parâmetro t discreto sob esse caso já que o evento nascimento/morte pode ser discretizado até um passo de tempo em que nasce/morre um único indivíduo).

Como a evolução produz novos indivíduos? Há duas formas opostas que conduzem a evolução: variação e seleção. Variação é o processo que produz novos alelos e, mais lentamente, genes. Variação também pode mudar os genes que são e não são expressos em um dado indivíduo, o método de fazer isso é a reprodução sexual com a sua interação de genes dominantes e recessivos. Seleção é o processo em que alguns alelos sobrevivem e outros não. Em resumo, a variação aumenta a diversidade e a seleção a diminui.

O processo de variação é complexo e ocorre a nível molecular. Biólogos aprendem novos sistemas para gerar variação a nível molecular. Seleção biológica é melhor entendida do que a variação biológica. Por outro lado, a seleção natural - sobrevivência dos mais adaptados ao meio em que vivem - tem sido o principal tipo de seleção biológica (Dawkins, 1991; Ridley, 2004).

Na computação evolutiva, as operações sobre a estrutura de dados (população) que geram a variação recebe o nome de *mutação* e *cruzamento* e atua por mudanças aleatórias nessa estrutura e misturando partes de estruturas diferentes, respectivamente. Já a seleção, pode ser interpretada como qualquer algoritmo que favoreça estruturas com maior adaptabilidade ao nosso problema. Há muitos métodos possíveis de seleção.

Mutações biológicas no DNA de algum organismo, sob um ponto de vista aleatório, são tipicamente neutras. Muitos trechos DNA não codificam informação útil (conhecida). A parte do DNA que codifica informação útil utiliza uma codificação robusta o que assegura que modificações de base simples não alteram a funcionalidade do gene. A rede de interação entre os genes é por si só robusta com multiplas cópias de alguns genes e múltiplos

genes que executam uma mesma tarefa específica (Lewin, 2000).

Será usado esse referencial para abordar os componentes de um algoritmo evolutivo.

3.4 Componentes de um algoritmo evolutivo

Um algoritmo evolutivo apresenta componentes bem característicos que devem ser bem definidos na elaboração da codificação do problema que será otimizado, são estes:

- Representação - Definição dos indivíduos;
- Função de avaliação (Função de ajuste ou Função de fitness);
- População;
- Mecanismo de seleção dos reprodutores;
- Operadores de variação: mutação e recombinação;
- Mecanismo de seleção dos sobreviventes (Substituição).

nesse passo, será discutido cada um desses componentes na tentativa de identificar cada particularidade associada a eles.

3.4.1 Representação

Essa é a etapa em que se deve configurar a conexão entre o cenário do problema e o espaço de solução do problema. Os objetos que formam possíveis soluções no contexto original do problema são referenciados como *fenótipos*, enquanto que as estruturas de dados, ou seja, indivíduos de um algoritmo evolutivo são chamados de *genótipos*.

Inicia-se definindo o mapeamento que relaciona os fenótipos (possíveis soluções) com os genótipos, de forma que cada genótipo corresponde a um fenótipo no cenário do problema. O espaço-fenótipo pode ser completamente diferente do espaço-genótipo, por exemplo, o problema em questão pode ser achar o ponto de máximo absoluto de uma função real de variáveis reais; como é uma função real, seu conjunto de fenótipo tem representação decimal e alguém pode codificar seus elementos genótipos em representação binária (completamente distinto do espaço-fenótipo), assim, toda a busca por soluções ótimas se dá no espaço genótipo, isto é, no espaço de representação binária.

Sempre que se encontra uma solução ótima (ou sub-ótima) no espaço-genótipo, é possível obter a solução associada no espaço-fenótipo através da decodificação da estrutura de dados ao final do processo de otimização.

A terminologia adotada para computação evolutiva denomina por: possível solução, fenótipo ou indivíduo os pontos do espaço de possíveis soluções, o próprio espaço de possíveis soluções é chamado de espaço-fenótipo. Já o genótipo, cromossomos e novamente indivíduo podem ser usados para pontos onde a busca evolucionária ocorre (espaço-genótipo).

Há também muitos termos relacionando os elementos de um indivíduo, cada componente do indivíduo é chamado de variável, locus, posição ou gene. Um objeto em tal lugar pode ser chamado de valor ou um alelo.

O trabalho de Nicol N. Schraudolph intitulado *Dynamic parameter encoding for genetic algorithms* apresenta uma interessante codificação dinâmica para evitar o dilema de sacrificar a precisão da representação numérica ou a eficiência na busca por uma solução ótima quando se trabalha com otimização de funções de valor real (Schraudolph, 1992).

3.4.2 Função de avaliação (Função de fitness)

A função de avaliação tem o papel de apresentar os requisitos para a adaptação, ela é o parâmetro para a seleção dos genótipos pais que darão origem a novos representantes mais adaptados ou não. Através da função de avaliação é possível medir o quanto um genótipo está resolvendo o problema associado, sua medida de qualidade está associada ao espaço dos genótipos.

Em geral, a função de fitness é qualquer função que capta a qualidade nos genótipos, qualidades estas que o tornam mais próximos da solução ótima ou sub-ótima para o problema. Porém, há casos em que a função de fitness pode vir a ser a própria função associada ao contexto do problema ou uma simples transformação desta.

No artigo *The Effects of Fitness Functions on Genetic Programming-Based Ranking Discovery For Web Search* escrito por Weiguo Fan, é apresentado o estudo a respeito da função de fitness com relação à eficiência e à efetividade de algoritmos evolutivos. O autor conclui que a adequação da função de fitness é fundamental para a melhoria no desempenho (Fan, 2004).

3.4.3 População

A população contém as possíveis soluções (representação destas), nesses termos, a população é uma coleção de genótipos. Ela é um passo da evolução, seus indivíduos são estáticos, isto é, não sofrem adaptação nem se modificam dentro da população sob uma geração fixa; em geral, a quantidade de indivíduos contidos na população se mantém fixa durante todo o processo de busca evolutiva.

Os operadores de variação atuam a nível de indivíduos, isto é, uma vez que o indivíduo é selecionado tais operadores atuam. Já os operadores de seleção agem a nível de população, daí, levam em conta a característica de todos os genótipos envolvidos na população e decide sua ação em função dessa.

A medida de diversidade está associada ao número de diferentes soluções que estão contidas em um dado momento da população, assim, é possível perceber que como a população é mutável e nesse contexto altera seus indivíduos ao longo do tempo, a diversidade tende a ser modificada durante toda a busca. É importante compreender que embora seja possível ter um único valor de fitness, ao fim de um processo de busca, não significa que se tenha um único genótipo; também, um único fenótipo não significa que se tenha um único genótipo. Por outro lado, o contrário não é verdadeiro, se houver um único genótipo terá um único valor de fitness e também um único fenótipo associado.

Alander J. T. em seu trabalho *On optimal population size of genetic algorithms* apresenta um estudo sobre o tamanho ótimo da população como função da complexidade do problema, o autor conclui que em máquinas sequenciais para problemas de complexidade moderada, o tamanho ótimo da população é função do tamanho do vetor que representa um genótipo. Já em arquitetura paralela, o tamanho ótimo é maior que o correspondente caso em arquitetura sequencial e que o tamanho exato é sensível a detalhes na implementação (Alander, 1992). Já o trabalho de Arabas J., intitulado *GAVaPS-a genetic algorithm with varying population size* apresenta uma abordagem de população variável que aumenta ou diminui em função de algumas características da pesquisa, ele justifica que se o tamanho da população for muito pequeno a convergência pode ser muito rápida e levaria a um máximo/mínimo local; e que se a população for muito grande, haverá desperdício dos recursos computacionais e o tempo de espera por uma melhoria pode ser demasiado longo (Arabas, 1994).

3.4.4 Mecanismo de seleção dos reprodutores

Os mecanismos adotados para selecionar os reprodutores são quaisquer algoritmos que beneficiam um genótipo proporcionalmente à sua qualidade na função de avaliação. Nesse sentido, os melhores genótipos terão maior chance de se tornarem os genótipos-pais (um genótipo será definido como genótipo-pai se o mesmo for selecionado através de algum mecanismo de seleção dos reprodutores). A composição do mecanismo de seleção dos reprodutores e o mecanismo de seleção dos sobreviventes são responsáveis pela melhoria na qualidade da população através da eliminação/inserção de indivíduos menos/mais qualificados.

Vale salientar que o algoritmo que terá o papel de selecionar os genótipos tem carac-

terísticas estocásticas, isto é, a seleção do indivíduo mais qualificado não é determinístico, e sim probabilisticamente, isto é, indivíduos mais qualificados terão maior probabilidade de serem selecionados. Probabilidade esta que é proporcional à sua qualidade (quão melhor for a qualidade, maior será a probabilidade de seleção associada a tal genótipo), embora todos os indivíduos tenham probabilidade positiva dando a possibilidade que o genótipo menos qualificado também possa ser selecionado. Essa condição garante que a seleção não sentencie que a função de avaliação tenda para um máximo/mínimo local, assegurando a diversidade dos alelos que se encontram espalhados por toda a população (basta imaginar o caso em que dado um momento da evolução, haja somente uma qualidade limitada de alelos comuns aos genótipos mais qualificados. Daí, se vier a selecionar deterministicamente os mais qualificados, está apenas combinando a mesma coleção de alelos, e assim, a cota superior da função de avaliação seria limitada a combinação dos alelos mais significativos desse subconjunto próprio dos alelos da população).

O problema do Caixeiro viajante - *Travelling Salesman Problem* (Applegate, 2006) consiste de um problema de otimização em que se busca a melhor rota passando uma única vez por diversos pontos em uma topologia qualquer, é apresentado por Larrañaga em seu trabalho *Genetic Algorithms for the Travelling Salesman Problem: A Review of Representations and Operators* (Larrañaga, 1999). Neste trabalho o autor apresenta os operadores de seleção, dentre outros operadores, em diversas representações (binária, caminho, adjacência, ordinal e matricial) verificando os desempenhos experimentais de tais operadores em função das diversas representações.

Já o trabalho de Goldberg e Deb, intitulado *A comparative analysis of selection schemes used in genetic algorithms* (Goldberg, 1991) apresenta uma análise sobre os operadores de seleção mais utilizados em computação evolutiva, entre eles temos: reprodução proporcional, ranking, torneio e genitor (estado estacionário). Como resultado, são fornecidas soluções aproximadas e exatas para simulações, bem como tempo de convergência útil e taxa de crescimento; a principal contribuição é dada por sugestão de caminhos para investigação analítica mais detalhada das técnicas de seleção.

3.4.5 Operadores de variação: mutação e recombinação

Os operadores de variação atuam nos genótipos modificando-os a fim de obter melhoramentos na qualidade do indivíduo, isto é, recebem genótipos selecionados na população através de operadores de seleção e os modifica seguindo algum algoritmo. Basicamente há dois tipos de operadores: Mutação e recombinação ou cruzamento.

3.4.5.1 Mutação

A mutação tem a tarefa de modificar o genótipo selecionado atuando diretamente na modificação de um ou mais alelos através de alguma rotina estocástica que define o alelo que sofrerá a mutação (Essa condição garante que as modificações ocorram em diferentes segmentos e assegure genótipos distintos a cada aplicação da mutação, mesmo que seja utilizado o mesmo genótipo como entrada para o operador).

Sob o espaço dos fenótipos, isto é, o espaço de soluções, o ato de aplicar o operador mutação é na verdade tomar uma nova solução para o problema em estudo. Assim, é possível perceber que para um tempo t , suficientemente grande, a mutação possibilitará que através de computação evolutiva se obtenha o ótimo (máximo ou mínimo) global (Eiben, 1991).

3.4.5.2 Recombinação

A recombinação (ou cruzamento) é a criação de genótipos. Assim, uma vez que através de um operador de seleção foram escolhidos os genótipos pais, a recombinação atua selecionando um (ou mais de um) ponto de corte de forma aleatória em que será permutado os alelos de um genótipo pai para outro. Em outras palavras, o processo de recombinação transmite parte da informação (alelos) dos pais para os filhos. Tal atitude, gera genótipos filhos que compõem parte dos alelos de cada um de seus genótipos pais.

É possível perceber que a escolha aleatória do ponto de corte (a partir de que momento a informação será misturada) garante que mesmo que, por ventura, escolha-se em um outro momento os mesmos genótipos pais para obter informação, terá pouca chance de que os atuais genótipos filhos sejam semelhantes (no quesito informação genética) de outros genótipos filhos gerados anteriormente por esses selecionados.

O espaço dos fenótipos encara tal modificação como uma nova solução obtida através de uma função de outras duas soluções, que naquele momento, se mostravam as melhores soluções para o problema.

No trabalho *Adaptive probabilities of crossover and mutation in genetic algorithms* (Srinivas, 1994) foi apresentada uma abordagem sobre operadores de mutação e recombinação sob o ponto de vista adaptativo, isto é, a probabilidade de mutação e recombinação varia dependendo do valor da função de avaliação dos genótipos envolvidos. Essa tentativa visa a proteger as soluções de alto valor de fitness. Ainda nesse trabalho, também é possível definir o valor ótimo para a probabilidade de mutação e recombinação para cada caso, não sendo necessário arbitrar uma probabilidade para esses operadores de forma geral. O autor justifica que essa abordagem se mostra satisfatória em dois sentidos: Man-

ter a diversidade da população e sustentar a capacidade de convergência do algoritmo.

3.4.6 Mecanismo de seleção dos sobreviventes (Substituição)

Uma vez que se tenha uma coleção de genótipos (população), e após a aplicação dos operadores de seleção - mutação - recombinação, terá também os genótipos filhos. A quantidade de elementos na coleção deve ser constante para não prejudicar o desempenho computacional no processo de busca, assim, devemos ter um mecanismo que substituirá tantos quantos forem os genótipos filhos na população. Porém, isso ocorre no melhor caso em que todos os genótipos filhos tem valor em sua função de avaliação maior que o de seus genótipos pais, caso contrário, sua inserção não contribuiria em nada para a convergência da função de avaliação para seu ótimo global (de fato, como todos os genótipos filhos tem alelos ofertados por seus genótipos pais a menos dos alelos alterados pelo processo mutacional, e que o valor de suas funções de avaliação não superaram o valor apresentado pelos genótipos pais, podemos concluir que os alelos pertencentes ao genótipo filho se distanciaram ainda mais do genótipo ótimo).

Com relação à característica, o mecanismo de seleção dos sobreviventes se assemelha muito ao procedimento apresentado pelo operador de seleção que fará seu julgamento através do valor da função de fitness, embora sua concepção (diferente do operador de seleção) é dada de forma determinística. Ou seja, o critério de substituição será feito na tentativa de melhorar a população. Com isso, uma forma é classificar através do fitness e a partir disso identificar os de menor valor (coleção de genótipos + genótipos filhos) e daí, eliminar tantos genótipos quanto forem necessários para manter a coleção com a mesma cardinalidade da geração anterior.

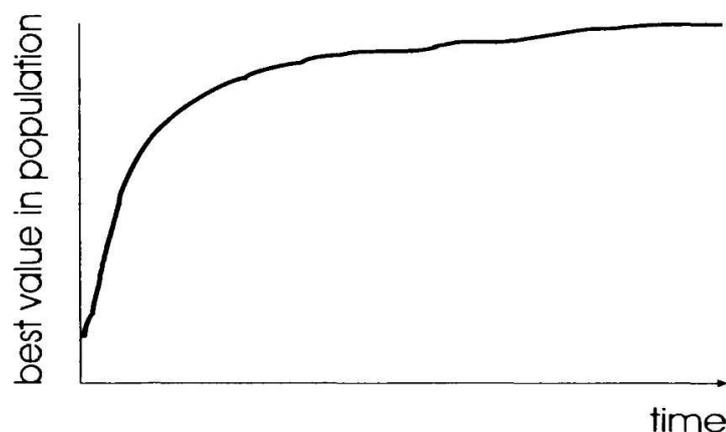


Figura 11: Ilustração da característica monótona e não negativa da função de avaliação (Eiben, 2008).

A cada passo teremos uma função de avaliação monótona e não negativo na Figura 11 apresentamos uma ilustração da curva da função de avaliação, também é possível perceber o ciclo que se forma entre selecionar - mutacionar - recombinar - substituir, Figura 12.

Por fim, uma vez o ciclo formado, é necessário que se defina como iniciar e finalizar o processo de busca. Geralmente o critério para iniciar a população é o preenchimento aleatório de todos os alelos. Assim, será obtida uma coleção de genótipo bem heterogênea.

Quanto à finalização, ao conhecer o valor da função de fitness ótimo, é possível estabelecer um critério de precisão (digamos $\epsilon > 0$), e com isso, obter um critério de parada para a busca tão logo se obtenha um valor para função de avaliação de algum candidato com distância menor que ϵ . Por outro lado, em um sentido mais realista, geralmente não é de conhecimento o ponto ótimo ou não há garantia de que se consiga chegar tão próximo do mesmo para satisfazer um critério de parada nesse contexto. Assim, cabe utilizar outra abordagem para finalizar um processo de busca:

- Estourar o tempo limite de CPU;
- Atingir uma quantidade de avaliações do fitness;
- Treinamento sem progresso, isto é, não ocorrer mudança por um tempo determinado;
- A diversidade de genótipos decaia até um limiar pré-estabelecido.

Basta que um desses critérios seja adotado para que se garanta a finalização do processo de busca.

3.5 Representação de um problema através da codificação de um algoritmo genético

Agora que são conhecidos os componentes necessários para trabalhar com computação evolutiva, a compreensão será direcionada para a forma de codificar um problema através da abordagem criada por Holland, isto é, algoritmo genético - AG. Não somente será abordada essa técnica, como será focado na abordagem adotada para modelagem do problema no próximo capítulo.

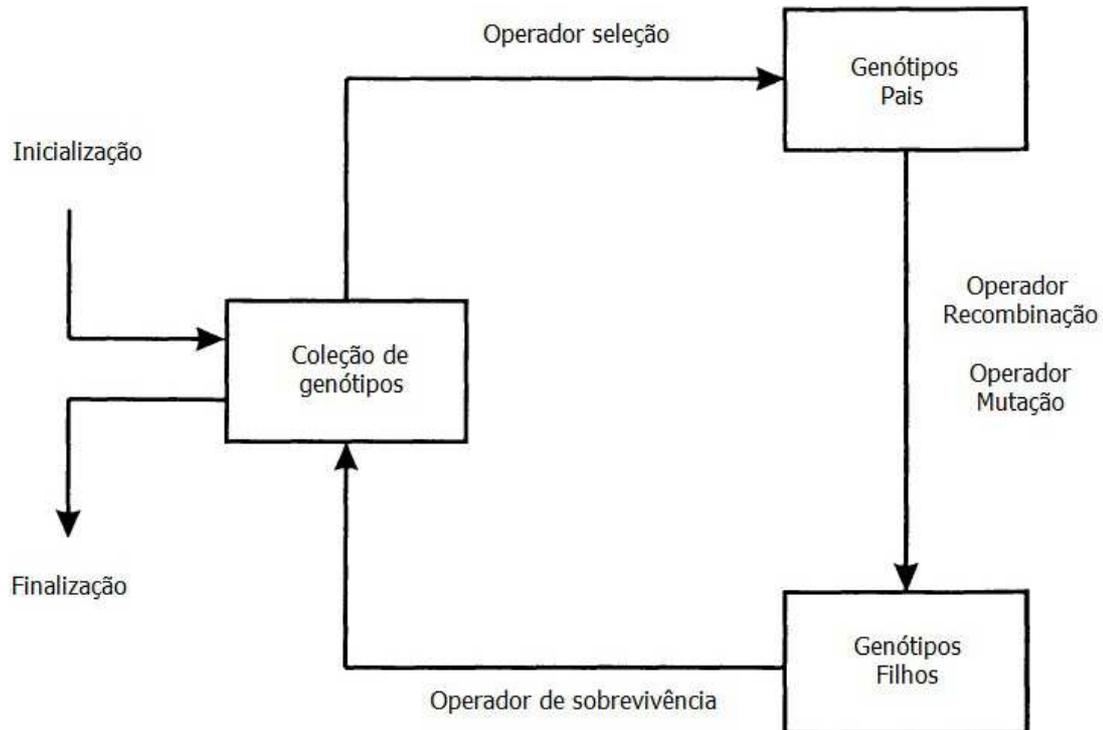


Figura 12: Ilustração do ciclo realizado pelo processo de busca; inicializando a coleção, passado por todos os operadores e avaliando se atende ou não o critério de parada.

