



B i o

Matemática

35



IMECC – Unicamp

Campinas

2025

ISSN 1679-365X



Loris molucano, Periquito arco-iris.

Foto: G.L.Diniz

Conselho Editorial

- Rodney C. Bassanezi – UNICAMP (Coordenador);
- Estevão E. Laureano – UNICAMP;
- Evandro E. Marquesone – UTFPR (Organizador);
- João Frederico C. A. Meyer – UNICAMP;
- Laécio C. Barros – UNICAMP;
- Laércio L. Vendite – UNICAMP;
- Marcos Eduardo R. V. Mesquita – UNICAMP;
- Wilson C. Ferreira Jr. – UNICAMP;
- Geraldo L. Diniz – UFMT (Organizador).

Acesso on-line:
<https://ime.unicamp.br/~biomat/revistas.htm>

Prefácio:

Neste trigésimo quinto volume da revista, há contribuições importantes para o conhecimento desta área interdisciplinar denominada Biomatemática.

Os artigos deste volume são fruto de um esforço conjunto e da dedicação dos pesquisadores que acreditam na solução dos problemas atuais e que vem exigindo uma maior cooperação entre pesquisadores de áreas diversificadas.

Ao percorrer as páginas desta revista, a pessoa irá descobrir diferentes abordagens, cujas aplicações se estendem a uma série de problemas em áreas do conhecimento ligadas à epidemiologia, oncologia, dinâmica populacional e questões ambientais.

Com o intuito de que tenham uma boa leitura e possam apreciar o que se apresenta no presente volume deste periódico, é que o corpo editorial selecionou estes artigos.

Comissão Editorial – IMECC-Unicamp.

Campinas, 24 de dezembro de 2025.

Índice

Estimativa do número de infectados por HIV via quadrados mínimos fuzzy e interatividade do tipo J_0 , A. L. P. Mendes, L. A. Nunes, Y. B. Shimizu & V. F. Wasques _____ 1

Estudo da epidemiologia matemática da tuberculose no Brasil, V. E. G. Barelli, T. M. Bigardi, D. V. P. Chaves, F. S. Minatogau & A. C. B. Loponi _____ 21

*Estudo de dinâmica populacional de *Brachionus plicatilis* em função da toxicidade da bituca de cigarro em diferentes concentrações*, A. A. Caetano, C. E. G. Santos, J. D. A. Sales & V. F. Wasques _____ 41

Avaliação de parâmetros em modelos de carga hidráulica para aquíferos confinados via equação de difusão, A. C. J. Coutinho & E. C. C. Poletti _____ 55

Influência dos diferentes métodos de defuzzificação nos resultados do estudo da qualidade da água usando Python, J. D. S. Oliveira & V. S. Oliveira _____ 75

Estimativa do número de infectados por HIV via quadrados mínimos fuzzy e interatividade do tipo J_0

Ana L. P. Mendes¹, Letícia A. Nunes², Yasmin B. Shimizu³, Vinícius F. Wasques⁴

Ilum Escola de Ciência, CNPEM, 13.087-548, Campinas/SP.

Resumo. Este trabalho fornece um estudo sobre a estimativa do número de infectados por HIV no Brasil de 2015 a 2024. Para isso é considerada a incerteza atrelada aos dados coletados na literatura, em que tal incerteza é modelada pela teoria de conjuntos fuzzy. Por meio do método de Quadrados Mínimos Fuzzy e da interatividade via J_0 é fornecido um ajuste de curvas, a fim de realizar uma previsão do número de infectados no país para os próximos anos. Os sistemas lineares fuzzy associados ao problema são resolvidos via decomposição LU e as operações computacionais são realizadas em Python 3.11.

Palavras-chave: Quadrados mínimos fuzzy, epidemiologia, HIV, interatividade fuzzy.

1. Introdução

O ajuste de curvas é uma técnica matemática fundamental em análise de dados e modelagem, que consiste em encontrar uma função que melhor descreve o comportamento de um conjunto de dados. Isto é, por meio de um gráfico de dispersão, os dados observados são comparados a funções conhecidas, como polinômios, exponenciais, logarítmicas ou modelos mais complexos, dependendo

¹ana24007@ilum.cnpem.com

²leticia24013@ilum.cnpem.br

³yasminbshimz@gmail.com

⁴vwasques@outlook.com - FAPESP n. 2023/03927-0

da natureza do problema. Em seguida, o objetivo se torna minimizar a discrepância entre os valores reais e os valores previstos pela função, geralmente por meio de métodos como quadrados mínimos, que é definido pelo problema $\|A\bar{x} - b\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2$, cuja $\|\cdot\|_2$ é a norma-2 (Ruggiero e Lopes, 2000).

Uma das aplicações mais comuns do ajuste de curvas é para realizar previsões, como estimar o número de infectados de uma certa doenças (Pinto, 2023). A partir de uma função ajustada, é possível extrapolar os dados além do intervalo observado, auxiliando na tomada de decisão como, por exemplo, na epidemiologia, cujas curvas ajustadas podem projetar a propagação de doenças.

