

**UNIVERSIDADE PRESBITERIANA MACKENZIE
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E COMPUTAÇÃO**

Deivison Marques Gandini

**Usando Algoritmo Genético na Identificação de um Modelo
Epidemiológico Baseado em Autômato Celular**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica e Computação.

Orientador: Prof. Dr. Luiz Henrique Alves Monteiro

São Paulo
2018

G195u

Gandini, Deivison Marques

Usando algoritmo genético na identificação de um modelo epidemiológico baseado em autômato celular. / Deivison Marques Gandini – São Paulo, 2018.

57 f.: il.; 30 cm.

Dissertação (Mestrado em Engenharia Elétrica e Computação) - Universidade Presbiteriana Mackenzie - São Paulo, 2018.

Orientador: Luiz Henrique Alves Monteiro.

Bibliografia: f. 44-47.

1. Algoritmo genético. 2. Autômato celular. 3. Epidemiologia. 4. Modelo SIR. 5. Varicela. I. Monteiro, Luiz Henrique Alves, *orientador*. II. Título.

CDD 006.31

Deivison Marques Gandini

Usando Algoritmo Genético na Identificação de um Modelo
Epidemiológico Baseado em Autômato Celular

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Computação da Universidade Presbiteriana Mackenzie como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica e Computação.

Aprovado em 21 de Fevereiro de 2018

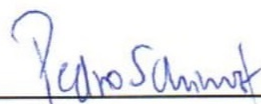
BANCA EXAMINADORA



Prof. Dr. Luiz Henrique Alves Monteiro
Universidade Presbiteriana Mackenzie



Prof. Dr. Pedro Paulo Balbi de Oliveira
Universidade Presbiteriana Mackenzie



Prof. Dr. Pedro Henrique Triguis Schimit
Universidade Nove de Julho

A meus pais, meus irmãos e minha mulher.

Agradecimentos

Aos meus pais, José Afonso e Maria Marques Gandini, que além da vida me darem, me ensinaram a vivê-la por meio de exemplos e trabalho árduo. Que em momento algum esmoreceram, mesmo na adversidade, e se desdoblaram por meio de muito suor, luta, perseverança e fé, para que eu tivesse condições que eles jamais tiveram. Sem vocês eu nada seria. Obrigado por não desistirem.

Aos meus irmãos, Douglas e Daniel, que me serviram de inspiração durante toda minha vida. Ensinaram-me a partilhar, ler, programar, onde estudar, a ouvir boa música, etc. Com orgulho digo que sou um amálgama de vocês.

À minha companheira Luiza Quaglio, para quem esta dissertação cobrou tanto tempo juntos, por todo seu amor, paciência, ouvidos e insights. Essa caminhada nem teria começado sem seu apoio, muito menos chegaria ao fim. Te amo hoje e sempre.

A todos os professores do programa pelo ensinamentos, dedicação e profissionalismo. O ato de ensinar com brilho no olhar é uma fonte de inspiração eterna.

Aos Prof. Dr. Pedro Paulo Balbi de Oliveira e Prof. Dr. Pedro Henrique Triguís Schimit por suas participações na banca de qualificação com entusiasmo, comentários e considerações, ajudando em muito a melhorar o nível deste trabalho.

Finalmente, ao Prof. Dr. Luiz Henrique Alves Monteiro, que me recebeu como seu aluno e com paciência e compreensão me orientou durante todas as fases dessa empreitada. Sua experiência, disponibilidade e liderança foram primordiais para que se extraísse o meu melhor mesmo durante as horas mais desesperadas. Muito obrigado.

*The fight is won or lost far away from witnesses – behind the lines, in the gym,
and out there on the road long before I dance under those lights.*

Muhammad Ali

RESUMO

Estudos epidemiológicos buscam entender e prever a propagação de doenças contagiosas. Em muitos desses estudos, empregam-se modelos do tipo SIR. Nesses modelos, cada indivíduo da população hospedeira está, a cada passo de tempo, em um de três estados: suscetível (S), infectado (I) ou recuperado (R). Nesta dissertação, um modelo SIR é implementado usando um autômato celular probabilista (ACP). Os parâmetros do ACP são identificados usando um algoritmo genético e dois conjuntos de dados: dados fictícios e dados realistas de casos de varicela na Alemanha antes da era da vacinação. Este trabalho mostra que conjuntos distintos de parâmetros do ACP podem levar a soluções em regime permanente similares. As limitações do procedimento de identificação são discutidas, principalmente no caso de dados realistas, em que a quantidade de infectados é muito pequena.

Palavras-chave: *algoritmo genético, autômato celular, epidemiologia, modelo SIR, varicela.*

ABSTRACT

Epidemiological studies aim to understand and predict the spreading of contagious diseases. In many of these studies, SIR-type models are employed. In these models, each individual of the host population is, at each time step, in one of three states: susceptible (S), infected (I) or recovered (R). In this dissertation, a SIR model is implemented by using a probabilistic cellular automaton (PCA). The PCA parameters are identified by using a genetic algorithm and two data sets: fictitious data and realistic data of varicella cases in Germany prior to the vaccination era. This work shows that different sets of PCA parameters can lead to similar steady-state solutions. The limitations of the identification procedure are discussed, especially in the case of realistic data, in which the amount of infected individuals is too small.

Keywords: *cellular automaton, epidemiology, genetic algorithm, SIR model, varicella.*

Sumário

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 1 |
| 2 | CONCEITOS PRELIMINARES | 4 |
| 2.1 | Modelo SIR | 5 |
| 2.1.1 | Um modelo SIR em EDO para varicela | 7 |
| 2.2 | Autômatos celulares | 8 |
| 2.2.1 | Um modelo SIR em ACP para varicela | 10 |
| 2.3 | Algoritmos genéticos | 12 |
| 2.4 | Sobre a implementação | 16 |
| 3 | IDENTIFICANDO ACP COM AG | 18 |
| 3.1 | O ACP | 18 |
| 3.2 | O AG | 20 |
| 4 | RESULTADOS | 22 |
| 4.1 | Escolha do tamanho do reticulado | 23 |
| 4.2 | Cenário fictício | 25 |
| 4.3 | Cenário realista | 30 |
| 5 | CONCLUSÕES | 39 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 45 |

Lista de Figuras

| | | |
|-----|--|----|
| 2.1 | Evolução temporal do número de crianças vivas com inoculação (crianças imunizadas) e sem inoculação (crianças não imunes), em relação à varíola. Adaptado de (GLOMSKI; OHANIAN, 2012). | 5 |
| 2.2 | Reticulado 9×9 evidenciando as vizinhanças de Moore de raio Γ , com Γ variando de 1 a 3. | 9 |
| 2.3 | A) Reticulado 3×3 . B) Condição de contorno periódica aplicada à célula $(1, 1)$, supondo que sua vizinhança é a de Moore com $\Gamma = 1$ | 10 |
| 2.4 | Fluxograma genérico de um AG. | 13 |
| 3.1 | Evolução temporal do ACP num reticulado 15×15 . Nessa figura, azul representa indivíduo recuperado; amarelo, suscetível; e vermelho, infectado. | 19 |
| 4.1 | Simulação do ACP com $K_a = 0,2$, $K_q = 0,1$, $K_p = 0,1$, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,09712125; 0,08874075)$ e $R_0 = 8,86$ | 25 |
| 4.2 | Variação da função-objetivo f durante 50 gerações. O AG possui a seguinte configuração: $C = 50$, $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$ e $\sigma = 8\%$. Mostram-se as variações relacionadas aos valores de f máximo, mínimo e médio. | 26 |
| 4.3 | Simulação do ACP com o cromossomo mais apto da geração 50, para o qual $K_a = 0,691677$, $K_q = 1,660371$ e $K_p = 0,110969$. Ainda, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,095365; 0,08842)$ e $R_0 = 18,4$ | 28 |
| 4.4 | Simulação do ACP com o cromossomo mais apto da geração 35, para o qual $K_a = 0,606023$, $K_q = 1,660371$ e $K_p = 0,121707$. Ainda, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,096325; 0,08879)$ e $R_0 = 8,86$ | 28 |
| 4.5 | Simulação do modelo em EDO com $a = 7,506961462$, $b = 0,401266682$, $c = 0,01391088$, $e = 0,101168197$, $q = 2,985121759$ e $p = 0,660041264$. Valores correspondentes ao melhor cromossomo da geração 50. | 29 |
| 4.6 | Variação de f durante 30 gerações do AG para Alemanha, com $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$ e população de $C = 200$. Mostram-se os valores mínimo, médio e máximo de f em cada uma das gerações. | 31 |

| | | |
|------|---|----|
| 4.7 | Evolução temporal do ACP com $P_{ic} = 0,14$, $P_{im} = P_{rm} = 0,000035$, $K_a = 0,281866$, $K_q = 0,082925$ e $K_p = 2,876278$. Com isso, atinge-se o estado estacionário com $(s^*, i^*) = (0,023089; 0)$, em que a doença desaparece. . . . | 32 |
| 4.8 | Evolução do AG utilizando f_{log} com $C = 200$, $\epsilon = 1$, $p_c = 0,7$, $p_m = 0,5$ e $\sigma = 8\%$ | 33 |
| 4.9 | Evolução temporal no ACP com o cromossomo mais apto encontrado pelo AG com f_{log} , para o qual $K_a = 6,968583$, $K_q = 3,465863$ e $K_p = 3,391813$. Com isso, converge-se para $(s^*, i^*) = (0,014816; 0)$ e erros $S_\xi = 35,6\%$ e $I_\xi = 100\%$ | 34 |
| 4.10 | Identificação com o AG no cenário fictício usando rearranjo no ACP. Configuração do AG: $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$. Na geração 53, obteve-se $f = 19495$ | 35 |
| 4.11 | O ACP no cenário fictício com rearranjo com $K_a = 0,338939$, $K_q = 0,072136$ e $K_p = 0,093755$ (além de $P_{ic} = 0,4$, $P_{im} = 0,2$, $P_{rm} = 0,1$). Converte-se para $(s^*, i^*) = (0,097125, 0,088775)$, com $S_\xi = 0,004\%$ e $I_\xi = 0,04\%$. Ainda, $R_0 = 12,7$ | 36 |
| 4.12 | Evolução do AG para o cenário da Alemanha com rearranjo durante 50 gerações, com $\epsilon = 1$, $p_c = 0,7$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$ | 37 |
| 4.13 | Evolução do ACP por quarenta passos de tempo para o cenário da Alemanha com rearranjo, com $K_a = 0,629071$, $K_q = 0,0649687$ e $K_p = 0,186962$. Em regime permanente, obtêm-se $(s^*, i^*) = (0,023091; 0)$ com erro $S_\xi = 0,4\%$ e $I_\xi = 100\%$ | 38 |

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Considere que, quanto maior o valor de f_i , maior a adaptação do cromossomo. Nesse exemplo, quatro cromossomos foram avaliados, sendo apresentado o valor de f_i de cada cromossomo. Somam-se esses valores de f_i e divide-se o círculo em quatro fatias, sendo o tamanho de cada fatia proporcional a f_i . Por exemplo, a fatia correspondente ao primeiro cromossomo é proporcional a $\frac{20}{20+13+23+16} = \frac{20}{72} = \frac{5}{18} = 27,8\%$. Ou seja, corresponde a um arco de circunferência de ângulo $360^\circ \times \frac{5}{18} = 100^\circ$ | 15 |
| 3.1 | Regras de transição do ACP por prioridade. As constantes $K_a, K_q, K_p, P_{ic}, P_{im}$ e P_{rm} são positivas. No cálculo das probabilidades $P_s(i), P_s(i, r)$ e $P_i(r)$, i denota o número de vizinhos infectados e r o número de vizinhos recuperados. | 20 |
| 4.1 | Influência da dimensão do reticulado na precisão dos resultados e no tempo de processamento do ACP e do AG. Testes executados com NVIDIA Tesla K80 com 2496 cuda cores. | 24 |
| 4.2 | Cromossomo mais apto encontrado na última geração (caso a , geração 50) e o mais apto dentre todas as gerações (caso b , geração 35). Mostra-se também como esse cromossomo mais apto se comporta na próxima geração (casos c e d , geração 36). Para cada cromossomo, apresentam-se os valores de s^* e i^* com os erros relativos associados, e o valor de f | 26 |
| 4.3 | Números correspondentes à solução-alvo (figura 4.1), ao melhor cromossomo da última geração (figura 4.3) e ao melhor cromossomo dentre todas as gerações (figura 4.4). | 30 |

Capítulo 1

INTRODUÇÃO

Epidemiologia é a ciência que estuda fatores ambientais, comportamentais, genéticos e sociais relacionados à incidência de doenças contagiosas em populações (ANDERSON; MAY, 1992). Estudos epidemiológicos são relevantes para que a humanidade possa ter maior qualidade de vida. Por exemplo, uma das maiores pandemias da história da humanidade, a da gripe espanhola, matou aproximadamente 100 milhões de pessoas entre os anos 1918 e 1919 (MURRAY, 2004), cerca de 5% da população mundial da época. Se uma epidemia dessa magnitude atingisse hoje o Brasil, com mesma taxa de mortalidade, mataria em torno de 10 milhões de pessoas, aproximadamente a população do estado de Pernambuco. Prever a ocorrência de epidemias, a fim de tomar medidas para evitá-las, ou ao menos controlá-las, certamente pode melhorar a qualidade de vida da nossa população.

Entender os mecanismos de disseminação de doenças contagiosas em uma população hospedeira, que é considerada um sistema complexo (MONTEIRO, 2014), não ajuda somente a combatê-las de forma eficiente, mas também auxilia a prever a gravidade de um surto epidêmico. Com isso, podem-se dimensionar as medidas de saneamento, as campanhas de vacinação e as campanhas de distribuição de medicamentos e, conseqüentemente, os custos aos cofres públicos (WHITE; REY; SÀNCHEZ, 2007).

O modelo epidemiológico SIR, baseado em equações diferenciais proposto por Kermack e McKendrick em 1927, e suas diversas variações são normalmente utilizados para modelar a propagação de doenças contagiosas (HETHCOTE, 2000). Num modelo do tipo SIR, a cada instante, cada indivíduo da população hospedeira está em um de três estados: S , que

representa o indivíduo suscetível; I , que é o infectado pela doença; e R , que corresponde ao recuperado ou removido do processo de propagação.

A utilidade prática de um modelo é determinada pela acurácia com que ele é capaz de prever, quantitativamente, o comportamento do sistema real em estudo. Dado que o sistema em estudo foi descrito em termos de um modelo matemático, pode-se então sintonizar os parâmetros desse modelo para que ele tente reproduzir as séries temporais obtidas a partir de medidas do sistema real. *Identificação de Sistemas* consiste de técnicas de criação e ajuste de modelos a partir de dados experimentais ou observacionais. Portanto, sintonizar os parâmetros de um modelo é um procedimento de identificação de sistemas (AGUIRRE, 2015). Algoritmo genético (AG), que é um método de busca de soluções inspirado na teoria darwiniana da evolução das espécies (MITCHELL, 1998; CASTRO, 2006), tem sido utilizado como uma ferramenta de identificação de sistemas modelados por autômatos celulares (MITCHELL; HRABER; CRUTCHFIELD, 1993; BILLINGS; YANG, 2003; GANGULY et al., 2004; OLIVEIRA; BORTOT; OLIVEIRA, 2006; GERRARDI; MONTEIRO, 2011; WHITSED; SMALLBONE, 2017).

Nesta dissertação, parte-se do modelo SIR de Moraes e Monteiro (MORAES; MONTEIRO, 2016). Este modelo foi descrito em termos de um autômato celular probabilista (ACP) (WOLFRAM, 2002) e em termos de equações diferenciais ordinárias (EDO) equivalentes a esse ACP. Ele foi proposto para estudar a disseminação de doenças contagiosas típicas de crianças como, por exemplo, a varicela, um tipo de herpes altamente contagiosa que concede imunidade permanente após a recuperação do indivíduo infectado (HEININGER; SEWARD, 2006; MINISTÉRIO da SAÚDE, 2014a). O diferencial desse modelo é considerar a influência dos recuperados tanto no processo de propagação da doença, facilitando o encontro entre suscetíveis e infectados, quanto no processo de cura, ajudando os infectados a se recuperarem mais rapidamente (MORAES, 2016). Quando se pretende usar esse modelo num cenário realista, alguns valores de parâmetros, como a expectativa de vida da população, são obtidos (facilmente) na literatura; já valores de outros parâmetros, como aqueles associados à propagação da doença, mediada ou não por recuperados, não são fáceis de se encontrar (na literatura ou por tentativa e erro). Esta dissertação tem por objetivo encontrar valores para estes parâmetros.