3.5.1 Representação do genótipo - AG

Uma representação em que o genótipo (indivíduo) é um string binário, isto é, um vetor (denomina-se vetor no sentido de informar que o posicionamento de cada bit tem importância) formado apenas por 0 ou 1.

Neste sentido, o vetor como um todo é um indivíduo e cada alelo é um agrupamento de bits, digamos 4 bits. Daí, é possível representar números que variam de 0 até 15 em cada alelo, neste exemplo.

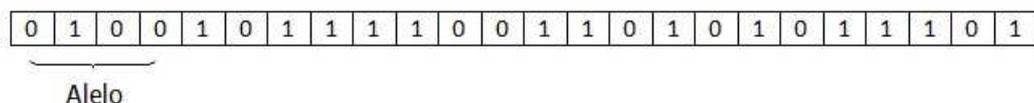


Figura 13: Ilustração de um genótipo em sua codificação binária. Note que, neste exemplo, a cada 4 bits formamos um alelo.

Na Figura 13 é ilustrado o que seria um genótipo na codificação binária para um vetor pertencente ao conjunto $\{0, 1\}^{24}$. Daí, fazendo a decodificação nota-se que se trata do

vetor fenótipo $\langle 4, 11, 12, 13, 5, 13 \rangle$, isto é, um vetor do conjunto $\{0, 1, \dots, 15\}^6$. Assim é possível perceber que o espaço fenótipo está univocamente correspondido com espaço genótipo tornando-o um isomorfismo do espaço que compõe nosso problema.

Outras formas de representação são possíveis como é o caso da representação por números inteiros ou representação de número real (ponto flutuante), entretanto, foge ao que será utilizado neste trabalho e por esse motivo não serão tratados aqui.

3.5.2 Operador de Mutação - AG

Na codificação binária o processo de mutação é realizado através de uma pequena probabilidade $p_m > 0$, assim, sob tal probabilidade o alelo selecionado tem seu valor alterado (de 0 para 1 ou de 1 para 0, também conhecido como *bit flip*). A quantidade de elementos que sofrem mutação geralmente é dado por $K \cdot p_m$, em que K é o comprimento do genótipo.

Por exemplo, suponha um genótipo formado por 20 bits, sendo cada alelo formado por 4 bits, sob uma probabilidade $p_m = 5\%$ terá mutação sob 1 bit $= 0,05 \cdot 20 = p_m \cdot K$. Na Figura 14 é ilustrado o processo de mutação, os pontos que sofrerão a mudança também são escolhidos aleatoriamente dentre todos os pontos que compõem o indivíduo.

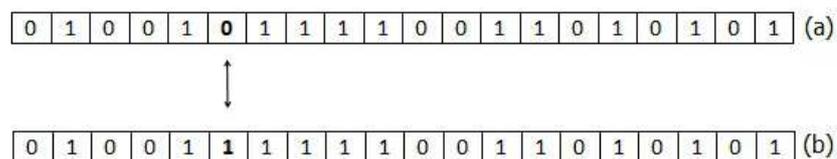


Figura 14: Ilustração do processo de mutação na codificação binária, através do processo bit flip. Em (a) o genótipo inicial e em (b) o mesmo genótipo após a mutação.

3.5.3 Operador de recombinação - AG

O operador de recombinação geralmente associado a uma probabilidade de ocorrência p_r que varia entre $[0, 5; 1, 0]$. Assim, se aplicado um teste o valor for menor ou igual a p_r , o operador de recombinação gera os fenótipos filhos através da recombinação de seus alelos; caso contrário, os fenótipos filhos são gerados de forma assexuada (através de uma cópia direta dos seus pais).

A recombinação na representação binária pode ocorrer sob o ponto de corte diretamente sobre 1 ponto entre os bits (em que a partir do ponto de corte toda a informação

do genótipo 1 é repassado para o genótipo 2 e vice-versa) ou através de n pontos através da permuta dos bits que se encontram entre dois pontos de corte distintos. A seleção do ponto ponto de corte é dado aleatoriamente seguindo algum processo de seleção.

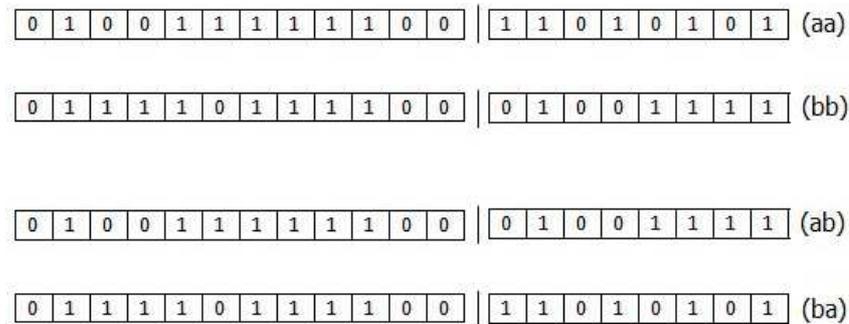


Figura 15: Ilustração do operador de recombinação atuando sobre dois genótipos pais através de uma recombinação de um único ponto.

Nas Figuras 15 e 16 ilustramos as duas situações, na primeira a atuação do operador de recombinação para 1 ponto e a atuação para um caso em que tomamos 3 pontos em um genótipo de comprimento $K = 20$.

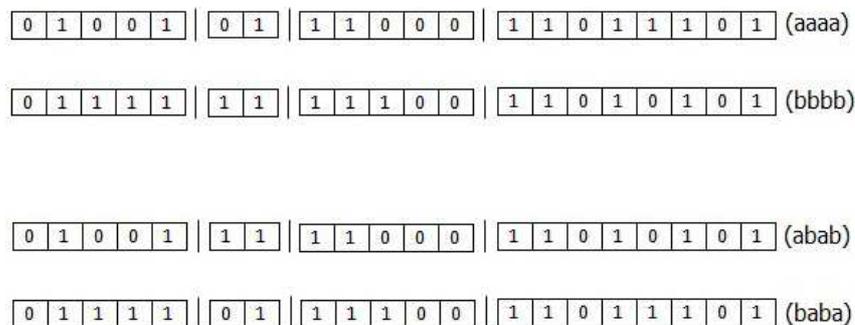


Figura 16: Ilustração do operador de recombinação atuando sobre dois genótipos pais através de uma recombinação de três pontos. Note que as partes 2 e 4 de cada fenótipo foram permutadas.

3.5.4 Modelo de população

Há dois tipos de modelos populacionais, o modelo geracional e o de estado estacionário (Eiben, 1991):

- Geracional - Neste modelo toda a coleção de genótipos é selecionada para receber os operadores de variação, com isso, são gerados genótipos filhos em quantidade igual à cardinalidade da coleção de genótipos pais. Após cada geração, toda a coleção é substituída pelos genótipos filhos gerados. Esse modelo apresenta uma situação indesejada quando a mutação modifica os alelos mais significativos e ao modificar toda população perde-se tal informação, daí, o indivíduo sofre uma perda em sua adaptabilidade nessa geração.
- Estado estacionário - Neste modelo não é a coleção que modifica a cada geração e sim parte dela. Assim uma quantidade q (menor que a cardinalidade da coleção) será substituída por q genótipos filhos; a porcentagem de elementos que serão selecionados é dada por q/K em que K é a cardinalidade da coleção. Para Whitney, em seu algoritmo Genitor, o valor de q é igual a 1, e com isso, o percentual de indivíduos selecionados é dado por $1/K$ (Whitney, 1988)

3.5.5 Mecanismo de seleção dos reprodutores

3.5.5.1 Seleção proporcional ao fitness

Consiste de selecionar os genótipos pais através da proporção gerada pela da comparação do valor função de avaliação com relação à soma dos valores da função de fitness da coleção de genótipo. A expressão proporcional é dada por:

$$fp_i = \frac{f_i}{\sum_{i=1}^K f_i} \quad (3.1)$$

em que f_i é o valor da função de avaliação para o genótipo i .

O principal problema desse método é a convergência prematura, em que um genótipo obtenha um valor muito alto para sua função de avaliação e com isso passe a dominar o processo de seleção forçando toda a coleção a ser uma combinação de seus alelos. Daí, a busca evolutiva ficaria restrita a um ótimo local que dependa dos alelos desse genótipo (Holland, 1992). Uma tentativa de diminuir a possibilidade dessa convergência seria a condição de que o indivíduo com valor de fitness muito alto já houver sido selecionado para geração de um filho, este terá um fator associado que temporariamente diminua seu valor de fitness dando possibilidade aos demais genótipos de distribuir sua informação genética.

3.5.5.2 Seleção por ranking

Nesse método, é feito um mapeamento do valor da função de fitness através de uma função que defina sua probabilidade de escolha. Em geral, é adotada uma expressão linear ou exponencial. Para o caso em que a expressão é linear usa-se como parâmetro um valor t ($1, 0 < t \leq 2, 0$), daí, uma possível expressão linear é dada por:

$$P_{ranking}(i) = \frac{2-t}{K} + \frac{2i(t-1)}{K(K-1)} \quad (3.2)$$

em que K é a cardinalidade da coleção de genótipos e o índice i se refere ao genótipo i .

Essa abordagem visa melhorar o critério de seleção nos momentos em que toda a coleção apresenta valor de fitness muito próximo, e sob o método de seleção anterior, viria a ser uma escolha praticamente aleatória sem evidenciar a qualidade que cada genótipo associe (mesmo que seja muito sutil tal qualidade) (Baker, 1987).

3.5.5.3 Seleção pelo algoritmo roleta

Esse algoritmo visa obter os indivíduos através de um procedimento semelhante a uma roleta. Inicialmente é utilizado o ranking ou o fitness proporcional para definirmos a probabilidade P_i para cada indivíduo i da coleção contendo K elementos. Esse procedimento compartilha o mesmo problema da seleção proporcional que pode gerar convergência prematura.

Uma vez definida a probabilidade de cada indivíduo proporcional ao valor de sua função de avaliação, são tomados os pontos acumulados:

$$a_j = \sum_{i=1}^j P_i \quad (3.3)$$

assim, haverá a lista $(0, a_1, a_2, \dots, a_K)$ que representam os pontos de acumulados das probabilidades dos indivíduos i . Note ainda que o intervalo $[a_{i-1}, a_i] = P_i$, isto é, cada partição dessa lista diz respeito ao intervalo proporcional ao valor de fitness para o indivíduo correspondente.

Sabendo que:

$$\sum_{i=1}^K P_i = 1 \quad (3.4)$$

basta que se selecione aleatoriamente um ponto entre o intervalo $[0, 1]$, e dado a partição

em que esse ponto pertença, será selecionado o indivíduo correspondente.

3.5.5.4 Seleção por torneio

O procedimento por torneio tem características mais simplistas, ideal para grandes populações ou coleções que não ocorra acesso a todas as informações naquele presente. Sua grande vantagem é não ter a necessidade de fazer uma avaliação com relação à toda a população, e sim a uma amostragem desta.

Inicialmente, define-se a quantidade de genótipos pais que serão selecionados, digamos k , daí, selecionam-se n elementos da população aleatoriamente (claramente $n < K$), e de posse desses genótipos será avaliado o de melhor valor de fitness. Chama-se esse indivíduo de g_1 , e se repete o procedimento até que se obtenha todos os k genótipos desejados.

Notoriamente, quanto maior for a quantidade de elementos no torneio, melhor será a chance de se obter genótipos com alto valor de fitness e menor será a chance de que ocorra um torneio completamente preenchido de indivíduos com baixo valor de fitness (Blickle, 1995).

3.5.6 Mecanismo de seleção dos sobreviventes (substituição)

3.5.6.1 Sobrevivência dos genótipos mais novos

Nesse contexto, não importa quão alto seja o valor de fitness obtido por um genótipo, ele só permanecerá na população uma quantidade fixa de gerações. Assim, os indivíduos mais velhos “morrem” ao completar seu ciclo de vida dando lugar para os mais novos.

Perceba que a lógica adotada não visa uma função do tipo serra, em que os valores de fitness atingem um patamar mais elevado e depois de uma quantidade fixa de gerações perdem esse efeito e voltam ao patamar anteriormente estabelecido; espera-se que os alelos significativos tiveram tempo suficiente para serem repassados para os demais genótipos mais novos, e a partir disso, assegurar que o ótimo local alcançado não seja perdido (geralmente).

É possível perceber que o modelo de população geracional, em que todos são substituídos, pode ser considerado a aplicação do caso trivial do mecanismo de seleção por sobrevivência, em que, o ciclo de vida de qualquer genótipo é de uma única geração.

3.5.6.2 Sobrevivência baseada no valor da função de fitness

O caso mais utilizado é sem dúvida o de *substituição do pior genótipo*. Daí, se tivermos k genótipos filho e K elementos na população, isto é $K + k$ elementos para que sejam selecionados K dentre estes; é feito um ranking com relação ao valor de fitness de todos os $K + k$ indivíduos, e por fim, selecionado os K primeiros elementos.

Esse tipo de estratégia tem o problema de convergir para ótimos locais, assim, são normalmente associados a grandes coleções de genótipos ou exigências de não haver mais de um genótipo com a mesma combinação de alelos.

Outra característica é o *elitismo*, nesse caso, os indivíduos de mais alto valor de fitness permanecem na população. Geralmente vem associado à estratégia de sobrevivência dos mais novos, garantindo assim que a condição de um indivíduo velho com alto valor de fitness permaneça na população nos casos em que nenhum dos indivíduos gerados melhoram ou estabilizam o valor da função de fitness.

3.6 Algoritmo genético utilizado no trabalho

Neste trabalho, utilizaremos uma variante melhorada do algoritmo genético, o algoritmo melhorado do Leung (Leung, 2003) utilizado para treinar os parâmetros da redes neurais. Resultados comparativos mostram a superioridade dessa variante com relação ao algoritmo genético tradicional.

3.7 Unificando os conhecimentos

Com base na codificação apresentada neste capítulo, a estratégia de busca por soluções ótimas ou sub-ótimas e com o conhecimento adquirido sobre a modelagem de cenários através de autômatos celulares, no próximo capítulo será definido o problema e codificado o mesmo atendendo as condições dessas duas teorias. Isto é, utilizar o autômato celular para simular o cenário do problema, e utilizar a estratégia de busca oferecida por computação evolutiva na tentativa de encontrar as regras de transição que melhor descrevem o problema.

4 Modelagem matemática

The “evolving cellular automata” framework is an idealized means for studying how evolution (natural or computational) can create systems in which “emergent computation” takes place.

Mitchell, 1996.

4.1 O problema

Modelar a evolução da infecção da população de uma região através do verme *Schistosoma* em face da dinâmica de movimentação da população e alterações dos cursos d'água.

Diante do complexo cenário de infecção em que há mais de uma fase, o problema será modelado através de camadas: A primeira delas, a camada de movimentação da população sadia por todo o sítio terrestre e meios alagados; A segunda, a camada de movimentação da população portadora da infecção; A terceira, a camada de movimentação da população que foi infectada, e mediante tratamento, se recuperou da infecção; A quarta, a camada de movimentação do molusco sadio através dos sítios terrestre e meios alagados; A quinta, a camada de movimentação dos moluscos que foram expostos a fezes infectadas pelo ovos do verme *Schistosoma* que deram origem aos miracídeos.

As camadas se fazem necessárias para que, por exemplo, os indivíduos infectado tenham trânsito através dos sítios de forma independente dos indivíduos infectados podendo ou não ocuparem as mesmas células.

4.1.1 Mapa

O mapa foi gerado através de uma malha de comprimento 10×10 de células. Cada célula simbolizará um sítio do mapa real, em teoria, estamos dividindo o mapa geográfico do evento numa malha 10×10 .

Foi utilizado uma malha 10×10 , sob o terreno de 1850×450 metros totalizando $832.500 m^2$ ou $83,25$ ha. Subdividindo pela quantidade de células teremos $0,8325$ ha por célula ou $8325 m^2$ que foi adotado como a área média de um quarteirão ($100 m \times 83 m$) que é a unidade utilizada neste trabalho para agrupamento familiar tendo em vista que há relação direta entre os habitantes de um mesmo quarteirão e seu contato com os demais quarteirões que o circundam define sua vizinhança. A ideia de diminuir a escala da célula faria com que o próximo agrupamento familiar fosse o de cada residência o que haveria a necessidade de modelar microdeslocamentos o que levaria a uma maior demanda de processamento por deslocamentos e iterações que não contribuiriam para a dinâmica de infecção. Sendo assim, a adoção da medida por quarteirão por praticidade de processamento e eficácia na observação dos deslocamentos significativos.

Um riacho foi incorporado ao modelo percorrendo as células:

$$(6, 1) \rightarrow (6, 2) \rightarrow (7, 2) \rightarrow (8, 3) \rightarrow (8, 4) \rightarrow (9, 5) \rightarrow (10, 6) \rightarrow (10, 7)$$

em que (i, j) é a coordenada da célula localizada na linha i e coluna j , com $1 \leq i, j \leq 10$.

Para simular as regiões no mapa em que há contato com água salgada, foram definidas células para se comportarem como mar aberto em que não haverá infecção pelo contato animal/humano. As células definidas para serem de água salgada são:

$$\begin{aligned} (1, 10) \leftrightarrow (2, 10) \leftrightarrow (3, 10) \leftrightarrow (4, 10) \leftrightarrow (5, 10) \leftrightarrow (6, 10) \leftrightarrow (7, 10) \leftrightarrow (8, 10) \leftrightarrow (9, 10) \leftrightarrow \\ (10, 10) \leftrightarrow (1, 9) \leftrightarrow (2, 9) \leftrightarrow (3, 9) \leftrightarrow (4, 9) \leftrightarrow (5, 9) \leftrightarrow (6, 9) \leftrightarrow (7, 9) \leftrightarrow (8, 9) \leftrightarrow \\ (1, 8) \leftrightarrow (2, 8) \leftrightarrow (3, 8) \leftrightarrow (4, 8) \end{aligned}$$

A figura 17 é uma ilustração da motivação para a malha e para a escolha das células que serão tratadas como água doce e água salgada.

Para as camadas seguintes, foram utilizados valores para a distribuições para preencher a população inicial, devido até o momento os dados de campo em formato de formulário com as quantidades de indivíduos e as classes de cada indivíduo não puderam ser entregues. Quanto aos valores utilizados, foram utilizados valores entre 0 e 30 por célula, para aproximar a população atual da praia de Carne de Vaca **aproximadamente** 1600 habitantes. A distribuição utilizada será a uniforme contínua para assegurar igual probabilidade para toda a amplitude dos parâmetros utilizados.



