

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

Estimando o número básico de reprodução de uma epidemia: Resolvendo o problema inverso de estimar a taxa de transmissão de uma epidemia utilizando o método *Simulated Annealing*

Guilherme S. da Silveira¹, Victoria de O. Spagiari², Olavo H. Menin³

¹Graduando em Engenharia Elétrica no IFSP, Campus Sertãozinho, silveira.s@aluno.ifsp.edu.br.

²Graduanda em Engenharia Elétrica no IFSP, Campus Sertãozinho, victoria.spagiari@aluno.ifsp.edu.br.

³Professor Doutor do IFSP, Campus Sertãozinho, olavohmenin@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): 1.03.02.02-6 (Modelos analíticos e de simulação).

RESUMO: Modelos matemáticos são amplamente utilizados para descrever epidemias e auxiliar as autoridades de saúde pública nas melhores tomadas de decisões. Em geral, esses modelos prevêem o comportamento da epidemia considerando-se determinados parâmetros, como taxa de transmissão, tamanho da população, taxa de nascimento/mortalidade e outros fatores. Porém, pode também ser interessante inverter o problema, ou seja, através dos dados epidemiológicos sobre uma determinada epidemia, estimar quais os parâmetros que levaram a sua evolução temporal, o que corresponde a um problema inverso. Apresentamos aqui a resolução de um problema desse tipo relacionado ao modelo epidêmico compartimental SIR (suscetível-infeccioso-recuperado) utilizando o método de otimização estocástico *Simulated Annealing*.

PALAVRAS-CHAVE: Modelo compartimental; Problema inverso; Simulações computacionais; Otimização; Equações diferenciais.

Estimating the basic reproduction number of an epidemic: Solving the inverse problem of estimating the transmission rate of an epidemic using the Simulated Annealing method.

ABSTRACT: Mathematical models are widely used to describe epidemics and assist public health authorities in making appropriate decisions. In general, these models predict the behavior of an epidemic by considering certain parameters, such as transmission rate, population size, birth/mortality rate, and other factors. However, it can also be interesting to reverse the problem, that is, using epidemiological data from a given epidemic to estimate the parameters that led to its temporal evolution, which corresponds to an inverse problem. Here, we present the solution to such a problem related to the SIR (susceptible-infectious-recovered) compartmental epidemic model using the stochastic optimization method Simulated Annealing.

KEYWORDS: Compartmental model; Inverse problem; Computational simulations; Optimization; Differential equations.

INTRODUÇÃO

Segundo Oleg M. Alifanov (2012), podemos associar um problema direto ao processo *causa* \rightarrow *efeito*, ou seja, determina-se os efeitos a partir do conhecimento das causas. O problema inverso, por sua vez, refere-se à reversão *efeito* \rightarrow *causa*, ou seja, determinar as causas a partir do conhecimento dos efeitos. Na área de modelagem epidêmica, é comum encontrarmos trabalhos que focam no problema direto, ou seja, prever a evolução temporal do número de infectados a partir de informações sobre o tipo de modelo empregado e os parâmetros e condições iniciais com os quais ele é alimentado (Kissler et al., 2020; Silva; Menin, 2021). Porém, pode ser necessário reverter a ordem, ou seja, resolver o problema inverso de encontrar os parâmetros e condições iniciais da epidemia a partir do conhecimento da série temporal do número de infectados.

Nesse contexto, este projeto se propõe a resolver o problema inverso de encontrar os parâmetros de uma epidemia, utilizando como dado de entrada a série temporal do número de infectados. Considerando que esse trabalho ainda está em andamento, aqui focamos especificamente em encontrar apenas a taxa de transmissão e, conseqüente, o número de reprodução básica de uma dada epidemia modelada pelo modelo SIR (suscetível-infeccioso-recuperado) (Keeling; Rohani, 2011; Kermack; McKendrick, 1927). Para isso, transformou-se o modelo em um problema de otimização, no qual deve-se minimizar uma função objetivo que mede a discrepância entre os dados reais e os dados simulados pelo modelo, que foi resolvido pelo método *Simulated Annealing* (Kirkpatrick; Jr; Vecchi, 1983; Menin; Martinez; Rolnik, 2016).

METODOLOGIA

Inicialmente realizou-se uma revisão de conceitos básicos de cálculo numérico (Chapra; Canale, 2016) e equações diferenciais (Boyce; DiPrima, 2010), bem como o estudo de modelos epidêmicos compartimentais (Keeling; Rohani, 2011), problemas inversos (Menin; Martinez; Rolnik, 2016) e, por fim, do método de otimização *Simulated Annealing* (Kirkpatrick; Jr; Vecchi, 1983). Em seguida, definiu-se o modelo epidêmico a ser estudado, que foi implementado, juntamente com o *Simulated Annealing*, em linguagem *Python*.