Esta dissertação está estruturada do seguinte modo. No capítulo 2, apresentam-se alguns conceitos básicos relacionados a modelo SIR, ACP e AG. No capítulo 3, descrevem-se o ACP usado para representar o modelo SIR de Moraes e Monteiro e o AG usado na identificação dos parâmetros do ACP. No capítulo 4, apresentam-se os resultados obtidos em simulações numéricas num cenário fictício (GANDINI; MONTEIRO, 2017) e num cenário realista usando dados da varicela na Alemanha. No capítulo 5, discutem-se os resultados e as limitações do modelo epidemiológico e do processo de identificação aqui empregados.

Capítulo 2

CONCEITOS PRELIMINARES

Observações de como as enfermidades são afetadas por fatores ambientais podem ser encontradas em notas escritas por Hipócrates, entre os séculos V e IV a.C. Porém, foi somente em 1854, durante um surto de cólera em Londres, que avanços mais significativos começaram a acontecer na área de saúde pública (BONITA; BEAGLEHOLE; KJELLSTROM, 2010). A investigação de John Snow feita naquela época, levando em consideração fatores físicos, políticos, químicos e sociológicos, evidenciou que a contaminação por cólera se dava por meio de água não tratada e usualmente em lares com carência de higiene básica (CAMERON; JONES, 1983).

Epidemiologia faz estudos qualitativos e quantitativos relacionados à propagação de doenças contagiosas por meio de modelos matemáticos (ANDERSON; MAY, 1992; BONITA; BEAGLEHOLE; KJELLSTROM, 2010). Entre os pioneiros e mais expressivos trabalhos na área, encontra-se o feito em 1766 pelo matemático suíço Daniel Bernoulli sobre a inoculação de varíola (GLOMSKI; OHANIAN, 2012). Bernoulli utilizou equações diferenciais para avaliar a hipótese do aumento da expectativa de vida, caso a varíola fosse eliminada como causa de morte (DIETZ; HEESTERBEEK, 2002). Em seu modelo, o primeiro relativo à propagação de uma doença contagiosa, dividiu-se a população em dois grupos (GLOMSKI; OHANIAN, 2012):

- o grupo suscetível, que é aquele que não foi exposto à doença;
- o grupo imune, formado por pessoas que foram contaminadas pela doença, mas sobreviveram e se tornaram livres dessa doença pelo resto de suas vidas.

Com seu modelo, Bernoulli projetou, partindo de 1300 crianças, quantas seriam salvas se fossem imunizadas contra varíola. A figura 2.1 ilustra essa projeção. Lembre que naquela época a vida média era de pouco mais de 25 anos.

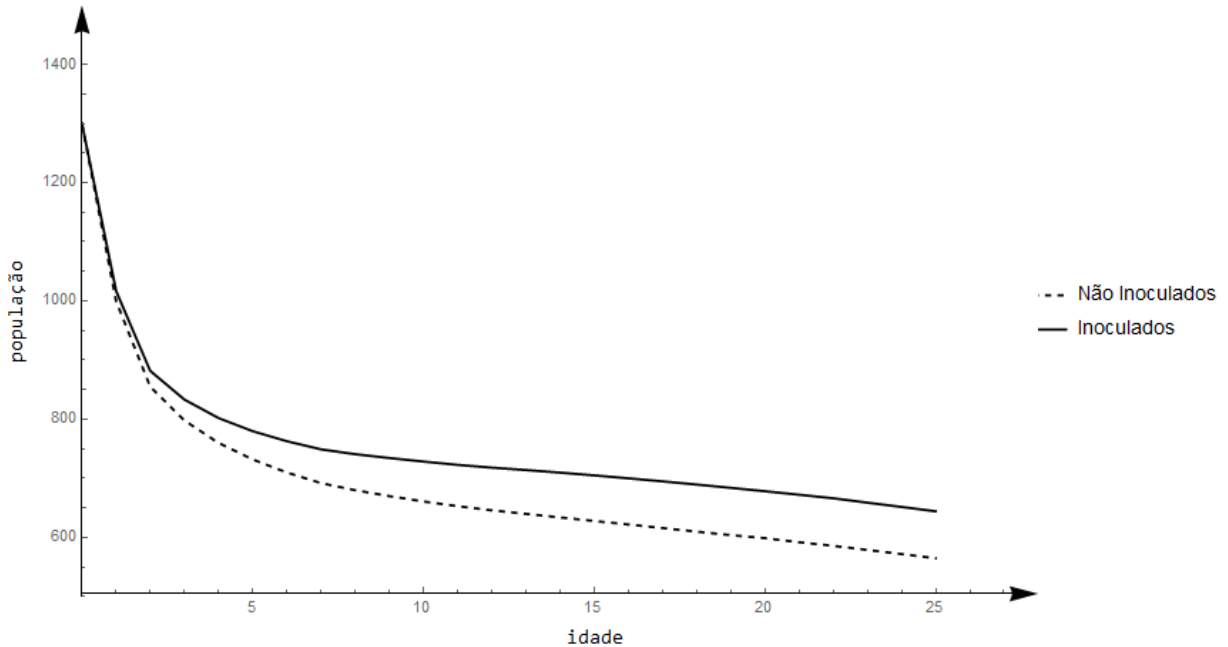


Figura 2.1: Evolução temporal do número de crianças vivas com inoculação (crianças imunizadas) e sem inoculação (crianças não imunes), em relação à varíola. Adaptado de (GLOMSKI; OHANIAN, 2012).

2.1 Modelo SIR

O modelo de Bernoulli não leva em conta a natureza dinâmica da infecção; isto é, que sua intensidade pode variar no tempo (DIETZ; HEESTERBEEK, 2002), nem seu modo de propagação. William Ogilvy Kermack e Anderson Gray McKendrick, no artigo “Uma Contribuição para a Teoria Matemática da Epidemias” de 1927, analisaram o que acontece quando uma população é afetada por uma infecção, supondo que a enfermidade se espalha devido à interação entre os doentes e os saudáveis, e que os convalescidos passam pelo ciclo da doença e são removidos desse estado porque ou faleceram ou se curaram. Com o modelo que propuseram, eles mostraram que, conforme a epidemia se espalha, o número de afetados inicialmente cresce, atinge um pico e depois diminui, até

que o surto chega ao seu fim. O modelo SIR proposto por eles para explicar o início e o fim de um surto epidêmico é tido como o modelo epidemiológico mais relevante. Ele é composto pelas seguintes equações diferenciais (KERMACK; MCKENDRICK, 1927; GLOMSKI; OHANIAN, 2012; MONTEIRO, 2014):

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S(t)I(t) \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \alpha I(t) \\ \frac{dR(t)}{dt} &= \alpha I(t)\end{aligned}$$

sendo $\beta > 0$ a constante de taxa de infecção, $\alpha > 0$ a constante de taxa de remoção (cura e/ou morte) e $S(t) + I(t) + R(t) = N$, em que N é o tamanho da população.

Esse modelo descreve adequadamente, por exemplo, a propagação de peste negra numa dada região (doença por eles analisada), em que, após um pico de casos de infectados, a doença some (KERMACK; MCKENDRICK, 1927). Nem toda doença, porém, desaparece naturalmente. Há muitas que se tornam crônicas na população hospedeira. Há outras que não conferem imunidade permanente. Por isso, variantes desse modelo SIR têm sido empregadas em estudos de doenças com diferentes características. Há, por exemplo, modelos SIRS (em que o recuperado volta a ser suscetível após um tempo, como gripe) e SIS (em que a cura não leva à imunidade, como hepatite C). Há, inclusive, modelos com outras classes como, SEIS, SEIR, MSIR, MSEIR e MSEIRS, em que a classe M (ou V) representa imunizado (ou vacinado) e E significa aquele que foi exposto à doença, mas ainda não está infeccioso (pois se encontra no período de incubação) (HETHCOTE, 2000).

As equações diferenciais dos modelos do tipo SIR permitem estudos analíticos, como, por exemplo, deduzir a fórmula para o fator de reprodutividade basal R_0 , que é a média de infectados decorrentes de um único indivíduo infectado introduzido numa população suscetível (ANDERSON; MAY, 1992). Se $R_0 > 1$, a doença tende a se espalhar na população; se $R_0 < 1$, ela tende a desaparecer naturalmente (HETHCOTE, 2000). No modelo SIR original, R_0 é dado por (KERMACK; MCKENDRICK, 1927; ANDERSON; MAY, 1992; MONTEIRO, 2014):

$$R_0 = \frac{\alpha S(0)}{\beta}$$

sendo $S(0)$ a população de suscetíveis no instante inicial ($t = 0$).

Entretanto, usar equações diferenciais pode dificultar a representação de alguns aspectos do sistema epidemiológico estudado, como a distribuição espacial da população hospedeira, o comportamento de cada um de seus indivíduos e a topologia da rede de contatos sociais (WHITE; REY; SÀNCHEZ, 2007). Por isso, muitos estudos epidemiológicos são formulados em termos de autômatos celulares, já que, com essa abordagem, aspectos espaciais e individuais podem ser levados em consideração de maneira mais natural.

2.1.1 Um modelo SIR em EDO para varicela

A varicela, conhecida popularmente como catapora, é causada pelo vírus Varicela-Zóster, da família dos herpesvírus. É mais comum em crianças. Sintomas típicos são: febre, erupção cutânea com cor avermelhada e mal-estar, raramente causando maiores complicações (HEININGER; SEWARD, 2006).

Sendo a varicela mais comum na tenra idade, os adultos, normalmente indivíduos recuperados (imunes), podem ter um papel de facilitador das interações sociais entre os suscetíveis e os infectados, pois crianças dependem de um responsável para irem a lugares públicos como parques, escolas, shoppings, restaurantes, etc. Entretanto, esses adultos também atuam no cuidado da criança doente e, assim, diminuem o tempo de recuperação da mesma. Baseado nessas características, foi proposto o seguinte modelo SIR em termos de EDO (MORAES; MONTEIRO, 2016):

$$\begin{aligned}
 \frac{dS(t)}{dt} &= -aS(t)I(t)[1 + qR(t)] + cI(t) + eR(t) \\
 \frac{dI(t)}{dt} &= aS(t)I(t)[1 + qR(t)] - bI(t)[1 + pR(t)] - cI(t) \\
 \frac{dR(t)}{dt} &= bI(t)[1 + pR(t)] - eR(t)
 \end{aligned} \tag{2.1}$$

As variáveis $S(t)$, $I(t)$ e $R(t)$ são as quantidades de suscetíveis, infectados e recuperados no instante t , sendo $S(t) + I(t) + R(t) = N$. Os parâmetros a , b , c , e , q e p são constantes positivas associadas a:

- a : infecção devido ao contato direto entre S e I ;
- b : recuperação espontânea de I ;

- c : morte de I devido à doença;
- e : morte de R por causas não relacionadas à doença;
- q : infecção devido ao contato entre S e I mediado por R ;
- p : recuperação de I mediada por R .

Como cada indivíduo infectado ou recuperado que falece é substituído por um (novo) suscetível, o tamanho da população permanece constante e igual a N .

Moraes (2016) mostrou que o fator de reprodutividade basal R_0 desse modelo é dado por:

$$R_0 = \frac{aN}{(b+c)}$$

Assim, se $R_0 > 1$, a doença permanece endemicamente na população hospedeira; se $R_0 < 1$, ela tende a desaparecer com o passar do tempo.

Esse mesmo modelo foi também implementado usando autômato celular, como mostrado ainda neste capítulo.

2.2 Autômatos celulares

Autômato celular (AC) é um modelo dinâmico discreto de computação, que vem sendo utilizado no estudo de diversos sistemas biológicos, físicos, químicos, sociais e de engenharia. Uma de suas características é a capacidade de paralelismo (WOLFRAM, 2002). Isso decorre do fato de que um autômato celular é um conjunto de células interligadas, de modo que cada célula está em um de k estados possíveis a cada passo de tempo t . Ainda, a cada passo de tempo, o estado s_t de uma célula é atualizado por meio da aplicação de uma regra ϕ que leva em consideração o estado s_t dessa célula e os estados dos vizinhos; ou seja, $s_{t+1} = \phi(s_t, s_t^1, s_t^2, \dots, s_t^v)$ sendo $s_{t+1} \in k$ e s_t^v o estado do v -ésimo vizinho em t . Nota-se, portanto, que um AC é um sistema de espaço discreto, tempo discreto e estado discreto.

As células do AC formam um reticulado. Aqui, supõe-se que os vizinhos são as células contidas numa matriz $(2\Gamma + 1) \times (2\Gamma + 1)$ centrada na célula cujo estado será atualizado.

Essa é a chamada vizinhança de Moore de raio Γ (WOLFRAM, 2002). Um exemplo de reticulado bidimensional com vizinhança de Moore é mostrado figura 2.2.

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------------------|
| 1,1 | 1,2 | 1,3 | 1,4 | 1,5 | 1,6 | 1,7 | 1,8 | 1,9 | |
| 2,1 | 2,2 | 2,3 | 2,4 | 2,5 | 2,6 | 2,7 | 2,8 | 2,9 | |
| 3,1 | 3,2 | 3,3 | 3,4 | 3,5 | 3,6 | 3,7 | 3,8 | 3,9 | |
| 4,1 | 4,2 | 4,3 | 4,4 | 4,5 | 4,6 | 4,7 | 4,8 | 4,9 | |
| 5,1 | 5,2 | 5,3 | 5,4 | 5,5 | 5,6 | 5,7 | 5,8 | 5,9 | |
| 6,1 | 6,2 | 6,3 | 6,4 | 6,5 | 6,6 | 6,7 | 6,8 | 6,9 | |
| 7,1 | 7,2 | 7,3 | 7,4 | 7,5 | 7,6 | 7,7 | 7,8 | 7,9 | |
| 8,1 | 8,2 | 8,3 | 8,4 | 8,5 | 8,6 | 8,7 | 8,8 | 8,9 | |
| 9,1 | 9,2 | 9,3 | 9,4 | 9,5 | 9,6 | 9,7 | 9,8 | 9,9 | |
| | | | | | | | | | Célula de referência |
| | | | | | | | | | Vizinhança de Raio 1 |
| | | | | | | | | | Vizinhança de Raio 2 |
| | | | | | | | | | Vizinhança de Raio 3 |

Figura 2.2: Reticulado 9×9 evidenciando as vizinhanças de Moore de raio Γ , com Γ variando de 1 a 3.

A quantidade de vizinhos varia com o raio Γ (WHITE; REY; SÀNCHEZ, 2007). No exemplo da figura 2.2, para $\Gamma = 1$, a vizinhança da célula (6, 5) é dada pelas células $V_1 = \{(5, 4), (5, 5), (5, 6), (6, 4), (6, 6), (7, 4), (7, 5), (7, 6)\}$. Para $\Gamma = 2$, a vizinhança é composta por $V_2 = V_1 + \{(4, 3), (4, 4), (4, 5), (4, 6), (4, 7), (5, 3), (5, 7), (6, 3), (6, 7), (7, 3), (7, 7), (8, 3), (8, 4), (8, 5), (8, 6), (8, 7)\}$.

Quando a distância entre a borda do reticulado e a célula cujo estado será atualizado é menor que o raio Γ , se faz necessário o uso de uma condição de contorno, que especifica a ação que o autômato toma quando necessita de um vizinho que estaria fora do reticulado. Aqui, usa-se condição de contorno periódica, em que se considera que a primeira linha do reticulado está em contato com a última linha, e a primeira coluna está em contato com a última coluna, como ilustrado na figura 2.3. Desse modo, cria-se um toro tridimensional a partir de um reticulado bidimensional, eliminando as bordas.

