

## DESENVOLVIMENTO DE SISTEMAS PARA TOMADA DE DECISÃO EM SAÚDE UTILIZANDO TÉCNICAS DE MODELAGEM MATEMÁTICA E COMPUTACIONAL

Eduardo Couto Ciscoto<sup>1</sup>

Lucio Mauro Cesca<sup>2</sup>

Rafael de Carvalho C. Oliveira<sup>3</sup>

Frâncila Weidt Neiva<sup>4</sup>

Bárbara de Melo Quintela<sup>5</sup>

### RESUMO

O presente trabalho propõe a utilização de três modelos distintos com aplicação à saúde: um modelo que utiliza mineração de dados para previsão de surtos; um modelo de epidemiologia do vírus da febre amarela; um modelo de infecção e tratamento de vírus da hepatite C com estrutura baseada em idade a ser aplicado a outros vírus da mesma família. É apresentado um referencial teórico de cada um dos temas e são propostos caminhos de pesquisa a serem seguidos. Espera-se que os modelos desenvolvidos sejam divulgados para a comunidade e tomadores de decisão em saúde para propor estratégias para prevenção de surtos e para melhor compreensão das dinâmicas dos vírus para que seja possível melhorar os tratamentos existentes.

**Palavras-chave:** Biologia computacional. Modelagem matemática. Mineração de Dados. Epidemiologia.

### 1 INTRODUÇÃO

---

<sup>1</sup> Discente do Curso de Engenharia de Software do Centro de Ensino Superior de Juiz de Fora – CES/JF. Email: eduardo.ciscoto@gmail.com

<sup>2</sup> Discente do Curso de Engenharia de Software do Centro de Ensino Superior de Juiz de Fora – CES/JF. Email: lucio.m.cesca@gmail.com

<sup>3</sup> Discente do Curso de Engenharia de Software do Centro de Ensino Superior de Juiz de Fora – CES/JF. Email: rafaelxeng@gmail.com

<sup>4</sup> Docente do Curso de Engenharia de Software do Centro de Ensino Superior de Juiz de Fora. Email: francilaneiva@cesjf.br

<sup>5</sup> Docente do Departamento de Ciência da Computação, Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora. Email: barbara@ice.ufjf.br

Uma das principais motivações do presente trabalho é o fato das doenças infecciosas estarem entre as principais causas de mortalidade, principalmente entre crianças e adolescentes em países menos desenvolvidos (WHO). Essas doenças são causadas por patógenos que podem ser vírus, bactérias, fungos ou parasitas (PARHAM, 2014). A compreensão dos modos de ação desses invasores no hospedeiro humano é essencial para desenvolvimento de prevenção, terapias e vacinas (ABBAS; LICHTMAN, 2010; SOMPAYRAC, 2008).

Os arbovírus (ARthropod BORne VIRUS), são vírus transmitidos por artrópodes (como os mosquitos hematófagos) que causam bastante preocupação em saúde pública, sendo os mais comuns no Brasil: Dengue (DENV), Zika (ZIKV), Chikungunya (CHIKV) e Febre Amarela (YFV) (DONALISIO et al, 2017). Surto recentes mostram que não estamos preparados para epidemias causadas por arbovírus (PINGEN et al, 2017).

Casos de Febre Amarela urbana não são reportados no Brasil desde 1942 (GOLDANI, 2017). No entanto, o YFV tem tido destaque na região pois somente no período de julho/2017 a Abril/2018 foram confirmados 446 casos de Febre Amarela do tipo Silvestre em Minas Gerais, sendo que 150 evoluíram para óbito conforme boletim de 04 de Abril da SES-MG (2018) e muitos casos de doenças causadas por arbovírus não são reportados corretamente (CAVALCANTI, 2017). Ainda que exista uma vacina segura e recomendada existem vários fatores que permitem que a doença ressurgisse periodicamente na África e Américas (JULANDER, 2013; GOLDANI, 2017).

Apesar de apresentar forma de transmissão bastante distinta, o vírus da *hepatite C* (HCV) é da mesma família (*flaviridae*) dos arbovírus mencionados e apresenta ciclo de vida semelhante dentro da célula do hospedeiro humano infectado (GARDNER e RYMAN, 2010; AFZAL et al, 2015). O HCV é um vírus de orientação positiva e infecta aproximadamente 170 milhões de pessoas no mundo (WHO).