Figura 17: Ilustração da motivação da configuração da malha

4.1.2 Camada de indivíduos sadios

Para o preenchimento dessa camada foi utilizada a distribuição uniforme discreta $U_d(5, 30)$, assim, cada célula habitável poderá ter de 5 até 30 indivíduos sadios inicialmente em sua construção.

4.1.3 Camada de indivíduos infectados

Para o preenchimento dessa camada foi utilizada a distribuição uniforme discreta $U_d(0, 3)$, assim, cada célula habitável poderá ter de 0 até 3 indivíduos infectados inicialmente em sua construção.

4.1.4 Camada de indivíduos recuperado

Inicialmente toda a camada de indivíduos recuperados está sem nenhum habitante, pois, espera-se que a ocorrência de recuperados seja originada da dinâmica da população configurada.

4.1.5 Camada de moluscos sadios

Para o preenchimento dessa camada foi utilizada a distribuição uniforme discreta $U_d(0, 10)$, assim, cada célula habitável poderá ter de 0 até 10 moluscos sadios em sua configuração inicial.

4.1.6 Camada de moluscos infectados

Para o preenchimento dessa camada foi utilizada a distribuição uniforme discreta $U_d(0, 3)$, assim, cada célula habitável poderá ter de 0 até 3 moluscos infectados em sua configuração inicial.

4.1.7 Alagamentos

A Tabela 1 apresenta o índice pluviométrico da praia de Carne de Vaca que será levado em consideração para modelar o alagamento e/ou aumento do fluxo de água doce pelo córrego definido em nossa malha.

Tabela 1: Índice pluviométrico da praia de Carne de Vaca.

Meses	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Índice (mm)	73,3	105,78	187,48	213,06	231,66	389,5	204,4	196,46	63,76	27,96	33,06	39,0

Esse índice será utilizado para avaliar a quantidade de células que serão consideradas passíveis de infecção, isto é, a partir do índice pluviométrico definiremos quais células se encontram com água doce (além do córrego) e serão levadas em consideração caso haja moluscos infectados neste sítio. Dessa forma, a dinâmica de infecção é função dos níveis de água doce ao longo do ano.

Para cada célula que contém o riacho, é testado o valor do *parâmetro de alagamento* com o referencial de um número aleatório obtido através de uma distribuição uniforme contínua entre 0 e 1 ($U_c(0, 1)$), caso o valor de referência for maior ou igual ao valor do parâmetro, é alterado o limite de alagamento. Os limites de alagamento foram definidos da seguinte maneira:

- Se o índice ficar entre 0 e 109 mm, então é feito o teste contra o valor de referência para célula que está imediatamente acima da célula em análise e decidido se a célula acima será definida como célula alagada;

- Se o índice ficar entre 109 e 179 mm, então a célula imediatamente acima da célula em estudo é definida como alagada e é feito novo teste contra o valor de referência para a célula que se encontra duas unidades acima da célula em estudo;
- Se o índice ficar entre 179 e 249 mm, então a célula que está imediatamente acima e a célula que está duas unidades acima da célula em estudo serão definidas como alagadas;
- Se o índice ficar entre 249 e 319 mm, então as duas células acima serão definidas como alagadas e é feito novo teste contra o valor de referência para a célula que se encontra 3 unidades acima da célula em estudo;
- Se o índice ficar entre 319 e 390 mm, então as três células acima da célula em estudo serão definidas como alagadas.

4.1.8 Mobilidade dos indivíduos

Cada célula tem até o momento uma quantidade de indivíduos que varia de 5 a 30 indivíduos. Assim, para cada indivíduo é feita a comparação entre um valor de referência obtido através de uma distribuição uniforme contínua entre 0 e 1 ($U_c(0, 1)$), esse valor de referência será comparado com o *parâmetro de mobilidade humana*. Caso o parâmetro seja maior ou igual ao valor de referência, então é gerado um par ordenado (x, y) formado por números aleatórios obtidos de uma distribuição uniforme discreta entre 1 e 10 ($U_d(1, 10)$) que representará o novo posicionamento desse indivíduo. A mobilidade dos indivíduos é sobre toda a malha, inclusive por: alagados, rios e oceanos.

Note que esse procedimento se repete em todas as células que contenham indivíduos e para cada indivíduo da célula, assim, sob um parâmetro de mobilidade todos os indivíduos tem a chance de se deslocar sobre a malha. O parâmetro de mobilidade humana é o mesmo para o deslocamento dos indivíduos sadios, infectados e recuperados; sob a premissa de não haver distinção na locomoção desses.

4.1.9 Mobilidade dos moluscos

O procedimento para trânsito dos moluscos é semelhante, para cada célula da malha que contém moluscos (inicialmente variando de 0 até 10) é feita a comparação do valor de referência obtido de uma distribuição uniforme contínua entre 0 e 1 ($U_c(0, 1)$) e comparado com o *parâmetro de mobilidade do molusco*. Caso o parâmetro seja maior ou igual ao valor de referência, então é gerado um par ordenado (x, y) formado por números aleatórios obtidos de uma distribuição uniforme discreta entre 1 e 10 ($U_d(1, 10)$) que representará o novo posicionamento do molusco.

4.1.10 Dinâmica de infecção

Foi considerado a cada passo do modelo como unidade mensal, daí, a cada 12 passos do modelo terá um ano de eventos ocorrido. Nessa concepção, a cada passo o valor de índice pluviométrico é alterado com referência a tabela 1 a fim de fidelizar os diferentes fluxos de água doce e/ou alagamentos nos sítios de estudo.

Modelling a real data cellular automaton to analyze the Schistosomiasis expansion process along the coastline of Brazil apresenta um modelo compartimental (SIRS - SI) para modelar a dinâmica de infecção humana (Silva *et al.*, 2012):

$$\begin{cases} \frac{dS}{dt} = -t_H S M_I + \alpha R \\ \frac{dI}{dt} = t_H S M_I - \chi I \\ \frac{dR}{dt} = \chi I - \alpha R \end{cases} \quad (4.1)$$

e a dinâmica de infecção do molusco:

$$\begin{cases} \frac{dM_S}{dt} = -t_M I M_S + r M_S \\ \frac{dM_I}{dt} = t_M I M_S - r M_I \end{cases} \quad (4.2)$$

em que são definidas três compartimentos para os indivíduos: S (Suscetível), I (Infectado) e R (Recuperado). Já para o molusco, é definido dois compartimentos: M_S (Molusco suscetível) e M_I (Molusco infectado). As taxas t_H e t_M são as *taxas de infecção humana* e do *molusco*, respectivamente, e por fim, as taxas α , χ e r são as taxas de: *suscetibilidade humana*, *recuperação humana* e *crescimento do molusco*.

A singularidade do problema exige que haja contato na mesma célula entre humano suscetível e molusco infectado (em $S M_I$) em meio alagado para que as cercárias expostas pelos moluscos entrem em contato com a pele humana. A recuperação é assegurada pelo tratamento e controle, sendo assim, um parâmetro flexível associado ao nível de controle de saúde da região. A infecção do molusco ocorre quando há contato na mesma célula de humano infectado em meio alagado e molusco suscetível (em $I M_S$), pois, nessa ocasião os ovos contidos nas fezes do humano infectado tem contato com a água eclodindo e dando origem a miracídios que ao se encontrarem com moluscos suscetíveis o penetram e passam a se desenvolver.

4.1.11 Modelagem para o saneamento

Em virtude do aumento do fluxo de água doce e/ou alagamentos nos sítios terrestres, há um aumento da quantidade de células em meio alagado. Diante disso, é preciso modelar o recuo dessas águas, pois, sem essa situação em um curto espaço de tempo é

possível obter todos os sítios terrestres alagados, fugindo da realidade. Para isso, é associado ao modelo o *parâmetro de saneamento* que atua através de uma comparação com um valor de referência obtido por distribuição uniforme contínua entre 0 e 1. Caso o valor de referência seja menor ou igual ao valor do *parâmetro de saneamento*, então a célula em análise volta a ser um sítio em meio seco.

Essa comparação é realizada a cada passo do modelo em todas as células da malha, preservando as células reservadas para o riacho e para a água salgada.

4.1.12 Parâmetros para a modelagem

Os parâmetros usados para a modelagem são:

- Mobilidade humana;
- Mobilidade molusco;
- Índice pluviométrico;
- Suscetibilidade humana;
- Recuperação humana;
- Infecção humana;
- Infecção molusco;
- Crescimento molusco;
- Saneamento em sítios terrestres.

Os níveis utilizados para esses parâmetros se encontram nos artigos: *Spatial pattern, water use and risk level associated with the transmission of schistosomiasis on the north coast of Pernambuco* para os parâmetros de suscetibilidade, recuperação e infecção humanas (0-23%), mobilidade humana (0-26%); *Aspectos ecológicos e levantamento malacológico para identificação de áreas de risco para a transmissão da esquistossomose no litoral norte de Pernambuco* para os parâmetros de mobilidade molusco (0-2%), infecção molusco (0-100%), saneamento sítios terrestres (0-93%) e *Laboratório de Meteorologia de Pernambuco (Lamepe)* para o parâmetro de índice pluviométrico (39-389 mm) (Lamepe, 2011).

Serão utilizados os valores de limiar inferior e superior para cada um desses parâmetros, além do valor médio dessas amplitudes. Assim, teremos três níveis: Baixo, Médio e Alto; para cada um desses parâmetros e poderemos inferir o comprometimento na evolução da infecção ao longo dos passos da modelagem para os diferentes níveis.

4.1.13 Quantificando a infecção

O procedimento se repete a cada passo da simulação, ao final de todos os passos de evolução na malha são quantificados em um grupo os indivíduos sadios e recuperados e em outro os indivíduos infectados. O índice de infecção I é dado então pela expressão:

$$I = \frac{\textit{Infectados}}{\textit{Infectados} + \textit{Sadios} + \textit{Recuperados}} \quad (4.3)$$

daí, o índice I varia no intervalo fechado entre 0 e 1, obtendo extremos nos casos em que ninguém permanece infectado ao final da simulação e no caso em que todos estão infectados ao final da simulação.

Consideraremos 6 níveis de infecção para a malha:

- $I = 100\%$, totalidade de infecção na malha;
- $80\% \leq I < 100\%$, infecção generalizada;
- $60\% \leq I < 80\%$, infecção severa;
- $40\% \leq I < 60\%$, infecção média;
- $20\% \leq I < 40\%$, infecção baixa;
- $I < 20\%$, infecção controlada.

5 Resultados

5.1 Simulações dos parâmetros

Para as simulações dos parâmetros foram consideradas 729 análises, a saber, a variação dos parâmetros:

- Taxa de infecção;
- Taxa de suscetibilidade;
- Taxa de recuperação;
- Taxa de infecção do molusco;
- Taxa de crescimento do molusco;
- Taxa de mobilidade humana.

é feita considerando seus limiares inferiores, superiores e média entre eles.

Para cada uma dessas análises foram considerados os passos temporais: 3, 6, 9, 12 e 15 meses; para garantir uma evolução no quadro de infecção em cada cenário. Ao final da simulação foi realizada uma análise de cluster, uma técnica de aprendizagem não supervisionada usada para classificação de dados, na tentativa de agrupar as coleções de parâmetros que apresentassem evolução correspondente. Ou seja, o ato de avaliarmos os três níveis para cada um dos parâmetros citados acima tem a finalidade de mapear discretamente o cenário evolutivo da infecção, e a partir deste, definir quais configurações de campo levariam a um mesmo nível de infecção.

A caracterização para a análise por clusters é dada pela avaliação do vetor do nível de infecção nos instantes: 3, 6, 9, 12 e 15 meses; para cada uma das 729 simulações:

$$sim_i = (n_3, n_6, n_9, n_{12}, n_{15}) \quad (5.1)$$

em que sim_i corresponde ao vetor de simulação i , com $1 \leq i \leq 729$, e n_t é o nível de infecção para a simulação i no passo temporal t , com $t \in \{3, 6, 9, 12, 15\}$. A distância adotada é a distância Euclidiana quadrada, que é definida por:

$$d_{i,j}^2 = \sum_{t \in \{3,6,9,12,15\}} (n_{t,j} - n_{t,i})^2 \quad (5.2)$$

em que i, j são vetores de simulação.

O método adotado para clusterização é o de otimização, isto é:

- 1- Define-se inicialmente um particionamento e calcula-se para cada partição sua centróide;
- 2- Calcula-se as distâncias entre cada indivíduo e os centróides de cada partição, daí, transfere-se o indivíduo para a partição com menor distância;
- 3- Recalcula o centróide de cada partição;
- 4- Repete-se os passos 2 e 3 até que todos os indivíduos se encontrem em partições estabilizadas e não seja possível efetuar transferências de uma partição para outra.

Foram encontrados 6 agrupamentos distintos indicando que embora haja inúmeras formas de configuração para os parâmetros do cenário em campo, estas configurações convergem para 6 resultados bem característicos. As Figuras 18 até 23 apresentam os gráficos boxplot para as simulações realizadas mostrando o comportamento de cada agrupamento. É possível notar um comportamento estacionário para o nível de infecção ao longo do tempo na Figura 18 e um crescimento linear no nível de infecção no agrupamento 3, o mesmo comportamento é apresentado na Figura 19 porém com um crescimento não linear. O agrupamento 4 apresentou diminuição no nível de infecção sugerindo que as configurações que correspondem a esse agrupamento levam a um melhoramento no cenário. Por fim, os agrupamentos 5 e 6 apresentaram um aumento mais significativo do nível de infecção sugerindo o descontrole da região culminado pela ocorrência das configurações dos parâmetros desses agrupamentos.

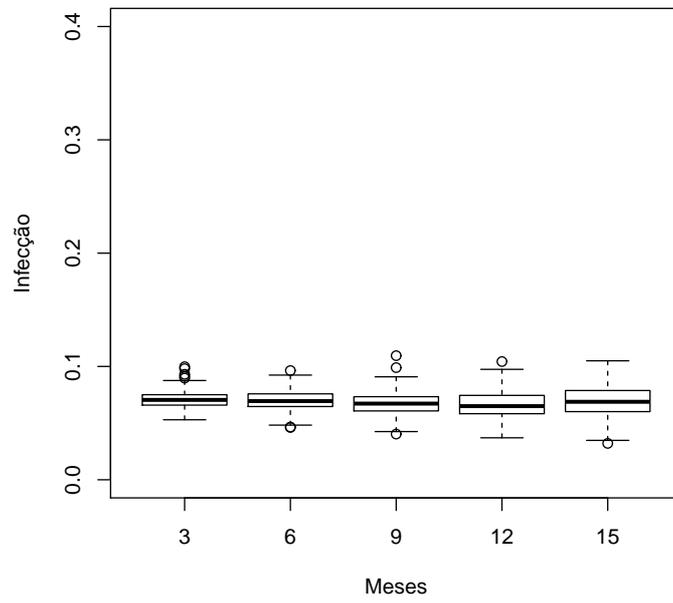


Figura 18: Boxplot para o agrupamento 1.

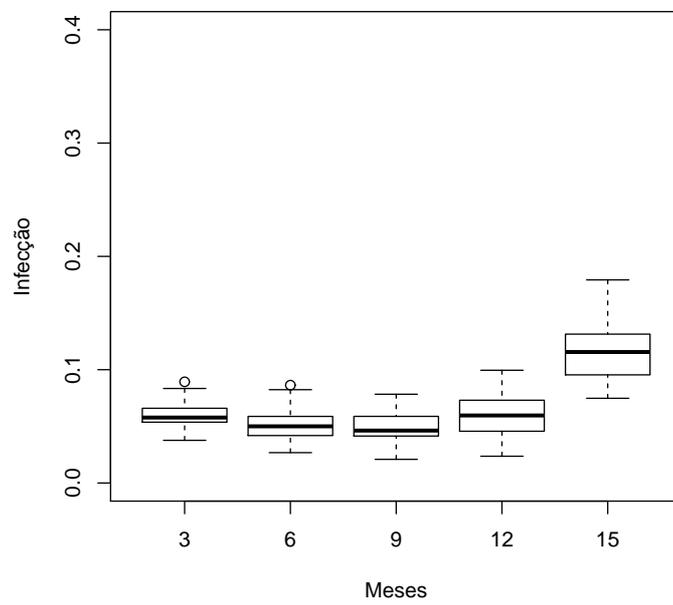


Figura 19: Boxplot para o agrupamento 2.

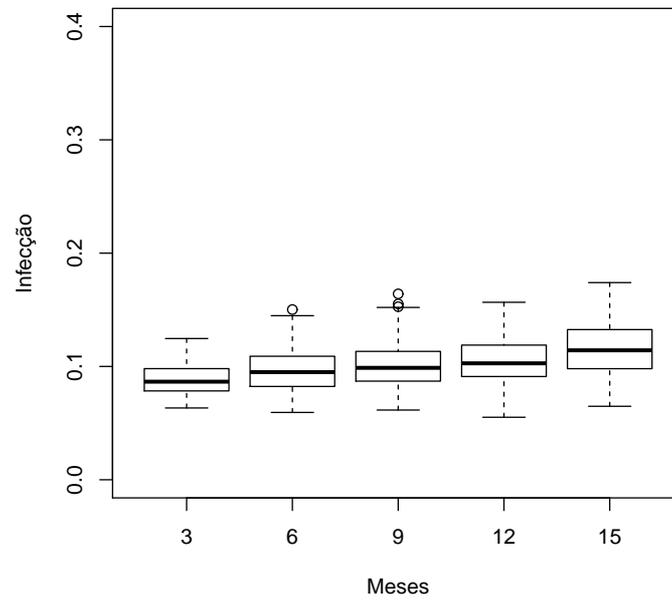


Figura 20: Boxplot para o agrupamento 3.

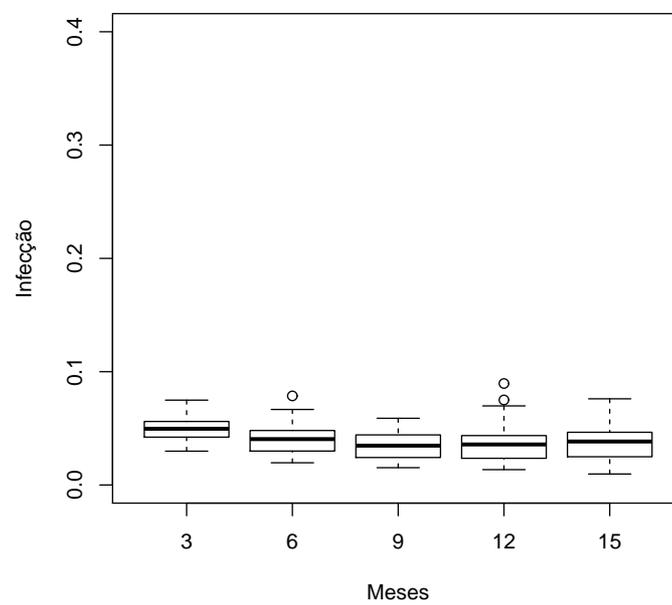


Figura 21: Boxplot para o agrupamento 4.

