

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Análise de Sensibilidade e Calibração de Modelos para o Crescimento Tumoral

Gustavo T. Naozuka¹Maurício Pessoa C. Menezes²Regina C. Almeida³

Laboratório Nacional de Computação Científica – LNCC

Resumo. O câncer é um conjunto de doenças cujos mecanismos não são completamente conhecidos. Existe na literatura uma grande diversidade de modelos matemáticos que objetivam contribuir para um melhor entendimento de fenômenos-chaves na dinâmica tumoral. Neste trabalho analisamos alguns modelos clássicos do crescimento de tumores sob o ponto de vista da quantificação de incertezas. Utilizamos um método simples de análise de sensibilidade e a abordagem Bayesiana para a estimação dos parâmetros dos modelos. Nossa análise mostra que as incertezas nos parâmetros impactam significativamente as estimativas do crescimento do tumor. A calibração dos modelos fornece, além do valor mais provável para os parâmetros, as suas distribuições de probabilidade. A análise sugere que uma abordagem abrangente para a seleção do melhor modelo é primordial para quantificação da qualidade da predição.

Palavras-chave. Oncologia Preditiva, Quantificação de Incertezas, Problema Inverso

1 Introdução

Sabe-se que o câncer é um conjunto de doenças que atualmente é uma das principais causas de morte no mundo [2, 7]. De forma simplificada, é caracterizado pelo crescimento celular descontrolado, frequentemente invadindo diferentes tecidos e órgãos do corpo. A modelagem matemática e computacional já é reconhecida como uma abordagem útil para contribuir no entendimento dos mecanismos envolvidos no crescimento tumoral, realizar predições e auxiliar na tomada de decisões. Entretanto, sem uma abordagem rigorosa para avaliar a confiabilidade de modelos sujeitos a diversos tipos de incertezas, os modelos computacionais são de pouca aplicabilidade na ciência médica. Em [3], por exemplo, os autores apontam diferenças relevantes nas predições obtidas com diversos modelos de crescimento tumoral, comumente usados na literatura para descrever o crescimento de tumores a partir de equações diferenciais ordinárias. Naquele estudo foram avaliados sete modelos, cujos parâmetros foram ajustados, através do método dos mínimos quadrados, aos dados

¹naozuka@lncc.br

²mpessoa@lncc.br

³rcca@lncc.br

publicados em [8] sobre o crescimento de tumor em camundongos submetidos à quimioterapia. Os autores apontam que os modelos melhores calibrados não necessariamente correspondem aos que conduzem às melhores previsões, enfatizando a importância de uma análise mais abrangente da que se baseia apenas no melhor ajuste. Este é um exemplo claro da importância de definições metodológicas para o tratamento de questões-chaves da ciência preditiva, como a quantificação de incertezas e seleção de modelos.

Neste trabalho, utilizando os mesmos dados e alguns dos modelos estudados em [3], realizamos um estudo do uso da abordagem Bayesiana e da análise de sensibilidade na quantificação das incertezas do cenário de interesse. Empregamos o método dos efeitos elementares (*elementary effects*) [6] para a análise de sensibilidade e o *software* livre QUESO (*Quantification of Uncertainty for Estimation, Simulation and Optimization*) [4] para a calibração dos modelos. As simulações realizadas demonstram como a combinação dessas duas estratégias contribui para o entendimento dos comportamentos dos modelos para a descrição do cenário de interesse.

2 Materiais e Métodos

2.1 Dados Experimentais

O conjunto de dados experimentais utilizado é constituído por 14 pontos de volume do tumor (mm^3), distribuídos em um período de 114 dias, em camundongos atômicos. Esse conjunto de dados foi extraído por Murphy *et al.* [3] de Worschech *et al.* [8]. Para a realização deste trabalho, empregamos a ferramenta *WebPlotDigitizer* [5] para capturar os dados do volume do tumor, nossa quantidade de interesse (QoI), presentes em [3].

