

Modelagem Matemática para Epidemia de COVID-19 e Estimativa de Casos no Brasil no Curto Prazo

Gabriela Moutinho de Souza Dias, Luiz Henrique da Costa Araújo

Instituto Militar de Engenharia, Seção de Ensino de Engenharia de Computação, Praça General Tibúrcio, 80, Praia Vermelha, 22290-270, Rio de Janeiro, RJ, Brasil

Resumo

Este artigo apresenta um método para a obtenção de um modelo matemático adaptativo do desenvolvimento da epidemia de COVID-19 na população brasileira. Para tal, foi utilizado como base o modelo matemático epidemiológico SIDR que permite acompanhar o desenvolvimento da doença, levando em consideração as medidas não farmacológicas de afastamento social adotadas pelas autoridades brasileiras. Resultados preliminares, com projeções de curto prazo, são apresentados. O presente estudo encontra-se em andamento, em constante atualização e aprimoramento.

Palavras-chave: covid-19, epidemic, pandemic, mathematical model, SIDR

1. Introdução

No final do ano de 2019, o mundo acendeu um sinal de alerta com o início de uma nova doença infecciosa viral chamada COVID-19 (Coronavirus Disease 2019) causada por um vírus chamado coronavírus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2 of the Genus Betacoronavirus) [17].

Conforme a epidemia foi se desenvolvendo no mundo, as autoridades médicas mundiais perceberam a rápida disseminação da doença e o seu potencial perigo para o colapso dos Sistemas de Saúde dos países afetados. Via de regra, as decisões tomadas pelas autoridades tiveram forte impacto nas sociedades afetadas, principalmente nos setores produtivos, causando efeitos nocivos além da doença propriamente dita.

A modelagem matemática epidemiológica é bastante eficiente no auxílio do processo de tomada de decisões, a aplicação de modelos analíticos na análise do desenvolvimento de epidemias pode representar importante ferramenta para decisores, permitindo a análise equilibrada e embasada do impacto da epidemia

E-mail: gabriela@ime.eb.br (Gabriela Moutinho de Souza Dias), lharaujo@ime.eb.br (Luiz Henrique da Costa Araújo)

na saúde, na economia, entre outras áreas da sociedade. Por essa razão, faz-se mister o desenvolvimento destes modelos objetivando fornecer embasamento para decisões acertadas.

Desde o surgimento do surto de COVID-19, diversos trabalhos científicos foram publicados com o objetivo de analisar o comportamento da epidemia na população.

Existem basicamente duas formas de analisar esse tipo de fenômeno, por meio de modelos matemáticos (analíticos) e por meio de modelos de simulação computacional. O presente trabalho está inserido na classe de modelos matemáticos.

Apesar de haverem diversos estudos já divulgados versando sobre a evolução da epidemia e a comparação entre possíveis cenários, cada estudo é direcionado para um país específico. Tendo em vista que cada país tem suas peculiaridades e particularidades locais, é importante que haja um estudo especificamente focado na população brasileira.

O trabalho mais recente direcionado para o Brasil é o estudo apresentado em [5], o qual apresenta uma análise da epidemia baseada em modelos epidemiológicos. O presente trabalho apresenta uma contribuição em relação ao anterior, tendo em vista que os dados utilizados para obtenção dos resultados desta modelagem foram atualizados até a data de 2 de abril de 2020. Como será discutido adiante, a constante atualização dos parâmetros do modelo é de extrema importância para se obter resultados com maior acurácia e mais próximos da real situação da epidemia.

O restante deste artigo está organizado da seguinte forma: na Seção 2 é apresentado de forma resumida o modelo SIDR aplicado nesta modelagem; na Seção 3 é apresentada a metodologia da modelagem utilizada; na Seção 4 os resultados preliminares obtidos com este estudo; e na Seção 5 as principais conclusões do trabalho.

2. Modelos Epidemiológicos

A modelagem matemática da propagação de doenças infecciosas começou por volta de 1760, com o trabalho de Daniel Bernoulli sobre a varíola. Em 1906, William Hamer foi o primeiro a prever o princípio da Lei de Ação das Massas para um modelo determinístico em tempo discreto [10]. Esse princípio, que incorpora o princípio da mistura homogênea, tem sido a base dos trabalhos subsequentes na área de epidemiologia matemática.