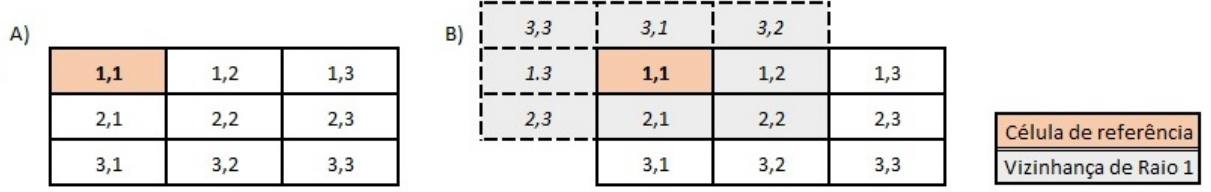


Figura 2.3: A) Reticulado 3×3 . B) Condição de contorno periódica aplicada à célula $(1, 1)$, supondo que sua vizinhança é a de Moore com $\Gamma = 1$.

2.2.1 Um modelo SIR em ACP para varicela

O modelo epidemiológico adotado aqui é aquele estudado por Moraes (2016). Em termos de ACP, ele é assim descrito. Usa-se um reticulado bidimensional $n \times n$, em que cada célula representa um indivíduo, com condição de contorno periódica, e vizinhança de Moore com raio $\Gamma = 2$. A transição $S \rightarrow I$ relacionada à infecção causada pelo contato direto entre S e I é determinada a partir da probabilidade $P_s(i) = 1 - e^{-K_a i}$, sendo i a quantidade de vizinhos infectados e K_a um parâmetro que expressa a infectividade do patógeno estudado. Note que $P_s(0) = 0$ e que $P_s(i)$ cresce monotonamente com i . Os processos de cura espontânea de infectado ($I \rightarrow R$), de morte de infectado ($I \rightarrow S$) e de morte de recuperado ($R \rightarrow S$) são determinados a partir das probabilidades constantes P_{ic} , P_{im} e P_{rm} . Lembre que, para cada falecimento, um novo indivíduo suscetível é adicionado, a fim de manter a população com tamanho constante e igual a $n^2 = N$ (MORAES; MONTEIRO, 2016).

Para que esse ACP seja equivalente ao modelo SIR para varicela descrito na subseção 2.1.1, consideram-se ainda as probabilidades $P_s(i, r) = 1 - e^{-K_q i r}$ e $P_i(r) = 1 - e^{-K_p r}$, sendo K_q uma constante relacionada à infecção de suscetível mediada por recuperado, K_p uma constante relacionada à cura de infectado mediada por recuperado, e r o número de recuperados na vizinhança da célula examinada. Os parâmetros K_a , K_q , K_p do ACP têm papel equivalente aos parâmetros a , q , p do modelo em EDO.

Para que os modelos formulados em termos de ACP e em termos de EDO produzam comportamentos dinâmicos similares, realiza-se o seguinte procedimento. A partir de uma simulação do ACP, contam-se, após o sistema atingir seu regime permanente (seu estado

estacionário), as quantidades de transições entre os estados do modelo, devido à aplicação das regras probabilistas acima mencionadas. Com isso, os valores dos parâmetros do modelo em EDO são determinados por (SCHIMIT; MONTEIRO, 2009; MORAES, 2016):

$$\begin{aligned}
 a &\approx \frac{\Delta I(t)_{S \rightarrow I}}{S(t)I(t)\Delta(t)} & e &\approx \frac{\Delta S(t)_{R \rightarrow S}}{R(t)\Delta(t)} \\
 b &\approx \frac{\Delta R(t)_{I \rightarrow R}}{I(t)\Delta(t)} & q &\approx \frac{\Delta I(t)_{qS \rightarrow I}}{aS(t)I(t)R(t)\Delta(t)} \\
 c &\approx \frac{\Delta I(t)_{I \rightarrow S}}{I(t)\Delta(t)} & p &\approx \frac{\Delta R(t)_{pI \rightarrow R}}{bI(t)R(t)\Delta(t)}
 \end{aligned}$$

Nessas expressões:

- $\Delta(t)$ representa o intervalo entre dois passos de tempo consecutivos (ou seja, $\Delta(t)=1$);
- $S(t)$, $I(t)$ e $R(t)$ são, respectivamente, as quantidades de suscetíveis, infectados e recuperados no passo de tempo t ;
- $\Delta I(t)_{S \rightarrow I}$ é número de transições $S \rightarrow I$ devido à regra que envolve a probabilidade $P_s(i)$;
- $\Delta I(t)_{qS \rightarrow I}$ é número de transições $S \rightarrow I$ devido à regra que envolve a probabilidade $P_s(i, r)$, após a aplicação da regra associada à probabilidade $P_s(i)$;
- $\Delta R(t)_{I \rightarrow R}$ é número de transições $I \rightarrow R$ devido à regra que envolve a probabilidade P_{ic} ;
- $\Delta R(t)_{pI \rightarrow R}$ é número de transições $I \rightarrow R$ devido à regra que envolve a probabilidade $P_i(r)$, após a aplicação da regra associada à probabilidade P_{ic} ;
- $\Delta S(t)_{I \rightarrow S}$ é número de transições $I \rightarrow S$ devido à regra que envolve a probabilidade P_{im} , após a aplicação da regra associada à probabilidade $P_i(r)$;
- $\Delta S(t)_{R \rightarrow S}$ é número de transições $R \rightarrow S$ devido à regra que envolve a probabilidade P_{rm} .

As quantidades de transições de estado no ACP acima mencionadas podem ser usadas para calcular o valor de R_0 . Como para esse modelo $R_0 = aN/(b + c)$, então (GANDINI; MONTEIRO, 2017):

$$R_0 = \frac{N\Delta I(t)_{S \rightarrow I}}{S(t)[\Delta R(t)_{I \rightarrow R} + \Delta S(t)_{I \rightarrow S}]} \quad (2.2)$$

Com esse procedimento, é possível determinar os valores dos parâmetros do modelo em EDO que tornam suas soluções numéricas semelhantes às evoluções temporais de $S(t)$, $I(t)$ e $R(t)$ encontradas no ACP (MORAES, 2016). O contrário, entretanto, não se sabe ainda fazer. Ou seja, dados os valores dos parâmetros do modelo em EDO e o resultado da integração numérica dessas equações, não se sabe como escolher os valores dos parâmetros do ACP para que este produza um resultado similar ao do modelo em EDO. Nesse processo de identificação, que é o assunto desta dissertação, podem-se usar algoritmos genéticos.

2.3 Algoritmos genéticos

Algoritmo genético (AG), programação evolutiva e programação genética são exemplos de algoritmos da computação evolucionária (CE). A principal característica que agrupa essas técnicas dentro da CE é que elas são inspiradas na teoria apresentada em “A Origem das Espécies” de Charles Darwin, em 1854. Nesse livro, é proposto que a evolução biológica é um processo de adaptação que pode ser didaticamente dividido em três etapas: reprodução com herança, variação aleatória e seleção natural (CASTRO, 2006). Um indivíduo herda as características de quem o concebeu; porém, variações podem ocorrer aleatoriamente, de geração em geração. As modificações que são benéficas para a adaptação do indivíduo tendem a ser incorporadas pela espécie no decorrer das gerações, pois indivíduos mais adaptados tendem a ter mais descendentes. Dessa maneira, a espécie vai se tornando cada vez mais/melhor adaptada ao meio em vive. Darwin chamou esse processo de seleção natural (CASTRO, 2006).

O AG é uma heurística, que se baseia no processo proposto por Darwin, para a resolução de problemas por meio de uma busca orientada de soluções. Inicialmente formalizado e usado por John Henry Holland em 1975, um AG parte de possíveis soluções,

aqui chamadas de cromossomos, que podem se modificar durante um ciclo, chamado de geração. Espera-se que, no final de cada ciclo, a (nova) geração seja formada por cromossomos (soluções) mais apropriados para resolver o problema em questão (SUMIDA et al., 1990).

Os termos utilizados na descrição de um AG são provenientes da Genética. Por exemplo, como já foi dito, a representação da solução é chamada de cromossomo e ele é dividido em partições, em genes (CASTRO, 2006).

Um AG pode ter implementações distintas, dependendo do seu objetivo e da representação do cromossomo. Por isso, conforme a especificidade do problema, alguns operadores genéticos podem ser incluídos ou suprimidos na criação de novos cromossomos a partir dos cromossomos já existentes. Um fluxograma é mostrado na figura 2.4.

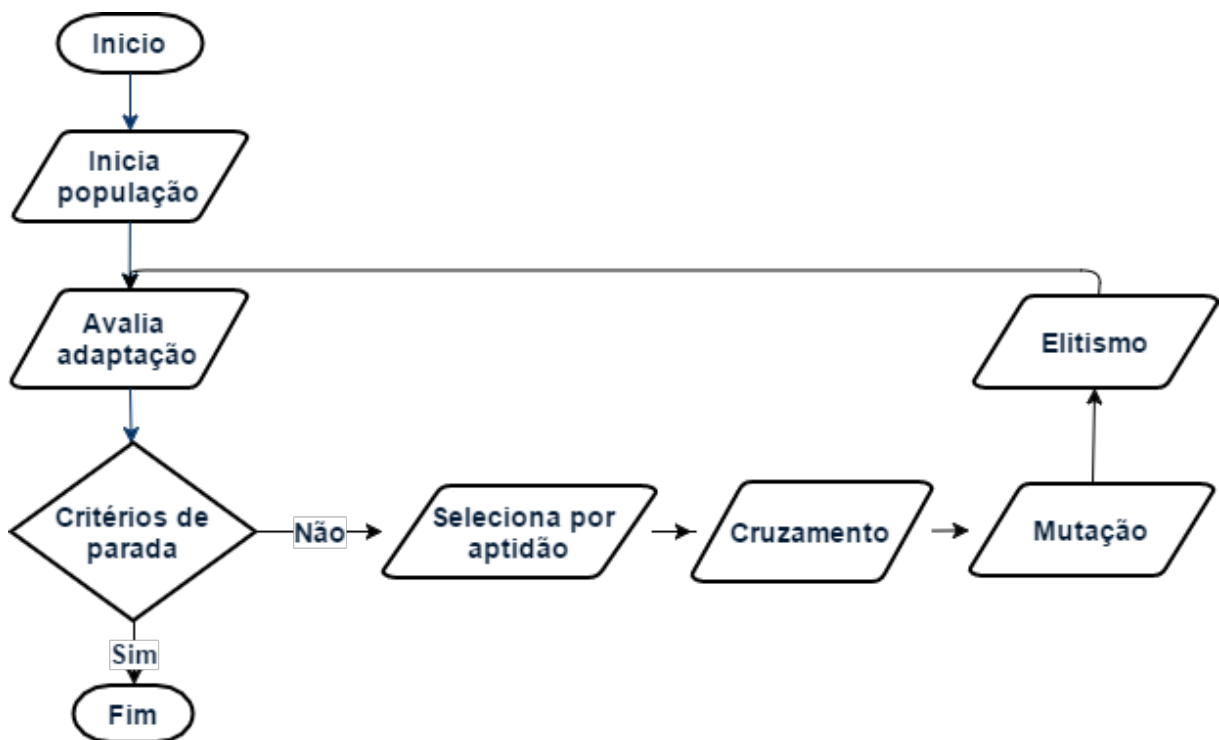


Figura 2.4: Fluxograma genérico de um AG.

As etapas desse fluxograma são explicadas a seguir.

Inicia população: O conjunto de cromossomos que forma cada geração é chamado de população. A primeira geração da população é criada com valores aleatórios, com cada gene respeitando o seu domínio permitido. Ao se criarem cromossomos de modo aleatório,

aumenta-se a abrangência do espaço de soluções percorrido pelo AG. Cada geração possui C cromossomos.

Avalia adaptação: Após a criação de uma geração, os cromossomos têm seu índice de adaptação mensurado. Essa aferição se dá por meio de uma função-objetivo f , existindo a possibilidade de variá-la com o tempo, representando uma mudança no ambiente e, conseqüentemente, no índice de adaptação do indivíduo (SUMIDA et al., 1990).

Crítérios de parada: O objetivo de um AG é encontrar uma solução satisfatória. Assim, uma das condições de parada é um valor pré-determinado para a função-objetivo. Porém, se o AG não encontra esse valor esperado, usa-se como critério de parada uma quantidade máxima de gerações a serem criadas.

Seleção: Análogo à seleção natural descrita por Darwin, esse processo leva em consideração a capacidade de adaptação do indivíduo, permitindo que características úteis à adaptação sejam fixadas na espécie. Quanto mais adaptado um indivíduo ou um cromossomo (no contexto do AG), maiores são as chances de ele se reproduzir. Em muitas aplicações, utiliza-se o algoritmo de seleção chamado de método da roleta (CASTRO, 2006).

O método da roleta particiona um círculo em segmentos, como fatias de uma pizza, de acordo com o resultado da função-objetivo. O tamanho do segmento correspondente ao i -ésimo cromossomo é proporcional ao índice de adaptação, denominado de f_i , que expressa o resultado da função-objetivo calculada com os genes correspondentes a esse i -ésimo cromossomo. Então, sorteiam-se os cromossomos que farão parte da nova geração em função da distribuição de probabilidades determinada por esse método. Claramente, o cromossomo mais adaptado (o de maior f_i) recebe um segmento maior do círculo (uma fatia maior da pizza) e tem maior chance de ser selecionado, por sorteio. A probabilidade do i -ésimo cromossomo ser sorteado é dada por $f_i / (\sum_{j=1}^C f_j)$. A quantidade de cromossomos selecionados aleatoriamente por esse método é a mesma da geração atual; assim, mantém-se constante o tamanho da população de cromossomos. Um mesmo cromossomo pode ser selecionado mais de uma vez. Um exemplo do método da roleta é mostrado na tabela 2.1.

| valor do índice aptidão | graus | tamanho da fatia |
|-------------------------|-------|------------------|
| 20 | 100° | 27,8% |
| 13 | 65° | 18,1% |
| 23 | 115° | 31,9% |
| 16 | 80° | 22,2% |

Tabela 2.1: Considere que, quanto maior o valor de f_i , maior a adaptação do cromossomo. Nesse exemplo, quatro cromossomos foram avaliados, sendo apresentado o valor de f_i de cada cromossomo. Somam-se esses valores de f_i e divide-se o círculo em quatro fatias, sendo o tamanho de cada fatia proporcional a f_i . Por exemplo, a fatia correspondente ao primeiro cromossomo é proporcional a $\frac{20}{20+13+23+16} = \frac{20}{72} = \frac{5}{18} = 27,8\%$. Ou seja, corresponde a um arco de circunferência de ângulo $360^\circ \times \frac{5}{18} = 100^\circ$.

Cruzamento: Os cromossomos selecionados são pareados (formam pares) e, então, podem gerar novos cromossomos por um processo chamado de cruzamento. Esses novos cromossomos são gerados trocando-se genes entre os cromossomos de cada par (CASTRO, 2006). Há diversas maneiras de realizar o cruzamento. Aqui, assume-se que cada par de genes do par de cromossomos são trocados com probabilidade p_c . Por exemplo, considere dois cromossomos com quatro genes cujos valores são binários. Suponha que esses cromossomos sejam representados por 0000 e 1111. Se para a primeira posição dos cromossomos (correspondente ao primeiro gene), ao se sortear um número aleatório x no intervalo $[0, 1]$, obtém-se um número menor que p_c , então os genes são trocados. Suponha que $x < p_c$ para a primeira posição, mas $x \geq p_c$ para as outras três posições. Então, o processo de cruzamento produz os seguintes cromossomos: 1000 e 0111.

Mutação: A fim de introduzir variabilidade, aplica-se, após o cruzamento, o operador genético denominado mutação. Em geral, é conveniente permitir uma “pequena” chance de alteração de um gene; uma alteração “elevada”, em muitos casos, deteriora o desempenho do AG (SUMIDA et al., 1990). Seja p_m a probabilidade do operador de mutação. Aqui, aplica-se esse operador assim. Inicialmente, sorteia-se um número aleatório x entre $[0, 1]$. Se $x < p_m$, aplica-se a mutação do seguinte modo. Toma-se o valor atual do gene como sendo a média de uma distribuição normal e o desvio padrão σ como sendo uma porcentagem dessa média. Então, sorteia-se o novo valor do gene a partir dessa distribuição. Por exemplo, se o valor do gene é 10 e considera-se um desvio padrão $\sigma = 20\%$,

então o valor do gene é obtido a partir de um sorteio que obedece a uma distribuição normal com média 10 e desvio padrão 2. Esse processo é aplicado a todos os genes de todos os cromossomos criados por cruzamento.