2.2 Modelos Matemáticos

Selecionamos dois modelos de crescimento ilimitado: o clássico modelo exponencial de Malthus (1766-1834) e sua generalização proposta por Mortimer L. Mendelsohn em 1963; e dois modelos que contemplam a limitação do crescimento pelos recursos: o também clássico modelo logístico de Verhulst (1804-1849) e o desenvolvido por Karl Ludwig von Bertalanffy em 1949, o qual considera que o aumento do volume do tumor é proporcional à área da superfície do tumor e que o decréscimo decorre da morte. Esses modelos, cujos desenvolvimentos têm fundamento experimental e vêm sendo usados para a modelagem do crescimento tumoral (ver referências em [3]), são apresentados na Tabela 1, que inclui as correspondentes soluções analíticas. Nesta descrição, N denota o volume do tumor (QoI adotada), N_0 representa seu valor inicial, e t indica tempo. O parâmetro a presente em todos os modelos representa a taxa de crescimento do tumor. Por outro lado, o parâmetro b tem diferentes significados: no modelo logístico, indica a capacidade suporte do meio; no de von Bertalanffy, representa a taxa de mortalidade do tumor; e no de Mendelsohn, a potência da população para a qual o crescimento é proporcional. Assumimos unidades apropriadas (u.a.) para todos os parâmetros, e o tumor é medido em mm^3 . Tanto N_0 quanto a e b são ajustados a partir dos dados experimentais de [3]. Por essa razão, N_0 é também denominado “parâmetro” daqui em diante. De forma genérica, definimos que a

cada modelo está associado um vetor de parâmetros $\theta = \{\theta_1, \dots, \theta_n\}$ a ser determinado pela calibração Bayesiana.

Tabela 1: Modelos matemáticos de crescimento tumoral utilizados neste trabalho.

Modelo	Equação	Solução Analítica
Exponencial	$\frac{dN}{dt} = aN$	$N(t) = N_0e^{at}$
Logístico	$\frac{dN}{dt} = aN \left(1 - \frac{N}{b}\right)$	$N(t) = \frac{N_0e^{at}}{1 - \frac{N_0}{b}(1 - e^{at})}$
von Bertalanffy	$\frac{dN}{dt} = aN^{\frac{2}{3}} - bN$	$N(t) = \left[\frac{a}{b} + \left(N_0^{\frac{1}{3}} - \frac{a}{b}\right) e^{-\frac{bt}{3}}\right]^3$
Mendelsohn	$\frac{dN}{dt} = aN^b$	$N(t) = \left[\left(at + \frac{N_0^{1-b}}{1-b}\right) \cdot (1-b)\right]^{\frac{1}{1-b}}$

2.3 Análise de Sensibilidade

Utilizamos o método dos efeitos elementares para identificar como incertezas nos parâmetros de entrada dos modelos influenciam o volume tumoral, que passamos a denotar por $N(\theta)$. O método define o efeito elementar EE para um parâmetro θ_i como:

$$EE_i = \frac{N(\theta_1, \dots, \theta_i + \delta, \dots, \theta_n) - N(\theta_1, \dots, \theta_i, \dots, \theta_n)}{\delta}, \quad (1)$$

em que $\delta \in \left\{\frac{1}{p-1}, \dots, 1 - \frac{1}{p-1}\right\}$, e p é o número de níveis de discretização do espaço paramétrico, definido pelo hipercubo unitário de dimensão n . Assim, cada direção do hipercubo está associada a um parâmetro, cuja variação é mapeada entre 0 e 1. Os seguintes índices de sensibilidade são calculados a partir dos $EE_i, i = 1, \dots, n$:

$$\mu_i = \frac{1}{r} \sum_{j=1}^r EE_i^j; \quad \mu_i^* = \frac{1}{r} \sum_{j=1}^r |EE_i^j|; \quad \sigma_i = \sqrt{\frac{1}{r-1} \sum_{j=1}^r (EE_i^j - \mu_i)^2}, \quad (2)$$

em que r é o número de trajetórias percorridas no espaço de parâmetros. Os índices μ_i^* e σ_i indicam a influência e a importância não linear do i -ésimo parâmetro sobre a QoI, respectivamente. Quanto mais influente for o parâmetro, pequenas incertezas no seu valor terão grande impacto sobre a estimativa da QoI, o que ressalta a necessidade de calibrá-los de forma mais precisa. Neste trabalho, a análise de sensibilidade foi realizada no tempo final dos dados experimentais (114 dias), e adotamos $p = 4, \delta = \frac{2}{3}$ e $r = 10$.