Desde então, diversos modelos matemáticos foram propostos para os mais diversos tipos de doenças infecciosas que afetam a população humana, seja em nível comunitário (epidemia local) ou mundial (pandemia), como a que o mundo enfrenta atualmente com a COVID-19.

O objetivo da modelagem matemática epidemiológica é analisar o comportamento e a evolução de uma doença dentre indivíduos de uma população ao longo do tempo. Essa análise tem o propósito de auxiliar no controle da propagação da doença, de modo a evitar o avanço de epidemias.

A literatura dessa área é bastante rica, principalmente por ser um assunto antigo e amplamente estudado. Existe uma grande quantidade de referências de trabalhos sobre modelos matemáticos para doenças infecciosas, tais como [1–4, 6, 9, 11, 12, 16].

Dentre os diversos modelos existentes na literatura, para a modelagem descrita neste documento foi escolhido o modelo SIDR (*Susceptible-Infected-Deceased-Removed*). Os principais motivos para tal escolha são a baixa complexidade e a eficiência do modelo. Em termos matemáticos, o modelo é relativamente simples, se comparado a outros modelos da área, pois o mesmo trata-se de um sistema de equações diferenciais. Em termos de eficiência, esse modelo já foi utilizado para modelar diversas doenças virais infecciosas no passado, mostrando-se altamente eficiente na estimação do andamento da epidemia na população, quando suas condições de aplicabilidade são satisfeitas.

O grande desafio em se trabalhar com modelos epidemiológicos é a estimação dos parâmetros. Essa etapa é de crucial importância, pois pequenas variações nos parâmetros são capazes de causar um grande impacto nos resultados finais. É importante ressaltar que, independente do método de estimação utilizado, um bom resultado só pode ser obtido se houver uma quantidade adequada de dados acerca da população e do comportamento da doença nos indivíduos em particular e na população como um todo.

2.1. Modelo SIDR

Esta seção apresenta o modelo SIDR utilizado no presente trabalho de forma compacta, com destaque para as principais características do mesmo.

Em 1927, Kermack e McKendrick desenvolveram o clássico modelo SIR (*Susceptible-Infected-Recovered*) para a propagação de doenças infecciosas [14, 15]. Trata-se de um modelo determinístico de tempo contínuo amplamente utilizado, tanto na área epidemiológica quanto em outras áreas da ciência [7, 8].

A partir do modelo SIR, diversas extensões foram propostas ao longo do tempo. Em termos gerais, esse modelo divide a população em três compartimentos: Suscetível, Infectado e Removido. O número de óbitos devido à doença está inserido no compartimento Removido. Porém, quando se torna relevante observar separadamente a evolução dos casos de óbitos, torna-se necessário criar um quarto compartimento, específico para tais casos. A literatura, em geral, se refere a esse compartimento como *Deceased*, e neste trabalho o modelo que incluir tal compartimento será chamado de **SIDR**.

Seja N o número total de indivíduos na população. Considera-se que a população é distribuída de forma homogênea na região. Além disso, o modelo também considera uma população homogênea, isto é, todos os indivíduos possuem mesma probabilidade de serem infectados e as taxas de contato de cada indivíduo são iguais. Esta consideração de homogeneidade é válida para grandes populações, no que tange a análise da modelagem.

Considere que o período de tempo em que uma epidemia ocorre seja muito menor que o tempo de existência da população, ou seja, o tamanho desta pode ser considerado constante no período, e os nascimentos e mortes não relacionadas à doença podem ser ignorados. A qualquer instante de tempo, cada

indivíduo encontra-se em um determinado estado com relação à doença, dividindo a população em quatro compartimentos:

- S – Suscetível: Indivíduos que estão suscetíveis a serem infectados;
- I – Infectado: Indivíduos que estão infectados com a doença e são capazes de transmiti-la a indivíduos suscetíveis;
- D – Óbito: Indivíduos que morreram devido à doença¹;
- R – Removido: Indivíduos que se recuperaram da doença ou foram isolados.