No entanto, os métodos clássicos de ajuste de curvas não incorporam incertezas inerentes ao fenômeno, como é o caso de infecção por doenças, no qual não é possível saber com exatidão o número de infectados em um determinado instante. Neste caso, outras teorias podem ser utilizadas como apoio, a fim de fornecer resultados mais próximos da realidade.

No presente trabalho a teoria de conjuntos fuzzy é utilizada para incorporar tal incerteza. Na literatura é possível encontrar vários trabalhos que utilizam a lógica fuzzy para a modelagem de problemas epidemiológicos. Por exemplo, Sacilotto (2017) realizou um estudo sobre a dinâmica do câncer utilizando sistemas baseados em regras fuzzy, Santos (2018) utilizou a mesma metodologia para realizar um estudo sobre estimativa de risco de endometriose, a partir de diagnósticos subjetivos. Pinto (2023) forneceu uma nova pseudo-norma para números fuzzy com aplicações em ajuste de curvas por meio de quadrados mínimos fuzzy.

A proposta deste artigo é na direção de estudos de Pinto (2023), isto é, será realizado um estudo de ajuste de curvas por meio do método de quadrados mínimos fuzzy, em que os dados considerados possuem incerteza. No entanto, diferentemente do trabalho proposto por Pinto (2023), em que é feito uso da relação de interatividade do tipo completamente correlacionado, aqui será utilizada a interatividade via distribuição de possibilidade conjunta J_γ , proposta por Wasques (2019).

A relação de interatividade é uma relação muito similar a de dependência para o caso de variáveis aleatórias. No entanto, não são equivalentes. Enquanto na teoria de probabilidade a relação de dependência está associada a uma distribuição de probabilidade, na teoria fuzzy a interatividade está associada a distribuição de possibilidade.

A interatividade completamente correlacionada está associada a uma correlação linear entre as variáveis fuzzy, isto é, as funções de pertinência dos números fuzzy envolvidos podem ser conectados por meio de uma reta (Simões, 2017). Isso obriga que os números fuzzy envolvidos tenham o mesmo formato, ou seja, se um número fuzzy for triangular e simétrico, todos os outros também devem ser. Apesar de gerar bons resultados, isso restringe o uso de tal proposta caso a modeladora queira utilizar mais de um tipo de conjunto fuzzy em sua abordagem. Por outro lado, a interatividade via J_γ , apesar de ser custosa computacionalmente, não obriga tal correlação. Na verdade, ela não possui restrições e pode ser usada para qualquer formato de número fuzzy.

Nessa família de distribuições J_γ , o parâmetro $\gamma \in [0, 1]$ está associado ao “nível” de interatividade, isto é, quanto mais próximo de 0 for o valor de γ , maior é o nível de interatividade (mais forte é a dependência entre as variáveis). Aqui o foco é estudar as consequências de um ajuste de curvas utilizando o nível mais forte de interatividade, isto é para $\gamma = 0$. Sendo assim, a proposta é utilizar o método de quadrados mínimos fuzzy sob aritmética herdada de J_0 para realizar a previsão de dados de HIV com incerteza e, a partir disso, realizar previsões do números de infectados pela doença no Brasil.

2. Preliminares

Esta seção traz alguns conceitos básicos para o entendimento sobre resolução de sistemas lineares via decomposição LU, o ajuste de curvas clássico e a teoria de conjuntos fuzzy e as diferentes álgebras definidas sobre o espaço fuzzy.

2.1. Decomposição LU para resolução de sistemas lineares

Considere um sistema linear de equações dados por

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nn}x_n = b_n \end{cases},$$

que pode ser representado na forma matricial $Ax = b$, sendo A a matriz dos coeficientes, x o vetor das incógnitas e b o vetor resposta. Acompanhe abaixo:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Partindo do pressuposto que $\det(A) \neq 0$, é possível dividir A em duas outras matrizes L e U , deixando as operações de resoluções mais simples. A matriz L se trata de uma matriz triangular inferior formada por multiplicadores (m_{ij}) da decomposição gaussiana e a diagonal é unitária, já a matriz U tem uma matriz triangular superior cujos seus componentes são os resultados da decomposição gaussiana feita em A (Ruggiero e Lopes, 2000).

Após a fatoração de A em duas matrizes tem-se $LUx = b$, em que pode ser entendida como $Ly = b$, se for admitido que $Ux = y$. Assim, é viável encontrar o vetor x a partir da determinação de y .

No presente trabalho, realizamos também a técnica de pivoteamento parcial para evitar divisões por zero e o problema da aritmética de ponto flutuante durante a eliminação gaussiana – uma das etapas da fatoração. Essa estratégia consiste em reorganizar as linhas de uma matriz A de forma que os valores da diagonal principal ($i = j$) sejam maiores que todos abaixo dela ($i > j$).