Elitismo: O operador genético chamado de elitismo assegura que os cromossomos mais adaptados passarão para a próxima geração de forma inalterada. Considere que, após os processos de cruzamento e mutação, tem-se uma população de cromossomos de tamanho igual ao que foi inicialmente criado. Então, descartam-se ϵ cromossomos, que são substituídos por aqueles escolhidos por elitismo. Por exemplo, se no início haviam 20 cromossomos, após cruzamento e mutação deve também haver 20 cromossomos. Se $\epsilon = 2$, então descartam-se 2 desses 20 cromossomos novos, que são substituídos pelos dois cromossomos mais bem avaliados (de maiores índices f) da geração atual.

Com isso, tem-se uma nova geração de cromossomos, que é avaliada e passa novamente pelos processos de seleção, cruzamento, mutação e elitismo, como mostrado na figura 2.4.

A seguir, apresentam-se alguns detalhes sobre a implementação computacional realizada neste trabalho.

2.4 Sobre a implementação

A lei de Moore, que previa que a capacidade de processamento dobraria a cada 18 meses, tem sido por décadas o porto seguro para a melhoria do desempenho de códigos sequenciais. Porém, algumas limitações físicas começaram a ser alcançadas pelas unidades de processamento central, do inglês CPU (Central Process Unit). Essas limitações se relacionam, por exemplo, ao consumo de energia e à dissipação de calor (CORMEN et al., 2009; COOK, 2013). Com o intuito de contornar tais limitações, computadores com mais unidades de processamento foram criados. Com isso, a forma de programar teve que evoluir e se adaptar: algoritmos tiveram que ser criados levando em conta que a máquina possui mais de um núcleo (CORMEN et al., 2009). Outra evolução veio da necessidade de jogos e interfaces gráficas exigirem cada vez mais recursos de hardware para o tratamento de imagens. As unidades gráficas de processamento, no inglês GPU (Graphics Processing Unit), mostraram desempenho melhor do que os processadores de múltiplo uso, e ainda têm custo mais baixo (COOK, 2013). Assim, CPUs têm sido projetadas para executar

tarefas sequenciais, enquanto GPUs têm melhor desempenho em tarefas paralelizáveis (COOK, 2013).

Compute Unified Device Architecture (CUDA), traduzido livremente como arquitetura de computação de dispositivo unificado, é a interface de aplicação programável disponibilizada pela fabricante NVIDIA para integração com a GPU das placas que fabrica. Pode-se, portanto, alcançar um alto nível de paralelismo, por exemplo, com a placa NVIDIA Tesla K80M, usada nas simulações desta dissertação, com a qual é possível executar 2496 threads em paralelo em seus núcleos de 562 MHz.

A implementação de um algoritmo utilizando GPU difere de sua implementação em código sequencial, e mesmo de implementações multithread que utilizem a CPU. Primeiramente, utilizou-se um reticulado extra para que não fosse necessário sincronizar os dados entre a leitura e a escrita durante o cálculo de um passo de tempo do ACP. Os laços de repetição alinhados para calcular os estados e as transições de estados de cada um dos N elementos do ACP foram substituídos pela configuração de lançamento da função, que é executada pela placa de vídeo, chamada de kernel. A configuração do lançamento do kernel é crucial no desempenho da implementação (COOK, 2013). Um exemplo: para um ACP com dimensão 500×500 , configurou-se o kernel com 250 threads em 1000 blocos de processamento, totalizando as 250000 células do reticulado.

No próximo capítulo, apresenta-se a técnica de identificação de sistemas utilizada neste trabalho.

Capítulo 3

IDENTIFICANDO ACP COM AG

Ao analisar séries históricas de epidemias, alguns parâmetros epidemiológicos podem ser determinados, como, por exemplo, o tempo médio de recuperação, o número de infectados e o número de óbitos associados à doença analisada. Ainda, a partir de dados demográficos, determina-se a média de vida da população. Por isso, suponha que se conheçam as probabilidades (constantes) P_{ic} , P_{im} e P_{rm} , relacionadas a cura de infectados, morte de doentes e morte de recuperados. Entretanto, suponha que não são conhecidos os valores dos parâmetros que expressam as interações entre as três classes de indivíduos que, no nosso modelo em ACP, são K_a , K_q e K_p . De fato, uma das maiores dificuldades num estudo epidemiológico é modelar como as classes S , I e R se afetam. Suponha também que se conhecem os valores normalizados de suscetíveis e infectados em regime permanente. Esses valores são chamados, respectivamente, de s_{alvo} e i_{alvo} (obviamente, $r_{alvo} = 1 - (s_{alvo} + i_{alvo})$). **Nosso objetivo** é, a partir dos valores de P_{ic} , P_{im} , P_{rm} , s_{alvo} e i_{alvo} , utilizar um AG para encontrar os valores dos parâmetros K_a , K_q e K_p , de modo que se obtenha esses mesmos valores normalizados s_{alvo} e i_{alvo} em regime permanente no ACP.

3.1 O ACP

O ACP usado possui 3 estados: suscetível S , infectado I e recuperado R . A figura 3.1 mostra uma visão espacial da evolução temporal desse ACP, conforme as regras descritas a seguir (já mencionadas no capítulo anterior).

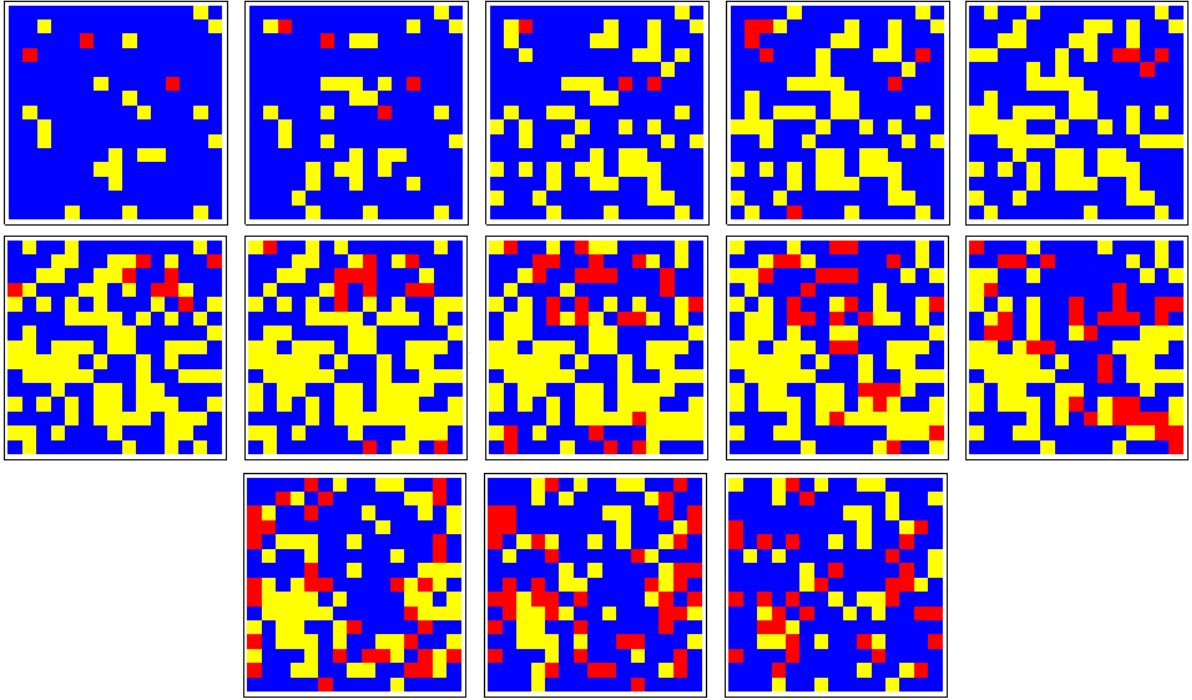


Figura 3.1: Evolução temporal do ACP num reticulado 15×15 . Nessa figura, azul representa indivíduo recuperado; amarelo, suscetível; e vermelho, infectado.

Inicialmente, identifica-se a célula que terá seu estado atualizado. Se a célula está no estado S no passo de tempo t , verifica-se se ela em $t+1$ se tornará uma célula I em função da probabilidade de infecção direta $P_s(i)$; se essa transição não ocorrer, verifica-se se ela se tornará I em função da probabilidade $P_s(i, r)$ de infecção catalisada por R (como descrito na seção 2.1.1). Se em t a célula está no estado I , verifica-se se em $t+1$ ela se tornará uma célula R em função da probabilidade P_{ic} de cura espontânea; se não, verifica-se se ela se tornará uma célula R em função da probabilidade $P_i(r)$ de cura assistida por R ; se não, verifica-se se ela se tornará uma célula S em função da probabilidade P_{im} de morte de I . Se em t a célula está no estado R , verifica-se se em $t+1$ ela se tornará uma célula S em função da probabilidade P_{rm} de morte de R . Esse procedimento é realizado para todas as células do reticulado, a cada passo de tempo. A tabela 3.1 mostra as transições, as probabilidades e a ordem com que as regras são aplicadas.

| estado | transição | função | prioridade |
|------------|-------------------|--------------------------------|------------|
| suscetível | $S \rightarrow I$ | $P_s(i) = 1 - e^{-K_a i}$ | 1 |
| | $S \rightarrow I$ | $P_s(i, r) = 1 - e^{-K_q i r}$ | 2 |
| infectado | $I \rightarrow R$ | P_{ic} | 1 |
| | $I \rightarrow R$ | $P_i(r) = 1 - e^{-K_p r}$ | 2 |
| | $I \rightarrow S$ | P_{im} | 3 |
| recuperado | $R \rightarrow S$ | P_{rm} | 1 |

Tabela 3.1: Regras de transição do ACP por prioridade. As constantes K_a , K_q , K_p , P_{ic} , P_{im} e P_{rm} são positivas. No cálculo das probabilidades $P_s(i)$, $P_s(i, r)$ e $P_i(r)$, i denota o número de vizinhos infectados e r o número de vizinhos recuperados.

O ACP é simulado com vizinhança de Moore de raio $\Gamma = 2$ por 40 passos de tempo, o que é suficiente para que ele atinja seu regime permanente. Com isso, calculam-se as médias das quantidades normalizadas de suscetíveis e infectados, considerando os últimos 5 passos de tempo, a fim de tentar eliminar os efeitos de oscilações de caráter aleatório. Obviamente, a normalização corresponde à soma das células num dado estado dividida por $N = n^2$ (o número total de células do reticulado).

3.2 O AG

O AG é parametrizado conforme explicado a seguir.

Representação: Cada cromossomo contém três genes, um para K_a , um para K_q e um para K_p , no domínio dos números reais variando no intervalo de $[0; 20]$.

Avalia adaptação: Cada geração do AG é composta de C cromossomos. Usa-se a função-objetivo f para avaliar a distância que o resultado de cada cromossomo (determinado pela evolução temporal do ACP) se encontra da solução-alvo. Para isso, primeiramente, simula-se o ACP com as configurações provenientes do cromossomo a ser avaliado. Depois, determinam-se os valores normalizados em regime permanente de suscetíveis, infectados e recuperados oriundos dessa simulação, chamados de s^* , i^* e r^* , respectivamente. Então, calcula-se f , que é dada pelo inverso da distância euclidiana entre o resultado do

ACP e o resultado esperado. No caso da distância ser igual a 0, substituí-se o denominador por 0,000001. Portanto, define-se f assim:

$$f = \frac{1}{\max(\sqrt{(s_{alvo} - s^*)^2 + (i_{alvo} - i^*)^2 + (r_{alvo} - r^*)^2}, 1 \times 10^{-6})}$$

Um cromossomo “bem” adaptado é aquele com “grande” valor de f , capaz de produzir quantidades de suscetíveis, infectados e recuperados em regime permanente “próximas” dos valores-alvo.

Seleção: A adaptação é proporcional ao valor da função-objetivo. Para realizar a seleção, usa-se o método da roleta já descrito na seção 2.3, em que a largura do segmento na roleta associada ao i -ésimo cromossomo é dada por $f_i / (\sum_{j=1}^C f_j)$.

Cruzamento: O cruzamento é feito como descrito na seção 2.3.

Mutação: A mutação é aplicada como apresentada na seção 2.3.

Elitismo: Aqui, toma-se $\epsilon = 1$; ou seja, apenas o cromossomo mais adaptado passa para a próxima geração de forma inalterada.

Condição de término: Caso se atinja o total pré-determinado de gerações ou se obtenha $f > 10000$, interrompe-se o AG.

Os resultados das simulações envolvendo AG e ACP são apresentados no próximo capítulo.

Capítulo 4

RESULTADOS

Este capítulo está dividido em três seções. Na primeira seção, descreve-se como o tamanho do reticulado influencia a precisão dos resultados e o tempo de processamento. Na segunda seção, analisa-se um cenário fictício, no qual se observa que conjuntos distintos de parâmetros do ACP (parte deles obtidos pelo AG) podem gerar resultados similares. Na terceira seção, usa-se o AG em um cenário com dados reais de incidência de varicela referentes à Alemanha em 1999, antes da era da vacinação.

O resultado de uma simulação com o ACP (a partir do cromossomo formado pelos genes K_a , K_q e K_p) é avaliado pela função-objetivo f e pelo erro relativo encontrado em cada uma das classes do modelo SIR. Por exemplo, o erro relacionado à classe dos suscetíveis, chamado de S_ξ , é determinado por:

$$S_\xi = \frac{|s_{alvo} - s^*|}{s_{alvo}}$$

Lembre que s_{alvo} e s^* são, respectivamente, o valor normalizado que se almeja atingir (que pode ter sido obtido, por exemplo, da literatura) e o valor normalizado gerado pelo ACP (usando os valores de K_a , K_q e K_p do cromossomo em questão) no estado estacionário. Os erros I_ξ e R_ξ são definidos de modo similar.

4.1 Escolha do tamanho do reticulado

O tamanho do reticulado tem influência direta em dois aspectos das simulações: tempo de processamento e precisão do resultado.

Numa simulação com o ACP, cada célula do reticulado de dimensão $n \times n$ é analisada, a fim de determinar sua possível transição de estado. As transições $I \rightarrow S$ e $R \rightarrow S$ independem da vizinhança; entretanto, as transições $S \rightarrow I$ e $I \rightarrow R$ dependem. Nas nossas simulações, o raio da vizinhança é constante e vale $\Gamma = 2$. Como tipicamente $\Gamma \ll n$, conclui-se que a complexidade da computação do ACP é $\mathcal{O}(n^2)$. Assim, a dimensão do reticulado têm um impacto quadrático (não linear) no tempo de processamento.

Valendo-se da natureza paralelizável do ACP, utilizou-se uma placa de vídeo para amortecer o custo computacional. Contudo, a quantidade de threads disponíveis é finita (COOK, 2013). A tabela 4.1 mostra como variam o desvio padrão de f e a duração da simulação de 40 passos de tempo no ACP, em função de n . Essa tabela foi assim construída: tomou-se um cromossomo e foram simuladas 50 vezes a evolução temporal desse cromossomo (com parâmetros K_a , K_q , K_p) no ACP. Então, foram calculados os valores médios de f e os correspondentes desvios, levando em conta os últimos 5 passos de tempo de cada simulação. Percebe-se que o desvio padrão diminui com o aumento de n e que o tempo de simulação de 40 passos de tempo do ACP aumenta com n . A última coluna representa o tempo de processamento de uma única geração do AG, caso ela fosse composta por 100 cromossomos (ou seja, é o valor da coluna relacionada à duração da simulação do ACP multiplicada por 100. O tempo associado aos operadores genéticos usados para criar a próxima geração de cromossomos é desprezível, se comparado com o tempo gasto em 40 passos do ACP).