O modelo epidêmico estudado foi baseado no clássico modelo SIR, no qual a população é dividida em três compartimentos, suscetíveis (S), infecciosos (I) e recuperados (R). Considere S , I e R como sendo, respectivamente, os números de indivíduos suscetíveis, infecciosos e recuperados e $N = S + I + R$ o tamanho da população. A transição de indivíduos do compartimento de suscetíveis para o de infecciosos ($S \rightarrow I$) acontece com uma taxa SI/N , sendo β a taxa de transmissão. Já a taxa de transição do compartimento de infecciosos para o de recuperados ($I \rightarrow R$) é dada por γI , sendo γ o inverso do período infeccioso. Nesse tipo de modelo, o número básico de reprodução é dado por $R_0 = \beta/\gamma$. (Keeling; Rohani, 2011). Os compartimentos e as transições entre eles são vistos na Fig. (1) e o modelo é dado pelo sistema de equações diferenciais

$$\frac{dS}{dt} = -\beta SI/N, \quad (1)$$

$$\frac{dI}{dt} = \beta SI/N - \gamma I, \quad (2)$$

$$\frac{dR}{dt} = \gamma I. \quad (3)$$

Considerando que esse sistema de equações não admite solução analítica, ele deve ser resolvido por meio de métodos numéricos (Chapra; Canale, 2016). Quando tratado de forma direta, o modelo recebe, como entrada, os valores dos parâmetros β e γ , bem como as condições iniciais (valores de S , I e R quando $t = 0$), e devolve, como saída, as séries temporais dos números de indivíduos suscetíveis, infecciosos e recuperados.

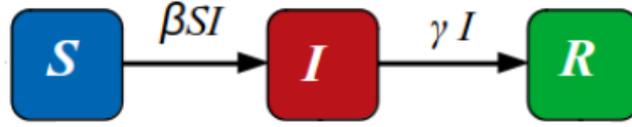


Figura 1: Diagrama com os compartimentos do modelo SIR e suas relações

Nesse projeto, queremos resolver um problema inverso do tipo *causal*, ou seja, conhecemos os dados de saída (no caso as séries temporais de cada compartimento) e desejamos determinar os parâmetros de entrada. Mais especificamente, tomamos como conhecida a série temporal diária *real* do número de infectados, $\mathbf{I}_{\text{real}} = [I_{\text{real}}^{(1)}, I_{\text{real}}^{(2)}, \dots, I_{\text{real}}^{(n)}]$, sendo $I_{\text{real}}^{(i)}$ o número real de infectados no dia $i = 1, 2, \dots, n$, e supomos um dado conjunto de parâmetros e condições iniciais x_1, x_2, \dots, x_m , expressados por meio do vetor $\mathbf{x} = (x_1, x_2, \dots, x_m)$. Com esse vetor resolvemos o modelo dado pelas Eqs. (1-3) para obtermos um série temporal *numérica* para o número de infectados, $\mathbf{I}_{\text{num}} = [I_{\text{num}}^{(1)}, I_{\text{num}}^{(2)}, \dots, I_{\text{num}}^{(n)}]$, sendo $I_{\text{num}}^{(i)}$ a previsão numérica para o número de infectados no dia i . A partir daí, constrói-se a função objetivo

$$f(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n [I_{\text{real}}^{(i)} - I_{\text{num}}^{(i)}]^2, \quad (4)$$

que mede a discrepância entre os dados reais e os previstos numericamente pelo modelo. Com isso, a solução do problema inverso será o conjunto de parâmetros $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_m^*)$ que minimize o funcional dado pela Eq. (4). Dessa forma, transformamos um problema inverso em um problema de otimização, o qual resolvemos por meio do método *Simulated Annealing* (SA).

Tal método faz analogia com o processo de “recozimento” (*annealing*), utilizado na metalurgia para obtenção de materiais com alto grau de dureza. Especificamente no algoritmo SA, parte-se de uma solução inicial arbitrária \mathbf{x}_0 a uma dada temperatura inicial T_0 pré-definida. Mantendo a temperatura fixa, realiza-se um processo iterativo de geração de novas soluções candidatas. Cada nova solução, \mathbf{x}_{new} , é confrontada com a solução anterior, \mathbf{x}_{old} , e aceita com probabilidade

$$p(\mathbf{x}_{\text{old}} \leftarrow \mathbf{x}_{\text{new}}) = \begin{cases} 1, & \text{se } \Delta f \leq 0, \\ e^{-\Delta f/T}, & \text{se } \Delta f > 0, \end{cases} \quad (5)$$

onde $\Delta f = f(\mathbf{x}_{\text{new}}) - f(\mathbf{x}_{\text{old}})$ é a variação da função objetivo e T é a temperatura.