2.4 Calibração Bayesiana

A abordagem Bayesiana permite considerar incertezas existentes tanto no modelo quanto nos dados experimentais. Essa técnica é baseada no teorema de Bayes [1]: dado um conhecimento inicial sobre os parâmetros, definido por uma distribuição de probabilidade

a priori $p(\theta)$, e uma função de verossimilhança $l(\mathbf{y}|\theta)$, podemos melhorar o conhecimento sobre os parâmetros obtendo a distribuição de probabilidade *a posteriori* $p(\theta|\mathbf{y})$, a partir da relação $p(\theta|\mathbf{y}) \propto l(\mathbf{y}|\theta)p(\theta)$. A função de verossimilhança $l(\mathbf{y}|\theta)$ é assumida aqui como gaussiana, define a dependência dos dados em relação aos parâmetros, permitindo atualizar o conhecimento sobre os parâmetros a partir dos dados experimentais. Assumimos que os parâmetros são independentes e seguem distribuições *a priori* uniformes, em virtude do conhecimento limitado sobre eles. Ressaltamos, entretanto, que diferentes formulações para a função de verossimilhança e distribuições *a priori* podem ser utilizadas. A distribuição de probabilidade *a posteriori* é obtida numericamente, usando procedimentos de amostragem do tipo *Markov Chain Monte Carlo*. Para isso, utilizamos a biblioteca de código aberto QUESO (*Quantification of Uncertainty for Estimation, Simulation and Optimization*) [4]. A calibração dos parâmetros de cada modelo é feita usando o método de Monte Carlo de Múltiplos níveis, com 20 mil números de amostras.

3 Resultados e Discussões

As distribuições uniformes *a priori* para os parâmetros dos modelos foram definidas, em sua maioria, a partir de intervalos de $\pm 20\%$ dos valores estimados em [3] para os sete modelos lá considerados. Assumimos que essa faixa é suficiente para caracterizar as incertezas dos parâmetros dos modelos aqui considerados, um subconjunto dos avaliados em [3]. Os intervalos correspondentes são apresentados na Tabela 2. Além disso, para todos os modelos assumimos $N_0 \sim \mathcal{U}[184, 286]mm^3$ e $\sigma \sim \mathcal{U}[0, 348]mm^3$, sendo que o hiperparâmetro σ inclui tanto incertezas nos modelos quanto nos dados.

Assumindo que as incertezas paramétricas são definidas pelas distribuições *a priori* definidas, realizamos primeiramente a análise de sensibilidade dos modelos. A

Tabela 2: Intervalos adotados para as distribuições *a priori*.

Modelo	a_{min}	a_{max}	b_{min}	b_{max}
Exponencial	0,01968 /d	0,02952 /d	–	–
Logístico	0,0236 /d	0,0354 /d	5536 mm ³	8304 mm ³
von Bertalanffy	0,1872 mm/d	0,2808 mm/d	0 /d	0,1 /d
Mendelsohn	0,084 /d	0,126 /d	0,628	0,942

Figura 1 ilustra os índices de sensibilidade obtidos para os diversos modelos. A análise mostra que de forma geral todos os parâmetros são importantes na dinâmica, mesmo que com intensidades diferentes. A taxa de crescimento a foi o parâmetro mais influente nos modelos exponencial e logístico, enquanto que o parâmetro b foi o mais influente nos modelos de von Bertalanffy e Mendelsohn. De forma similar, todos os parâmetros influenciam a QoI de forma não linear, mesmo que com intensidades diferentes. Exceção para a condição inicial N_0 no modelo de von Bertalanffy, que apresenta valores bem pequenos para os índices de sensibilidade, demonstrando ser pouco influente na dinâmica do crescimento em relação aos demais parâmetros do modelo. De forma geral, a análise mostra que, para os modelos e dados considerados, as incertezas nos parâmetros impactam significativamente as estimativas do crescimento do tumor, sendo fundamental calibrá-los precisamente para a obtenção de predições precisas.

A calibração dos parâmetros dos modelos matemáticos foi realizada empregando a técnica Bayesiana, utilizando os 14 pontos de volume do tumor em um período de 114 dias apresentados em [3]. A Tabela 3 contém os histogramas de frequências referentes à distribuição *a posteriori* dos parâmetros *a* e *b*, mais influentes nos modelos considerados. Não apresentamos os histogramas dos parâmetros N_0 e σ em virtude da limitação de espaço. Para as figuras ilustradas na tabela, decidimos manter os rótulos dos eixos, apesar da escala reduzida.

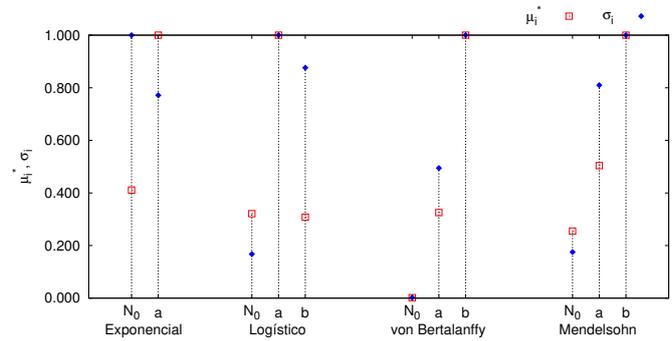


Figura 1: Índices de sensibilidade μ_i^* e σ_i , normalizados para cada modelo.