A transmissão da doença ocorre por contato entre os indivíduos². Um indivíduo suscetível se torna infectado após contato com um indivíduo Infectado. O termo contato deve ser interpretado de forma ampla, pois refere-se ao evento responsável pela transmissão do vírus, podendo ser um contato físico propriamente dito (como aperto de mãos ou a proximidade física no momento de espirro ou tosse do infectado) ou outro tipo de situação de transmissão, como quando um suscetível toca em um local contaminado pelo vírus, por exemplo.

Considera-se que um indivíduo recuperado torna-se imune durante o período de duração da epidemia. Indivíduos infectados que são isolados do contato com indivíduos suscetíveis (não sendo assim capazes de transmitir a doença) são inseridos no compartimento R.

Sejam $S(t)$, $I(t)$, $D(t)$, $R(t)$ o número de Suscetíveis, Infectados, Óbitos e Removidos, respectivamente, a cada instante de tempo $t \geq 0$. Assim, para todo $t \geq 0$, temos que:

$$S(t) + I(t) + D(t) + R(t) = N \quad (1)$$

A evolução do processo infeccioso de acordo com o modelo SIDR é definida pelo seguinte sistema de equações diferenciais [13]:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \frac{\gamma}{1-\rho} I(t) \quad (3)$$

$$\frac{dD(t)}{dt} = \frac{\rho}{1-\rho} \gamma I(t) \quad (4)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (5)$$

¹Utiliza-se a letra D para este compartimento para acompanhar a notação da literatura, do inglês *Deceased*.

²Na literatura, há outros modelos para tipos de doença não transmissíveis por contato direto entre indivíduos, como a dengue, por exemplo.

Condição inicial: $(S(0), I(0), D(0), R(0)) = (S_0, I_0, 0, 0)$, $S_0, I_0 > 0$. $\beta, \gamma, \rho > 0$.

O parâmetro β é o **parâmetro de infecção** do processo. Sendo a doença transmitida por contato, β corresponde à taxa com que ocorrem **contatos** entre dois indivíduos quaisquer, isto é, o número de contatos entre dois indivíduos quaisquer por unidade de tempo. Considera-se **encontro** o evento no qual há um contato entre dois indivíduos quaisquer. Muitos autores também se referem a β como “taxa de contato”, porém esse termo pode se confundir com a taxa de contato de cada indivíduo, isto é, número de contatos que um indivíduo faz com outros indivíduos na população por unidade de tempo.

O parâmetro γ é chamado **taxa de recuperação**, e representa a taxa com que um indivíduo infectado é recuperado (taxa de recuperação por indivíduo infectado).

O parâmetro ρ é chamado **probabilidade de mortalidade**, isto é, a probabilidade de um indivíduo infectado morrer devido à doença antes de se recuperar.

3. Metodologia da Modelagem

Como mencionado anteriormente, o desafio da modelagem epidemiológica é a estimação dos parâmetros. Nesta seção é apresentada a metodologia de estimação utilizada.

Os parâmetros do modelo que precisam ser estimados são β, γ, ρ . O método utilizado para estimação neste trabalho é a minimização do erro médio quadrático para as variáveis $I(t)$ e $D(t)$. Esse método é o mais sugerido pela literatura da área, conforme exposto em [6], sendo também aplicado por [5].

Assim, para estimar os parâmetros β, γ, ρ do modelo, o objetivo é minimizar a seguinte expressão para o erro e_n :

$$e_n = \frac{1}{2} \left(\sum_{t=t_0}^{t_0+n} (I_t - \hat{I}_t)^2 + \sum_t (D_t - \hat{D}_t)^2 \right), \quad (6)$$

onde \hat{I}_t, \hat{D}_t são os valores estimados para os valores reais I_t, D_t , respectivamente, no tempo t . O erro e_n é obtido considerando um conjunto de dados de tamanho $n + 1$, ou seja, correspondendo a $n + 1$ valores das variáveis, a partir do tempo t_0 .

Os dados reais foram obtidos no painel oficial de casos consolidados de COVID-19 disponibilizado pelo Ministério da Saúde do Brasil³.