Para realizar esse método, se parte de uma matriz inicial U que será uma cópia da matriz que será fatorada $A_{n \times n}$ e a matriz inicial L será uma matriz identidade com o mesmo tamanho que $A_{n \times n}$. Com a variável m sendo um contador começando em 0, teremos o seguinte algoritmo para a obtenção das matrizes L e U :

Enquanto $m < n$:

<table style="border-collapse: collapse; margin-left: 1em;"> <tr> <td style="border-left: 1px solid black; padding-left: 0.5em;"> Para todo $i \in \{m + 1, n - 1\}$: <table style="border-collapse: collapse; margin-left: 1em;"> <tr> <td style="padding-left: 0.5em;">$multiplicador = U_{im}/U_{mm}$</td> </tr> <tr> <td style="padding-left: 0.5em;">$L_{im} = multiplicador$</td> </tr> <tr> <td style="padding-left: 0.5em;">Para todo $j \in \{m, n\}$:</td> </tr> <tr> <td style="padding-left: 1em;">$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$</td> </tr> </table> </td> </tr> <tr> <td style="border-left: 1px solid black; padding-left: 0.5em;">$m = m + 1$</td> </tr> </table>	Para todo $i \in \{m + 1, n - 1\}$: <table style="border-collapse: collapse; margin-left: 1em;"> <tr> <td style="padding-left: 0.5em;">$multiplicador = U_{im}/U_{mm}$</td> </tr> <tr> <td style="padding-left: 0.5em;">$L_{im} = multiplicador$</td> </tr> <tr> <td style="padding-left: 0.5em;">Para todo $j \in \{m, n\}$:</td> </tr> <tr> <td style="padding-left: 1em;">$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$</td> </tr> </table>	$multiplicador = U_{im}/U_{mm}$	$L_{im} = multiplicador$	Para todo $j \in \{m, n\}$:	$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$	$m = m + 1$
Para todo $i \in \{m + 1, n - 1\}$: <table style="border-collapse: collapse; margin-left: 1em;"> <tr> <td style="padding-left: 0.5em;">$multiplicador = U_{im}/U_{mm}$</td> </tr> <tr> <td style="padding-left: 0.5em;">$L_{im} = multiplicador$</td> </tr> <tr> <td style="padding-left: 0.5em;">Para todo $j \in \{m, n\}$:</td> </tr> <tr> <td style="padding-left: 1em;">$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$</td> </tr> </table>	$multiplicador = U_{im}/U_{mm}$	$L_{im} = multiplicador$	Para todo $j \in \{m, n\}$:	$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$		
$multiplicador = U_{im}/U_{mm}$						
$L_{im} = multiplicador$						
Para todo $j \in \{m, n\}$:						
$U_{ij} = U_{ij} - (multiplicador \times U_{mj})$						
$m = m + 1$						

Retorne L, U .

Ao fim desse processo, é possível obter as seguintes matrizes:

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ m_{21} & 1 & 0 & \cdots & 0 \\ m_{31} & m_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \cdots & 1 \end{bmatrix} \quad \text{e} \quad U = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn}^{(k)} \end{bmatrix},$$

sendo k o número de eliminações gaussianas que foram realizadas sobre a matriz A até que esta se tornasse uma matriz triangular superior.

É importante entender que a escolha desse método se baseia no fato de que, ao aplicar o método de mínimos quadrados com abordagem fuzzy descrito abaixo, será necessário resolver vários sistemas lineares da forma $Ax = b$, no qual A permanece igual e apenas o vetor b varia. Portanto, com apenas uma decomposição de A em LU é possível resolver vários sistemas.

2.2. Ajuste de curva

Quando é feita uma coleta de dados, geralmente, temos uma distribuição descontínua, mas esse conjunto de dados pode apresentar certos padrões ou relações entre suas variáveis. Nesses casos, é possível encontrar uma função matemática que melhor se ajuste aos dados, ou seja, que passa ou exatamente sobre os pontos do gráfico ou próxima deles. Encontrar essa função é fazer um ajuste de curva. A partir do ajuste, podemos modelar fenômenos ou realizar previsões. A seguir, para visualizarmos e analisarmos os dados, vamos usar o método dos quadrados mínimos.

2.3. Quadrados mínimos

O método dos quadrados mínimos é uma técnica de otimização matemática utilizada com a finalidade de preencher lacunas de dados através da aproximação de uma curva que melhor descreva a dispersão dos dados conhecidos. Nesse método, são escolhidas funções de base $g_1(x), g_2(x), \dots, g_n(x)$ e busca-se encontrar os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_n$, de modo que a função $\varphi(x_i) = \alpha_1 g_1(x_i) + \alpha_2 g_2(x_i) + \dots + \alpha_n g_n(x_i)$ se aproxime ao máximo dos dados conhecidos $(x_i, f(x_i))$ (Ruggiero e Lopes, 2000).

Há preferência para o método dos quadrados mínimos, em comparação com a interpolação, nos casos em que se deseja fazer uma previsão de dados

futuros, gerando um ajuste com menor taxa de erro, uma vez que o método do quadrados mínimos minimiza o desvio (Ruggiero e Lopes, 2000). Aqui, é considerado o erro absoluto, ou desvio, $d_i = f(x_i) - \varphi(x_i)$, determinando a função $\varphi(x)$ de forma que o desvio d_i seja mínimo. Sendo assim, , como abordado por Ruggiero e Lopes (2000), ao buscar os desvios mínimos, opta-se pela abordagem de buscar os quadrados mínimos dos desvios, isto é

$$\begin{aligned} \min & \sum_{i=1}^m d_i^2 \\ \min & \sum_{i=1}^m (f(x_i) - \varphi(x_i))^2 \\ \min & \sum_{i=1}^m (f(x_i) - \alpha_1 g_1(x_i) - \alpha_2 g_2(x_i) - \dots - \alpha_n g_n(x_i))^2 \end{aligned}$$