O presente trabalho está dividido de forma que o referencial teórico para cada um dos tipos de modelos é apresentado na próxima Seção, seguida da explicação da metodologia adotada para cada tema. Os principais objetivos são apresentados a seguir:

## 1.1 OBJETIVOS

O objetivo geral do presente projeto de pesquisa é aplicar o conhecimento de desenvolvimento de sistemas para atender demandas das áreas de saúde. Para atingir tal objetivo propõe-se como objetivos específicos:

- i. Utilizar técnicas de mineração de dados e aprendizado de máquina para realizar previsões de ocorrência de futuros surtos de doenças transmitidas por arbovírus.
- ii. Implementar um modelo matemático epidemiológico do YFV baseado nos trabalhos de Raimundo et al (2015) e Ribeiro et al (2015), para análise do impacto da vacinação e, evolução temporal das epidemias, considerando os dados disponíveis para a região de Juiz de Fora-MG.
- iii. Utilizar o modelo matemático já validado para o HCV (Quintela et al, 2018) para identificar a eficácia de medicamentos de ação direta para tratamento do YFV, baseado em dados disponibilizados nos trabalhos de Guo et al (2016) e Julander (2013).

## 2 REFERENCIAL TEÓRICO

A Epidemiologia é uma disciplina científica quantitativa baseada em dados e métodos bem estabelecidos para análises de dados relacionados à saúde pública (CDC, 2012). Uma ferramenta que tem sido bastante aplicada a Epidemiologia é a modelagem matemática (NOKES; ANDERSON, 1988; HUPPERT; KATRIEL, 2013).

Em Murray (2002) são apresentados vários exemplos de modelos matemáticos que incorporam aspectos da epidemiologia como transmissão de doenças, evolução temporal de epidemias, resistência a infecção adquirida, estratégias de vacinação, entre outros. Um tipo de modelo matemático bastante utilizado para realizar previsões é o proposto por Kermack e McKendrick (1927) *apud* (HUPPERT; KATRIEL, 2013) em que as populações são divididas entre suscetíveis, infectadas e recuperados e representadas como funções no tempo  $S(t)$ ,  $I(t)$ , e  $R(t)$ , conforme mostrado a seguir:

$$\begin{aligned}\frac{dS}{dt} &= -\beta S I, \\ \frac{dI}{dt} &= \beta S I - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}$$

onde as derivadas  $dS/dt$ ,  $dI/dt$  e  $dR/dt$  medem as taxas de mudança das quantidades  $S(t)$ ,  $I(t)$ , e  $R(t)$ . O parâmetro  $\beta$  representa o número médio de indivíduos que um indivíduo infectado irá infectar por unidade de tempo, assumindo que todos os contatos desse indivíduo serão feitos com indivíduos suscetíveis. Dessa forma, uma doença altamente infecciosa terá um valor de  $\beta$  maior. O parâmetro  $\gamma$  é a taxa de recuperação. A razão  $\beta/\gamma$  é também conhecida como número básico de reprodução,  $R_0$ , que é um índice importante para quantificar transmissão de patógenos. Esse modelo simples é base para várias elaborações e fornece possibilidade para previsões (HUPPERT; KATRIEL, 2013).

Exemplos de modelos matemáticos um pouco mais detalhados baseados nesse modelo simples são apresentados nos trabalhos de Guo et al (2016) e Julander (2013), que representam os modos de transmissão do vírus da Febre Amarela nas populações de mosquitos e humanos.

Devido à complexidade e multiplicidade de fatores envolvidos a epidemiologia tem se beneficiado também de tecnologias para permitir a

análises de grandes bases de dados como mineração de dados e aprendizado de máquina (LYNCH; MOORE, 2016).

A mineração de dados biológicos é uma ferramenta de grande valor para o auxílio a definição de padrões em epidemiologia. No entanto, é preciso cautela pois, a partir da análise dos dados, podem ser levantados padrões inconsistentes talvez devido a ausência de fatores influenciadores na doença ou outros “pontos fora da curva”, que se sugere serem considerados (Stricker, 2017).