A unidade de tempo considerada é de um dia, tendo em vista que os dados reais divulgados são diários.

Cabe ressaltar que o valor do erro e_n e o resultado do processo de minimização dependem fortemente da quantidade de dados reais disponíveis n . Quanto maior esse número, melhor será o ajuste da modelagem. Consequentemente, o valor final encontrado para cada parâmetro também é dependente de n .

³Disponível em <https://covid.saude.gov.br/>. Acesso em 02/04/20.

Os resultados apresentados neste trabalho foram obtidos com base nos dados reais divulgados até a data de 02/04/20 e duas estimativas diferentes foram realizadas, denotadas por E1 e E2. A **estimativa E1** considera o período de 19/03/20 a 28/03/20, ou seja, $n = 10$. A **estimativa E2** considera o período de 19/03/20 a 02/04/20, ou seja, $n = 15$. Para cada estimativa, os valores dos parâmetros β, γ, ρ , obtidos pelo método explicado anteriormente, são diferentes.

O valor para o tamanho da população do Brasil foi obtido na página do Instituto Brasileiro de Geografia e Estatística (IBGE) na Internet⁴.

O modelo como apresentado na Seção 2.1 não considera nenhuma medida de controle exógena para reduzir a propagação da doença na população. Porém, na prática, no período referente aos dados reais utilizados para estimação dos parâmetros neste trabalho, existe uma medida de controle que é o isolamento social. Tal tipo de medida de controle afeta o parâmetro de infecção do modelo, e por isso, o mesmo deve ser ajustado. Com o isolamento social, o número de contatos realizados por cada indivíduo diminui, e em consequência também diminui o contato com indivíduos infectados. Logo, a velocidade de propagação da doença na população diminui e isso se reflete na redução do valor do parâmetro β .

Sendo assim, é necessário definir um parâmetro para representar a aplicação de medidas de controle de propagação, que neste caso específico é o isolamento social⁵. Para a obtenção dos resultados apresentados neste documento, foi considerado para tal o parâmetro $\psi = 0,67$ utilizado em [5], conforme notação e valor estimado no citado estudo. Logo, considerando o isolamento social, o parâmetro de infecção a ser aplicado nas equações do modelo, Eq. 2 e Eq. 3, deve ser ajustado para $\beta\psi$.

Com relação à condição inicial, o processo de propagação inicia com 1 indivíduo infectado e $N - 1$ suscetíveis, ou seja, $(S(0), I(0), D(0), R(0)) = (N - 1, 1, 0, 0)$.

4. Resultados

A análise de modelagem de que trata este documento é um trabalho em andamento, em constante atualização e refinamento. Nesta seção são apresentados alguns resultados preliminares obtidos até a data de 02/04/20 para a população do Brasil.

A Figura 1 apresenta o número de indivíduos infectados ao longo do tempo no intervalo entre os dias 19/03/20 e 06/04/20. Nessa figura são indicadas as curvas para as estimativas E1 e E2, obtidas conforme explicado na seção anterior. Os dados reais foram consolidados até a data de 02/04/20. Pode-se observar na figura o ajuste fino entre os resultados da modelagem e os dados reais do

⁴Disponível em <https://www.ibge.gov.br/apps/populacao/projecao/>. Acesso em 02/04/20.

⁵Existem outras medidas de controle de propagação de doenças infecciosas não consideradas na obtenções dos resultados preliminares deste estudo, como adoção de medicamentos específicos ou vacinas, por exemplo.

histórico da epidemia, validando, assim, o modelo implementado. Nesse gráfico é indicado também uma projeção de curtíssimo prazo, mostrando a tendência de crescimento do número de infectados.