Por exemplo, um reticulado em que cada célula representa um indivíduo da população de 82 milhões da Alemanha em 1999 (WORLDBANK, 2017), deve ter $n = 9055$. Esse seria o valor máximo de n que poderia ser escolhido nesse caso. Com essa escolha, porém, 50 gerações do AG consumiriam dias de processamento com o equipamento disponível, conforme sugere a tabela 4.1.

Como limitante inferior da escolha da dimensão do reticulado, tem-se a precisão dos resultados. O ACP é um modelo probabilista e, quanto menor seu reticulado, maior

o desvio padrão dos resultados, pois menos eventos são considerados nos cálculos dos valores médios. Ainda, n determina o valor mínimo de indivíduos pertencentes a uma classe que podem ser representados no reticulado do ACP. Por exemplo, para $n = 10$, tem-se um reticulado com 100 indivíduos. Se a doença em regime permanente apresenta $s_{alvo} = 0,5\%$, então esse valor de n é inapropriado, pois o menor valor não nulo para a quantidade normalizada de suscetíveis que pode ser encontrado nesse reticulado é 1%. Para se encontrar $s_{alvo} = 0,5\%$, tem-se que escolher $n \geq 15$.

| $n \times n$ | desvio padrão | duração da simulação no ACP | tempo do AG por geração |
|--------------------|-----------------------|-----------------------------|-------------------------|
| 50×50 | $67,1 \times 10^{-5}$ | 31,8 ms | 3,18 seg |
| 100×100 | $32,0 \times 10^{-5}$ | 43,5 ms | 4,35 seg |
| 200×200 | $17,7 \times 10^{-5}$ | 435 ms | 43,5 seg |
| 300×300 | $11,2 \times 10^{-5}$ | 1,46 seg | 2,44 min |
| 400×400 | $8,85 \times 10^{-5}$ | 3,01 seg | 5,01 min |
| 500×500 | $6,34 \times 10^{-5}$ | 6,15 seg | 10,3 min |
| 600×600 | $6,34 \times 10^{-5}$ | 42,3 seg | 1,17 horas |
| 700×700 | $5,10 \times 10^{-5}$ | 1,37 min | 2,28 horas |
| 800×800 | $3,83 \times 10^{-5}$ | 3,59 min | 5,97 horas |
| 900×900 | $3,38 \times 10^{-5}$ | 3,27 min | 5,44 horas |
| 1000×1000 | $2,91 \times 10^{-5}$ | 4,19 min | 6,99 horas |

Tabela 4.1: Influência da dimensão do reticulado na precisão dos resultados e no tempo de processamento do ACP e do AG. Testes executados com NVIDIA Tesla K80 com 2496 cuda cores.

Para o cenário fictício, foi escolhido um reticulado de dimensão de 200×200 ; para a Alemanha, a escolha foi 500×500 ; assim, cada célula equivale a 324 cidadãos.

Em todas as simulações mostradas a seguir, a condição inicial normalizada é $s(0) = 0,09$, $i(0) = 0,01$ e $r(0) = 0,90$. A condição inicial afeta o tempo em que a solução estacionária é atingida, mas não afeta as coordenadas dessa solução (ou seja, não afeta o atrator para o qual o sistema converge).

4.2 Cenário fictício

Considere o ACP com vizinhança de Moore de raio $\Gamma = 2$, $n = 200$ (população total de 40000 indivíduos), $K_a = 0,2$, $K_q = 0,1$, $K_p = 0,1$, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, atinge-se $(s^*, i^*) = (0,09712125; 0,08874075)$, sendo esses dois números as quantidades normalizadas de suscetíveis e infectados em regime permanente. Ainda, obtêm-se $R_0 = 8,86$, calculado usando a expressão (2.4). Como $R_0 > 1$, a doença permanece endemicamente, como mostra a figura 4.1.

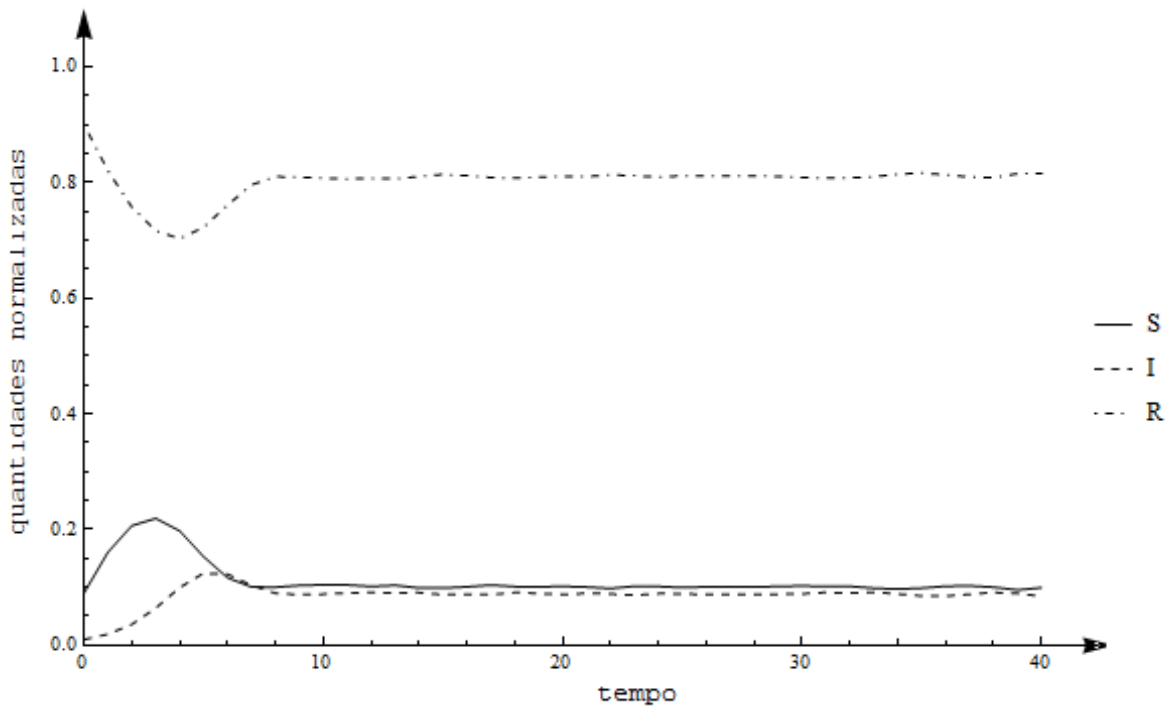


Figura 4.1: Simulação do ACP com $K_a = 0,2$, $K_q = 0,1$, $K_p = 0,1$, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,09712125; 0,08874075)$ e $R_0 = 8,86$.

Identificação via AG

Considere agora que sejam conhecidos $P_{ic} = 0,4$, $P_{im} = 0,2$, $P_{rm} = 0,1$, $s_{alvo} = 0,09712125$ e $i_{alvo} = 0,08874075$ (números obtidos na simulação da figura 4.1). Deseja-se identificar o ACP (ou seja, determinar os valores de K_a , K_q e K_p) a partir desses números, que levem o ACP a convergir para as quantidades s_{alvo} e i_{alvo} mencionadas. Para isso, empregou-se um AG com a seguinte configuração: $C = 50$ cromossomos em cada geração,

elitismo $\epsilon = 1$, probabilidade de cruzamento $p_c = 0,5$, probabilidade de mutação $p_m = 0,1$ e desvio $\sigma = 8\%$. A variação temporal de f ao longo de 50 gerações é mostrada na figura 4.2.

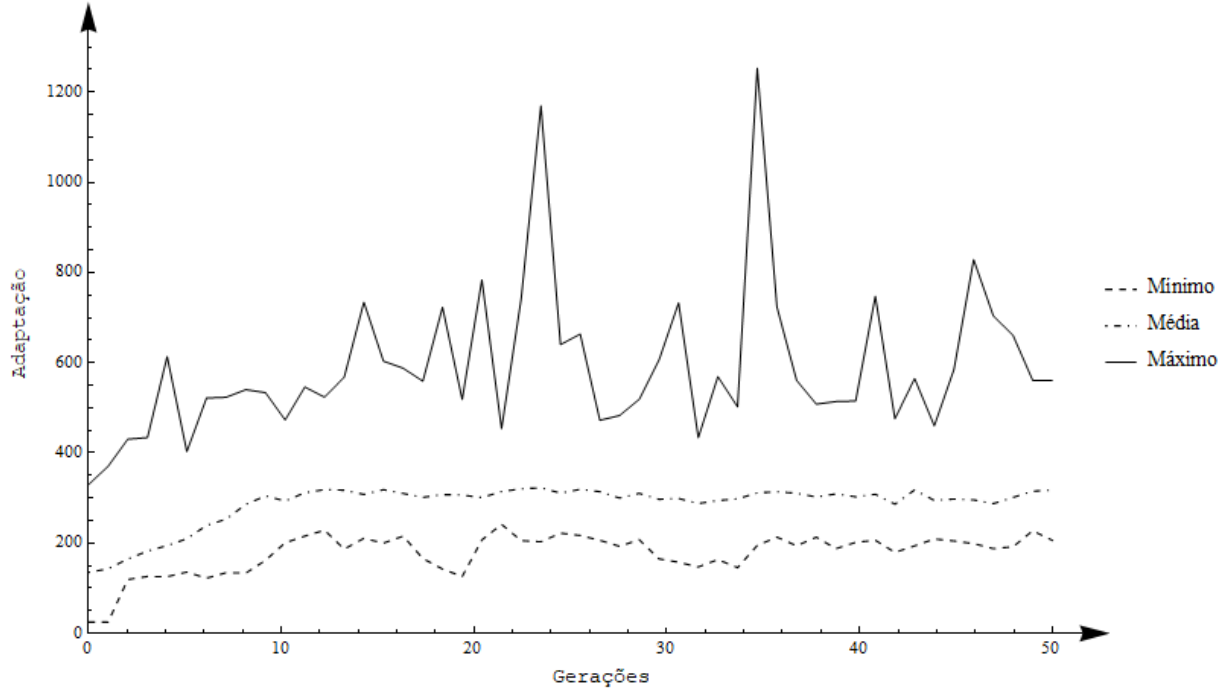


Figura 4.2: Variação da função-objetivo f durante 50 gerações. O AG possui a seguinte configuração: $C = 50$, $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$ e $\sigma = 8\%$. Mostram-se as variações relacionadas aos valores de f máximo, mínimo e médio.

| | geração | K_a | K_q | K_p | s^* | i^* | S_ξ | I_ξ | f |
|---|---------|----------|----------|----------|----------|---------|---------|---------|------|
| a | 50 | 0,691677 | 1,660371 | 0,110969 | 0,095365 | 0,08842 | 1,80 | 0,36 | 560 |
| b | 35 | 0,606023 | 1,660371 | 0,121707 | 0,096325 | 0,08879 | 0,82 | 0,05 | 1253 |
| c | 36 | 0,606023 | 1,660371 | 0,121707 | 0,094845 | 0,08667 | 2,34 | 2,34 | 324 |
| d | 36 | 0,606023 | 1,660371 | 0,121707 | 0,094460 | 0,08581 | 2,74 | 3,30 | 252 |

Tabela 4.2: Cromossomo mais apto encontrado na última geração (caso a , geração 50) e o mais apto dentre todas as gerações (caso b , geração 35). Mostra-se também como esse cromossomo mais apto se comporta na próxima geração (casos c e d , geração 36). Para cada cromossomo, apresentam-se os valores de s^* e i^* com os erros relativos associados, e o valor de f .

As buscas nesse cenário efetuadas pelo AG foram realizadas com $\epsilon = 1$; ou seja, garante-se que uma cópia do cromossomo mais apto da geração atual é levada para a próxima geração sem alterações. Na tabela 4.2, percebe-se que o cromossomo mais apto (encontrado na geração 35) tem desempenho bem inferior na geração seguinte (geração 36). Devido a esse tipo de flutuação decorrente do ACP ser não determinista, o resultado final da busca, correspondente ao caso a , teve índice de adaptação menor que o índice máximo encontrado (na geração 35). Entretanto, os resultados do melhor cromossomo da última geração e do melhor cromossomo de todas as gerações são bastante próximos em termos dos erros S_ξ , I_ξ e R_ξ , como mostrado a seguir.

Resultados no ACP

O cromossomo melhor adaptado da última geração do AG apresenta os seguintes parâmetros: $K_a = 0,691677$, $K_q = 1,660371$ e $K_p = 0,110969$. Simulando o ACP com tais parâmetros e com $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$, obtém-se $(s^*, i^*, r^*) = (0,095365; 0,08842; 0,816215)$. Portanto, os erros em cada classe são $S_\xi = 1,80\%$, $I_\xi = 0,36\%$ e $R_\xi = 0,26\%$. Ainda, obtém-se $R_0 = 18,4$. A evolução temporal do ACP é mostrada na figura 4.3.

Com o melhor cromossomo encontrado no processo de identificação (na geração 35), constituído dos genes $K_a = 0,606023$, $K_q = 1,660371$ e $K_p = 0,121707$, atingiu-se o estado estacionário $(s^*, i^*, r^*) = (0,096325; 0,08879; 0,814885)$; conseqüentemente, $S_\xi = 0,82\%$, $I_\xi = 0,05\%$ e $R_\xi = 0,09\%$. Ainda, $R_0 = 8,86$. A evolução temporal do ACP é mostrada na figura 4.4.

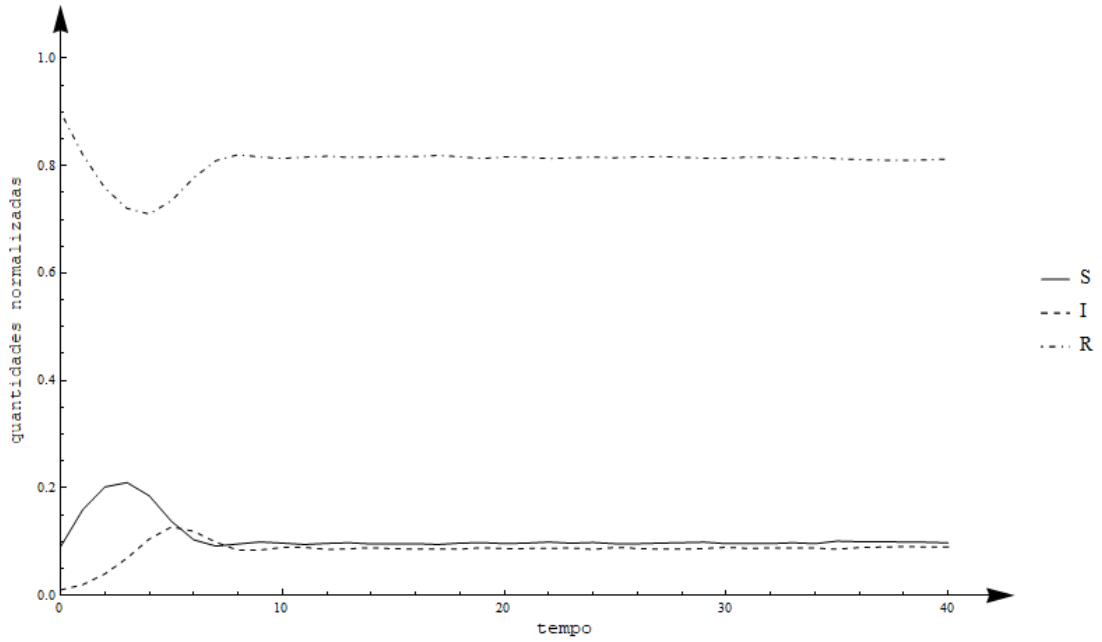


Figura 4.3: Simulação do ACP com o cromossomo mais apto da geração 50, para o qual $K_a = 0,691677$, $K_q = 1,660371$ e $K_p = 0,110969$. Ainda, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,095365; 0,08842)$ e $R_0 = 18,4$.