Uma ferramenta de mineração de dados existente que apoia a tomada de decisão na formulação de novas hipóteses é a MELODI. Essa ferramenta é capaz de fazer a leitura de dois artigos e identificar termos que sejam comuns (ELSWORTH, 2018). Em seu trabalho, Figge, (2018) aborda vários aspectos do uso de inteligência artificial aplicada a melhor compreensão das doenças infecciosas.

Modelos matemáticos também têm sido aplicados ao estudo das dinâmicas dos patógenos intra-hospedeiro. Mais especificamente relacionado à dinâmica viral, o trabalho de Perelson (2002) demonstrou o uso de um modelo simples que pode ser aplicado aos vírus da imunodeficiência humana (HIV), vírus da hepatite C (HCV), vírus da hepatite B (HCB) e citomegalovírus (CMV) para compreensão de infecção. O modelo básico consiste de três equações e vários modelos incluindo um dos que é proposto no presente trabalho derivam desse modelo simples de infecção mostrado a seguir:

$$\frac{dT}{dt} = \lambda - dT - \beta VT,$$

$$\frac{dI}{dt} = \beta VT - \delta I,$$

$$\frac{dV}{dt} = \rho I - cV,$$

onde T representa as células suscetíveis a serem infectadas e I, representa as células infectadas pelos vírus, V. Assume-se nesse modelo que as células alvo,

possuem uma fonte de regeneração constante,  $\lambda$ , e decaem a uma taxa  $d$ . A taxa em que são infectadas é representada pelo parâmetro  $\beta$ . Assume-se ainda que as células infectadas são capazes de produzir novos vírus a uma taxa  $\rho$  e morrem a uma taxa  $\delta$ . Além disso os vírus podem ser eliminados pela ação do sistema imune a uma taxa  $c$ .

O presente trabalho parte da continuidade de estudos relacionados a dinâmica do vírus da hepatite C, em que foi definido um modelo baseado em estrutura etária, derivado da epidemiologia, para estudar a replicação do HCV dentro de um célula infectada (JEREMIE et al, 2011; RONG et al, 2013; QUINTELA et al, 2018). Guo et al (2016) comentam sobre o uso dos medicamentos antivirais de ação direta (Direct Acting Antiviral, DAA), que são moléculas pequenas que agem diretamente no encapsulamento dos vírus, e ainda reforçam a importância de estudar medicamentos antivirais para os casos não resolvidos pela vacina.

A partir de resultados positivos da ação de DAAs sobre o vírus da hepatite C (IOANNOU et al, 2017; VACHON; DIETERICH, 2017; KOIZUMI et al, 2017; CHATTERJEE et al, 2013), modelos *in vitro* e *in vivo* de YFV estão sendo utilizados para testar eficácia de DAAs sobre o vírus da Febre Amarela (GUO et al, 2016; JULANDER, 2013) e Zika (SACRAMENTO et al, 2017), Chikungunya (FERREIRA et al 2018).

### 3 METODOLOGIA

Para atender o objetivo proposto de aplicar o conhecimento de desenvolvimento de sistemas para atender demandas da área de saúde, foram definidos três objetivos específicos. A metodologia proposta para cada um é descrita nas subseções a seguir.

#### 3.1 MINERAÇÃO DE DADOS BIOLÓGICOS

Para o desenvolvimento de um modelo preditivo propõe-se a utilização de uma ferramenta gratuita, chamada WEKA, *Waikato Environment for Knowledge Analysis* (WITTEN *et al*, 2016). A ferramenta proposta oferece possibilidade de desenvolvimento de rotinas na linguagem de programação Java. A vantagem da linguagem Java é a possibilidade de ser executada em diversas plataformas através da máquina virtual (Java).

Foi realizada inicialmente uma análise dos dados disponibilizados pela Secretaria de Saúde e pelo DataSus; Os passos seguinte incluem: pré-processamento dos dados; separação dos dados em conjuntos de treinamento e testes na proporção de (70/30); definição e desenvolvimento de um modelo de previsão de surtos utilizando a ferramenta Weka; treinamento do modelo com pelo menos três algoritmos de aprendizado de máquina distintos disponíveis na ferramenta; comparação dos resultados e escolha do algoritmo que apresente melhor desempenho para o conjunto de dados utilizado.