Considerando $F(\alpha_1, \alpha_2, \dots, \alpha_n) = (f(x_i) - \alpha_1 g_1(x_i) - \dots - \alpha_n g_n(x_i))^2$, para determinar os valores de forma que $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ seja mínima, basta que $\frac{\partial F}{\partial \alpha_k} = 0$, sendo $k = 1, \dots, n$. Assim,

$$\begin{aligned} \frac{\partial F}{\partial \alpha_k} &= 2 \sum_{i=1}^m (f(x_i) - \alpha_1 g_1(x_i) - \dots - \alpha_n g_n(x_i))(-g_k(x_i)) \\ 0 &= 2 \sum_{i=1}^m (f(x_i) - \alpha_1 g_1(x_i) - \dots - \alpha_n g_n(x_i))(-g_k(x_i)) \\ 0 &= \sum_{i=1}^m f(x_i)g_k(x_i) - \alpha_1 g_1(x_i)g_k(x_i) - \dots - \alpha_n g_n(x_i)g_k(x_i) \end{aligned}$$

Abrindo o somatório para cada valor de k , obtemos um sistema linear com n incógnitas e n variáveis, cujas equações são chamadas de equações normais:

$$\left\{ \begin{array}{l} \sum_{i=1}^m f(x_i)g_1(x_i) - \alpha_1 g_1(x_i)g_1(x_i) - \alpha_2 g_2(x_i)g_1(x_i) - \dots - \alpha_n g_n(x_i)g_1(x_i) = 0 \\ \sum_{i=1}^m f(x_i)g_2(x_i) - \alpha_1 g_1(x_i)g_2(x_i) - \alpha_2 g_2(x_i)g_2(x_i) - \dots - \alpha_n g_n(x_i)g_2(x_i) = 0 \\ \vdots \\ \sum_{i=1}^m f(x_i)g_n(x_i) - \alpha_1 g_1(x_i)g_n(x_i) - \alpha_2 g_2(x_i)g_n(x_i) - \dots - \alpha_n g_n(x_i)g_n(x_i) = 0 \end{array} \right. ,$$

sendo equivalente a

$$\begin{cases} \sum_{i=1}^m \alpha_1 g_1(x_i) g_1(x_i) + \alpha_2 g_2(x_i) g_1(x_i) + \dots + \alpha_n g_n(x_i) g_1(x_i) = \sum_{i=1}^m f(x_i) g_1(x_i) \\ \sum_{i=1}^m \alpha_1 g_1(x_i) g_2(x_i) + \alpha_2 g_2(x_i) g_2(x_i) + \dots + \alpha_n g_n(x_i) g_2(x_i) = \sum_{i=1}^m f(x_i) g_2(x_i) \\ \vdots \\ \sum_{i=1}^m \alpha_1 g_1(x_i) g_n(x_i) + \alpha_2 g_2(x_i) g_n(x_i) + \dots + \alpha_n g_n(x_i) g_n(x_i) = \sum_{i=1}^m f(x_i) g_n(x_i) \end{cases}$$

Portanto, representando tal sistema matricialmente, o método de quadrados mínimos pode ser escrito pelo sistema $A\alpha = b$ - sendo a matriz A uma matriz quadrada $n \times n$ cujos elementos a_{ij} são obtidos pelo cálculo do produto interno das funções de base $\langle g_i, g_j \rangle$, a matriz α uma matriz coluna $n \times 1$ cujos elementos α_i são os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_n$ que se deseja encontrar para minimizar a função, e a matriz b uma matriz coluna $n \times 1$ cujos elementos b_i são obtidos pelo produto interno $\langle g_i, f \rangle$ entre a função de base e os dados conhecidos, isto é

$$\begin{bmatrix} \langle g_1, g_1 \rangle & \langle g_1, g_2 \rangle & \dots & \langle g_1, g_n \rangle \\ \langle g_2, g_1 \rangle & \langle g_2, g_2 \rangle & \dots & \langle g_2, g_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_n, g_1 \rangle & \langle g_n, g_2 \rangle & \dots & \langle g_n, g_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \langle g_1, f \rangle \\ \langle g_2, f \rangle \\ \vdots \\ \langle g_n, f \rangle \end{bmatrix}.$$

A seguir, vamos definir alguns conceitos da teoria de conjuntos fuzzy.

2.4. Teoria de conjuntos fuzzy e álgebra fuzzy

Um conjunto fuzzy A de um universo X é caracterizado por uma função de pertinência $\mu_A : X \rightarrow [0, 1]$, na qual $\mu_A(x)$ - que aqui será denotada como $A(x)$ - representa o grau de pertinência de x em A para todo x em X (Barros e Bassanezi, 2006). Os subconjuntos fuzzy de X são denotados por $\mathcal{F}(X)$. Os α -níveis de um conjunto fuzzy $A \subseteq X$, dados por $[A]^\alpha$, são definidos da seguinte forma (Klir e Yuan, 1995):

$$[A]^\alpha = \{x \in X : A(x) \geq \alpha\}, \quad \forall \alpha \in (0, 1].$$

Já para o 0-nível é necessário ter em mãos a definição de fecho, e consequentemente, espaço topológico. Caso o conjunto X seja um espaço topológico (Bede, 2013), o 0-nível é definido por

$$[A]^0 = cl(\{x \in X : A(x) > 0\}),$$

sendo que $cl(Y)$ representa o fecho do conjunto $Y \subseteq X$.