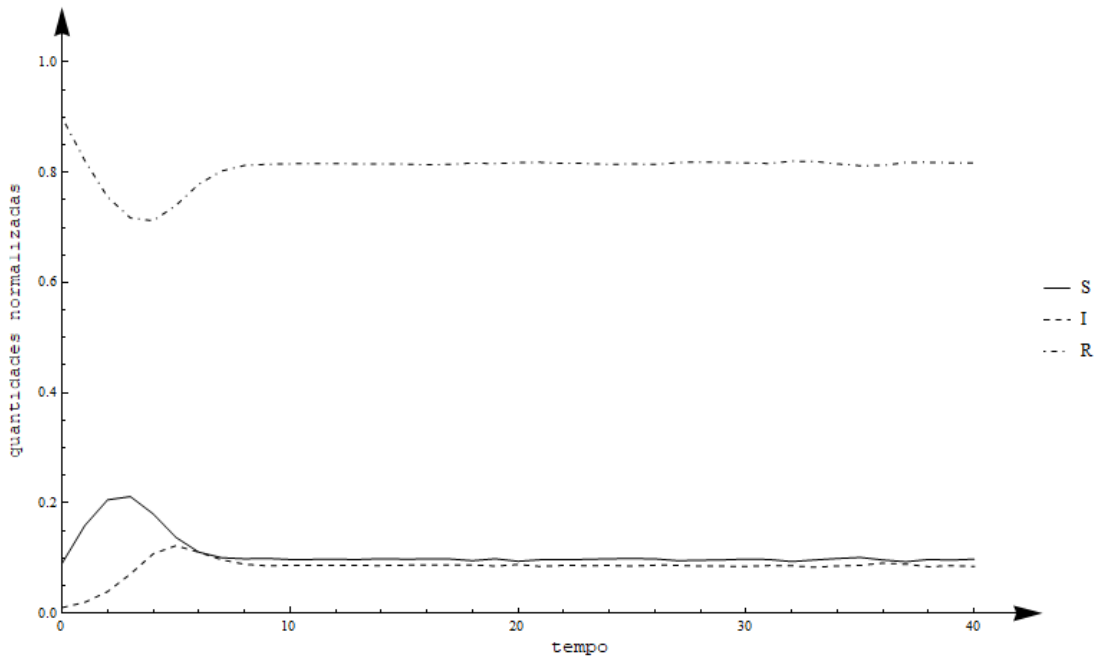


Figura 4.4: Simulação do ACP com o cromossomo mais apto da geração 35, para o qual $K_a = 0,606023$, $K_q = 1,660371$ e $K_p = 0,121707$. Ainda, $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$. Com isso, obtêm-se $(s^*, i^*) = (0,096325; 0,08879)$ e $R_0 = 8,86$.

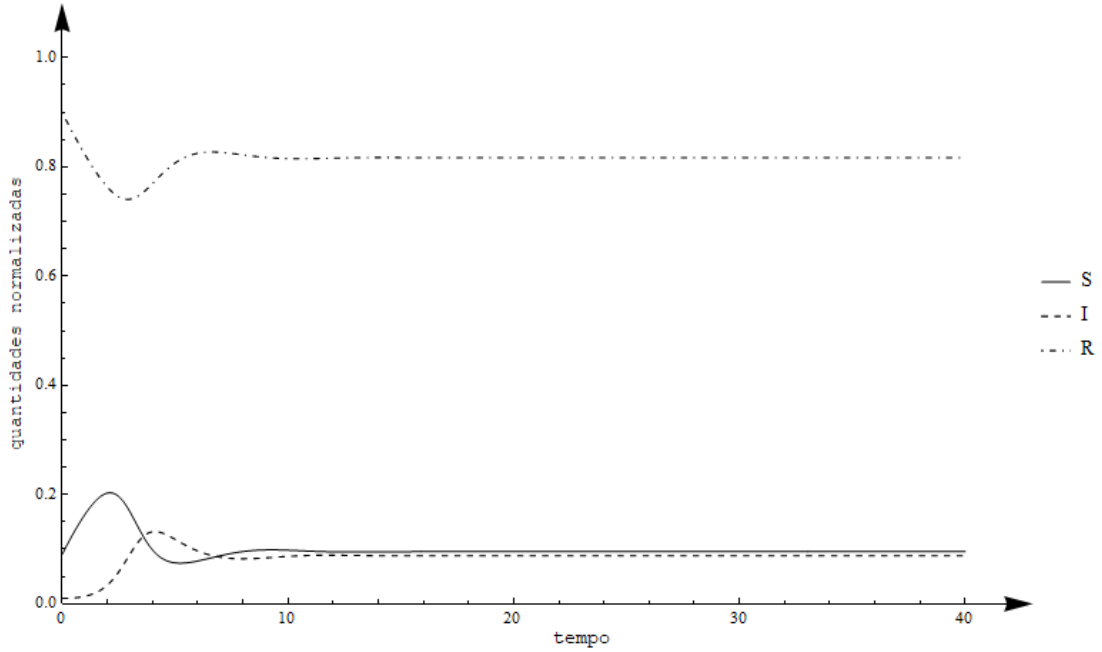


Figura 4.5: Simulação do modelo em EDO com $a = 7,506961462$, $b = 0,401266682$, $c = 0,01391088$, $e = 0,101168197$, $q = 2,985121759$ e $p = 0,660041264$. Valores correspondentes ao melhor cromossomo da geração 50.

Observe que o valor de R_0 para a solução-alvo é de 8,86, enquanto que para o melhor cromossomo da última geração, tem-se $R_0 = 18,4$. Apesar dos erros relativos nas quantidades normalizadas de S , I e R serem inferiores a 2%, o R_0 alvo é metade daquele correspondente a esse cromossomo. Esse tipo de variação “grande” de R_0 com erros relativos “pequenos” foi encontrado em muitas outras buscas do AG. Observe também a semelhança entre as figuras 4.1, 4.3 e 4.4, essa última correspondente ao melhor cromossomo encontrado (na geração 35). Para esse cromossomo, o valor de R_0 coincide com o da solução-alvo, apesar dos valores de K_a , K_q e K_p encontrados serem bem diferentes daqueles usados para gerar essa solução. Esses resultados estão resumidos na tabela 4.3. A relevância desses resultados é discutida no próximo capítulo.

| figura | K_a | K_q | K_p | s^* | i^* | r^* | R_0 |
|--------|----------|----------|----------|----------|---------|----------|-------|
| 4.1 | 0,2 | 0,1 | 0,2 | 0,097121 | 0,08874 | 0,814139 | 8,86 |
| 4.3 | 0,691677 | 1,660371 | 0,110969 | 0,095365 | 0,08842 | 0,816215 | 18,4 |
| 4.4 | 0,606023 | 1,660371 | 0,121707 | 0,096325 | 0,08879 | 0,814885 | 8,86 |

Tabela 4.3: Números correspondentes à solução-alvo (figura 4.1), ao melhor cromossomo da última geração (figura 4.3) e ao melhor cromossomo dentre todas as gerações (figura 4.4).

A figura 4.5 exibe a evolução temporal obtida pela resolução numérica das equações (2.1), com os valores dos parâmetros determinados usando as expressões apresentadas na subseção 2.2.1, para o melhor cromossomo da última geração. Repare a semelhança dessa figura com as figuras 4.1, 4.3 e 4.4.

4.3 Cenário realista

Nesse cenário, foram utilizados dados epidemiológicos da incidência de varicela (HEININGER; SEWARD, 2006) na Alemanha no ano de 1999 (MORAES, 2016; NARDONE et al., 2007), antes da era de vacinação (pois nosso modelo não considera a vacinação de indivíduos, ou seja, a transição $S \rightarrow R$). Assumiu-se que um passo de tempo da simulação corresponde a um dia de tempo real. A infecção de varicela dura em média uma semana; portanto, $P_{ic} = \frac{1}{7}$ (ANDERSON; MAY, 1992). Há casos de mortalidade, porém são raros; por isso, assumiu-se que a probabilidade de morte pela doença é igual à probabilidade de morte por outras causas, ou seja, $P_{im} = P_{rm}$. Como a expectativa de vida era de 78 anos em 1999, então $P_{im} = P_{rm} = \frac{1}{365 \times 78}$. A população total foi estimada em 82 milhões (WORLDBANK, 2017), tendo 760 mil ocorrências de catapora na Alemanha no ano de 1999. Para esse cenário, $(s_{alvo}, i_{alvo}, r_{alvo}) = (0,023; 0,00018; 0,97682)$ (MORAES, 2016; MORAES; MONTEIRO, 2016). O valor de R_0 foi estimado em 5,5 (NARDONE et al., 2007).

A busca com o AG para esse cenário foi executada durante 30 gerações, usando a mesma configuração empregada no cenário anterior ($\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$), com $C = 200$ cromossomos em cada geração. Como mostra a figura 4.6, na geração 18,

o valor de f se estabilizou em torno de 4535. O cromossomo mais apto levou o ACP de reticulado 500×500 a convergir para $(s^*, i^*, r^*) = (0,023089; 0; 0,976911)$, sendo formado pelos genes $K_a = 0,281866$, $K_q = 0,082925$ e $K_p = 2,876278$.

Os erros em relação às quantidades de suscetíveis ($S_\xi = 0,4\%$) e de recuperados ($R_\xi = 0,009\%$) são bastante pequenos; entretanto, o erro em relação à classe dos infectados é de $I_\xi = 100\%$, pois $i^* = 0$. Além disso, obtém-se $R_0 = 0$. Esse resultado não condiz com os dados reais.

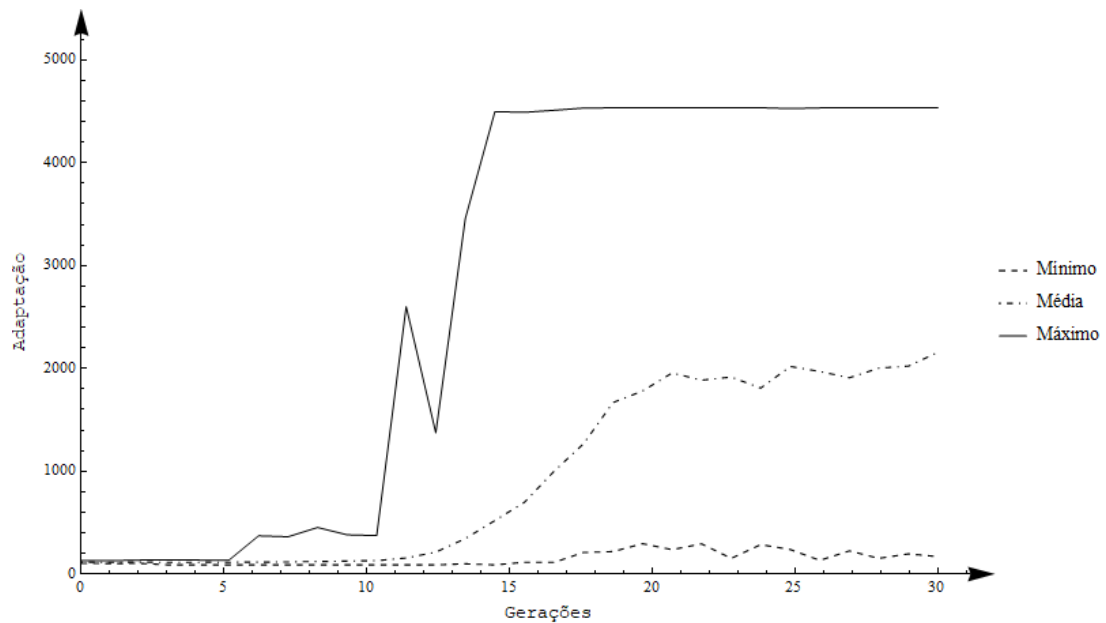


Figura 4.6: Variação de f durante 30 gerações do AG para Alemanha, com $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$ e população de $C = 200$. Mostram-se os valores mínimo, médio e máximo de f em cada uma das gerações.

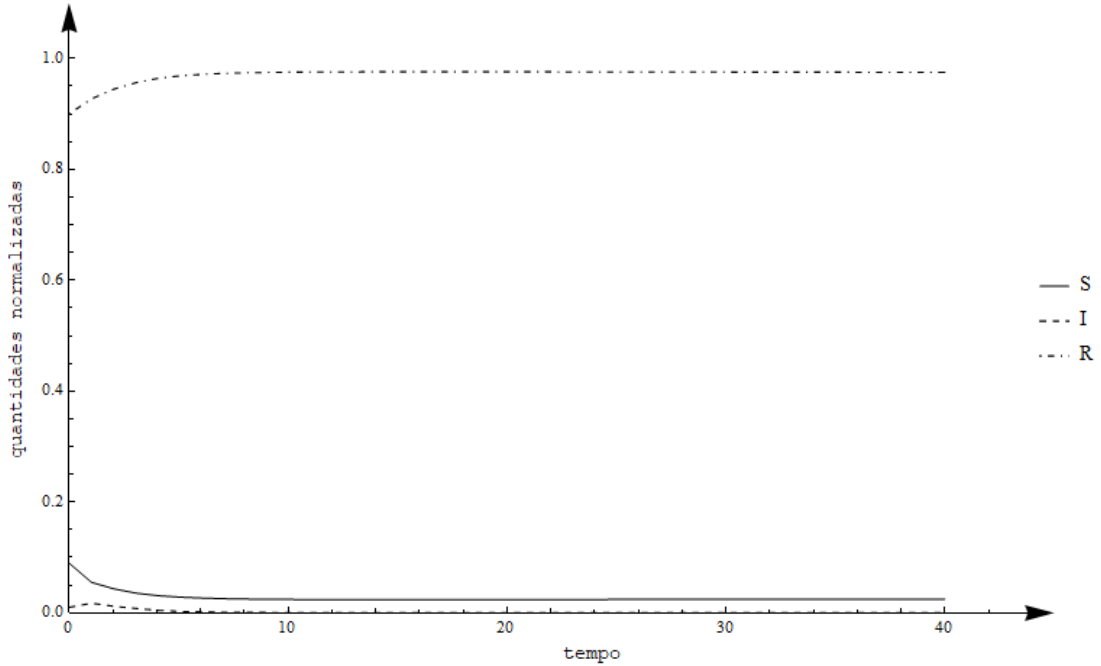


Figura 4.7: Evolução temporal do ACP com $P_{ic} = 0,14$, $P_{im} = P_{rm} = 0,000035$, $K_a = 0,281866$, $K_q = 0,082925$ e $K_p = 2,876278$. Com isso, atinge-se o estado estacionário com $(s^*, i^*) = (0,023089; 0)$, em que a doença desaparece.

Ao se observar a evolução temporal no ACP, mostrada na figura 4.7, nota-se que a doença desaparece na população, em média, após uma semana. Isso ocorre devido à existência de um número “pequeno” de indivíduos suscetíveis. Tais indivíduos são necessários para dar fôlego à preservação da doença na população. De fato, quando os infectados perdem contato com os suscetíveis, a doença naturalmente desaparece. No cenário fictício, as “altas” probabilidades de morte de recuperados e de infectados assegura a preservação da doença na população hospedeira, pois sempre há novos suscetíveis para serem infectados.

Com o propósito de encontrar resultados com menor erro I_ξ , foram realizadas diversas alterações no AG e no ACP. Algumas dessas alterações são listadas a seguir, assim como seus resultados.