Tabela 3: Histogramas de frequências referentes à distribuição *a posteriori* dos parâmetros *a* e *b* resultantes do processo de calibração Bayesiana.

Par.	Exponencial	Logístico	von Bertalanffy	Mendelsohn
<i>a</i>				
<i>b</i>	—			

A inspeção visual dos histogramas permite identificar a qualidade das calibrações realizadas. De forma geral, podemos observar que apenas para os modelos exponencial e logístico houve uma melhora substancial do conhecimento sobre os parâmetros com a aplicação da inferência Bayesiana, apesar da distribuição *a posteriori* do parâmetro *b* do modelo logístico ser praticamente uniforme. Isso decorre do fato de os dados não apresentarem qualquer comportamento de saturação do crescimento, não sendo portanto adequados para uma boa calibração de *b*. Ressaltamos que as demais distribuições associadas aos modelos de von Bertalanffy e Mendelsohn indicam a necessidade de rever os intervalos de definição das distribuições *a priori* e de dados mais informativos.

A Tabela 4 apresenta a comparação entre os valores mais prováveis (VMPs) dos parâmetros calibrados neste trabalho com os calculados em [3], e inclui os VMPs obtidos para σ . Podemos observar a similaridade entre os valores dos parâmetros, com exceção do parâmetro *b* do modelo de von Bertalanffy. Apesar disso, devemos destacar que os histogramas apresentados na Tabela 3 indicam que as estimativas realizadas estão acompanhadas de incertezas significativas, como mencionado anteriormente. A estratégia

de calibração usada em [3] não possibilita análise similar.

Tabela 4: Valores dos parâmetros de Murphy *et al.* [3] e VMPs dos parâmetros calibrados dos modelos matemáticos de crescimento tumoral.

Modelo	Murphy <i>et al.</i> [3]		Parâmetros calibrados (VMPs)			
	a	b	N_0	a	b	σ
Exponencial	0,0246 /d	–	275,837 mm ³	0,023 /d	–	143,157 mm ³
Logístico	0,0295 /d	6920 mm ³	236,991 mm ³	0,028 /d	7987,085 mm ³	107,689 mm ³
von Bertalanffy	0,234 mm/d	$3,46 \times 10^{-19}$ /d	184,750 mm ³	0,250 mm/d	$3,436 \times 10^{-4}$ /d	125,963 mm ³
Mendelsohn	0,105 /d	0,785	234,093 mm ³	0,092 /d	0,803	113,535 mm ³

Finalmente, a Figura 2 exibe a evolução do crescimento do tumor para cada um dos modelos matemáticos calibrados. De forma geral, todos os modelos representaram os dados experimentais de maneira satisfatória. Observamos que o intervalo de credibilidade é bem estreito, apesar das incertezas apontadas anteriormente. Os VMPs dos hiperparâmetros exibidos na Tabela 4 são bastante similares, sendo que o modelo logístico apresentou o menor valor.