Analogamente ao processo de “recozimento” da siderurgia, após o algoritmo tentar um número suficiente de soluções candidatas, a temperatura é resfriada e um novo processo iterativo de busca de novas soluções se inicia.

Existem diferentes formas de gerar novas soluções candidatas bem como de resfriar a temperatura. Em geral, cada nova solução, \mathbf{x}_{new} , é gerada a partir da solução atual, \mathbf{x}_{old} , de acordo com a equação

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{old}} + \sigma \xi, \quad (6)$$

onde ξ é um vetor de números aleatórios que seguem uma distribuição de probabilidade pré-definida (no caso desse trabalho, a distribuição normal) e σ é a amplitude do deslocamento da nova solução em relação à anterior. Já a temperatura parte de um valor inicial T_0 (interação $t = 0$) e cai exponencialmente a cada iteração t , tal que

$$T_t = \lambda^t T_0, \quad (7)$$

onde $\lambda \in (0, 1)$ é a taxa de resfriamento.

RESULTADOS E DISCUSSÃO

Tanto o modelo SIR proposto nas Eqs. (1-3), quanto o método de otimização foram implementados computacionalmente através da linguagem de programação *Python*. O código do SIR recebe, como parâmetro de entrada, uma estimativa para a taxa de transmissão β e retorna, como saída, a série temporal do número de infectados. Todos os demais parâmetros e condições iniciais utilizados foram considerados conhecidos, e seus valores são observados na Tabela 1.

Tabela 1: Parâmetros utilizados para a simulação do modelo SIR.

Parâmetro	Descrição	Valor
N	Tamanho da população	100000
γ	Inverso do período infeccioso	$0,2 \text{ dia}^{-1}$
t_i	Tempo inicial	0
t_f	Tempo final	100 dia^{-1}
I_0	Número de indivíduos infectados inicialmente	1

Para a simulação, chamou-se de I_{real} a série temporal obtida ao executar o código considerando $\beta_{\text{real}} = 0,5 \text{ dia}^{-1}$, seja, um número básico de reprodução $R_{0_{\text{real}}} = 0,5/0,2 = 2,5$. Com isso executou-se o código do *Simulated Annealing*, cuja função foi de gerar diferentes valores de β e obter, com base neles, diferentes I_{num} . Essas diferentes séries numéricas foram comparadas a série real, I_{real} , por meio da Eq. (4), minimizada com o passar das iterações. Os parâmetros para o código do *Simulated Annealing* estão descritos na Tabela 2.

Tabela 2: Parâmetros utilizados para o *Simulated Annealing*.

Parâmetro	Descrição	Valor
r	Número de iterações antes do resfriamento	10
n_{it}	Número de vezes que o sistema é resfriado	100
λ	taxa de resfriamento da temperatura	0.9

A série temporal real e numérica final são observados na Fig. (2-a), onde percebe-se que o *Simulated Annealing* foi capaz de encontrar corretamente a taxa β que descreve o I_{real} . O valor encontrado foi $\beta = 0,49999$, que corresponde a um número básico de reprodução $R_0 = 0,49999/0,2 \approx 2,5$.

Para explorar um cenário diferente, que simulasse um problema nos dados reais, adicionou-se um ruído gaussiano à serie temporal I_{real} . Os parâmetros da simulação foram mantidos em conformidade

com as Tabelas 1 e 2 e os resultados para a série temporal ruidosa são vistos na Fig. (2-b). Mais uma vez nota-se a capacidade do método de encontrar a taxa β que melhor descreva uma série temporal similar a real antes da adição do ruído. O valor encontrado nesse caso foi de $\beta = 0,49856$, correspondendo a $R_0 = 0,49856/0,2 \approx 2,5$.