As Tabelas 2 até 7 apresentam as estatísticas descritivas para cada um dos agrupamentos

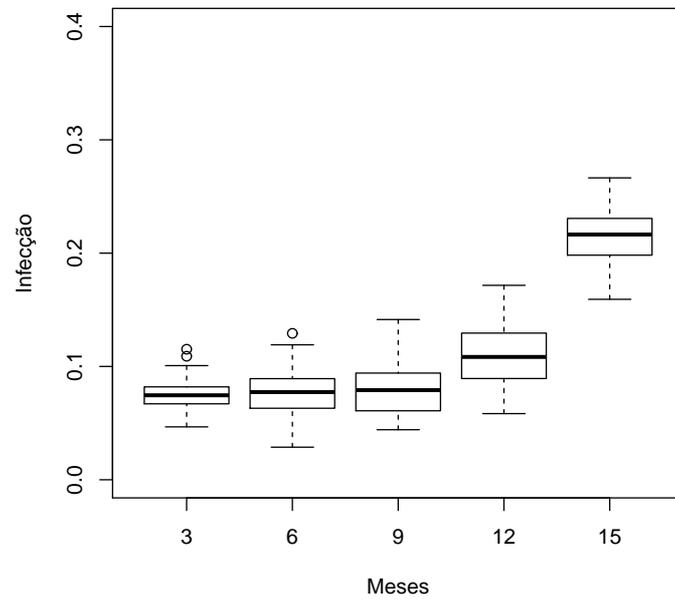


Figura 22: Boxplot para o agrupamento 5.

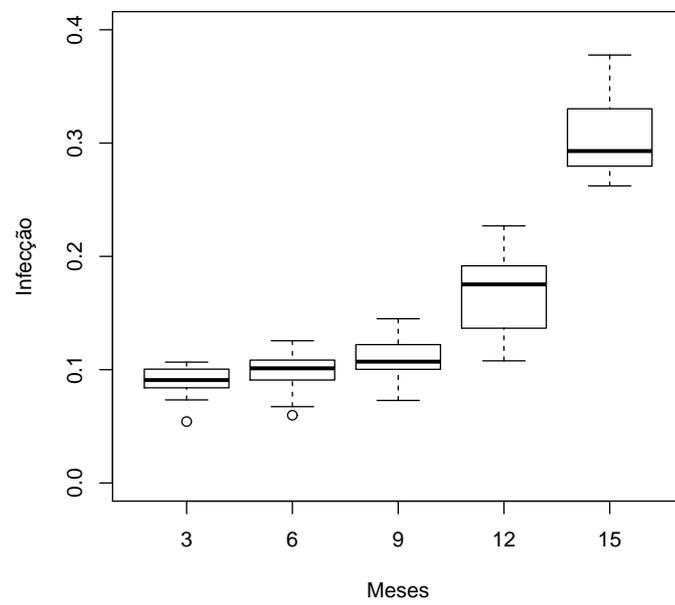


Figura 23: Boxplot para o agrupamento 6.

Tabela 2: Estatísticas descritivas para o agrupamento 1.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,07075252	0,07002256	0,06740603	0,06575762	0,06940249
Percentil 0%	0,05293790	0,04615450	0,04032840	0,03702290	0,03208600
Percentil 25%	0,06587025	0,06463778	0,06076873	0,05840102	0,06017420
Percentil 50%	0,07047770	0,06942645	0,06720035	0,06497250	0,06881735
Percentil 75%	0,07506800	0,07581875	0,07320570	0,07445515	0,07872300
Percentil 100%	0,09977260	0,09636860	0,10951200	0,10433700	0,10497900
Variância	6,19723e-05	8,014044e-05	9,737001e-05	0,0001649811	0,0002151363
CV	11,12646	12,78462	14,63908	19,53309	21,134
Amplitude	0,0468347	0,0502141	0,0691836	0,0673141	0,072893

Tabela 3: Estatísticas descritivas para o agrupamento 2.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,05996852	0,05137273	0,04955347	0,06047746	0,1151833
Percentil 0%	0,0376463	0,0267751	0,0208725	0,0236095	0,0747364
Percentil 25%	0,0536559	0,0418625	0,0414157	0,0456797	0,0954341
Percentil 50%	0,0577634	0,0500219	0,0462617	0,0595934	0,1155980
Percentil 75%	0,0659462	0,0587739	0,0588118	0,0730401	0,1313910
Percentil 100%	0,0892907	0,0863387	0,0783320	0,0994954	0,1792320
Variância	9,769843e-05	0,0001585966	0,0001692433	0,0003394641	0,0005333544
CV	16,4824	24,51401	26,25317	30,46515	20,05019
Amplitude	0,0516444	0,0595636	0,0574595	0,0758859	0,1044956

Tabela 4: Estatísticas descritivas para o agrupamento 3.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,08814439	0,09653487	0,1019754	0,1052163	0,1154459
Percentil 0%	0,06335240	0,05940610	0,06152560	0,0236095	0,0747364
Percentil 25%	0,07850670	0,08237915	0,08740598	0,0456797	0,0954341
Percentil 50%	0,08655845	0,09497005	0,09876520	0,0595934	0,1155980
Percentil 75%	0,09800935	0,10900075	0,11315275	0,0730401	0,1313910
Percentil 100%	0,12462800	0,15021100	0,16405100	0,0994954	0,1792320
Variância	0,0001771286	0,0003329728	0,0004352506	0,0003196986	0,0005497746
CV	15,09905	18,90254	20,45853	16,99368	20,31018
Amplitude	0,0612756	0,0908049	0,1025254	0,1014722	0,1091441

Nas Figuras 24 até 29 são apresentados os histogramas para as médias dividida pelo período de simulação para cada agrupamento. Apresentando uma distribuição normal das médias, em seguida, foi realizada uma comparação através da plotagem dos quantis (Q-Q

Tabela 5: Estatísticas descritivas para o agrupamento 4.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,04962301	0,03980409	0,03441571	0,03450884	0,03817586
Percentil 0%	0,02985490	0,01960390	0,01526620	0,01354970	0,00970140
Percentil 25%	0,04222968	0,02997453	0,02426595	0,02359827	0,02495890
Percentil 50%	0,04963645	0,04048900	0,03471575	0,03577850	0,03837695
Percentil 75%	0,05604602	0,04800145	0,04419343	0,04358582	0,04651047
Percentil 100%	0,07490220	0,07864470	0,05887650	0,08956970	0,07607410
Variância	7,832931e-05	0,0001210986	0,0001228752	0,0001672867	0,0002246253
CV	17,83524	27,64661	32,20886	37,48006	39,25912
Amplitude	0,0450473	0,0590408	0,0436103	0,07602	0,0663727

Tabela 6: Estatísticas descritivas para o agrupamento 5.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,07496963	0,07634866	0,07948588	0,1099251	0,2153482
Percentil 0%	0,04668250	0,02875070	0,04416610	0,05838160	0,1592560
Percentil 25%	0,06717147	0,06335132	0,06098385	0,08945213	0,1989327
Percentil 50%	0,07455070	0,07734095	0,07908135	0,10838000	0,2164075
Percentil 75%	0,08196328	0,08920692	0,09399933	0,12908750	0,2304983
Percentil 100%	0,11519700	0,12920900	0,14139100	0,17158800	0,2664280
Variância	0,0001672737	0,0004009403	0,0004595815	0,0006987406	0,0006362098
CV	17,25157	26,22638	26,97064	24,047	11,71275
Amplitude	0,0685145	0,1004583	0,0972249	0,1132064	0,107172

Tabela 7: Estatísticas descritivas para o agrupamento 6.

Informação	Tempo (meses)				
	3	6	9	12	15
Média	0,09048901	0,09902282	0,1094773	0,1687147	0,302245
Percentil 0%	0,05423320	0,05980570	0,0728636	0,1077930	0,2621750
Percentil 25%	0,08413455	0,09113073	0,1007865	0,1367195	0,2797142
Percentil 50%	0,09083245	0,10120200	0,1071035	0,1752640	0,2929580
Percentil 75%	0,10029650	0,10807725	0,1213583	0,1911760	0,3282030
Percentil 100%	0,10670200	0,12554300	0,1449510	0,2269440	0,3777140
Variância	0,000135493	0,0002420093	0,0003165014	0,001081354	0,0009966537
CV	12,8636	15,71016	16,25039	19,49086	10,44511
Amplitude	0,0524688	0,0657373	0,0720874	0,119151	0,115539

plot, uma amostra do método é apresetado na Figura 30) assegurando a boa representação para cada uma desses agrupamentos através da distribuição normal. Nas tabelas 8 e 9 são apresentados os ajustes dos parâmetros para distribuições normais para cada um dos casos através do método de máxima verossimilhança.

A partir desse resultado é possível inferir as coleções de parâmetros que levam uma região ao descontrole sobre a infecção. Mesmo assim, o método não se mostra prático

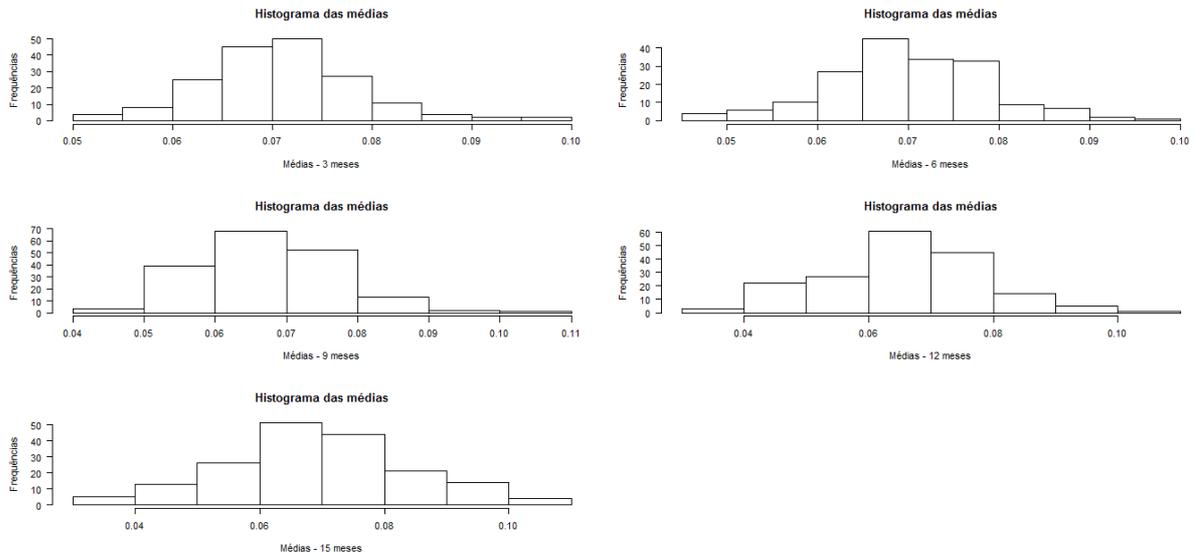


Figura 24: Histograma para as médias de infecção do agrupamento 1.

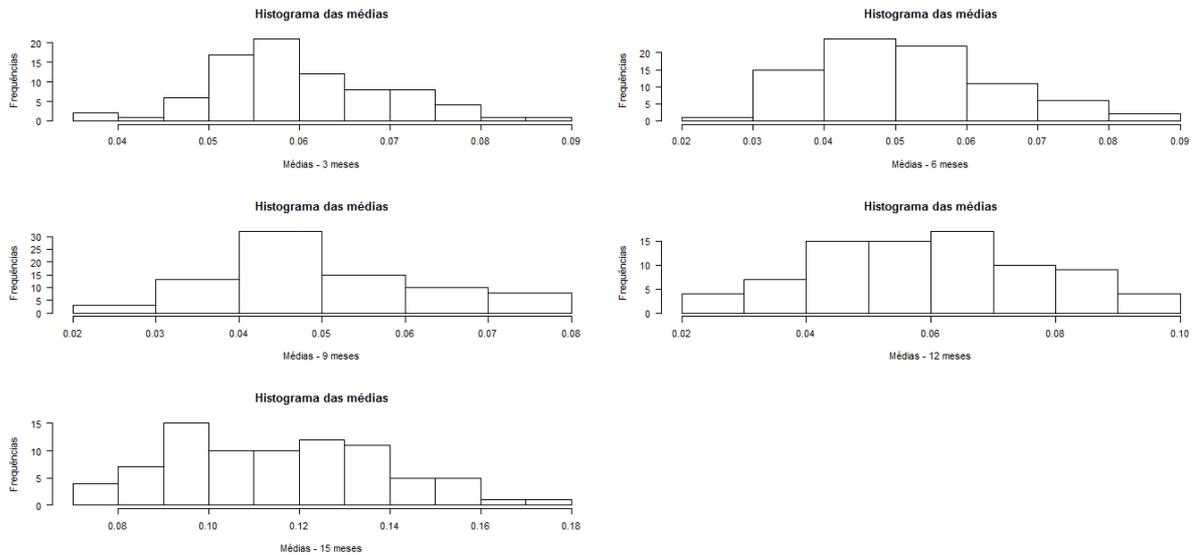


Figura 25: Histograma para as médias de infecção do agrupamento 2.

na tomada de decisão se estiver em um cenário qualquer na região em estudo, sendo necessário examinar mais profundamente o resultado a fim de obter critérios para uma intervenção para que seja possível prever o cenário futuro (e mais importante, modificar os parâmetros de forma a não se chegue ao resultado futuro que seja indesejável). Em outras palavras, busca-se um esquema que mediante a informação atual de campo seja possível antecipar o cenário à frente, e a partir desse esquema, seja possível identificar qual parâmetro merece maior/menor atenção para que o nível de infecção futuro seja modificado.

Para esse fim, será utilizada a árvore de decisão, um diagrama em forma de esquema de árvore usado para determinar um curso de ação ou mostrar uma probabilidade estatística. Cada ramo da árvore de decisão representa uma possível decisão ou ocorrência.

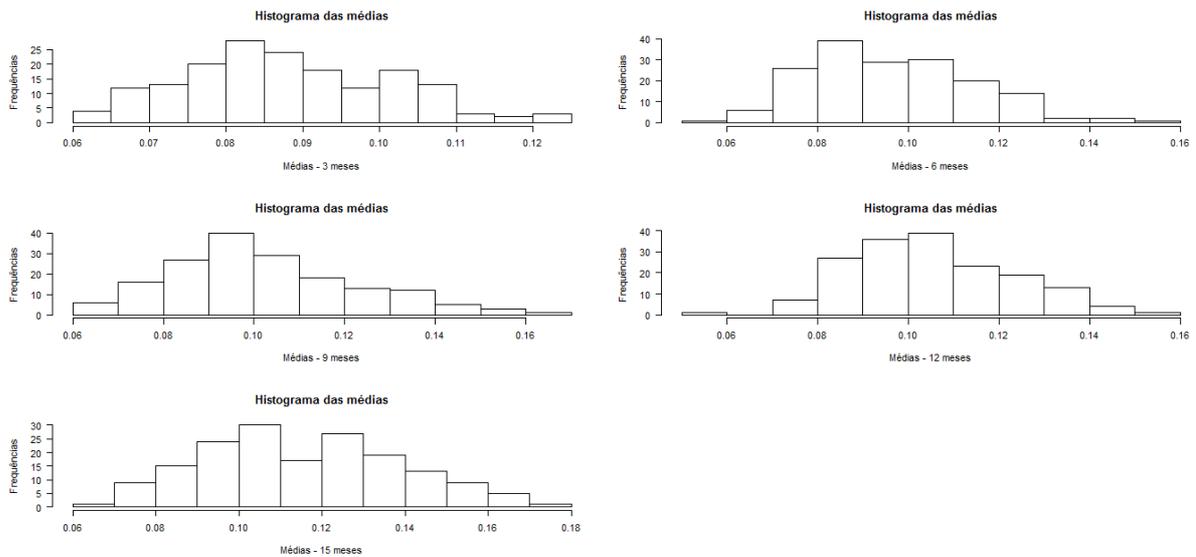


Figura 26: Histograma para as médias de infecção do agrupamento 3.

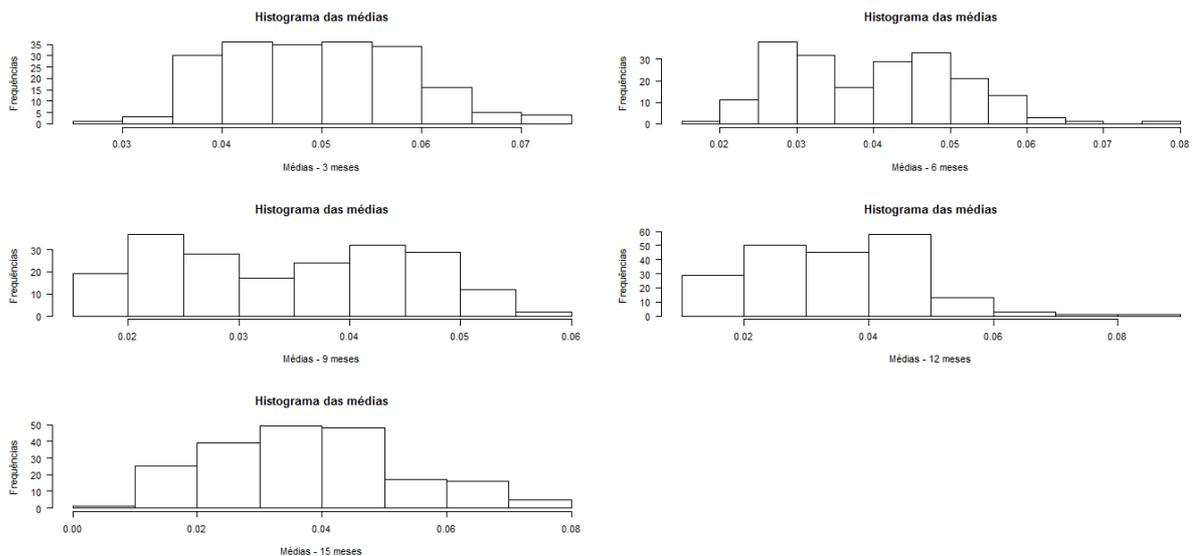


Figura 27: Histograma para as médias de infecção do agrupamento 4.

A estrutura em árvore mostra como uma escolha leva para o próximo, e o uso de ramos indica que cada opção é mutuamente exclusivo.

5.2 Árvore de decisão

O método de reconhecimento de padrões adotado a seguir visa definir as decisões para os valores dos parâmetros, com a finalidade de mudar um quadro futuro e/ou preservar uma evolução desejada no cenário de infecção da região. Para este trabalho foi utilizado o algoritmo (C4.5) as premissas em que este algoritmo se baseia são (Quilan, 1992):

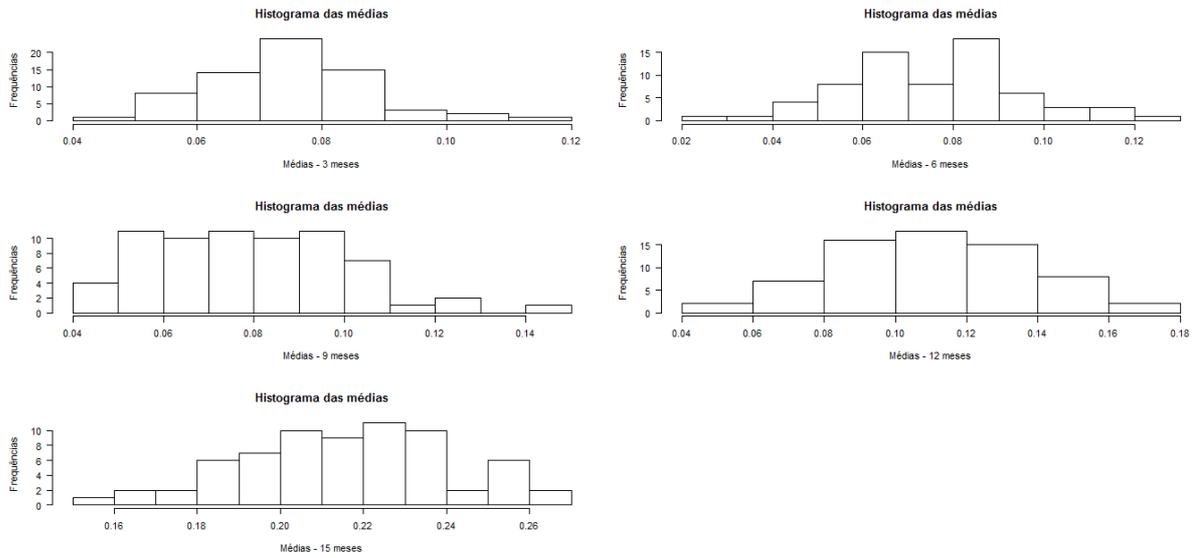


Figura 28: Histograma para as médias de infecção do agrupamento 5.