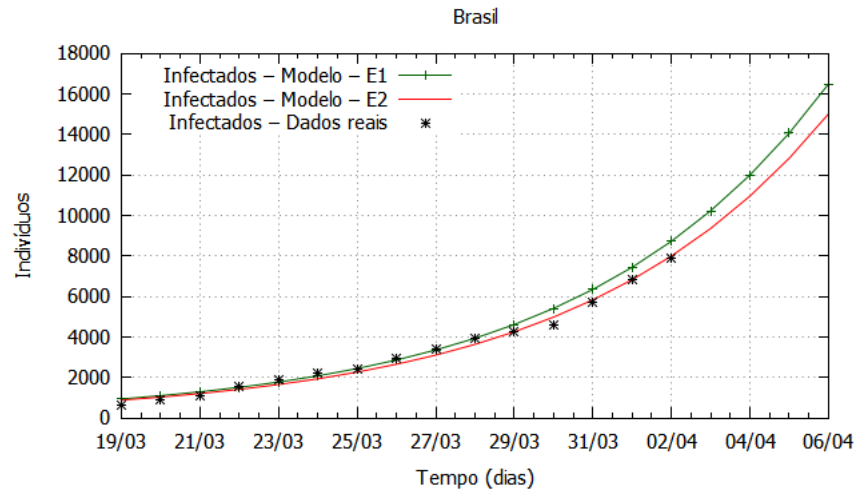


Figura 1: Ajuste do modelo aos dados reais. Número de indivíduos infectados no período de 19/03/20 até 06/04/20, com estimativa do modelo no período de 03/04/20 a 06/04/20 e dados reais até a data de 02/04/20.

De forma semelhante à figura anterior, a Figura 2 apresenta o número de óbitos ao longo do tempo no intervalo entre os dias 19/03/20 e 06/04/20.

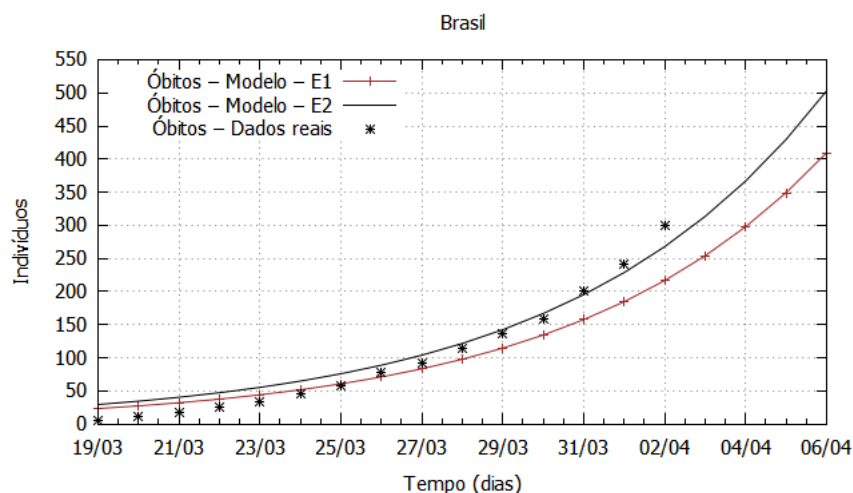


Figura 2: Ajuste do modelo aos dados reais. Número de óbitos no período de 19/03/20 até 06/04/20, com estimativa do modelo no período de 03/04/20 a 06/04/20 e dados reais até a data de 02/04/20.

É importante ressaltar que a estimativa E1 se ajusta melhor aos dados até a data de 28/03, pois essa estimativa considera apenas os valores divulgados até essa data. A estimativa E2, por sua vez, apresenta melhor ajuste aos dados a partir dessa data em diante. A estimativa E2 é a mais atualizada, considerando todos os valores consolidados até a data de 02/04/20, sendo a mais adequada para previsões futuras até a data de escrita deste documento⁶. O erro médio quadrático normalizado na estimativa E2, relativo aos dados reais apresentados para o número de infectados, é de 6,37%, enquanto que para a estimativa E1 o erro é de 12,13%.

As Figuras 3 e 4 apresentam uma projeção de curto prazo, período de 02/04/20 até 20/04/20, para o número de indivíduos infectados e número de óbitos, respectivamente. Os resultados do modelo foram obtidos de acordo com a estimativa E2 pelos motivos explicados anteriormente. De acordo com o modelo, o número de infectados atinge a marca de 100 mil indivíduos próximo ao dia 18/04/20. Nessa data, a estimativa para o número de óbitos ultrapassa 3000 ocorrências. Esses valores aparentemente altos podem surpreender o leitor em uma primeira análise, porém cabe destacar que os mesmos são coerentes com o número de casos em outros países. A título de exemplo, na Itália o número de infectados ultrapassou a marca de 124 mil, com mais de 15 mil óbitos⁷.