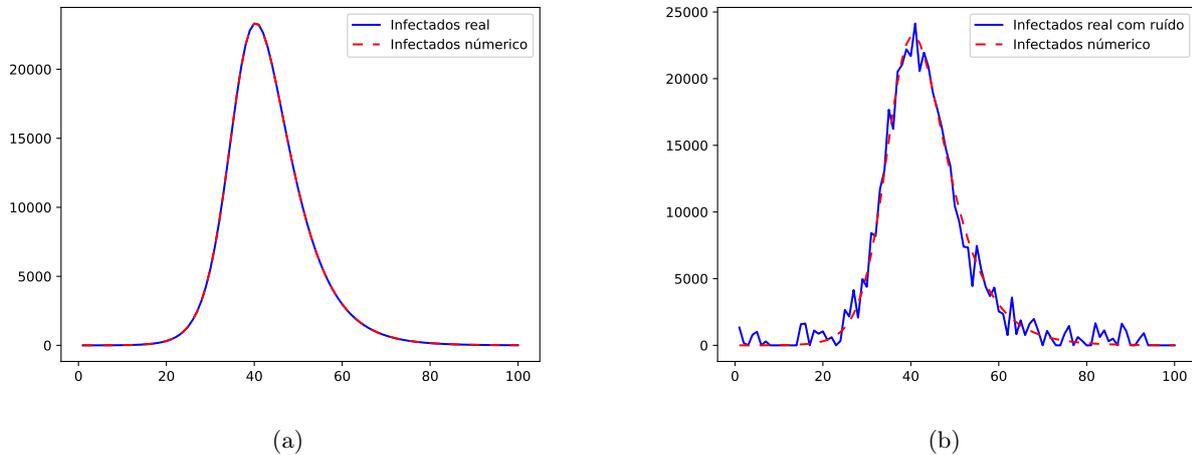


Figura 2: Séries temporais do número de infectados real (azul sólido), obtida com $\beta_{\text{real}} = 0,5$, e numérica (vermelho tracejado), obtida pelo método *Simulated Annealing*, considerando a série temporal real (a) sem ruído e (b) adicionada de um ruído gaussiano. Em ambos os casos, o método encontrou uma taxa de transmissão $\beta \approx 0,5$, correspondendo a um número básico de reprodução $R_0 \approx 2,5$.

CONCLUSÕES

Apresentamos aqui uma possibilidade de encontrar parâmetros de uma epidemia com base em sua evolução temporal, utilizando o método de otimização *Simulated Annealing*. Através dos resultados obtidos, seria possível entender melhor um surto epidêmico ocorrido, o que possibilitaria, por sua vez, ajudar entidades responsáveis no estudo das melhores decisões a serem tomadas caso uma epidemia similar voltasse a ocorrer. É importante destacar que modelos matemáticos possuem limitações, pois em muitos casos não consideram todos os parâmetros que influenciam uma simulação. No caso desse trabalho, por exemplo, apesar de considerarmos um ruído na série temporal de infectados, é difícil saber quais são todos os parâmetros que podem perturbar a série, o que torna a simulação da mesma inexata. Apesar disso espera-se que os resultados obtidos sejam capazes de orientar trabalhos futuros no desenvolvimento de modelos auxiliares no estudo de epidemias.

CONTRIBUIÇÕES DOS AUTORES

Os autores contribuíram igualmente no estudo, concepção e execução do projeto de pesquisa, bem como na revisão e aprovaram a versão final submetida.

AGRADECIMENTOS

Os autores agradecem ao Centro de Pesquisa e Inovação (CEPIN) do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - Campus Sertãozinho e ao Programa Institucional Voluntário de Iniciação Científica e Tecnológica (PIVICT).

REFERÊNCIAS

- Alifanov, O. M. *Inverse heat transfer problems*. [S.l.]: Springer Science & Business Media, 2012.
- Boyce, W. E.; DiPrima, R. C. *Equações diferenciais elementares e problemas de valores de contorno*. [S.l.]: LTC Rio de Janeiro, 2010. v. 10.
- Chapra, S. C.; Canale, R. P. *Métodos Numéricos para Engenharia-7ª Edição*. [S.l.]: McGraw Hill Brasil, 2016.
- Keeling, M. J.; Rohani, P. *Modeling infectious diseases in humans and animals*. [S.l.]: Princeton university press, 2011.
- Kermack, W. O.; McKendrick, A. G. 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*, The Royal Society London, v. 115, n. 772, p. 700–721, 1927.
- Kirkpatrick, S.; Jr, C. D. G.; Vecchi, M. P. Optimization by simulated annealing. *science*, American association for the advancement of science, v. 220, n. 4598, p. 671–680, 1983.
- Kissler, S. M. et al. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science*, American Association for the Advancement of Science, v. 368, n. 6493, p. 860–868, 2020.
- Menin, O. H.; Martinez, A. S.; Rolnik, V. *Tomografia de impedância elétrica: métodos computacionais*. [S.l.]: Livraria da Física, 2016. v. 1.
- Silva, V. R. D.; Menin, O. H. Modified sir compartmental epidemic model with social distancing and hospital saturation applied to the covid-19 pandemic. *Russian Journal of Nonlinear Dynamics*, v. 17, n. 3, p. 275–287, 2021.