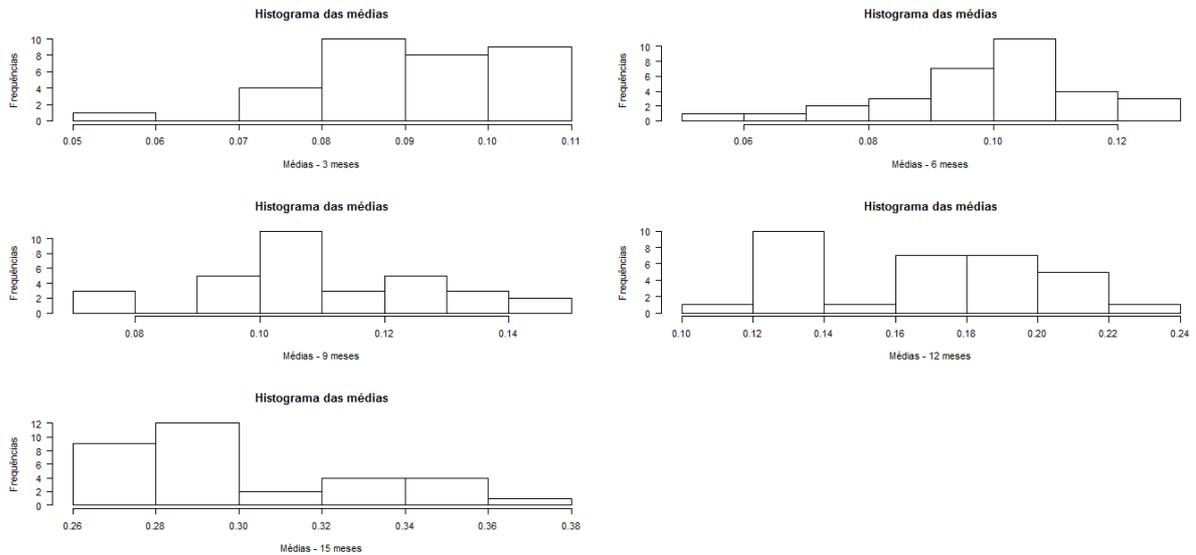


Figura 29: Histograma para as médias de infecção do agrupamento 6.

- Se todos os casos estão em uma mesma classe, então a árvore construída terá uma única folha e esta folha indicará a classe em que todos os casos se encontram;
- Para cada atributo, calcula-se a informação potencial obtida pelo teste do atributo (baseado na probabilidade de cada caso mantendo um valor particular do atributo). Também se calcula o ganho de informação que resulta de um teste sobre o atributo (baseado na probabilidade de cada caso com um valor particular sendo de uma classe particular);
- Dependendo do critério de seleção adotado, encontra-se o melhor atributo para ramificar.

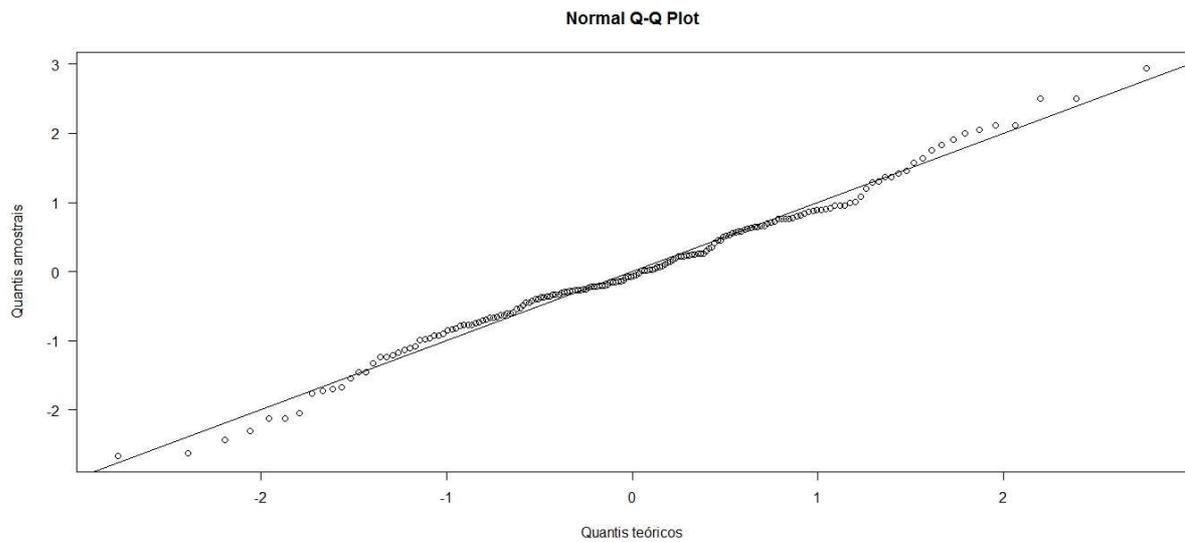


Figura 30: Q-Q plot para o agrupamento 1 na simulação para 6 meses.

Tabela 8: Ajuste dos parâmetros para distribuição normal.

Tempo (meses)	Agrupamento 1	
	Média	Desvio Padrão
3	0,0707525174	0,0078501048
6	0,0700225635	0,0089269377
9	0,0674060298	0,0098398670
12	0,0657576247	0,0128083654
15	0,0694024927	0,0146262668
Tempo (meses)	Agrupamento 2	
	Média	Desvio Padrão
3	0,0599685185	0,0098230482
6	0,0513727321	0,0125155353
9	0,049553465	0,012928802
12	0,060477456	0,018310465
15	0,115183274	0,022951466
Tempo (meses)	Agrupamento 3	
	Média	Desvio Padrão
3	0,088144389	0,013269765
6	0,096534871	0,018193794
9	0,101975362	0,020801210
12	0,1052162818	0,0178274515
15	0,115445904	0,023378209

Para o calcular a informação, usa-se entropia, isto é, a medida de incerteza sobre os dados. Seja $Y = \{y_1, y_2, \dots, y_n\}$ um conjunto de dados observados separadamente em classes, a entropia dos dados Y é dada por:

$$Entropia(Y) = - \sum_{j=1}^n \frac{|y_j|}{|y|} \log \frac{|y_j|}{|y|} \quad (5.3)$$

Tabela 9: Ajuste dos parâmetros para distribuição normal.

Tempo (meses)	Agrupamento 4	
	Média	Desvio Padrão
3	0,0496230120	0,0088282310
6	0,0398040940	0,0109769352
9	0,0344157055	0,0110571607
12	0,0345088440	0,0129015589
15	0,0381758575	0,0149499884
Tempo (meses)	Agrupamento 5	
	Média	Desvio Padrão
3	0,074969631	0,012837984
6	0,076348660	0,019875716
9	0,079485879	0,021279638
12	0,109925131	0,026238616
15	0,215348235	0,025037049
Tempo (meses)	Agrupamento 6	
	Média	Desvio Padrão
3	0,090489009	0,011456825
6	0,099022819	0,015311645
9	0,109477319	0,017510303
12	0,168714687	0,032366053
15	0,302245000	0,031072629

em que n define a quantidade de classes em estudo, $|y_j|$ a quantidade de vetores na classe j , y é o vetor de médias e $|y|$ é a quantidade total de vetores. Já a informação condicional a uma classe particular j é dado por:

$$Entropia(j|Y) = -\frac{|y_j|}{|y|} \log \frac{|y_j|}{|y|} \quad (5.4)$$

Assim, o ganho de informação é dado por:

$$Ganho(j|Y) = Entropia(y) - Entropia(j|y) \quad (5.5)$$

Busca-se maximizar o ganho de informação. Neste trabalho foi usado o ganho de informação normalizado (Quilan, 1986) definido por:

$$GanhoNormalizado(j|Y) = \frac{Entropia(y) - Entropia(j|y)}{Entropia(Y)} \quad (5.6)$$

o procedimento é repetido a cada avanço sobre a criação da árvore e avaliado a cada ramo o maior ganho de informação para tomar a decisão.

Após essa etapa, é feita a poda na árvore. Isto é, depois de feita a árvore que deve classificar todas as instâncias no conjunto de treinamento, esta é podada. Essa atitude visa reduzir o erro de classificação oriundo dos *outliers* (pequeno conjunto de instâncias que não são bem definidas e difere dos demais em sua vizinhança). Tal atitude torna a

árvore final (árvore podada) mais geral evitando o super ajustamento.

As vantagens e desvantagens do método são listadas a seguir:

- Vantagem 1 - As árvores de decisão são auto-exploratórias e quando compactas são fáceis de seguir, o que facilita sua compreensão por pessoas de outra área;
- Vantagem 2 - Flexibilidade em se trabalhar com atributos (classes) nominais e numéricas;
- Vantagem 3 - A representação da árvore é ideal para classificadores discretos;
- Vantagem 4 - Habilidade de se trabalhar com dados que podem conter erros;
- Vantagem 5 - Habilidade de se trabalhar com dados censurados;
- Vantagem 6 - É considerado um método não paramétrico, isto é, não há suposições sobre a distribuição do espaço e a estrutura do classificador;
- Desvantagem 1 - Alguns algoritmos (como ID3 e C4.5) requerem um atributo alvo que assume apenas valores discretos;
- Desvantagem 2 - A árvore utiliza o método de "dividir para conquistar", o que tende a ter boa performance se houver poucos atributos altamente significantes;
- Desvantagem 3 - É altamente sensível ao conjunto de treinamento para atributos irrelevantes e ruído (Quilan, 1992).

Na Figura 31 é apresentada a árvore de decisão com relação ao atributo agrupamento. É possível perceber as legendas V1, V3, V4, V5 que são, nessa ordem: Taxa de infecção humana, Taxa de recuperação humana, Taxa de infecção do molusco e Taxa de crescimento do molusco. Esses foram os atributos significativos que tem a possibilidade de modificar o agrupamento final que a coleção de parâmetros faz o ambiente convergir. Ainda, notamos que a taxa de crescimento do molusco é a primeira decisão que deve ser tomada quanto ao controle da infecção, na análise, o crescimento ser menor que 0,75 (isto é, taxa de crescimento baixa ou média) fará com que se siga para os ramos direita (se for maior) ou esquerda (se for menor). Apenas com esse controle já limitamos os casos no ramo da esquerda em agrupamento 1 com 168 casos (estabilidade na infecção), agrupamento 3 com 146 casos (crescimento linear na infecção) e agrupamento 4 com 166 casos (decaimento no nível de infecção). Já os demais, com essa escolha os casos presentes do agrupamento 2 foram 4 casos, do agrupamento 5 foram 2 casos e do agrupamento 6 não houveram casos.

Se continuarmos com as escolhas ainda no ramo esquerdo teremos que definir o nível do parâmetro de recuperação humana: Se for baixo, seguiremos para o ramo esquerdo

e se for médio ou alto para o ramo direito. Note que ao se seguir para o lado esquerdo novamente os agrupamentos atingidos são 1 com 55 ocorrências e 3 com apenas 104 ocorrências e resíduos sobre os agrupamentos 2 com 1 caso e agrupamento 5 com 2 casos. Seguindo o ramo direito seguem como opção de agrupamento dominantes 1,3 e 4; tendo 113 ocorrências para o agrupamento 1, 42 ocorrências para o agrupamento 3, 166 ocorrências para o agrupamento 4 e os resíduos 3 ocorrências para o agrupamento 2 e nenhuma ocorrência para os agrupamentos 5 e 6.

Para continuar a análise no ramo esquerdo a um terceiro nível devemos examinar a taxa de infecção humana que for baixo se segue para o ramo da esquerda e se for médio ou alto se segue para o ramo da direita. Seguindo pela direita no terceiro nível se obtém 53 ocorrências do agrupamento 1 e 1 ocorrência do agrupamento 2 definindo assim que (CreMol=Baixo ou médio, RecHum=Baixo e InfHum=Baixo) resultará num cenário de agrupamento 1 que tende a estabilidade. Seguindo para a esquerda, isto é, com infecção humana média ou alta terá 104 ocorrências do agrupamento 3, 2 ocorrências do agrupamento 1 e 2 ocorrências do agrupamento 5 indicando que (CreMol=Baixo ou médio, RecHum=Baixo e InfHum=Médio ou alto) resultará num cenário de agrupamento 3 que tende a um crescimento linear no nível da infecção ao longo do tempo.

Por outro lado, se no segundo nível fosse optado por uma taxa de recuperação humana entre médio e alto teremos dois ramos a seguir novamente sob uma avaliação através do parâmetro de infecção humana, se baixo seguindo o ramo da esquerda e se médio ou alto seguindo o ramo da direita. Optando por médio ou alto teremos 107 ocorrências do agrupamento 4 e 1 caso do agrupamento 1 definindo que (CreMol=Baixo ou médio, RecHum=Médio ou alto e InfHum=Médio ou alto) resultará num cenário de agrupamento 4 que tende a um decaimento do nível de infecção. Já se for escolhido uma taxa de infecção humana baixa é levado novamente a uma avaliação quanto à taxa de recuperação humana, se média tomando o ramo da esquerda e se alta tomando da direita. Seguindo o ramo da esquerda, volta a decisão em função da taxa de infecção humana, sendo ela média ocorreram 53 casos do agrupamento 1 e 1 caso do agrupamento 4; se ela for alta, ocorreram 11 casos do agrupamento 1, 1 caso do agrupamento 2 e 42 casos para o agrupamento 3 definindo que (CreMol=Baixo ou médio, RecHum=Médio e InfHum=Médio) resultará num cenário de agrupamento 1 e (CreMol=Baixo ou médio, RecHum=Médio e InfHum=Alto) resultará num cenário de agrupamento 3. Em contrapartida, definindo a taxa de recuperação em alta e a taxa de infecção média teremos 48 casos do agrupamento 1, 1 caso do agrupamento 2 e 5 casos do agrupamento 4; e se a taxa de infecção for alta se obtém 1 caso do agrupamento 2 e 53 casos do agrupamento 4, assegurando que (CreMol=Baixo ou médio, RecHum=Alto e InfHum=Médio) resultará num cenário de agrupamento 1 e (CreMol=Baixo ou médio, RecHum=Alto e InfHum=Alto) resultará num cenário de agrupamento 4.

A análise se faz de forma análoga no ramo direito a partir do primeiro nível, obtendo (CreMol=Alto, RecHum=Alto e InfHum=Baixo) haverá 38 casos do agrupamento 2, 3 casos do agrupamento 3, 1 caso do agrupamento 1, 1 caso do agrupamento 4 e 11 casos do agrupamento 5, definindo que essa configuração converge o cenário para um crescimento não linear do nível de infecção; e em (CreMol=Alto, RecHum=Médio e InfHum=Médio) apresenta 22 casos do agrupamento 2 e 5 casos do agrupamento 4 assegurando que essa configuração também converge para um crescimento não linear no nível de infecção; na configuração (CreMol=Alto, RecHum=Alto e InfHum=Médio) temos apenas 27 casos do agrupamento 4 que apresenta decaimento no nível da infecção do cenário.

Se for seguido o outro ramo tem-se as configurações (CreMol=Alto, RecHum=Médio ou Alto, InfHum=Baixo e InfMol=Médio ou Baixo) há 7 casos do agrupamento 1 e 2 casos do agrupamento 2, garantindo que o esquema converge para uma estabilidade; a configuração (CreMol=Alto, RecHum=Médio ou Alto, InfHum=Baixo e InfMol=Alto) tem-se 1 caso de agrupamento 1, 8 casos de agrupamento 2 e 9 casos de agrupamento 3 assegurando que o esquema tende a um crescimento linear; a configuração (CreMol=Alto, RecHum=Médio ou Alto, InfHum= Médio) tem-se 1 caso de agrupamento 1, 6 casos de agrupamento 3 e 20 casos de agrupamento 5 mostrando que este esquema leva o cenário a um crescimento alarmante do nível de infecção. Já a configuração (CreMol=Alto, RecHum=Baixo ou Médio, InfHum= Alto e InfMol= Baixo ou Médio) apresenta 7 casos do agrupamento 2, 1 caso do agrupamento 4 e 1 caso do agrupamento 5 garantindo que a configuração leva a um crescimento não linear no nível de infecção do cenário; a configuração (CreMol=Alto, RecHum=Baixo ou Médio, InfHum= Alto e InfMol= Alto) tem apenas 18 casos de agrupamento 5 que leva o cenário a um crescimento alarmante do nível de infecção; a configuração (CreMol=Alto, RecHum=Alto, InfHum= Alto e InfMol= Baixo e Médio) apresenta 6 casos de agrupamento 3, 1 caso de agrupamento 6 e 11 casos de agrupamento 5, o que leva novamente a uma cenário com nível de infecção alarmante. Por fim, a configuração (CreMol=Alto, RecHum=Alto, InfHum= Alto e InfMol= Alto) apresenta 5 casos de agrupamento 5 e 31 casos de agrupamento 6 que indica um crescimento sem controle no cenário em estudo, e assim, modelando o pior caso.