## 3.2 MODELO EPIDEMIOLÓGICO

Para o modelo matemático de epidemiologia foi escolhida a linguagem de programação Python por ser livre e por já possuir bibliotecas para resolução de sistemas de equações diferenciais.

Com relação aos métodos numéricos para resolução das equações serão utilizados métodos estabelecidos dando preferência a bibliotecas otimizadas. O solver será implementado de forma a facilitar a análise de sensibilidade do modelo aos parâmetros e ajuste de parâmetros.

Foi proposto primeiramente a implementação do modelo de epidemiologia do vírus causador da febre amarela disponível em Raimundo *et al* (2015) e replicação dos resultados, para a em seguida identificar necessidades de adaptação para representar o cenário epidemiológico de Minas Gerais. Propôs-se utilizar como base para os parâmetros os dados dos boletins da SES-MG no período de 2017-2018.



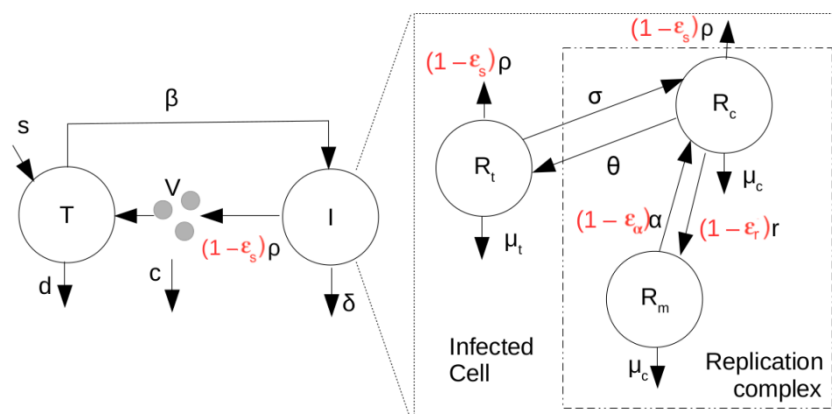
## 3.3 MODELO MULTIESCALA DE EQUAÇÕES DIFERENCIAIS COM ESTRUTURA ETÁRIA

O modelo matemático multiescala com estrutura etária de infecção e tratamento já possui um solver validado para equações diferenciais ordinárias e parciais implementado em Matlab® (QUINTELA, 2015; QUINTELA *et al*, 2018).

Foi proposto o desenvolvimento do modelo do HCV na linguagem de programação Python utilizando o método das linhas, como primeira etapa para comparação de resultados com o modelo de HCV implementado em Matlab®. A partir da validação do solver em Python espera-se prosseguir para ajustar os parâmetros do modelo a partir dados disponibilizados na literatura sobre YFV.

A Figura 1 mostra um esquema representativo do modelo de equações diferenciais. O modelo é baseado no modelo simples de infecção de vírus apresentado na seção anterior, complementado pela estrutura etária para representar a dinâmica dentro de uma célula. Maiores detalhes do modelo podem ser encontrados nos trabalhos de Quintela (2015); Quintela *et al* (2018).

Esquema representativo do modelo matemático de equações diferenciais ordinárias e parciais. O modelo representa as células alvo (T), os vírus no meio extracelular (V) e a partir do contato a célula alvo torna-se infectada (I). Dentro da célula infectada são representados os estados do RNA viral de orientação positiva ( $R_c$ ), negativa ( $R_m$ ) e durante o período em que está sendo traduzido ( $R_t$ ). Notações em vermelho indicam locais que podem sofrer efeitos da terapia.





Fonte: Quintela et al, 2018.

Foram realizados contatos com institutos de pesquisa para obtenção de dados experimentais de tratamento de Febre Amarela, Zika e Chikungunya para ajustar os parâmetros do modelo. No entanto, os dados não foram disponibilizados o que impediu realização de simulações e ajustes de parâmetros até o momento para o modelo de infecção e tratamento.