1) Função de adaptação com log

Os valores-alvo $s^* = 0,023$ e $i^* = 0,00018$ variam em duas ordens de grandeza. Já $r^* = 0,97682$ e $i^* = 0,00018$ variam em três ordens de grandeza. A fim de suavizar essa discrepância, de modo que s^* , i^* e r^* tenham influência similar na determinação do índice

de adaptação, modificou-se a função f utilizando o logaritmo de base 10 no cálculo da distância de cada uma das classes do modelo SIR do seguinte modo:

$$f_{log} = \frac{1}{\max(\sqrt{g(s_{alvo}, s^*) + g(i_{alvo}, i^*) + g(r_{alvo}, r^*)}, 1 \times 10^{-6})}$$

sendo:

$$g(x_{alvo}, x^*) = [\log_{10}(x_{alvo} - x^*)]^2$$

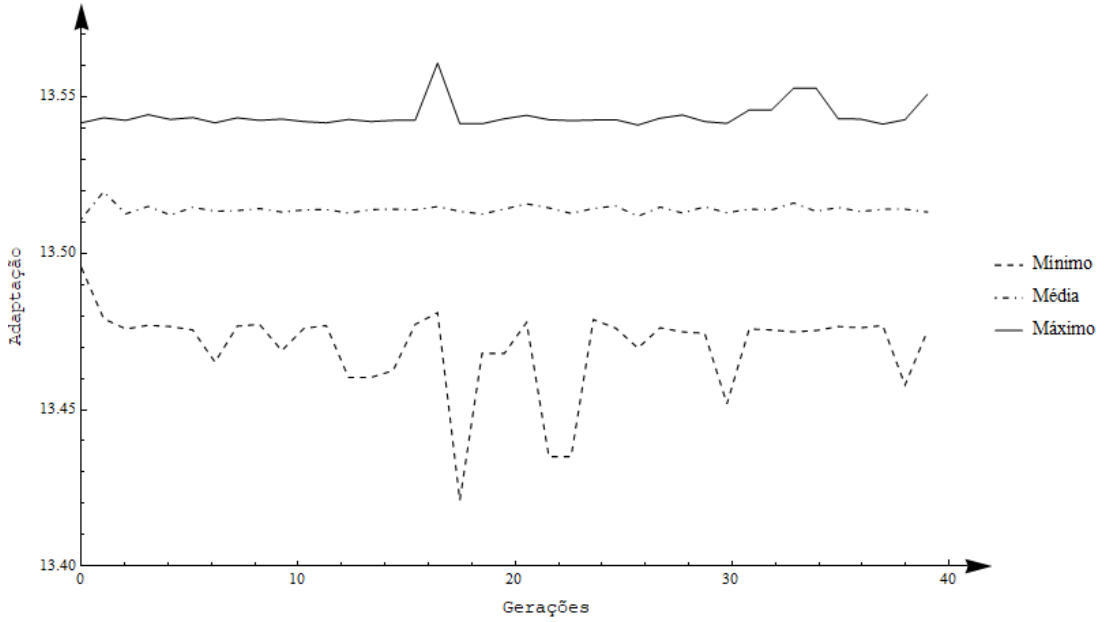


Figura 4.8: Evolução do AG utilizando f_{log} com $C = 200$, $\epsilon = 1$, $p_c = 0,7$, $p_m = 0,5$ e $\sigma = 8\%$.

Como mostra a figura 4.8, o AG não evoluiu, mantendo os índices de adaptação máximo e médio praticamente constantes. Os valores de p_c e p_m usados anteriormente não levam a resultados melhores do que aqueles com $p_c = 0,7$ e $p_m = 0,5$, utilizados nessa figura. O melhor cromossomo foi encontrado na geração 17 para $K_a = 6,968583$, $K_q = 3,465863$ e $K_p = 3,391813$. Com isso, obtém-se $(s^*, i^*, r^*) = (0,014816; 0; 0,985184)$, como mostra a figura 4.9, e erros $S_\xi = 35,6\%$, $I_\xi = 100\%$ e $R_\xi = 0,85\%$.

A substituição da função f por f_{log} fez com que a diferença absoluta da adaptação de cada cromossomo ficasse menor. Com isso, o método de seleção do AG teve que escolher os cromossomos com notas mais homogêneas, o que, por sua vez, acabou diminuindo a pressão seletiva. Talvez, por isso, o AG tenha convergido para soluções piores do que aquelas encontradas usando a função original f .

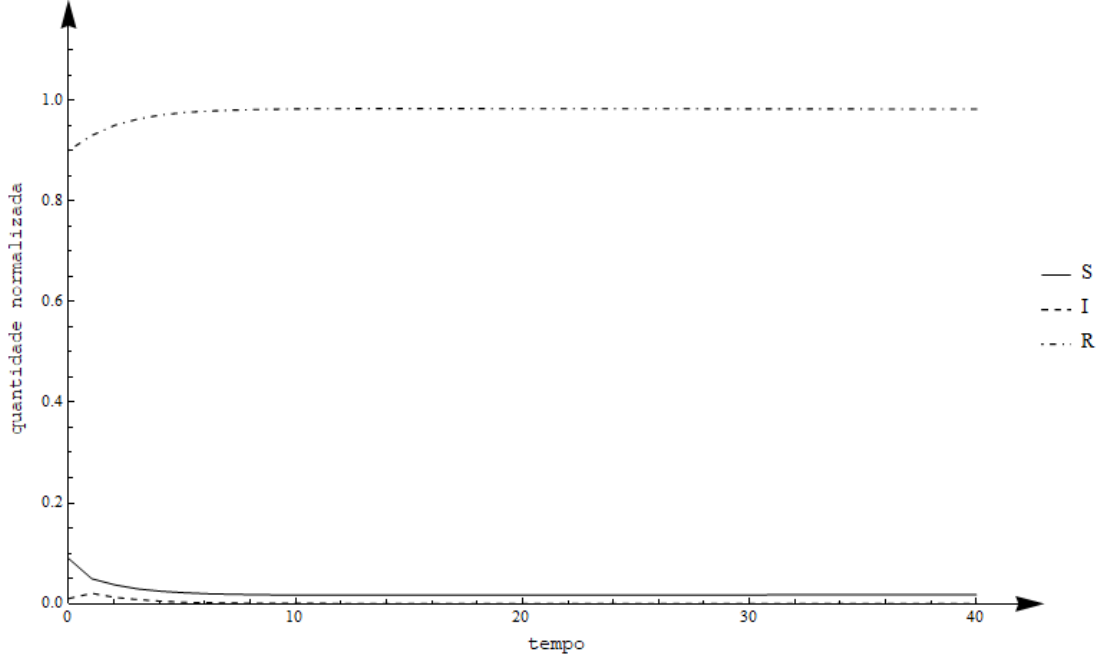


Figura 4.9: Evolução temporal no ACP com o cromossomo mais apto encontrado pelo AG com f_{log} , para o qual $K_a = 6,968583$, $K_q = 3,465863$ e $K_p = 3,391813$. Com isso, converge-se para $(s^*, i^*) = (0,014816; 0)$ e erros $S_\xi = 35,6\%$ e $I_\xi = 100\%$.

2) Função de adaptação f proporcional a i^*

A fim de tentar forçar a seleção de cromossomos que levem os infectados a serem preservados na população hospedeira, usou-se a seguinte nova função de adaptação:

$$f_{prop} = \frac{i^* N}{\max(\sqrt{(s_{alvo} - s^*)^2 + (i_{alvo} - i^*)^2 + (r_{alvo} - r^*)^2}, 1 \times 10^{-6})}$$

que é a função f usada originalmente multiplicada pela quantidade de infectados no reticulado. Assim, $f_{prop} = 0$ se $i^* = 0$.

Nas execuções do AG com $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,9$, $\sigma = 8\%$ e com f_{prop} , houve apenas dois cromossomos tais que $i^* \neq 0$, dentre 6200 (31 gerações do AG, $C = 200$ cromossomos por geração). Esses cromossomos são dados por $(K_a, K_q, K_p) = (3,369974; 1,154762; 7,187220)$ e $(0,299626; 0,082925; 4,823841)$, para quais se obtêm $(s^*, i^*) = (0,016973; 0,000001)$ e $(s^*, i^*) = (0,023943; 0,000002)$, respectivamente. Obviamente, a preservação dos infectados no reticulado do ACP não é encontrada de forma robusta usando f_{prop} .

3) Aumentar a exposição dos suscetíveis aos infectados

Como mudanças nos valores dos parâmetros e na escolha da função de avaliação f do AG não surtiram o efeito desejado, tentou-se modificar o modelo em ACP. Os infectados desaparecem da população hospedeira quando não mais conseguem infectar (encontrar) suscetíveis. Para aumentar o contato entre S e I , após cada passo de tempo na simulação do ACP, decidiu-se mover, de modo aleatório, todos os indivíduos, colocando-os em posições diferentes das ocupadas no passo de tempo anterior. Isso foi feito preservando as quantidades de suscetíveis, infectados e recuperados daquele passo de tempo, além das regras e das probabilidades usadas na atualização dos estados. Portanto, o modelo em EDO continua sendo aquele expresso pelas equações (2.1). A diferença é que a localização espacial dos vizinhos de cada célula do reticulado se modifica a cada passo de tempo.

Primeiro, incorporou-se essa modificação no ACP no cenário fictício, tomando o AG com $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$. Conforme observado na figura 4.10, no final atinge-se um índice de adaptação $f = 19495$, cerca de 10 vezes maior do que o melhor resultado encontrado sem o rearranjo dos indivíduos (mostrado na figura 4.2).

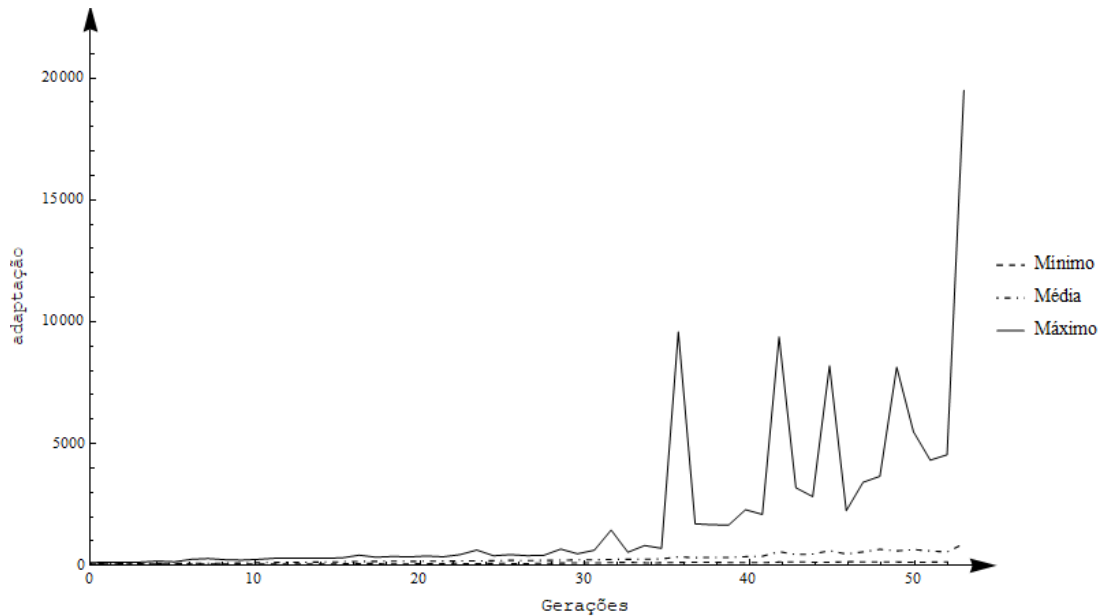


Figura 4.10: Identificação com o AG no cenário fictício usando rearranjo no ACP. Configuração do AG: $\epsilon = 1$, $p_c = 0,5$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$. Na geração 53, obteve-se $f = 19495$.

O melhor resultado foi encontrado na geração 53, a última, na qual se ultrapassou o limiar da condição de término da simulação (que era $f = 10000$). O ACP identificado tem como parâmetros $K_a = 0,338939$, $K_q = 0,072136$ e $K_p = 0,093755$. Com isso e com $P_{ic} = 0,4$, $P_{im} = 0,2$ e $P_{rm} = 0,1$, obtém-se no ACP o seguinte estado estacionário $(s^*, i^*, r^*) = (0,097125, 0,088775, 0,8141)$, com $R_0 = 12,7$, cujos erros valem $S_\xi = 0,004\%$, $I_\xi = 0,04\%$ e $R_\xi = 0,005\%$, como mostra a figura 4.11. Note que esses erros são bem menores do que aqueles da tabela 4.2.

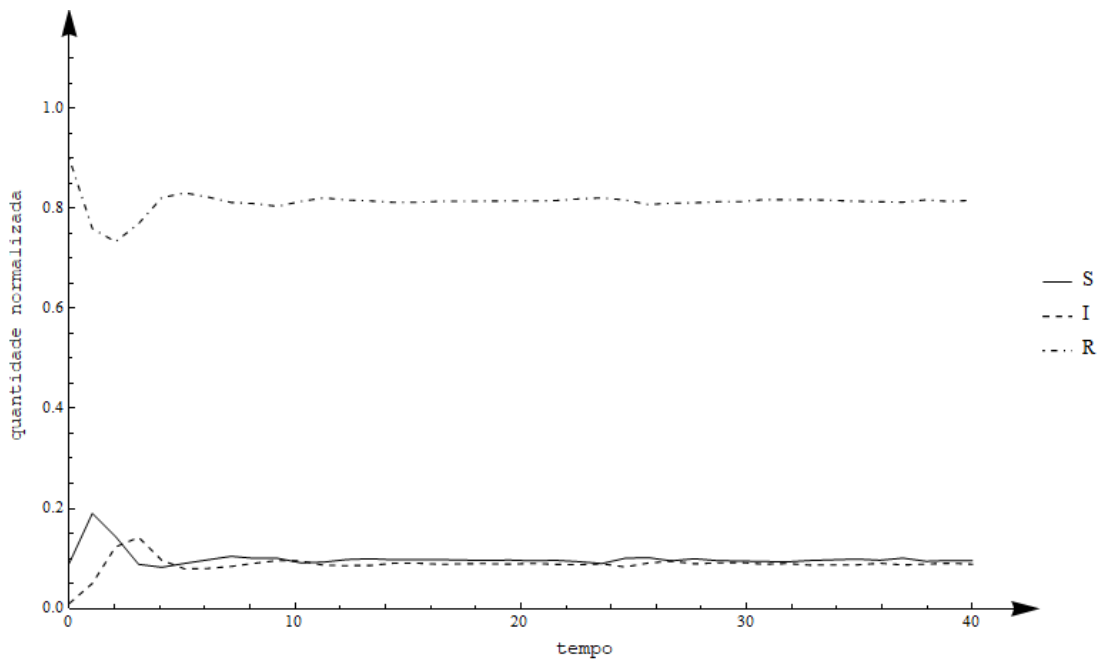


Figura 4.11: O ACP no cenário fictício com rearranjo com $K_a = 0,338939$, $K_q = 0,072136$ e $K_p = 0,093755$ (além de $P_{ic} = 0,4$, $P_{im} = 0,2$, $P_{rm} = 0,1$). Converte-se para $(s^*, i^*) = (0,097125, 0,088775)$, com $S_\xi = 0,004\%$ e $I_\xi = 0,04\%$. Ainda, $R_0 = 12,7$.

Como o rearranjo surtiu efeito positivo na identificação de cenário fictício, ele foi também empregado no cenário da Alemanha. O melhor cromossomo foi encontrado na geração 18, com $\epsilon = 1$, $p_c = 0,7$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$.

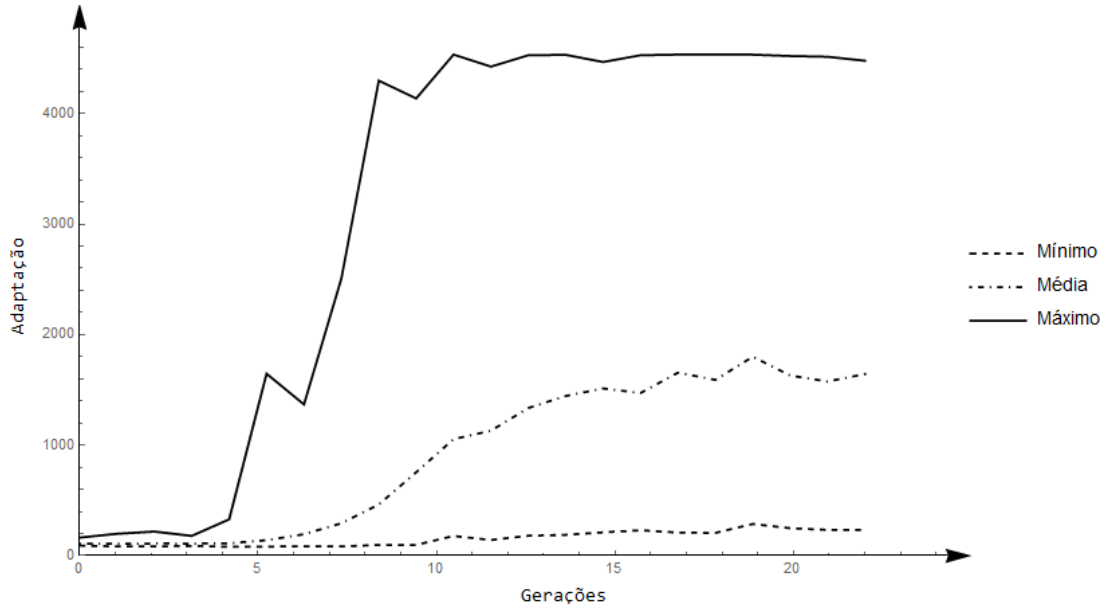


Figura 4.12: Evolução do AG para o cenário da Alemanha com rearranjo durante 50 gerações, com $\epsilon = 1$, $p_c = 0,7$, $p_m = 0,1$, $\sigma = 8\%$ e $C = 100$.