A classe de números fuzzy $\mathbb{R}_{\mathcal{F}}$, subclasse de $\mathcal{F}(\mathbb{R})$, inclui os conjuntos de números reais assim como os conjunto dos intervalos fechados limitados de \mathbb{R} . Os elementos dessa classe são caracterizados pelos subconjuntos fuzzy de $X = \mathbb{R}$, cujos α -níveis são intervalos encaixantes limitados, fechados e não vazios para todo $\alpha \in [0, 1]$ (Esmi et al., 2021). A partir disso, podemos reescrever os α -níveis de um número fuzzy A como $[A]^\alpha = [a_\alpha^-, a_\alpha^+]$.

Se a_α^- e a_α^+ forem funções contínuas com respeito a α – como os números fuzzy triangulares –, essa classe de números fuzzy passa a ser denotada por $\mathbb{R}_{\mathcal{F}_c}$. Dado que todo número fuzzy triangular pode ser definido por uma tripla do tipo $(a; b; c)$, para qualquer $a \leq b \leq c$, a partir dos α -níveis temos que $[A]^\alpha = [a + \alpha(b - a), c - \alpha(c - b)]$, $\forall \alpha \in [0, 1]$.

Se $A(x - y) = A(x + y)$, $\forall y \in \mathbb{R}$, o número fuzzy A é dito simétrico com respeito a $x \in \mathbb{R}$. Se não há qualquer x que satisfaça essa condição, então A é não-simétrico (Esmi et al., 2021). A relação fuzzy R entre dois universos X e Y é dada pela função $R : X \times Y \rightarrow [0, 1]$, na qual $R(x, y) \in [0, 1]$ representa o grau de relação entre $x \in X$ e $y \in Y$.

Uma relação fuzzy $J \in \mathcal{F}(\mathbb{R}^n)$ é classificada como distribuição de possibilidade conjunta (DPC) entre os números fuzzy pertencentes à $\mathbb{R}_{\mathcal{F}}$ se

$$A_i(y) = \bigvee_{x: x_i=y} J(x), \quad \forall y \in \mathbb{R}, \quad (2.1)$$

para todo $i = 1, \dots, n$, em que o símbolo \bigvee representa o supremo.

Geometricamente, a equação (2.1) diz que cada número fuzzy A_i , também chamado de marginal, pode ser obtido pela projeção da relação J na i -ésima direção.

Um exemplo de DPC é a baseada em t -normas. Lembrando que uma t -norma é um operador $t : [0, 1]^2 \rightarrow [0, 1]$ que admite as propriedades de associatividade, comutatividade, monotonicidade e tem o valor 1 como elemento neutro (Barros e Bassanezi, 2006). A partir dessa definição é possível estabelecer a seguinte DPC

$$J_t(x_1, \dots, x_n) = A_1(x_1)t \dots t A_n(x_n).$$

Por exemplo, duas DPCs que podem ser obtidas por meio desse tipo de distribuição são

$$J_\wedge(x_1, \dots, x_n) = \bigwedge_{i=1}^n A_i(x_i)$$

e

$$J(x_1, \dots, x_n) = \prod_{i=1}^n A_i(x_i)$$

sendo que os símbolos \wedge e \prod representam o mínimo e produtório, respectivamente, as quais são baseadas nas t -normas do mínimo $t_\wedge(x, y) = \min\{x, y\}$ e do produto $t_\cdot(x, y) = x \cdot y$, respectivamente.

A partir da noção de interatividade, apresenta-se a seguinte definição para números fuzzy interativos.

Definição 1 Os números fuzzy A_1, \dots, A_n são denominados interativos se J satisfaz a equação (2.1) e $J \neq J_\wedge$.

Por exemplo, os números fuzzy A_1, \dots, A_n são interativos se a DPC considerada entre eles for a J . Por outro lado, os mesmos números fuzzy são não interativos se a DPC considerada entre eles for a J_\wedge . Isso ilustra que a interatividade entre os números fuzzy A_1, \dots, A_n é definida a partir de uma distribuição de possibilidade conjunta, se assemelhando a uma relação de dependência entre variáveis aleatórias.

Para lidar com tal relação, de um ponto de vista de operação aritmética, é preciso estender o princípio de extensão de Zadeh, uma vez que esta ferramenta é adaptada para lidar com aritmética voltada para números fuzzy não interativos. O princípio de extensão sup- J , estende as funções clássicas para funções fuzzy levando em consideração a interatividade entre os números fuzzy.