## 4 CONSIDERAÇÕES FINAIS

O presente trabalho apresenta uma breve revisão sobre modelos computacionais aplicados à saúde. Foram apresentados três tipos de modelos e as etapas de estudo dos temas por parte dos alunos envolvidos no projeto de pesquisa realizado no CES/JF e preparação dos ambientes que utilizaram para desenvolvimento.

Espera-se que com a continuidade do presente trabalho seja possível concluir o desenvolvimento de ferramentas que auxiliem aos tomadores de decisão em saúde na elaboração de melhores práticas de prevenção e tratamento de doenças infecciosas causadas por vírus.

Entre as limitações do trabalho se encontram a disponibilidade e formato dos dados. Muitas vezes os dados disponíveis precisam ser tratados para que possam ser utilizados. Em outros casos, devido a burocracia há uma demora no acesso aos dados que atrasa todo o projeto de desenvolvimento que depende destes. Mesmo quando há o acesso aos dados, outra limitação é a variação dos resultados experimentais que dificultam bastante os ajustes de parâmetros dos modelos.

Além do desenvolvimento de ferramentas de software para auxiliar a tomada de decisão por profissionais da área de saúde pública, o presente

projeto visa capacitação dos alunos de graduação envolvidos para atuarem também na área de pesquisa.

## DEVELOPMENT OF HEALTH DECISION MAKING SYSTEMS USING MATHEMATICAL AND COMPUTATIONAL MODELING TECHNIQUES

### ABSTRACT

The present work proposes the use of three distinct models applied to health studies: a model that uses data mining to predict outbreaks; a model of the epidemiology of the yellow fever; an age-based model of infection and therapy of the hepatitis C virus to be applied to other viruses of the same family *flaviridae*. It is presented a background of each research theme and future work is proposed. It is expected that the models to be developed are made available to the community and the decision makers to propose strategies to prevent outbreaks from affecting so much the population and to better understanding of viral dynamics to improve current therapy.

Keywords: Computational biology. Mathematical modeling. Data Mining. Epidemiology.

## REFERÊNCIAS

ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. H. I. V. **Imunologia celular e molecular**. 7. ed. Rio de Janeiro: Elsevier, 2012.

AFZAL, M. S. et al. Regulation of core expression during the hepatitis C virus life cycle. **Journal of General Virology**, Microbiology Society, 96, 311-321. **2015**. <https://doi.org/10.1099/vir.0.070433-0>.

CAVALCANTI, L.P.G. et al. **Surveillance of deaths caused by arboviruses in Brazil: from dengue to chikungunya**. *Mem Inst Oswaldo Cruz*. Rio de Janeiro. Vol. 112(8): 583-585, August, 2017

CDC, Centers for Disease Control and Prevention. **Principles of Epidemiology in Public Health Practice**. U.S. Department of Health and Human Services, 2012.

DONALISIO, MR et al. Arboviruses emerging in Brazil: challenges for clinic and implications for public health. **Revista de Saúde Pública**, 51-30, 2017.

ELSWORTH, B. et al. **MELODI: Mining Enriched Literature Objects to Derive Intermediates**. *International Journal of Epidemiology*, Volume 47, Issue 2, 1 April 2018, Pages 369–379, <https://doi.org/10.1093/ije/dyx251>.

FERREIRA, A. C. et al. **Beyond members of the Flaviviridae family, sofosbuvir also inhibits chikungunya virus replication**. November 2018. *Antimicrobial Agents and Chemotherapy*. DOI: 10.1128/AAC.01389-18

FIGGE, H.; **Deploying Artificial Intelligence Against Infectious Diseases**. *US Pharm*. 43 (3), p. 21-24. 2018.

GARDNER, C. L.; RYMAN, K. D. **Yellow Fever: A Reemerging Threat**. *Clin Lab Med*, 30(1), 237-60, 2010.

GOLDANI, L. Z. Yellow fever outbreak in Brazil. **The Brazilian Journal of Infectious Diseases**. 21, 123-124, 2017.

GUO, F. et al. **A Novel Benzodiazepine Compound Inhibits Yellow Fever Virus Infection by Specifically Targeting NS4B Protein**. *Journal of Virology*, 90, 2016.