Diante dessa abordagem, pode-se montar estratégias para modificar um cenário atual. Por exemplo, feita uma avaliação de campo foi constatado que os níveis dos parâmetros: crescimento do molusco, recuperação humana, infecção humana e infecção do molusco estavam com configurações (Alta, Média, Alta, Média) indicando que o cenário em campo convergirá, se permanecer no controle que está, a um crescimento não linear no nível de infecção. Daí, estando com crescimento não linear, as opções para melhoramento são: crescimento linear (o que seria uma sensível melhora na evolução da infecção) que é agru-

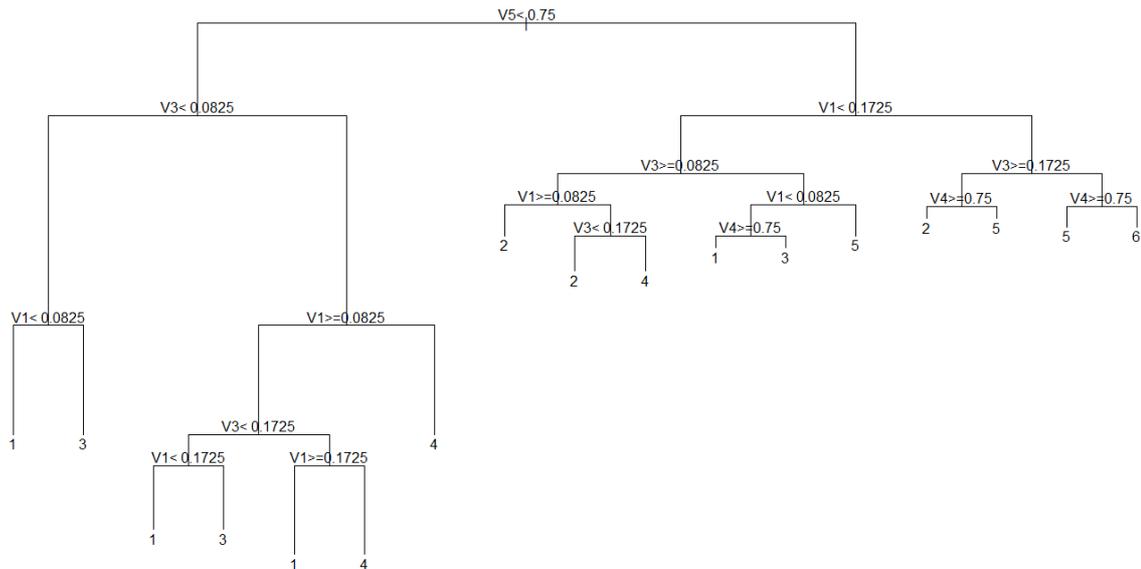


Figura 31: Árvore de decisão para os agrupamentos.

pamento 3, estabilidade que é agrupamento 1 ou decaimento no nível de infecção que é agrupamento 4. Para que haja mudança do agrupamento 2 para o agrupamento o agrupamento 3 é necessário alterar para configuração (CreMol=Baixo ou médio, RecHum=Baixo e InfHum=Médio ou alto) que deve mudar a taxa de recuperação de médio para baixa (o que significaria um controle maior dos casos infectados e tratados para não haver ou diminuir possíveis reinfecções) e alterar a taxa de crescimento do molusco de alta para baixa ou média (através de coleta dos moluscos ou tratamento químico). Outras opções são (CreMol=Baixo ou médio, RecHum=Médio e InfHum=Alto), mudança de CreMol de alta para baixa ou médio; e (CreMol=Alto, RecHum=Médio ou Alto, InfHum=Baixo e InfMol=Alto), mudança de InfHum de alta para baixa e InfMol de média para alta.

Para fazer a mudança de agrupamento 2 para agrupamento 1 temos as configurações: (CreMol=Alto, RecHum=Médio ou Alto, InfHum=Baixo e InfMol=Médio ou Baixo), mudança em InfHum de alto para baixo; (CreMol=Baixo ou médio, RecHum=Baixo e InfHum=Baixo) mudança de CreMol de alto para baixo ou médio, RecHum de médio para baixo e InfHum de alta para baixa; (CreMol=Baixo ou médio, RecHum=Médio e InfHum=Médio) mudança de CreMol de alto para baixo ou médio, InfHum de alta para média; (CreMol=Baixo ou médio, RecHum=Alto e InfHum=Médio) mudança de CreMol de alta para baixo ou médio, RecHum de médio para alta, InfHum de alta para médio.

Por fim, para uma mudança do agrupamento 2 para o agrupamento 4 temos as configurações: (CreMol=Baixo ou médio, RecHum=Médio ou alto e InfHum=Médio ou alto) mudança de CreMol alta para baixo ou médio; (CreMol=Baixo ou médio, RecHum=Alto e InfHum=Alto) mudança de CreMol alta para baixo ou médio, RecHum de médio para alto; (CreMol=Alto, RecHum=Alto e InfHum=Médio) mudança de RecHum de médio para alto e

InfHum de alta para médio. A configuração de estratégias para os demais agrupamentos ocorrem de forma análoga.

Embora se tenha um esquema para elaborar estratégias, há a necessidade de se obter o valor dos parâmetros em campo. Para essa necessidade é apresentada uma alternativa de ajuste dos parâmetros com base em algoritmo genético a fim de aproximar a situação do ambiente e a partir disso elaborar a melhor estratégia para reverter o quadro de infecção.

5.3 Ajuste dos parâmetros

A estratégia adotada para ajustar os parâmetros do autômato celular faz uso de algoritmo genético para identificar a coleção de parâmetros que explica o nível de infecção do cenário. A abordagem pressupõe a medição do nível de infecção em dois momentos: O primeiro, o que será definido como ponto de controle para início das simulações, e o segundo, que será considerado o alvo e usado para aferir o modelo.

Em campo, a definição dos níveis de infecção iniciais, vem da utilização de questionários e de sua interpretação nas quantidades de indivíduos e caramujos que serão inseridos no modelo, bem como o nível de saneamento através de observação do ambiente, índice de chuvas e índice de infectados e/ou tratados no ambiente. Já o nível de infecção de controle é obtido através de nova avaliação nos mesmos moldes a fim de identificar o avanço no quadro após alguns meses.

Para o ajuste dos parâmetros, será adotado o algoritmo genético melhorado publicado em *Tuning of the structure and parameters of neural network using an improved genetic algorithm* de autoria de Frank H. F. Leung *et al.* (Leung, 2003).

O algoritmo 1 apresenta uma rotina melhorada destacando cada uma de suas etapas. Como primeiro passo deve-se inicializar a população. Foram utilizados valores reais variando entre 0 e o limiar de cada parâmetro definido na literatura e apresentado na seção 4.1. Foram utilizadas populações com 10 indivíduos.

$$taxa_i = X \cdot \frac{limiar_i}{100} \quad (5.7)$$

em que X é uma amostra aleatória de uma distribuição uniforme contínua entre 0 e 1 ($U_c(0, 1)$).

Em sequência, é necessário avaliar cada um dos indivíduos (coleção de parâmetros) da população. Para a avaliação, utiliza-se o nível de infecção de cada célula da malha do indivíduo, isto é, utilizada a configuração da malha inicial e é aplicada a coleção de parâmetros do indivíduo em análise, e por fim evoluída a quantidade de gerações que

levaram a malha inicial até a malha de controle. Essa abordagem é realizada para cada indivíduo da população no momento.

```

begin
   $\tau \leftarrow 0$            //  $\tau$ : número de iterações;
  Inicializa  $\mathbf{P}(\tau)$    //  $\mathbf{P}(\tau)$ : população para iteração  $\tau$  ;
  Avalia  $f(\mathbf{P}(\tau))$    //  $f(\mathbf{P}(\tau))$ : função de fitness ;
  while (Não atingir condição de parada) do
    begin
       $\tau \leftarrow \tau + 1$  ;
      Seleciona 2 pais  $p_1$  e  $p_2$  de  $\mathbf{P}(\tau - 1)$ ;
      Realiza a operação de crossover pelas expressões 5.10 a 5.13 ;
      Realiza a operação de mutation pela expressão 5.14 para gerar 3 filhos
       $nos_1, nos_2$  e  $nos_3$  ;
      // reproduz um novo  $\mathbf{P}(\tau)$ 
      if (Número aleatório  $< p_a$  //  $p_a$ : probabilidade de aceitação) then
        O filho de maior valor da função de fitness entre  $nos_1, nos_2$  e  $nos_3$ 
        substitui o indivíduo com o menor valor da função de fitness da
        população;
      else
        begin
          if ( $f(nos_1) >$  menor valor de fitness na população  $\mathbf{P}(\tau - 1)$ ) then
             $nos_1$  substitui o indivíduo com o menor valor de fitness;
          end
          if ( $f(nos_2) >$  menor valor de fitness na população  $\mathbf{P}(\tau - 1)$ ) then
             $nos_2$  substitui o indivíduo com o menor valor de fitness;
          end
          if ( $f(nos_3) >$  menor valor de fitness na população  $\mathbf{P}(\tau - 1)$ ) then
             $nos_3$  substitui o indivíduo com o menor valor de fitness;
          end
        end
      Avalia  $f(\mathbf{P}(\tau))$ 
    end
  end
end

```

Algorithm 1: Rotina do algoritmo genético melhorado (Leung, 2003).

A avaliação é realizada por:

$$A = \sum_{i=1}^{10} \sum_{j=1}^{10} (M_{inicial}(i, j) - M_{controle}(i, j))^2 \quad (5.8)$$

em que $M(i, j)$ é o valor da infecção da célula de localização (i, j) da malha. É possível perceber que cada célula pode assumir valor $0 \leq M(i, j) \leq 1$ o que garante $0 \leq A \leq 100$, daí, a avaliação se resume a apresentar o valor mais próximo de zero possível.

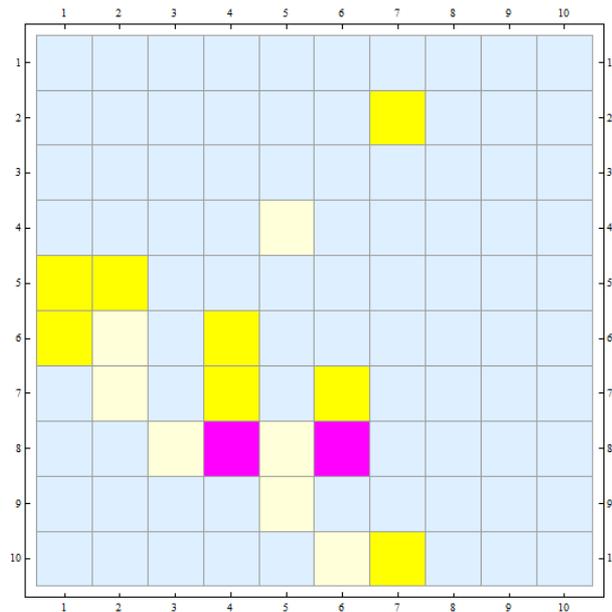


Figura 32: Ilustração da malha com as suas taxas de infecção convertidas em cores.

A Figura 32 apresenta uma malha com suas células apresentando diferentes níveis de infecção com tonalidades que vão de azul claro (infecção entre 0% e 20%) até preto (infecção entre 80% e 100%) (Silva, 2012).

A função de fitness utilizada para avaliar o desempenho de cada indivíduo da população é dada por:

$$Fitness = \frac{1}{1 + A} \quad (5.9)$$

dessa forma, ao minimizar-se o valor de A (O que leva a malha de infecção do indivíduo a um quadro de infecção próximo ao da malha de controle) ocorrerá a maximização da função de fitness, $0 \leq Fitness \leq 1$. Sendo assim, o algoritmo visará o valor de fitness igual a 1 o que resultaria em $A = 0$ e um perfeita previsão entre a malha do melhor indivíduo e a malha de controle.

Em sequência, é realizada a seleção de dois indivíduos, T_a e T_b , da população inicialmente avaliada através do procedimento de roleta (Eiben, 2008) e dado início ao procedimento de cruzamento, seja T_i o indivíduo i (coleção de parâmetros) da população:

$$C_1 = \frac{T_a + T_b}{2} = \frac{[t_{a1}, t_{a2}, \dots, t_{a9}] + [t_{b1}, t_{b2}, \dots, t_{b9}]}{2} = \left[\frac{t_{a1} + t_{b1}}{2}, \frac{t_{a2} + t_{b2}}{2}, \dots, \frac{t_{a9} + t_{b9}}{2} \right] \quad (5.10)$$

em que t_i é a média aritmética entre as taxas i dos indivíduos selecionados no passo anterior, $0 \leq i \leq 9$;

$$C_2 = t_{MAX} \cdot (1 - p) + \max([t_{a1}, t_{a2}, \dots, t_{a9}], [t_{b1}, t_{b2}, \dots, t_{b9}]) \cdot p \quad (5.11)$$

em que t_{MAX} é a coleção dos valores de limiar máximo para cada parâmetro avaliado e p é o peso associado, que em questão foi utilizado o valor $p = 0,3$;

$$C_3 = t_{MIN} \cdot (1 - p) + \min([t_{a1}, t_{a2}, \dots, t_{a9}], [t_{b1}, t_{b2}, \dots, t_{b9}]) \cdot p \quad (5.12)$$

em que t_{MIN} é a coleção dos valores de limiar mínimos para cada parâmetros avaliado;

$$C_4 = \frac{(t_{MAX} + t_{MIN})(1 - p) + ([t_{a1}, t_{a2}, \dots, t_{a9}] + [t_{b1}, t_{b2}, \dots, t_{b9}]) \cdot p}{2} \quad (5.13)$$

uma vez gerados os quatro candidatos a indivíduo de cruzamento, é feita a avaliação de cada um dos candidatos e é escolhido aquele que apresentar maior valor de fitness.

Após escolhermos o indivíduo filho gerado pelo cruzamento, T_c , é dado início ao processo de mutação:

$$T_{mi} = T_c + T_d = [t_{c1}, t_{c2}, \dots, t_{c9}] + [b_1 \Delta_1, b_2 \Delta_2, \dots, b_9 \Delta_9] \quad (5.14)$$

em que b_j , $1 \leq j \leq 9$, assume valor 0 ou 1. Δ_k , $1 \leq k \leq 9$ são números aleatoriamente gerados tais a soma do parâmetro t_k e o Δ_k podem assumir valores entre o limiar inferior e superior do parâmetro k . Assim, são geradas 3 mutações: A primeira, com apenas um b_j assumindo valor 1 e todos os demais igual a 0 (a escolha de qual assumirá o valor 1 é dada através de uma amostra aleatória de uma distribuição uniforme discreta, $U_d(1, 9)$); a segunda, alguns dos b_j escolhidos para assumir valor 1 e os demais valor 0; a terceira, todos os b_j assumem valor 1. Por fim, cada uma das três mutações são avaliadas através da função de fitness.

Em seguida, uma amostra aleatória de uma distribuição uniforme contínua entre 0 e 1 é testada contra a probabilidade de aceitação (que para o trabalho foi definida como 10%), se a amostra for menor que a probabilidade de aceitação, então T_{mi} , $1 \leq i \leq 3$ com maior valor de fitness irá substituir o indivíduo com pior valor de fitness na população. Se a amostra for maior ou igual do que a probabilidade de aceitação, então cada T_{mi} é testado contra o pior indivíduo da população, substituindo-o se tiver maior valor de fitness.

O procedimento se repete a cada geração do algoritmo. Note ainda, que a quantidade de passos do automato é fixa em 24 meses e difere da geração do algoritmo que foi fixado em: 5.000, 7.500 e 10.000 gerações.

A Tabela 10 apresenta os resultados para o melhor indivíduo para as gerações 5.000, 7.500 e 10.000. É possível perceber a comparação entre o ajuste inicial dos parâmetros e o valor do fitness em comparação aos ajustes encontrados nas gerações citadas, ao passo que mesmo aumentando o número de gerações o melhoramento na função de fitness se mostra sem alterações significativas.

O procedimento se mostra interessante para ajustar os parâmetros para um estudo preliminar e para uma tomada de decisão fazendo uso da árvore de decisão na tentativa de identificar os parâmetros mais significativos para focar, e daí, melhorar o cenário do ambiente estudado.

Tabela 10: Ajuste dos parâmetros para o autômato celular

Controle	Parâmetros								Fitness		
	0,23	0,23	0,23	0,26	0,15	0,02	0,20	0,45	0,05	N/A	N/A
Inicial	0,0232355	0,10218	0,179976	0,0450214	0,0890116	0,0169945	0,041819	0,0849354	0,048673	0,040692	23,574855
5.000 gerações	0,177529	0,164081	0,014614	0,181132	0,245337	0,017830	0,248747	0,209564	0,138794	0,107153	8,332450
7.500 gerações	0,171166	0,172342	0,014104	0,181152	0,245321	0,017412	0,247449	0,209721	0,133865	0,107778	8,278331
10.000 gerações	0,178836	0,20681	0,0142612	0,19938	0,243043	0,0183966	0,254661	0,210592	0,135143	0,109092	8,166575

O uso do algoritmo genético vem para automatizar o processo e retirar a subjetividade da seleção dos parâmetros. Enquanto que o pesquisador precisaria ter um conhecimento profundo quanto a implicação de cada componente do modelo baseado em autômatos celulares, o que demandaria conhecimento do modelo afim de conhecer a função de cada parâmetro e conhecimento na área para saber o valor utilizado para que o modelo se comportasse conforme a avaliação de campo sem que o valor do parâmetro fugisse a média conhecida na literatura.

Sendo assim, a aplicabilidade do ajuste feito através do AG elimina esse procedimento uma vez que a população otimizada ao final da simulação garantirá as coleções de parâmetros que mimetizam o comportamento encontrado em campo, sendo por fim apenas necessária a conferência quanto aos valores apresentados para cada parâmetro com relação às médias consagradas na literatura.

Na prática, é uma solução imediata ainda em campo logo após o processamento dos questionários da segunda visita ao local de estudo ou único questionário aplicado e conhe-

cidos os valores de referência para a região. Diante disso, há economia de capital humano para execução da tarefa, financeiro devido a diminuição da quantidade de campanhas de coleta de informação de 2 ou até 1 visita, de acompanhamento já que é sempre possível voltar a campo e reajustar o modelo na tentativa de simulações mais próximas ao cenário real.

6 Conclusão

Nosso objetivo era a modelagem da dinâmica de infecção do hemoparasita *S. mansoni*, através do uso de modelo baseado autômatos celulares. Para tanto, foi necessário o conhecimento prévio das particularidades do ciclo do trematodo, bem como sua forma de contato e o meio em que tal contato ocorre. Diante dessas informações foi modelada a dinâmica da população de um sítio artificial com características semelhantes às encontradas em nosso litoral. Foram realizadas simulações com três níveis para cada parâmetros: baixo, médio e alto; de acordo com a amplitude que cada parâmetro é compreendido na literatura corrente e analisado o nível de infecção da região para 5 passos temporais bem definidos. Foram obtidos os mais diversos cenários, desde quadros de diminuição do nível de infecção do sítio, até cenários em que o nível de infecção subiu na ordem de 334 %.

Através do presente trabalho foi possível identificar os pontos marcantes que contribuem para uma atenuação do quadro de infecção de uma dada região. Com base em coleções de parâmetros é provável a diminuição do nível de infecção sob um intervalo de tempo curto, essa é uma forte estratégia no combate a regiões endêmicas.

Conseguimos constatar que as diversas combinações de parâmetros estudados se converteram em seis casos bem característicos no avanço da infecção: Crescimento exponencia 1, crescimento exponencial 2, crescimento não linear, crescimento linear, estabilidade e decaimento. Essas coleções de parâmetros, apresentam uma ferramenta de controle comparativo com os parâmetros estimados no local do estudo.

Posterior a essa análise, foi realizado um aprofundamento quanto aos parâmetros que tinham maior significância na determinação do avanço da infecção. Para isso, utilizamos a árvore de decisão para determinarmos tais parâmetros: Taxa de infecção humana, Taxa de recuperação humana, Taxa de infecção do molusco, Taxa de crescimento do molusco. Através do quadro apresentado, o pesquisador terá total conhecimento sobre o cenário que apresenta maior probabilidade de ocorrer comparando os parâmetros de campo. Além disso, de posse da estimativa dos parâmetros de campo, poderá confrontar o quadro de infecção e através dessa informação identificar os principais parâmetros que merecem atenção para eliminar ou atenuar o quadro atual.