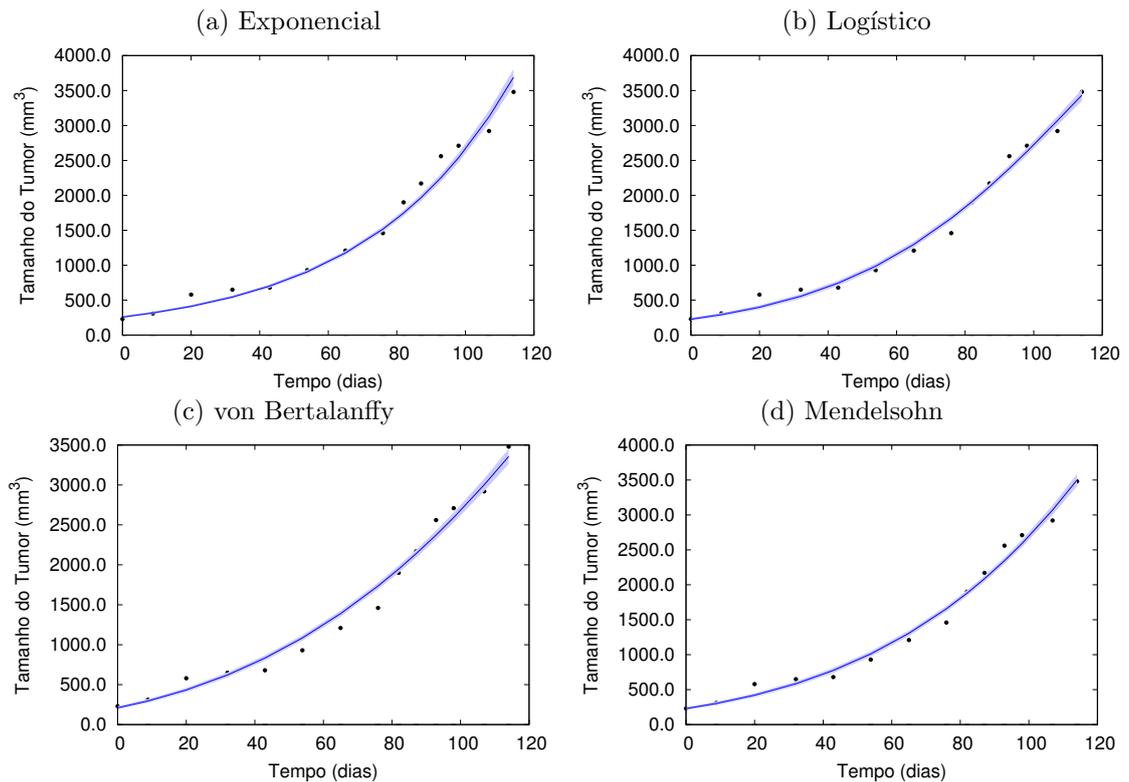


Figura 2: Simulação dos modelos matemáticos considerados em azul, com o intervalo de 95% de credibilidade sobreposto. Os pontos em preto representam os dados experimentais.

4 Considerações Finais

Neste trabalho analisamos alguns modelos clássicos da dinâmica de evolução do câncer a partir de dados experimentais em murinos. Essa análise permitiu a identificação da ordem de influência entre os parâmetros de cada modelo com relação à evolução do volume tumoral. A calibração dos modelos foi feita utilizando a abordagem Bayesiana, que fornece não apenas o valor mais provável para os parâmetros quanto suas distribuições de probabilidade, em contraste com os valores determinísticos obtidos por Murphy *et al.* [3]. Identificamos que em apenas dois modelos, o exponencial e o logístico, os dados foram informativos, aprimorando significativamente o conhecimento *a priori* sobre os parâmetros. Entretanto, a quantificação mais precisa desse comportamento requer o uso de técnicas apropriadas para esse fim, como o critério de informação de Fisher, que captura a sensibilidade da função de verossimilhança às variações nos parâmetros.

Agradecimentos: Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

Referências

- [1] G. Box and G. Tiao. *Bayesian inference in statistical analysis*. Addison-Wesley series in behavioral science: Quantitative methods. Addison-Wesley Pub. Co., 1973. doi: 10.1002/9781118033197.
- [2] INCA. Instituto Nacional de câncer. Estatísticas de câncer. www.inca.gov.br/numeros-de-cancer, Nov 2018.
- [3] H. Murphy, H. Jaafari, and H. M. Dobrovolny. Differences in predictions of ODE models of tumor growth: a cautionary example. *BMC Cancer*, 16(1):163, 2016. doi: 10.1186/s12885-016-2164-x.
- [4] E. E. Prudêncio and K. W. Schulz. The parallel C++ statistical library ‘QUESO’: Quantification of Uncertainty for Estimation, Simulation and Optimization. In *EuroPar 2011: Parallel Processing Workshops*, pages 398–407. Springer, 2012. doi: 10.1007/978-3-642-29737-3_44.
- [5] A. Rohatgi. WebPlotDigitizer version: 4.1. <https://automeris.io/WebPlotDigitizer/>, Jan 2018.
- [6] A. Saltelli. *Global sensitivity analysis: the primer*. John Wiley, 2008. ISBN 9780470059975. doi: 10.1002/9780470725184.
- [7] WHO. World Health Organization. Cancer. www.who.int/cancer/en/, Feb 2019.
- [8] A. Worschech, N. Chen, Y. A. Yu, and et al. Systemic treatment of xenografts with vaccinia virus GLV-1h68 reveals the immunologic facet of oncolytic therapy. *BMC Genomics*, 10:301–301, 2009. doi: 10.1186/1471-2164-10-301.