HUPPERT, A.; KATRIEL, G. **Mathematical Modelling and Prediction in Infectious Disease Epidemiology.** *Clinical Microbiology and Infection*, 19, no. 11:999–1005. 2013. <https://doi.org/10.1111/1469-0691.12308>.

IOANNOU, George N. et al. **HCV Eradication Induced by Direct-Acting Antiviral Agents Reduces the Risk of Hepatocellular Carcinoma.** *Journal of Hepatology*, 68, no. 1 (January): 25–32, 2017.

JULANDER, J. G. **Experimental therapies for yellow fever.** *Antiviral Research*, 97, 169-79, 2013.

KOIZUMI, Yoshiki et al. Quantifying Antiviral Activity Optimizes Drug Combinations against Hepatitis C Virus Infection. **Proceedings of the National Academy of Sciences**, 114, no. 8 (February 2017): 1922–1927. <https://doi.org/10.1073/pnas.1610197114>.

LYNCH, S. M.; MOORE, J. H. **A call for biological data mining approaches in epidemiology.** *BioData Mining*, 9:1, 2016.

NOKES, D. J.; ANDERSON, R. M. **The Use of Mathematical Models in the Epidemiological Study of Infectious Diseases and in the Design of Mass Immunization Programmes.** *Epidemiology and Infection* 101, no. 01:1–20. 2010. <https://doi.org/10.1017/S0950268800029186>.

MATLAB and Statistics Toolbox Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States, 2015.

MURRAY, J. D. **Mathematical Biology.** I. An Introduction. Vol. I. Springer, 2002.

PINGEN, M.; SCHMID, M.A.; HARRIS, E.; MCKIMMIE, C.S. **Mosquito Biting Modulates Skin Response to Virus Infection.** *Trends in Parasitology*. 33 (8), p. 645-57, August, 2017. doi: <https://doi.org/10.1016/j.pt.2017.04.003>.

SOMPAYRAC, L. **How the immune system works.** 3th Edition, Blackwell Publishing, Oxford, UK, 130 pages, 2008.

STRICKER, B. H. **Epidemiology and Big Data.** *Eur J Epidemiol.* 32 (7). P. 535-536. 2017. doi: [10.1007/s10654-017-0294-3]

QUINTELA, B M. **Aoplamento de Modelos Computacionais de Doenças Infecciosas.** Tese (Doutorado). Programa de Pós-graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora. 2015.

QUINTELA, B. M. *et al.* **A New Age-Structured Multiscale Model of the Hepatitis C Virus Life-Cycle During Infection and Therapy With Direct-Acting Antiviral Agents.** *Frontiers in Microbiology*, 9, 601, 2018.  
<https://doi.org/10.3389/fmicb.2018.00601>.

RAIMUNDO, S. M. *et al.* **Equilibrium Analysis of a Yellow Fever Dynamical Model with Vaccination.** *Computational and Mathematical Methods in Medicine*, 2015. <https://doi.org/10.1155/2015/482091>

RIBEIRO, A. F. *et al.* **A Public Health Risk Assessment for Yellow Fever Vaccination: A Model Exemplified by an Outbreak in the State of São Paulo, Brazil.** *Memórias Do Instituto Oswaldo Cruz*, 110, no. 2, 230–234. 2015.  
<https://doi.org/10.1590/0074-02760140345>.

RONG, L. *et al.* **Analysis of hepatitis c virus decline during treatment with the protease inhibitor danoprevir using a multiscale model.** *PLOS Computational Biology*, v. 9, n. 3, p. e1002959, 2013.

SACRAMENTO, C. Q. *et al.* **Corrigendum:** The clinically approved antiviral drug sofosbuvir inhibits Zika virus replication. April 2017. *Scientific Reports* 7(1):46772. DOI: 10.1038/srep46772

VACHON, Marie-Louise; DIETERICH, D. T. The Era of Direct-Acting Antivirals Has Begun: The Beginning of the End for HCV?. **Seminars in Liver Disease**, 31, no. 4. 399–409, 2011.

WITTEN I.H. *et al.* Online Appendix on the WEKA Workbench. **Data Mining: Practical Machine Learning Tools and Techniques**, 4 ed. Morgan Kaufmann, 2016.