O melhor cromossomo encontrado é formado por $K_a = 0,629071$, $K_q = 0,0649687$ e $K_p = 0,186962$. O atrator no ACP tem como coordenadas normalizadas $(s^*, i^*, r^*) = (0,023091; 0; 0,976909)$, sendo $R_0 = 0$ e erros $S_\xi = 0,4\%$, $I_\xi = 100\%$ e $R_\xi = 0,009\%$. Tais erros são idênticos aos da figura 4.7, com f original e sem rearranjo. Como nos casos anteriores, atrator com $i^* > 0$ não foi encontrado em praticamente nenhuma amostra durante os testes, mostrando que rearranjar o reticulado não surtiu o efeito esperado no caso realista. A evolução do melhor cromossomo no ACP é mostrada na figura 4.13.

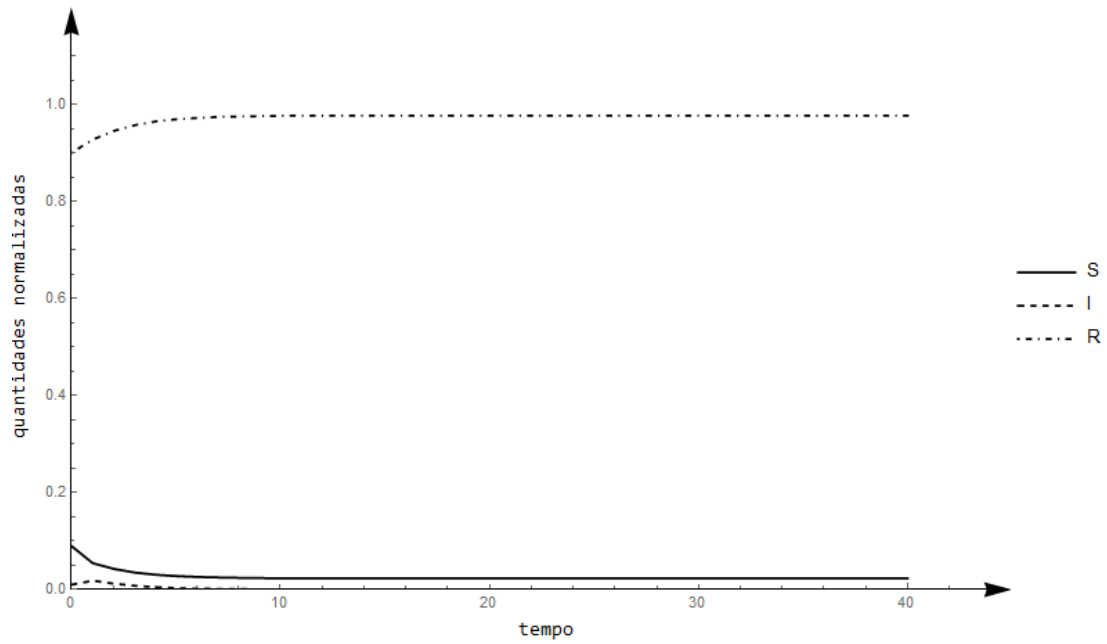


Figura 4.13: Evolução do ACP por quarenta passos de tempo para o cenário da Alemanha com rearranjo, com $K_a = 0,629071$, $K_q = 0,0649687$ e $K_p = 0,186962$. Em regime permanente, obtêm-se $(s^*, i^*) = (0,023091; 0)$ com erro $S_\xi = 0,4\%$ e $I_\xi = 100\%$.

No próximo capítulo, discutem-se os resultados apresentados nesta dissertação.

Capítulo 5

CONCLUSÕES

Nesta dissertação, buscou-se identificar os parâmetros de um modelo SIR formulado em termos de ACP usando AG. O modelo em ACP possui seis parâmetros (MORAES, 2016). Os valores de três podem ser encontrados na literatura. Esses parâmetros são: probabilidade de cura (P_{ic}), probabilidade de óbito devido à doença (P_{im}) e probabilidade de óbito por outras causas (P_{rm}). Os outros três parâmetros, chamados de K_a , K_q e K_p , relacionam-se, respectivamente, a infecção de S por I , infecção de S por I catalisada por R , cura de I influenciada por R . Valores para tais parâmetros são difíceis de serem encontrados na literatura.

Algumas lições foram aprendidas com este trabalho:

- No cenário fictício, em que as quantidades s_{alvo} , i_{alvo} e r_{alvo} são de mesma ordem de grandeza, o processo de identificação usando AG foi realizado de modo satisfatório, no sentido de que o AG consegue encontrar cromossomos (formados pelos genes K_a , K_q e K_p) que levam o ACP a convergir para as quantidades desejadas. Entretanto, cromossomos com valores muito diferentes entre si podem levar a resultados muito próximos em termos dos erros S_ξ , I_ξ e R_ξ , como mostrado na tabela 4.3.
- O valor de R_0 é usualmente empregado para dimensionar estratégias de controle de um ponto de vista prático (por exemplo, estimar a quantidade de vacinas a serem distribuídas para combater um surto epidêmico numa região (ANDERSON; MAY, 1992)). Quanto maior R_0 , maior o esforço necessário para controlar/erradicar a doença. No cenário fictício, verificou-se que a solução-alvo (mostrada na figura 4.1)

pode ser reproduzida por um mesmo modelo com valores de parâmetros bastante distintos, levando a números de R_0 próximos (figura 4.4) ou distantes (figura 4.3) dessa solução-alvo. Isso talvez explique por que uma mesma doença, no caso, a varicela, apresenta valores estimados de R_0 tão diferentes para países que são vizinhos geograficamente. De fato, para varicela, $R_0 = 16,9$ na Holanda; já na Bélgica, $R_0 = 6,5$ (NARDONE et al., 2007; GANDINI; MONTEIRO, 2017).

- Para o cenário realista, o AG usado no cenário fictício foi capaz de encontrar soluções com erros diminutos nas quantidades de suscetíveis e de recuperados, mas, nesses casos, os infectados desapareciam da população hospedeira (figura 4.7). No cenário realista, a quantidade-alvo de infectados é, na verdade, muito pequena (da ordem de 0,02%), mas não é nula. Diversas modificações no AG foram feitas, sem sucesso efetivo. Talvez, o que tenha dificultado o processo de identificação não seja a configuração do AG usado, mas sim o modelo em ACP.
- Sabe-se que, quanto menor uma população hospedeira, mais facilmente uma doença é naturalmente erradicada (MONTEIRO; CHIMARA; CHAUI-BERLINCK, 2006). Considere, por exemplo, uma cidade pequena. Suponha que ninguém, num dado instante, tenha varicela, até que chega um visitante com essa doença e contamina alguns dos moradores. Esses se curam e/ou contaminam outros, mas, após um certo tempo, a doença desaparece daquela cidade, porque todos aqueles que foram infectados ou se curaram ou morreram. A doença só reaparece nessa cidade se for reintroduzida. Nossa suspeita é que o modelo SIR aqui usado para representar a propagação de varicela na Alemanha está incompleto. Ele carece de um termo que leve em conta a reintrodução da doença, de tempos em tempos, por turistas infectados ou por moradores que saem do país sadios, mas voltam contaminados. Ou, então, pode-se tomar a condição de contorno do reticulado como sendo aberta, considerando a influência das regiões vizinhas no processo de propagação da doença. A influência de regiões vizinhas ou de turistas doentes pode não ser relevante para que uma doença se torne endêmica numa região em que a mortalidade de infectados e recuperados é “grande” (como no cenário fictício), e quando se garante que eles serão substituídos por suscetíveis. Essa garantia, porém, não existe no mundo real.

- Tentou-se contornar a falta do termo associado à reintrodução de infectados promovendo um rearranjo na posição dos indivíduos a cada passo de tempo. Por um lado, isso levou a resultados muito bons no cenário fictício, com menores erros em S_ξ , I_ξ e R_ξ , mas com R_0 cerca de 40% acima do esperado ($R_0 = 12,7$ para a figura 4.11, enquanto $R_0 = 8,86$ para a figura 4.1, que mostra a solução-alvo). Por outro lado, a doença não foi preservada na população hospedeira no caso da Alemanha (como ilustrado pela figura 4.13). Ainda, esse rearranjo soa pouco aderente à realidade, pois normalmente doentes convalescem permanecendo em repouso, em vez de passearem a cada passo de tempo pela região que habitam.

Em geral, séries epidemiológicas estão distorcidas por diferentes fatores, como diagnóstico incorreto e sub-notificação. Por exemplo, no Brasil, a varicela não é notificada compulsoriamente ao SINAN, Sistema Nacional de Agravos de Notificação (MINISTÉRIO da SAÚDE, 2014b), o que diminui a precisão dos dados e dificulta identificar corretamente os parâmetros de modelos epidemiológicos. Este trabalho mostrou que: 1) os mesmos dados, dentro de uma certa precisão (da ordem de 1-2%), podem ser reproduzidos por conjuntos de parâmetros bem distintos, levando a valores de R_0 também bem distintos; 2) doenças com grande quantidade de recuperados (imunes) talvez precisem ser reintroduzidas de tempos em tempos para serem preservadas, mesmo numa população hospedeira “grande” como a de um país como a Alemanha. Analisar um modelo SIR com essa característica é tema para trabalho futuro.

Referências Bibliográficas

- AGUIRRE, L. A. *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. 3. ed. Belo Horizonte: Editora UFMG, 2015.
- ANDERSON, R. M.; MAY, R. M. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1992.
- BILLINGS, S. A.; YANG, Y. Identification of probabilistic cellular automata. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, v. 33, p. 225–236, 2003.
- BONITA, R.; BEAGLEHOLE, R.; KJELLSTROM, T. *Introdução à Epidemiologia*. 2. ed. São Paulo: Grupo Editorial Nacional, 2010.
- CAMERON, D.; JONES, I. G. John Snow, the broad street pump and modern epidemiology. *International Journal of Epidemiology*, v. 12, p. 393–396, 1983.
- CASTRO, L. N. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. 1. ed. New York: Chapman and Hall, 2006.
- COOK, S. *CUDA Programming: A Developer’s Guide to Parallel Computing with GPUs*. 1. ed. San Francisco: Morgan Kaufmann Publishers, 2013.
- CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. *Introduction to Algorithms*. 3. ed. Cambridge: The MIT Press, 2009.
- DIETZ, K.; HEESTERBEEK, J. Daniel Bernoulli’s epidemiological model revisited. *Mathematical Biosciences*, v. 180, p. 1–21, 2002.
- GANDINI, D. M.; MONTEIRO, L. H. A. Identificando parâmetros de um modelo epidemiológico usando algoritmo genético. *Proceedings of the IV World Congress on Systems Engineering and Information Technology*, Guimarães, Portugal, 2017.

- GANGULY, N.; MAJI, P.; SIKDAR, B. K.; CHAUDHURI, P. P. Design and characterization of cellular automata based associative memory for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, v. 34, p. 672–678, 2004.
- GERARDI, D. O.; MONTEIRO, L. H. A. System identification and prediction of dengue fever incidence in Rio de Janeiro. *Mathematical Problems in Engineering*, v. 2011, p. 720304, 2011.
- GLOMSKI, M.; OHANIAN, E. Eradicating a disease: Lessons from mathematical epidemiology. *The College Mathematics Journal*, v. 43, p. 123–132, 2012.
- HEININGER, U.; SEWARD, J. F. Varicella. *The Lancet*, v. 368, p. 1365–1376, 2006.
- HETHCOTE, H. W. The mathematics of infectious diseases. *SIAM Review*, v. 42, p. 599–653, 2000.
- KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, v. 115, p. 700–721, 1927.
- MINISTÉRIO da SAÚDE. *Informações Técnicas*. 2014. Disponível em: <<http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/776-secretaria-svs/vigilancia-de-a-a-z/varicela-herpes-zoster/11494-informacoes-tecnicas>>.
- MINISTÉRIO da SAÚDE. *Lista Nacional de Doenças de Notificação Compulsória*. 2014. Disponível em: <http://www.saude.sp.gov.br/resources/ses/perfil/profissional-da-saude/servicos/lista_nacional_de_doencas_de_notificacao_compulsoria_.pdf>.
- MITCHELL, M. *An Introduction to Genetic Algorithms*. Cambridge: The MIT Press, 1998.
- MITCHELL, M.; HRABER, P.; CRUTCHFIELD, J. P. Revisiting the edge of chaos: Evolving cellular automata to perform computations. *Complex Systems*, v. 7, p. 89–130, 1993.
- MONTEIRO, L. H. A. *Sistemas Dinâmicos Complexos*. 2. ed. São Paulo: Editora Livraria da Física, 2014.

- MONTEIRO, L. H. A.; CHIMARA, H. D. B.; CHAUI-BERLINCK, J. G. Big cities: shelters for contagious diseases. *Ecological Modelling*, v. 197, p. 258–262, 2006.
- MORAES, A. L. S. *Avaliando a influência de indivíduos imunes na propagação de doenças contagiosas*. Dissertação (Mestrado), 2016. Disponível em: <<http://tede.mackenzie.br/jspui/handle/tede/1471>>.
- MORAES, A. L. S.; MONTEIRO, L. H. A. On considering the influence of recovered individuals in disease propagations. *Communications in Nonlinear Science and Numerical Simulation*, v. 34, p. 224–230, 2016.
- MURRAY, J. D. *Mathematical Biology. I. An Introduction*. 3. ed. New York: Springer, 2004.
- NARDONE, A.; ORY, F. de; CARTON, M.; COHEN, D.; DAMME, P. van; DAVIDKIN, I.; ROTA, M.; MELKER, H. de; MOSSONG, J.; SLACIKOVA, M.; TISCHER, A.; ANDREWS, N.; BERBERS, G.; GABUTTI, G.; GAY, N.; JONES, L.; JOKINEN, S.; KAFATOS, G.; ARAGÓN, M. M. de; SCHNEIDER, F.; SMETANA, Z.; VARGOVA, B.; VRANCKX, R.; MILLER, E. The comparative sero-epidemiology of varicella zoster virus in 11 countries in the european region. *Vaccine*, v. 25, p. 7866–7872, 2007.
- OLIVEIRA, P. P. de; BORTOT, J. C.; OLIVEIRA, G. M. The best currently known class of dynamically equivalent cellular automata rules for density classification. *Neurocomputing*, v. 70, p. 35–43, 2006.
- SCHIMIT, P. H. T.; MONTEIRO, L. H. A. On the basic reproduction number and the topological properties of the contact network: An epidemiological study in mainly locally connected cellular automata. *Ecological Modelling*, v. 220, p. 1034–1042, 2009.
- SUMIDA, B.; HOUSTON, A.; MCNAMARA, J.; HAMILTON, W. Genetic algorithms and evolution. *Journal of Theoretical Biology*, v. 147, p. 59–84, 1990.
- WHITE, S. H.; REY, A. M. del; SÀNCHEZ, G. R. Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, v. 186, p. 193–202, 2007.
- WHITSED, R.; SMALLBONE, L. T. A hybrid genetic algorithm with local optimiser improves calibration of a vegetation change cellular automata model. *International Journal of Geographical Information Science*, v. 31, p. 717–737, 2017.

WOLFRAM, S. *A New Kind of Science*. Champaign: Wolfram Media, 2002.

WORLDBANK. *Life Expectancy and Population*. 2017. Disponível em: <http://databank.worldbank.org/data/reports.aspx?Report_Name=Life-Expectancy-And-Population&Id=cfa9bb3d>.