Para os parâmetros, através da coleta dos moluscos, questionários aplicados a popu-

lação do sítio em estudo e exames laboratoriais, é possível aproximar o nível de infecção da localidade. De posse dessa informação, e realizada nova consulta a população sob um intervalo de tempo definido e estimado o novo nível de infecção, foi possível estimar a coleção de parâmetros (ou uma coleção de parâmetros equivalentes que culminam num mesmo quadro) através de uma estratégia de otimização baseada em algoritmo genético.

Obtivemos **bons** resultados em sua estimação e validando como um método alternativo para diagnosticar em que classe de evolução se encontra a região estudada, pois, através da estimativa é possível definir a classe de evolução da infecção encontrada, e a partir desta determinar quais os parâmetros que foram mais importantes através da árvore de decisão, e por fim, montar estratégias de combate a esses parâmetros na tentativa de alterá-los, e daí, modificar o quadro da região.

A grande dificuldade nesse tipo de pesquisa se deve **a** grande quantidade de pessoas envolvidas na coleta do material para análise laboratorial, questionários e acompanhamento médico pós pesquisa assegurado em lei, bem como grande processamento de dados obtidos através de formulários, análises laboratoriais que são utilizados para definir os valores dos parâmetros. Sendo assim, qualquer iniciativa que venha a diminuir ou erradicar as visitas a campo é uma grande contribuição para a logística e efetiva execução de políticas públicas de saúde coletiva. Este trabalho não somente aborda uma **propostas** factível de execução e logística dos dados de campo para ajustar os parâmetros do modelo, como propõe um novo modelo baseado em autômatos celulares para previsão de cenários futuros e dessa forma poder antecipar a ação efetuada sobre cada componente modelado.

Diante disso, **considero proveitoso** o estudo realizado e bem sucedida a busca ao objetivo deste trabalho.

Trabalhos futuros visam utilizar dados reais de campos para confrontar com o modelo estimado. Isto é, de posse das informações de campo simular e ajustar os parâmetros e aplicá-los no modelo baseado em autômatos celulares e montar um cenário futuro e diante de nova pesquisa em campo confirmar a proximidade do cenário modelado com o que evoluiu na prática. Outra abordagem imediata, é a adequação do modelo para outros ciclos de infecção correlatos como malária ou leishmaniose afim de verificar o impacto no controle do nível de seus parâmetros na diminuição de contágio das enfermidades.

Referências

- ABBOUD, J. A., HANNA, L. S., RAGAB, H. A. A., **Experimental ocular schistosomiasis**. The british Journal of Ophthalmology, 55: 106-115, 1971.
- AL SALAMI, N. M. A., Ant Colony Optimization Algorithm, **UbiCC Journal**, v. 4, n. 3, p. 823-826, 2009
- AL-MAGALEH, B. M., SHAHBAZKIA, A.-M. H., A Genetic Algorithm for Discovering Classification Rules in Data Mining, **International Journal of Computer Applications**, v. 41, n. 18, p. 40-44, 2012.
- AL-SHAMERI, W. F. H., Dynamical Properties of the Hénon Mapping, **Int. Journal of Math. Analysis**, v. 6, n. 49, p. 2419-2430, 2012.
- ALANDER, J. T., On optimal population size of genetic algorithms, In: CompEuro '92 - COMPUTER SYSTEMS AND SOFTWARE. The Hague, Netherlands, **Proceedings**, 1992, p. 65-70.
- ALMEIDA, R. M., MACAU, E. N., Stochastic cellular automata model for wildland fire spread dynamics. **Journal of Physics: Conference Series**, 285 (2011).
- ALBUQUERQUE, J.O., A two years simulation using a real data cellular automaton. A predictive case study with the schistosomiasis expansion process along the coastline of Brazil, **Systema: connecting matter, life, culture and technology**, v. 1, n. 3, 2013.
- ANDRE, D., BRENNETT III, F. H., KOZA, J. R., Discovery by genetic programming of a cellular automata rule that is better than any known rule for the majority classification problem. In: **Proceeding GECCO '96 Proceedings of the First Annual Conference on Genetic Programming**, MIT Press Cambridge, 1996, p. 3-11.
- APPLEGATE, D. L.; BIXLY, R. E.; CHVÁTAL, V.; COOK, W. J., **The Traveling Salesman Problem: A Computational Study**, Princeton University Press, 2006, 606 p.
- ARABAS, J., MICHALEWICZ, Z., MULAWKA, J., GAVaPS-a genetic algorithm with varying population size, In: Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence. Orlando, FL, **Proceedings of the First IEEE Conference on Evolutionary Computation**, 1994, p. 73-78 v.1.
- ARIFOVIC, J., GENÇAY, R., Using genetic algorithms to select architecture of a feedforward artificial neural network, **Physica A**, v. 289, p. 574-594, 2001.
- BAKER, J. E., Reducing bias and inefficiency in the selection algorithm. In. GREFENS-TETTE, J. J. Eds., **Proceedings of the 2nd International Conference on Genetic Algorithms and Their Applications**. Lawrence Erlbaum, Hillsdale, New Jersey, 1987, p. 14-21.

- BARTOLOZZI, M., THOMAS, A., Stochastic cellular automata model for stock market dynamics. **Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)**, 69 (2004).
- BERAN, J., **Statistics for Long-Memory Processes**. New York: Chapman & Hall, 1994.
- BERTOLDO, J. G., ROCHA, F., COIMBRA, J. L. M., ZITTERELL, D., GRAH, V. F., Teste de comparação de médias: Dificuldades e acertos em artigos científicos, **R. Bras. Agrocência**, v. 13, n. 4, p. 441-447, 2007.
- BLICKLE, T., THIELE, L., A comparison of selection schemes used in genetic algorithms. **Tech. Rep. TIK Report 11**, December 1995, Computer Engineering and Communication Networks Lab, Swiss Federal Institute of Technology, 1995.
- BREMERMANN, H.J. Optimization Through Evolution and Recombination, In M.C. Yovits et al, Editors, **Self-Organizing Systems**, Spartan Books, Washington, DC: 1962, p. 93-106.
- BRUCE, J. I. et al. **Schistosoma mansoni: pulmonary phase of schistosomule migration studied by electron microscopy**. *Experimental Parasitology*, 35: 150-160, 1974.
- CADENAS, E., JARAMILLO, O. A., RIVERA, W., Analysis and forecasting of wind velocity in Chetumal, Quintana Roo, using the single exponential smoothing method, **Renewable Energy**, v. 35, n. 5, p. 925-930, 2010.
- CAMPBELL, G. et al., **Molecular evidence supports an African affinity of the Neotropical freshwater gastropod, *Biomphalaria glabrata*, Say 1818, an intermediate host for *Schistosoma mansoni***. *Proceedings of the Royal Society of London*, 267: 2. 351- 2.358, 2000.
- CARVALHO, O. S. (Org.), COELHO, P. M. Z. (Org.), LENZI, H. L. (Org.), **Schistosoma mansoni e esquistossomose: uma visão multidisciplinar.**, Rio de Janeiro, Editora Fiocruz, 1124 p., 2008.
- CHANG, J. H., IBARRA, O. H., VERGIS, A. On the power of one-way communication. **J. ACM**, v. 35, n. 3, p. 697-726, 1988.
- CHAIA, G., **Técnica para concentração de miracídios**. *Revista Brasileira de Malariologia e Doenças Tropicais*. 8: 355-357, 1956.
- CHEN, C.-F., HO, W.-H., CHOU, H.-Y., YANG, S.-M., CHEN, I.-T., SHI, H.-Y. C., Long-Term Prediction of Emergency Department Revenue and Visitor Volume Using Autoregressive Integrated Moving Average Model, **Computational and Mathematical Methods in Medicine**, v. 2011, p. 7, 2011.
- CHEN, Z., ZOU, L., ZHANG, W., RUAN, S., Mathematical modelling and control of schistosomiasis in Hubei Province, China., **Acta Trop.**, v. 115, n. 1-2, p. 119-125, 2010.
- CHERNIN, E.; ANTOLICS, V. M., **Penetrative capacity of *Schistosoma mansoni* miracidia**. *Journal of Parasitology*, 63: 560-561, 1975.

- CHUA, L. H. C., WONG, T. S. W., Runoff forecasting for an asphalt plane by Artificial Neural Networks and comparisons with kinematic wave and autoregressive moving average models, **Journal of Hydrology**, v. 397, n. 3, p. 191-201, 2011.
- COBBOLD, T.S. **On some new forms of Entozoa**. Transactions of the Linnean Society of London, 22: 363-366, 1859.
- COBBOLD, T.S., **Remarks on Dr. Harley's paper "Haematuria of the Cape of Good Hope"**. Lancet, i: 156-157, 1864.
- CODD, E. F., **Cellular automata**, 1st Edition, Academic Press, 1968, 122 p.
- COELHO, M. V., **Aspectos do desenvolvimento de formas larvárias de *Schistosoma mansoni* em *Australorbis nigricans***. Revista Brasileira de Biologia, 17: 325-337, 1957.
- COELHO, M. V. et al., **Parasito: *Schistosoma mansoni*** In: CUNHA, A. S. *Esquistossomose mansônica*. São Paulo: Sarvier, 1970.
- COLES, G. C., **The carbohydrate metabolism of larval *Schistosoma mansoni***. International Journal for Parasitology, 2: 341-352, 1972.
- CORONADO, A. V., CARPENA, P., Size effects on correlation measures, **Journal of Biological Physics**, v. 31, p. 121-133, 2005.
- COSTA, R. L.; VASCONCELOS, G. L. Long-range correlations and nonstationarity in Brazilian stock market. **Physica A**. v.329, n. 1-2, p. 231-248, 2003.
- CULIK, K., HURD, L. P., YU, S. Computation theoretic aspects of cellular automata. **Physica D**, v. 45, n. 1-3, p. 357-378, 1990.
- DAWKINS, R. **O relojoeiro cego. A teoria da evolução contra o desígnio divino**. Editora Companhia das Letras. 1991, 488p.
- DAVIS, G. M., **Snail hosts of Asian *Schistosoma* infecting man: evolution and coevolution**. In: BRUCE, J.I.; SORNMANI, S., The Mekong Schistosome. Malacological Review, supl. 2: 195-238, 1980.
- DAVIS, G. M., **Evolution of prosobranch snails transmitting Asian *Schistosoma*; coevolution with *Schistosoma*: a review**. Progress in Clinical Parasitology, 3: 145-204, 1992.
- DESPRÈS, L. et al., **Molecular evidence linking hominid evolution to recent radiation of schistosomes (*Platyhelminthes: Trematoda*)**. Molecular Phylogenetics and Evolution, 1: 295-304, 1992.
- DEJONG, R.J., et al., **Evolutionary relationships and biogeography of Biomphalaria (Gastropoda: Planorbidae) with implications regarding its role as host of the human blood fluke *Schistosoma mansoni***. Molecular Biology and Evolution, 18: 2.225 - 2.239, 2001.
- DE JONG, K. A., **An Analysis of the Behaviour of a Class of Genetic Adaptive Systems**, Tese (Doctor of Philosophy), University of Michigan, 1975.

- DEELDER, A. M. et al., **Detection of schistosome antigen in Mummies**. *Lancet*, 335: 724-725, 1990.
- EIBEN, A. E., AARTS, E. H. L., VAN HEE, K. M., Global convergence of genetic algorithms: a Markov chain analysis, In: SCHWEFEL, H. -P., MÄNNER, R., Eds., **Proceedings of the 1st Conference on Parallel Problem Solving from Nature n. 496 in Lecture Notes in Computer Science**, Springer, Berlin, Heidelberg, New York, 1991.
- EIBEN, A. E., SMITH, J. E., **Introduction to Evolutionary Computing** (Natural Computing Series), Springer, 2008, 315 p.
- FAUST, E. C.; HOFFMANN, W. A., **Studies on *Schistosomiasis mansoni* in Puerto Rico. I The extramammalian phases on life-cycle**. *Puerto Rico Journal of Public Health and Tropical Medicine*, 10: 1-47, 1934.
- FAN, W., FOX, E. A., PATHAK, P., WU, H., The Effects of Fitness Functions on Genetic Programming-Based Ranking Discovery For Web Search, **Journal of the American Society for Information Science and Technology**, v. 55, n. 7, p. 628-636, 2004.
- FEDER, J., **Fractals**. New York: Plenum Press, 1988.
- FILES, V.S., **A study of the vector-parasite relationships in *Schistosoma mansoni***. *Parasitology*, 41: 264-269, 1951.
- FOGEL, L. J., OWENS, A. J., WALSH, M. J., Artificial intelligence through a simulation of evolution. In: CALLAHAN, A., MAXFIELD, M., FOGEL, L.J., Eds., **Biophysics and Cybernetic Systems**. Spartan, Washington DC, p. 131-156, 1965.
- FOGEL, L. J., OWENS, A. J., WALSH, M. J., **Artificial Intelligence through Simulated Evolution**. Wiley, Chichester, UK, 1966.
- FOGEL, D. B., **Evolutionary Computation: The Fossil Record**, IEEE press, Piscataway, NJ, 1998, 656 p.
- FRANSES, P. H., Forecasting 1 to h steps ahead using partial least squares, **Econometric Institute**, Erasmus University Rotterdam, Econometric Institute Report 2006-47, Disponível em: <<http://repub.eur.nl/res/pub/8093/ei2006-47.pdf>>. Acesso em: 21 abr. 2013.
- FUJII, D. An account of a journey to Katayama. *Chugai Iji Shimpo*, 691: 55-56. Em japonês, tradução inglesa em Kean et al. (1978) *Tropical Medicine and Parasitology, Classic Investigations*. v. II. Cornell Univ. Press: Ithaca, 1847 (republicado em 1909).
- GABAIX, X., GOPIKRISHNAN, P., PLEUROU, V., STANLEY, H.E., A Theory of powerlaw distributions in financial market fluctuations, **Nature**, v. 423, p. 267 - 270, 2003.
- GABAIX, X., GOPIKRISHNAN, P., STANLEY, H.E., Institutional investors and stock market volatility, **Quarterly Journal of Economics**, v. 121, p. 461 - 504, 2006.
- GAREY, M. R., JOHNSON, D. S., **Computers and intractability: a guide to NP-completeness**. New York: H. Freeman, 1979.
- GÉRARD, Y. V., Simulating physics with cellular automata. **Physica D**, v. 10, p. 96-116, 1984.

- GIFANI, P., RABIEE, H. R., HASHEMI, M. H., et al., Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantification, **Journal of the Franklin Institute**, v. 344, p. 212-229, 2007.
- GOLDBERG, D. E., DEB, K., A Comparative Analysis of Selection Schemes Used in Genetic Algorithms, **Foundations of Genetic Algorithms**, p 69-93. 1991.
- GOPIKRISHNAN, P. et al., Scaling of the distribution of fluctuations of financial market indices. **Physical Review E**, v. 60, n. 5, p. 5305-5316, 1999.
- GRIFFITHS, R. B., **Further observations on the penetration of mammalian skin by the cercariae of *Schistosoma mansoni*, with special reference to the effects of mass invasion.**, Annals of Tropical Medicine and Parasitology, 47: 86-94, 1953.
- GUI, M. et al., ***Schistosoma japonicum* e *S. mansoni* comparison of larval migration patterns in mice.** Journal of Helminthology, 69: 19-25, 1995.
- HAAS, W. **Physiological analyses of cercarial behaviour.**, The Journal of Parasitology, 2: 243-255, 1992.
- HAUSDORFF, J. M. et al. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. **Journal of Applied Physiology**, v. 80, n. 5, p. 1448-1457, 1996.
- HE, X.-Z., ZHENG, M., Dynamics of moving average rules in a continuous-time financial market model, **Journal of Economic Behavior and Organization**, v. 76, n. 3, p. 615-634, 2010.
- HOLLAND, J. H., Genetic algorithms and the optimal allocation of trials, **SIAM J. of Computing**, v. 2, p. 88-105, 1973.
- HOLLAND, J. H., **Adaption in Natural and Artificial Systems**, MIT Press, Cambridge, MA, 1992, 211 p.
- ITURBE, J; GONZÁLEZ, E., **El Huésped Intermediario del *Schistosoma mansoni* em Venezuela.** Edição Especial. Caracas: Academia Nacional de Medicina, 1916.
- ITURBE, J; **Distribución y profilaxia de la *Bilharzia mansoni* en Caracas.** Gaceta Médica de Caracas, 24: 70, 1917a.
- ITURBE, J; **Anatomía de la cercaria del *Schistosomum mansoni*.** Gaceta Médica de Caracas, 24: 81, 1917b.
- IVANOV, P. Ch. Scaling behaviors of Heartbeat Intervals obtained by WaveletBased Time-Series Analysis. **Nature**, v.383, p.323-327, 1996.
- IVANOVA, K.; AUSLOOS, M. Application of the detrended fluctuation analysis (DFA) method for describing cloud breaking. **Physica A**, v. 274, n. 1, p.349-354, 1999.
- JOURDANE, J.; THERON, A.; COMBES, C., **Demonstration of several sporocysts generations as a normal pattern of reproduction of *Schistosoma mansoni*.** Acta Tropica, 37(2): 177-182, 1980.

- JURBERG, A., **Fotografias por microscopia a laser confocal, a imagem exhibe larvas do S. mansoni (miracídio)** Disponível em: <<http://www.fiocruz.br/ioc/cgi/cgilua.exe/sys/start.htm?infoid=511&sid=32&tpl=printerview>>. Acesso em: 27 jan. 2013.
- KALISKY, T., ASHKENAZY, Y., HAVLIN, S., Volatility of linear and nonlinear time series, **Physical Review E**, v.72, p. 011913-1 - 011913-8, 2005.
- KALBE, M.; HABERL, B.; HAAS, W. **Schistosoma mansoni Miracidial host-finding: species specificity of an Egyptian strain**. Parasitology Research, 82: 8, 1996.
- KARABOGA, D., BASTURK, B., A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, **Journal of Global Optimization**, v. 39, n. 3, p. 459-471, 2007.
- KENNEDY, J., EBERHART, R. Particle swarm optimization. In Proc. of the **IEEE Int. Conf. on Neural Networks**, Piscataway, NJ., 1995. p. 1942-1948.
- KOZA, J. R., **Genetic Programming**, MIT Press, Cambridge, MA, 1992, 840 p.
- KOZA, J. R., **Genetic Programming II**, MIT Press, Cambridge, MA, 1994, 768 p.
- KRISTIANSEN, T., Forecasting Nord Pool day-ahead prices with an autoregressive model, **Energy policy**, v. 49, p. 328-332, 2012.
- KUMAR, R., JYOTISHREE, Blending Roulette Wheel Selection & Rank Selection in Genetic Algorithms, **International Journal of Machine Learning and Computing**, v. 2, n. 4, p. 365-370, 2012.
- KRUGER, S. P. et al., **The route of migration of Schistosoma mattheei from lungs to the liver in sheep**. Journal of the South African Veterinary Medical Association, 40: 39-43, 1969.
- LAMEPE, **Laboratório de Meteorologia de Pernambuco (Lamepe)**. Acessado em: December 17, 2011, from <http://www.itep.br/index.php/laboratorios-lamepe/meteorologia-lamepe/links-lamepe>.
- LARRAÑAGA, P., KUIJPRES, C. M. H., MURGA, R. H., INZA, I., DIZDAREVIC, S., Genetic Algorithms for the Travelling Salesman Problem: A Review of Representations and Operators, **Artificial Intelligence Review**, v. 13, p. 129-170, 1999.
- LEUNG, F. H. F. **Tuning of the structure and parameters of a neural network using an improved genetic algorithm**, IEEE transactions on neural networks, V. 14, N. 1, pp. 79-88, 2003.
- LEE, J. M., KIM, D. J., KIM, I. Y., et al., Detrended fluctuation analysis of EEG in sleep apnea using MIT / BIH polysomnography data, **Computers in Biology and Medicine**, v. 32, p. 37-47, 2002.
- LEIPER, R.T.; ATKINSON, E.L., **Observation on the spread of Asiatic Schistosomiasis**. British Medical Journal, i: 201-203, 1915.
- LEWIN, B., **Genes VII**. Oxford University Press, New York, 2000, 990 p.