⁶Documento finalizado em 04/04/20.

⁷Dados disponíveis em <https://www.worldometers.info/coronavirus/>. Acesso em 04/04/2020 indica 124632 infectados e 15362 óbitos.

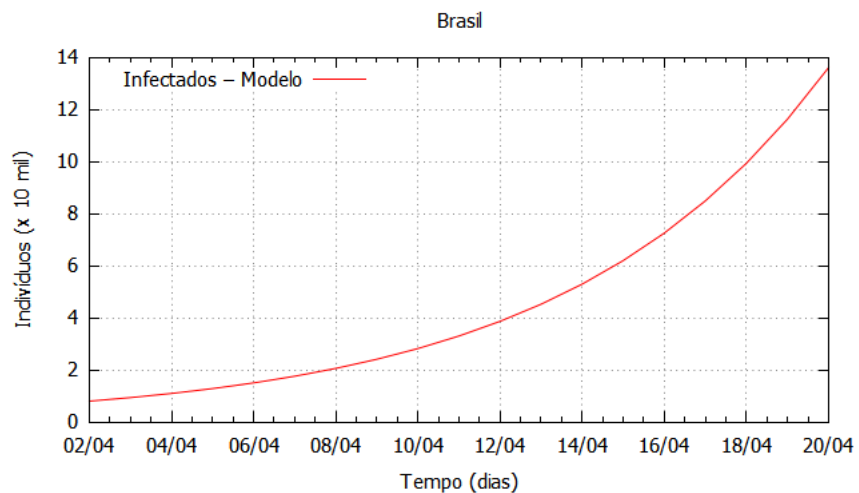


Figura 3: Projeção de curto prazo. Estimativa do modelo para o número de indivíduos infectados no período de 02/04/20 até 20/04/20.

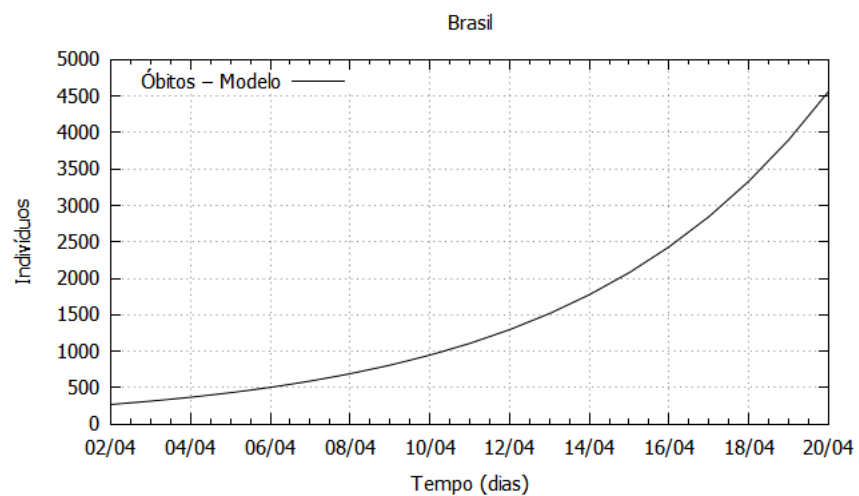


Figura 4: Projeção de curto prazo. Estimativa do modelo para o número de óbitos no período de 02/04/20 até 20/04/20.

5. Conclusão

O presente trabalho apresentou um método capaz de, a partir de dados oficiais coletados do sistema de atendimento ao COVID-19, prever o desen-

volvimento da doença em um cenário específico, população Brasileira, com as medidas de mitigação de propagação da doença não farmacológicas adotadas até o momento de escrita deste documento. Os dados reais utilizados para estimação dos parâmetros da modelagem são atualizados até a data de 2 de abril de 2020.