Definição 2 Seja $J \in \mathcal{F}(\mathbb{R}^n)$ uma distribuição de possibilidade conjunta de $(A_1, \dots, A_n) \in \mathbb{R}_{\mathcal{F}}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$. A extensão sup- J de f em $A_1, \dots, A_n \in \mathbb{R}_{\mathcal{F}}^n$, denotado por $f_J(A_1, \dots, A_n)$ é o conjunto fuzzy definido por

$$f_J(A_1, \dots, A_n)(y) = \bigvee_{(x_1, \dots, x_n) \in f^{-1}(y)} J(x_1, \dots, x_n),$$

sendo $f^{-1}(y) = \{(x_1, \dots, x_n) \in \mathbb{R}^n : f(x_1, \dots, x_n) = y\}$.

A partir do princípio de extensão de sup- J , é possível também estabelecer uma aritmética para os números fuzzy interativos, como, por exemplo, a soma interativa. A soma interativa é dada por

$$(A_1 +_J A_2)(y) = \bigvee_{x_1 + x_2 = y} J(x_1, x_2)$$

sendo J uma JPD arbitrária de A_1 e A_2 .

Wasques (2019) propôs uma DPC, denotada por J_0 também chamada de distribuição de possibilidade conjunta otimista, que é muito útil para o controle de incerteza ao longo de dinâmicas temporais. A partir dessa conjunta é definida uma nova aritmética, que no caso da soma é denotada por $A_1 +_0 A_2$.

O teorema abaixo caracteriza a soma interativa $+_0$ entre dois números fuzzy triangulares.

Teorema 1 (Pinto et al. (2022)) *Sejam os números fuzzy triangulares $A = (a; b; c)$ e $B = (d; e; f)$ e J_0 a distribuição de possibilidade conjunta entre A e B . Então a soma interativa $A +_0 B$ é dada por*

$$A +_0 B = \begin{cases} ((a + f) \wedge (b + e); b + e; (b + e) \vee (c + d)), & \text{se } \text{diam}(A) \geq \text{diam}(B) \\ ((c + d) \wedge (b + e); b + e; (b + e) \vee (a + f)), & \text{se } \text{diam}(A) \leq \text{diam}(B) \end{cases},$$

em que \wedge e \vee representam o mínimo e o máximo, respectivamente.

A soma interativa zero ($+_0$) será a utilizada no presente trabalho para a resolução de sistemas lineares fuzzy do tipo

$$\begin{cases} \lambda_1 X_1 +_0 \lambda_2 X_2 = r_1 \\ \lambda_3 X_1 +_0 \lambda_4 X_2 = r_2 \end{cases},$$

nos quais, temos os coeficientes $\lambda_1, \lambda_2, \lambda_3, \lambda_4, r_1, r_2 \in \mathbb{R}$ com variáveis fuzzy $X_1 = (a; b; c)$ e $X_2 = (d; e; f)$ a serem determinados.

É importante ressaltar que equações do tipo $\lambda_1 X_1 + \lambda_2 X_2 = r$, em que $\lambda_1, \lambda_2, r \in \mathbb{R}$, não admite solução para o caso em que a soma $+$ é a usual. De fato, se X_1 e X_2 são números fuzzy, então $\text{diam}(X_1) > 0$ e $\text{diam}(X_2) > 0$. Assim, $\text{diam}(\lambda_1 X_1) > 0$ e $\text{diam}(\lambda_2 X_2) > 0$, e portanto, $\text{diam}(\lambda_1 X_1 + \lambda_2 X_2) \geq \text{diam}(\lambda_1 X_1) > 0 = \text{diam}(r)$. Logo, $\text{diam}(\lambda_1 X_1 + \lambda_2 X_2) \neq \text{diam}(r)$, e por consequência $\lambda_1 X_1 + \lambda_2 X_2 \neq r$ para qualquer $X_1, X_2 \in \mathbb{R}_{\mathcal{F}}$.

Para ilustrar que a escolha da soma interfere na obtenção de solução para o problema, é dado o exemplo abaixo.

Exemplo 1 *Sejam os números fuzzy $A_1 = (1; 2; 3)$ e $A_2 = (2; 3; 4)$. Tem-se que*

$$A_1 +_0 A_2 = 5,$$

sendo o resultado o número crisp 5, o qual possui função de pertinência descrita pela função característica $\chi_{\{5\}}$.

Embora a soma $+_0$ entre números fuzzy possa resultar em um valor real, como mostrado acima, o exemplo abaixo mostra que nem sempre isso ocorre.

Exemplo 2 *Sejam os números fuzzy $A_1 = (1; 2; 3)$ e $A_2 = (4; 6; 8)$. Tem-se que*

$$A_1 +_0 A_2 = (7; 8; 9)$$

Uma outra boa propriedade que a soma $+_0$ admite é a propriedade do elemento oposto, isto é, para cada número fuzzy A existe um único número fuzzy $-A$ de tal modo que $A +_0 (-A) = 0$, como ilustrado no exemplo abaixo.

Exemplo 3 *Tendo os números fuzzy $A_1 = A_2 = (4; 5; 6)$*

$$A_1 +_0 (-A_2) = A_1 -_0 A_2 = (4; 5; 6) +_0 (-6; -5; -4) = 0$$

Vale ressaltar que se no Exemplo 3 tivesse sido utilizada a operação de soma usual, o resultado seria $(-2; 0; 2)$ e não 0 (zero).