LEWIS, H., PAPADIMITRIOU, C., **Elements of the Theory of Computation**. Prentice Hall. 2a Edição. 1997.

LI, W., PACKARD, N. H., LANGTON, C. G., Transition phenomena in cellular automata rule space, **Physica D**, v. 45, n. 1-3, p. 77-94, 1990.

LINDENMAYER, A., Mathematical models for cellular interactions in development. parts I and II. **J. Theor. Biol.**, v. 18, p. 280-315, 1968.

LIU Y. et al. Quantification of correlations in economic time series. **Physica A**, v.245, p.437-440, 1997.

LIU, Y. et al., Statistical properties of the volatility of price fluctuations. **Physical Review E**, v. 60, n. 2, p. 1390-1400, 1999.

LIU, Y., CIZEAU, P., GOPIKRISHNAN, P., MEYER, M., PENG, C.-K., STANLEY, H.E., **Volatility studies of the S&P500 index in Econophysics: An Emerging Science**. In: Proc. 1997 Budapest Conference, edited by Y. Kertesz and I. Kondor, Kluwer, Dordrecht, 2000.

LIU, Y. (Eds.) **The Han Tombs of Mawangdui Changsha**. House, China: Human People's Publishing, 1978.

LIMA, L. M., **Fotografias em microscópio: Ovos de Schistosoma intercalatum** Disponível em: <<http://www.parasitologiaclinica.ufsc.br/index.php/info/conteudo/fotografias/ovos-sintercalatum/>>. Acesso em: 27 jan. 2013.

LUTZ, A. **Observações sobre a evolução do Schistosomum mansoni**. Nota prévia. Brazil-Medico, 30: 385-387, 1916.

LUTZ, A. **Observações sobre a evolução do Schistosomum mansoni**. Segunda nota prévia. Brazil-Medico, 31: 81-82, 89-90, 1917a.

LUTZ, A. **Observações sobre a evolução do Schistosomum mansoni**. Revista da Sociedade Brasileira de Ciências, 1: 41-48, 1917b.

LUTZ, A. **O Schistosomum mansoni e a schistosomatose segundo observações feitas no Brazil**. Memórias do Instituto Oswaldo Cruz, 11: 121-155, 1919.

MALDONADO, J.F.; ACOSTA-MATIENZO, J., **Evolution del Schistosoma mansoni dentro de su hosped intermediário, el caracol Australorbis glabratus**. Puerto Rico Journal of Public Health and Tropical Medicine, 22: 374-404, 1947.

MALKIEL, B. G., The efficient market hypothesis and its critics, **Journal of Economic Perspectives**, v. 17, n. 1, p. 59-82, 2003.

MARUOKA, A., KIMURA, M. Completeness problem of one-dimensional binary scope-3 tessellation automata. **J. Comput. Syst. Sci**, v. 9, n. 1, p. 31-47, 1974.

MANSON, P., **Report of a case of bilharzia from the West Indies**. The Journal of Tropical Medicine, 5: 384-385, 1902.

MANSON, P., **Tropical Diseases: a manual of the diseases of warm climates**. 3. ed. London: Cassell & Co, 1903.

- MARUOKA, A., KIMURA, M. Completeness problem of multi-dimensional tessellation automata. **Inf. Control**, v. 35, n. 1, 52-86, 1977.
- MATTOS NETO, P. S. G. ; SILVA, D. A. ; FERREIRA, T. A. E.; CAVALCANTI, G. D. C. ; Market volatility modeling for short time window. **Physica A**, v. 390, p. 3444-3453, 2011.
- McMULLIN, B., John von Neumann and the Evolutionary Growth of Complexity: Looking Backwards, Looking Forwards..., **Artificial Life**, v.6, n. 4, p. 347-361, 2000.
- MEHDI, S., MORTEZA, I., ARABSHAH, H., REZA, S. M., Monitoring of the genetic algorithm operators in application to the GaAs_{0.7}Sb_{0.3}/GaAs single quantum well nanostructure, **International Journal of Physical Sciences**, v. 7 (39), p. 5485-5492, 2012.
- MIRZAYOF, D., ASHKENAZY, Y., Preservation of long range temporal correlations under extreme random dilution, **Physica A**, v. 389, p. 5573-5580, 2010.
- MITCHELL, M., CRUTCHFIELD, J. P., DAS, R., Evolving Cellular Automata with Genetic Algorithms: A Review of Recent Work. In: **Proceedings of the First International Conference on Evolutionary Computation and Its Applications (EvCA'96)**, Russian Academy of Sciences, Moscow, Russia, 1996.
- MIYAGAWA, Y. **Über den Wanderungsweg des *Schistosmum japonicum* von der Haut bis zum Pfortadersystem, und Über die körperkonstitution der jüngsten würmer zur zeit der Hautinvasion.** Zentralblatt für Bakteriologie, I(66): 406-417, 1912.
- MILLER, R. M., WILSON, R. A., **Migration of schistosomula of *Schistosoma mansoni* from skin to lungs.** Parasitology, 77: 281-302, 1978.
- MONTROLL, E. W., SHLESINGER, M. F., The wonderful world of random walks. In: **Nonequilibrium Phenomena II. From Stochastics to Hydrodynamics.** Lebowitz, J. L., Montroll, E. W., eds. Amsterdam: North-Holland, 1984, pp. 1-121.
- MORETTIN, P. A., TOLOI, C. **Análise de Séries Temporais.** 2^a Edição. Brasil: Edgard Blucher, 2006. 564 p.
- MORITA, K. Computation-universality of one-dimensional one-way reversible cellular automata. **Inf. Process. Lett.**, v. 42, n. 6, p. 325-329, 1992.
- NEUMANN, J. V. (1936) Em BURKS, A. (Ed.), **Theory of self-reproducing Automata**, University Illinois Press, 1966. 403p.
- NIETO, P., MALONE, J.B., BAVIA, M.E., Ecological niche modeling for visceral leishmaniasis in the state of Bahia, Brazil, using genetic algorithm for rule-set prediction and growing degree day-water budget analysis, **Geospatial Health**, v. 8, n. 1, p. 115-126, 2006.
- OTTOLINA, C., **El milagro del *Schistosoma mansoni*. Anatomía, Citología, Fisiología.**, Revista de Sanidad y Asistencia Social, 22 (1-4): 1-411, 1957.
- PATTANTYÚS-ÁBRAHÁM, M.; KIRÁLY, A.; JÁNOSI, I. M. Nonuniversal atmospheric persistence: Different scaling of daily minimum and maximum temperatures. **Physical Review E**, United States, v. 69, n. 2, p. 021110-1-021110-7, 2004.

PALAN, S., **The Efficient Market Hypothesis and its Validity in Today's Markets**. 2004. 67 p., Dissertação de mestrado - Faculty of Social and Economic Sciences, Karl-Franzens University Graz, Áustria, 2004.

PAN, C., **Studies on the host-parasite relationship between *Schistosoma mansoni* and the snail *Australorbis glabratus***. American Journal of Tropical Medicine and Hygiene, 14: 931-976, 1965.

PAREDES, H., SOUZA-SANTOS, R., RESENDES, A.P., ALBUQUERQUE, J.O., BOCANEGRA, S., GOMES, E.C., BARBOSA, C.S., Spatial pattern, water use and risk level associated with the transmission of schistosomiasis on the north coast of Pernambuco, Brazil. **Cadernos de Saúde Pública** (ENSP. Impresso), v. 26,1013-1023, 2010.

PEARCE, E. J., MCLAREN, D. J., ***Schistosoma mansoni*: the cutaneous response to cercarial challenge in naive guinea pigs and guinea pigs vaccinated with highly irradiated cercariae**. International Journal for Parasitology, 16: 491-510, 1986.

PENG, C-K. et al. Mosaic organization of DNA nucleotides. **Physical Review E**, v. 49, n. 2, p. 1685-1989, 1994.

PENG, C-K. et. al., Quantification of scaling exponents and crossover heartbeat time series. **Chaos**, v. 5, n. 1, p. 82-87, 1995.

PESAVENTO, U. An Implementation of von Neumann's Self-Reproducing Machine, **Artificial Life**, v. 2, n. 4, p. 337-354, 1995.

PRATA, A.; ANDRADE, Z.A., **Chronic glomerulonephritis associated with hepatosplenic schistosomiasis**. Revista do Instituto de Medicina Tropical de São Paulo, 29: 162-167, 1987.

PIRAJÁ DA SILVA, M., **Cercaire brésilienne (*Cercaria blanchardi*) à queue bifurquée**. Archives de Parasitologie, 15: 398-400, 1912.

PIMENTEL, F. et al., **Comportamento elementar dp caramujo *Biomphalaria glabrata*. II. Localização e tamanho da massa cefalopodal exposta**. Revista Brasileira de Pesquisas Médicas e Biológicas, 9: 123-127, 1976.

PINTO, C., ALMEIDA, A. F., **Penetração das cercárias de *Schistosoma mansoni* na pele de *Canis familiaris* e do homem**. Revista Brasileira de Biologia, 5: 219-229, 1945.

QUILAN, J.R., C4.5: **Programs for machine learning**. Morgan Kaufmann, 1a Edição, 302 p., 1992.

QUILAN, J.R., **Induction of decision trees**, Machine Learning 1, 81-106, 1986.

RIDLEY, M. **Evolution**. 3rd edition, Blackwell Publishing, 2004, 751p.

SA, P. G., MAES, C., The Gacs-Kurdyumov-Levin automaton revised, **Journal of Statistical Physics**, v. 67, n. 3/4, 1992.

SANTOS, J. P. O., MELLO, M. P., MURANI, I. T., **Introdução à análise combinatória**, 4ª Edição, Rio de Janeiro: Editora Ciência Moderna, 2007, 390 p.

- SALADIN, K. S., ***Schistosoma mansoni*: cercarial responses to irradiance changes**. The Journal of Parasitology, 68: 120-124, 1982.
- SARKAR, P., BARUA, R. . The set of reversible 90/150 cellular automata is regular. **Discrete Appl. Math**, v. 84, n. 1-3, p. 199-213, 1998.
- SAFAK, H., SAHIN, M., GÜLVEREN, B., TOMAK, M., Efficiency of genetic algorithm and determination of ground state energy of impurity in a spherical quantum dot, **Int. J. Mod. Phys. C**, v. 14, p. 775-784, 2003.
- SCHRAUDOLPH, N. N., BELEW, R. K., Dynamic Parameter Encoding for Genetic Algorithms, **Machine Learning**, v. 9, p. 9-21, 1992.
- SCHWEFEL, H. -P., **Numerical Optimisation of Computer Models**, Wiley, New York, 1981, 398 p.
- SHALIZI, C. R., **Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata**. Tese (Doctor of Philosophy) - Physics Department, University of Wisconsin, Madison, 2001.
- SHANG, R., BAI, J., JIAO, L., JIN, C., Community detection based on modularity and an improved genetic algorithm. **Physica A**, v. 392, p. 1215-1231, 2013.
- SIEBOLD, C. T., **Ein Beitrag zur Helminthographia humana, aus brieflichen Mittheilungen des Dr. Bilharz in Cairo, nebs Bemerkungen von Prof. C. Th. v. Siebold in Breslau**. Zeitschrift für wissenschaftlich Zoologie, 4: 53-76, 1852.
- SIPSER, M., **Introduction to the Theory of Computation**. Brooks/Cole Pub Co, 1a Edição, 1996; 2a Edição 2005.
- SIQUEIRA JÚNIOR, E. L., FERREIRA, T. A. E., SILVA, M. G., Discovering the Rules of a Elementary One-Dimensional Automaton. In: YIN, H., COSTA, J. A. F., BARRETO, G., **Intelligent Data Engineering and Automated Learning - IDEAL 2012**, Lecture Notes in Computer Science, Natal: Springer Berlin Heidelberg, 2012, v. 7435, p. 312-319.
- SILVA, G., ALBUQUERQUE, J., SILVA, R., BOCANEGRA, S., FERREIRA, T., FERRER-SAVALL, J., LÓPEZ-CORDINA, D., SOUZA, M., SOUZA-SANTOS, R., BARBOSA, C., Modelling a real data cellular automaton to analyze the schistosomiasis expansion process along the coastline of Brazil. In: **European Meetings on Cybernetics and Systems Research**, 2012.
- SIWY, Z.; AUSLOOS, M.; IVANOVA, K. Correlation studies of open and closed state fluctuations in an ion channel: Analysis of ion current through a large-conductance locust potassium channel. **Physical Review E**, v. 65, n. 1, p. 031907-1-031907-6, 2002.
- SOUZA, M.A.A., BARBOSA, V. S., ALBUQUERQUE, J. O., BOCANEGRA, S., SOUZA-SANTOS, R., PAREDES, H., BARBOSA, C., Aspectos ecológicos e levantamento malacológico para identificação de áreas de risco para a transmissão da esquistossomose no litoral norte de Pernambuco, Brasil. **Iheringia. Série Zoologia (Impresso)**, v. 100, 1-6, 2010.

SRINIVAS, M., PATNAIK, L. M., Adaptive probabilities of crossover and mutation in genetic algorithms, **IEEE Transactions on Systems, Man and Cybernetics**, v. 24, n. 4, p. 656-667, 1994.

STIREWALT, M. A., **Penetration stimuli for schistosome cercariae.**, In: CHENG, T. C. Aspects of the Biology of Symbiosis. Baltimore: University Park Press, 1971.

STIREWALT, M. A., HACKEY, J. R., **Penetration of host skin by cercariae of *Schistosoma mansoni***. I. Observed entry into skin of mouse, hamster, rat, monkey, man. The Journal of Parasitology, 42: 565-580, 1956.

STIREWALT, M. A., ***Schistosoma mansoni*: cercaria to schistosomule.** Advances in Parasitology, 12: 115-182, 1974.

STENSGAARD, A.S., JORGENSEN, A., KABATEREINE, N.B., RAHBK, C., KRISTENSEN, T.K., Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda, **Geospatial Health**, v. 1, n. 1, p. 93-104, 2006.

STANLEY, H.E., PLEROU, V., Scaling and universality in economics: Empirical results and theoretical implications, **Quantitative finance**, v.1, p. 563 - 567, 2001.

STANDEN, O. D., **The effect of temperature, light, and salinity upon the hatching of the ova of *Schistosoma mansoni***. Transactions of the Royal Society of Tropical Medicine and Hygiene, 45: 225-241, 1951.

STOICO, C. O., RENZI, D. G., VERICAT, F., A genetic algorithm for the 1D electron gas. **Physica A**, v. 387, p. 159-166, 2008.

TELLES, M. P. C., DINIZ-FILHO, J. A. F., COELHO, A. S. G., CHAVES, L. J., Autocorrelação espacial das frequências alélicas em subpopulações de cagaiteira (*Eugenia dysenterica* DC., *Myrtaceae*) no sudeste de Goiás, **Revista brasil. Bot.**, v. 24, n. 2, p. 145-154, 2001.

THÉRON, A.; ROGNAN, A.; PAGES, J. R., **Host choice by larval parasites: a study of *Biomphalaria glabrata* snails and *Schistosoma mansoni* miracidia related to host size.** Parasitology Research, 84: 727-732, 1998.

TOOM, A., **Cellular Automata with Errors: Problems for Students of Probability.** Topics in Contemporary Probability and its Applications. Ed. J. Laurie Snell. Series Probability and Stochastics ed. by Richard Durrett and Mark Pinsky. CRC Press, 1995, pp. 117-157.

TOOM, A., **Contours, convex sets, and cellular automata.** Text of a course at the 23-th Colloquium of Brazilian mathematicians. Translated from English into Portuguese by Andréa Vanessa Rocha and Pedro Ferreira de Lima. IMPA, 2001.

TURING, A. M. On Computable Numbers, with an Application to the Entscheidungs problem, **Proceedings of the London Mathematical Society**, v. 2, n. 47, p. 230-265, 1937.

TURING, A. M. (1948), Intelligent Machinery, In: Mechanical Intelligence, Collected Works, D. C. Ince ed., Amsterdam : North-Holland, 1992, p. 49. As to initiative, it is exemplified by the research of algorithms for open arithmetical problems.

TURING, A.M. Computing machinery and intelligence. **Mind**, New Series, v. 59, p. 433-460, 1950.

VARSHAVSKY, V., MARAKHOVSKY, V., PECHANSKY, V. Synchronization of interacting automata, **Math. Syst. Theory**, v. 4, n. 3, p. 212-230, 1969.

WARREN, K. S., The pathobiology and pathogenesis of schistosomiasis. **Nature**, 273: 609-612, 1978.

WATSON, J. M. **Medical Helminthology**. London: Baillière Tindall and Cox, 1960.

WAN, C., LIU, Y., ZHANG, Y.Y., XU, J.M., LIN, D.D., Luo, J.P., Chen, F., WU, H.W., A cellular automaton model of Schistosoma japonicum infection., **Acta Trop**, v. 126, n. 3, p. 256-264, 2013.

WHITNEY, L. D., KAUTH, J., Genitor: A different genetic algorithm. In: **Proceedings of the Rocky Mountain Conference on Artificial Intelligence**. 1988, p. 118-130.

WILLIAMS, P., COELHO, P. M., **The attraction of snail to betalights**. Transactions of the Royal Society of Tropical Medicine and Hygiene, 67: 616-617, 1975.

WILSON, R. A., **Cercarie to river worms: development and migration in the mammalian host**. In: ROLLINSON, D., SIMPSON. A. J. G., The Biology of Schistosomes: from genes to latrines. London: Academic Press, 1987.

WOLFRAM, S., Statistical mechanics of cellular automata, **Rev. Mod. Phys.**, v. 55, n. 3, p. 601-644, 1983.

WOOLDRIDGE, J. M., **Introductory Econometrics: a Modern Approach**. 5 edition. Gengage Learning, 2012. 912 p.

WUCHERER, O. **Noticia preliminar sobre vermes de uma espécie ainda não descrita, encontrados na urina de doentes de hematuria intertropical no Brazil**. Gazeta Médica da Bahia, 3: 97-99, 1868.

XIAN-JUN, S., LEI, H., A Genetic Algorithm-Based Approach for Classification Rule Discovery. In: **ICIII '08. International Conference on Information Management, Innovation Management and Industrial Engineering**, Taipei, 2008, v. 1, p. 175-178.

XIANG, J., CHEN, H., ISHIKAWA, H., A mathematical model for the transmission of Schistosoma japonicum in consideration of seasonal water level fluctuations of Poyang Lake in Jiangxi, China., **Parasitol Int**, v. 62, n. 2, p. 118-126, 2013.

YAMADA, H., AMOROSO, S. Structural and behavioural equivalences of tessellation automata. **Inf. Control**, v. 18, p. 1-31, 1971.

ZHOU, T., ZHOU, P.-L., WANG, B.-H., TANG, Z.-N., LIU, J., Modeling stock market based on genetic cellular automata, **Int. J. Mod. Phys. B**, v. 18, p. 2697, 2004.