A qualidade da previsão do desenvolvimento da epidemia está relacionada com a qualidade dos índices de notificação da doença pelo Sistema de Saúde, isto é, quanto mais eficiente for a testagem da doença em casos suspeitos e notificação para o Sistema de Saúde, melhor serão os resultados da modelagem, uma vez que a curva de desenvolvimento da epidemia ajusta-se aos números informados no processo de modelagem.

Na Seção 4 deste trabalho, foram apresentados os resultados preliminares do presente estudo, com projeções de curto prazo para a evolução da epidemia no Brasil. Outros resultados de médio e longo prazo, além da comparação de cenários considerando diferentes medidas de controle de propagação, foram obtidos e encontram-se em fase de consolidação e análise.

O presente estudo encontra-se em andamento e as próximas etapas serão direcionadas para o aprimoramento da modelagem e a análise de outros cenários, em curto, médio e longo prazos.

Como proposta de trabalhos futuros, sugere-se a comparação da evolução não somente dos casos confirmados mas também dos casos suspeitos, uma vez que grande parte desses casos não são testados.

É importante ressaltar que os resultados apresentados neste documento são teóricos, obtidos de forma técnica e objetiva, com base simplesmente no resultado matemático que a modelagem apresenta. Este estudo não tem o objetivo de analisar impactos econômicos ou sociais de qualquer medida adotada pelas autoridades governamentais, e não reflete a opinião pessoal dos autores acerca da adoção ou não de qualquer medida.

Referências

- [1] L. J. S. Allen, A. M. Burgin, Comparison of deterministic and stochastic sis and sir models in discrete time, *Mathematical Biosciences* 163 (1) (2000) 1–33.
URL <http://www.ncbi.nlm.nih.gov/pubmed/10652843>
- [2] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Prentice Hall, 2003.
URL <http://www.hit.ac.il/staff/leonidM/stochastic-processes/Chapter-2.pdf>
- [3] N. T. J. Bailey, A simple stochastic epidemic, *Biometrika* 37 (3-4) (1950) 193–202.
- [4] N. T. J. Bailey, The total size of a general stochastic epidemic, *Biometrika* 40 (1-2) (1953) 177–185.

URL <http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/40.1-2.177>

- [5] S. B. Bastos, D. O. Cajueiro, Modeling and forecasting the Covid-19 pandemic in Brazil, Universidade de Brasília, 2020.
- [6] D. J. Daley, J. Gani, Epidemic Modelling: An Introduction, Cambridge Studies in Mathematical Biology, 2001.
- [7] G. M. S. Dias, Modelo epidemiológico sir aplicado a redes tolerantes a atrasos e desconexões, Master's thesis, Instituto Militar de Engenharia (2013).
- [8] G. M. S. Dias, R. M. Salles, Epidemic sir model applied to delay-tolerant networks, in: 30th Brazilian Telecommunications Symposium, SBrT'12, SBrT, 2012.
- [9] N. C. Grassly, C. Fraser, Mathematical models of infectious disease transmission, *Nature Reviews Microbiology* 6 (6) (2008) 477–487.
URL <http://www.ncbi.nlm.nih.gov/pubmed/18533288>
- [10] W. Hamer, The Milroy lectures on epidemic disease in England: the evidence of variability and of persistency of type, Bedford Press, 1906.
- [11] H. Hethcote, The mathematics of infectious diseases, *SIAM review* (2000) 599–653.
- [12] A. Kaddar, A. Abta, H. T. Alaoui, A comparison of delayed sir and seir epidemic models, *Nonlinear Analysis: Modelling and Control* 16 (2) (2011) 181–190.
- [13] M. J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2011.
- [14] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London Series A Containing Papers of a Mathematical and Physical Character* 19051934 115 (772) (1927) 700–721.
URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1927.0118>
- [15] W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics. ii. the problem of endemicity, *Proceedings of the Royal Society of London Series A* 138 (834) (1932) 55–83.
URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1932.0171>
- [16] M. G. Roberts, J. A. P. Heesterbeek, Mathematical models in epidemiology, *Mathematical models. UNESCO-EOLSS* (2004).

- [17] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in china, *Nature* 579 (2020) 265–269.