2.5. Construção dos dados fuzzy

Neste trabalho, analisamos o número de infectados por HIV no Brasil nos últimos anos, baseados nos dados disponibilizados no Boletim Epidemiológico da Aids/HIV pelo Ministério da Saúde (2021). Entretanto, tais dados estão contidos no conjunto dos reais, de modo que sua rigidez numérica não considere imprecisões intrínsecas à sua coleta. Dessa forma, foi feita uma construção de dados triangulares fuzzy partindo de tais dados, criando uma faixa de incerteza associada à uma função de pertinência.

Para isso, primeiramente, foi feita uma regressão dos dados reais y obtidos – aqui, escolhemos utilizar um polinômio de segundo grau; em seguida, calculamos a distância δ_i entre cada dado real y_i e a curva encontrada; por fim, calculamos o dobro de cada distância $2\delta_i$ e usamos como amplitude de cada valor, de modo que a distância entre os extremos de cada dado fuzzy e o ajuste realizado seja a mesma e igual a δ_i . Assim, a pertinência em y_i é 1, enquanto em $y_i \pm \delta_i$ a pertinência é 0. Tal construção está representada graficamente na Figura 1. Então, os dados utilizados na modelagem subsequente encontram-se na Tabela 1.

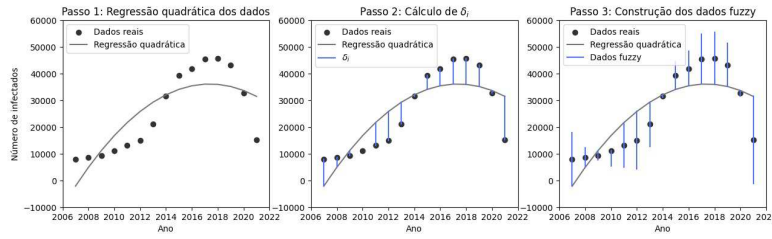


Figura 1: Método utilizado para construção de dados fuzzy.

Tabela 1: Construção de dados triangulares fuzzy para o número de infectados por HIV no Brasil de 2015 a 2021.

Ano	Dado Real	Dado Fuzzy
2007	7872	(-2109, 7872, 17853)
2008	8617	(4902, 8617, 12332)
2009	9430	(7657, 9430, 11203)
2010	11099	(5403, 11099, 16795)
2011	13289	(4901, 13289, 21677)
2012	15096	(4342, 15096, 25850)
2013	21071	(12828, 21071, 29314)
2014	31620	(31172, 31620, 32068)
2015	39375	(34112, 39375, 44638)
2016	41927	(35448, 4927, 48406)
2017	45421	(36073, 45421, 54769)
2018	45685	(35990, 45685, 55380)
2019	43292	(35196, 43292, 51388)
2020	32687	(31680, 32687, 33694)
2021	15212	(-1058, 15212, 31482)

2.6. Sistemas lineares fuzzy

Os sistemas lineares fuzzy utilizados neste trabalho têm como entrada números fuzzy triangulares que são multiplicados por coeficientes reais e retornam números fuzzy triangulares. As abordagens e notações usadas nessa e nas próximas seções são as mesmas introduzidas por Pinto et al. (2022).

Assim, busca-se resolver o sistema linear

$$\begin{cases} \alpha_{11}(a; b; c) +_0 \alpha_{12}(d; e; f) = (r; s; t) \\ \alpha_{21}(a; b; c) +_0 \alpha_{22}(d; e; f) = (u; v; w) \end{cases}$$

também representado por $Ux = V$, sendo

$$U = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} (a; b; c) \\ (d; e; f) \end{bmatrix} \text{ e } V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} (r; s; t) \\ (u; v; w) \end{bmatrix}$$

de modo que os valores de $X = [X_1 \ X_2]^t$ satisfaçam $a \leq b \leq c$ e $d \leq e \leq f$.

Assim, 4 passos precisam ser seguidos para que as soluções do sistema sejam encontradas. Nesse processo, uma única matriz de entrada A é transformada em outras 4 matrizes e, em decorrência disso, um único sistema pode apresentar uma, múltiplas ou nenhuma solução coerente (Pinto et al., 2022).

PASSO 1: Para que sejam definidos os picos do ajuste, resolve-se o sistema:

$$U \begin{bmatrix} b \\ e \end{bmatrix} = \begin{bmatrix} s \\ v \end{bmatrix}$$

PASSO 2: usando como critério os sinais dos elementos da matriz U , divididos 2 a 2, resolve-se cada um dos 8 sistemas da tabela abaixo para definir as bases das soluções;

Sinais de α_{ij}	Sistemas a serem resolvidos				
$\alpha_{ij} \geq 0; i, j = 1, 2$ ou	U	$\begin{bmatrix} a \\ f \end{bmatrix} = \begin{bmatrix} r \\ u \end{bmatrix}$	e	U	$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} t \\ w \end{bmatrix}$; e
$\alpha_{ij} \leq 0; i, j = 1, 2$ ou	U	$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} t \\ u \end{bmatrix}$	e	U	$\begin{bmatrix} a \\ f \end{bmatrix} = \begin{bmatrix} r \\ w \end{bmatrix}$; e
$\alpha_{11}, \alpha_{12} \geq 0$ e $\alpha_{21}, \alpha_{22} \leq 0$ ou	U	$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} r \\ w \end{bmatrix}$	e	U	$\begin{bmatrix} a \\ f \end{bmatrix} = \begin{bmatrix} t \\ u \end{bmatrix}$; e
$\alpha_{21}, \alpha_{22} \geq 0$ e $\alpha_{11}, \alpha_{12} \leq 0$	U	$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} r \\ u \end{bmatrix}$	e	U	$\begin{bmatrix} a \\ f \end{bmatrix} = \begin{bmatrix} t \\ w \end{bmatrix}$
$\alpha_{11}, \alpha_{21} \geq 0$ e $\alpha_{12}, \alpha_{22} \leq 0$ ou	U	$\begin{bmatrix} a \\ d \end{bmatrix} = \begin{bmatrix} r \\ u \end{bmatrix}$	e	U	$\begin{bmatrix} c \\ f \end{bmatrix} = \begin{bmatrix} t \\ w \end{bmatrix}$; e
$\alpha_{11}, \alpha_{22} \geq 0$ e $\alpha_{12}, \alpha_{21} \leq 0$ ou	U	$\begin{bmatrix} a \\ d \end{bmatrix} = \begin{bmatrix} r \\ w \end{bmatrix}$	e	U	$\begin{bmatrix} c \\ f \end{bmatrix} = \begin{bmatrix} t \\ u \end{bmatrix}$; e
$\alpha_{12}, \alpha_{21} \geq 0$ e $\alpha_{11}, \alpha_{22} \leq 0$ ou	U	$\begin{bmatrix} c \\ f \end{bmatrix} = \begin{bmatrix} r \\ w \end{bmatrix}$	e	U	$\begin{bmatrix} a \\ d \end{bmatrix} = \begin{bmatrix} t \\ u \end{bmatrix}$; e
$\alpha_{12}, \alpha_{22} \geq 0$ e $\alpha_{11}, \alpha_{21} \leq 0$	U	$\begin{bmatrix} c \\ f \end{bmatrix} = \begin{bmatrix} r \\ u \end{bmatrix}$	e	U	$\begin{bmatrix} a \\ d \end{bmatrix} = \begin{bmatrix} t \\ w \end{bmatrix}$

PASSO 3: usando como critério os sinais dos elementos da matriz U , sendo apenas 1 diferente dos demais, resolve-se cada um dos 4 sistemas do tipo $Z_k x = z$ da tabela abaixo para definir os extremos das soluções, sendo $x = [a \ c \ d \ f]^t$, $z = [r \ u \ t \ w]^t$ e Z_k cada uma das 4 matrizes;

Sinal de α_{ij}	Matrizes dos sistemas a serem resolvidos
$\alpha_{11}, \alpha_{12}, \alpha_{21} \geq 0$ e $\alpha_{22} \leq 0$ ou	$Z_1 = \begin{bmatrix} \alpha_{11} & 0 & 0 & \alpha_{12} \\ \alpha_{21} & 0 & \alpha_{22} & 0 \\ 0 & \alpha_{11} & \alpha_{12} & 0 \\ 0 & \alpha_{21} & 0 & \alpha_{22} \end{bmatrix}$,
$\alpha_{11}, \alpha_{12}, \alpha_{22} \geq 0$ e $\alpha_{21} \leq 0$ ou	$Z_2 = \begin{bmatrix} \alpha_{11} & 0 & 0 & \alpha_{12} \\ 0 & \alpha_{21} & 0 & \alpha_{22} \\ 0 & \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & 0 & \alpha_{22} & 0 \end{bmatrix}$,
$\alpha_{21} \geq 0$ e $\alpha_{12}, \alpha_{21}, \alpha_{22} \leq 0$ ou	$Z_3 = \begin{bmatrix} 0 & \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & 0 & \alpha_{22} & 0 \\ \alpha_{11} & 0 & 0 & \alpha_{12} \\ 0 & \alpha_{21} & 0 & \alpha_{22} \end{bmatrix}$,
$\alpha_{22} \geq 0$ e $\alpha_{11}, \alpha_{21}, \alpha_{22} \leq 0$	$Z_4 = \begin{bmatrix} 0 & \alpha_{11} & \alpha_{12} & 0 \\ 0 & \alpha_{21} & 0 & \alpha_{22} \\ \alpha_{11} & 0 & 0 & \alpha_{12} \\ \alpha_{21} & 0 & \alpha_{22} & 0 \end{bmatrix}$.
$\alpha_{11}, \alpha_{21}, \alpha_{22} \geq 0$ e $\alpha_{12} \leq 0$ ou	$Z_5 = \begin{bmatrix} 0 & \alpha_{11} & 0 & \alpha_{12} \\ \alpha_{21} & 0 & 0 & \alpha_{22} \\ \alpha_{11} & 0 & \alpha_{12} & 0 \\ 0 & \alpha_{21} & \alpha_{22} & 0 \end{bmatrix}$,
$\alpha_{12}, \alpha_{21}, \alpha_{22} \geq 0$ e $\alpha_{11} \leq 0$ ou	$Z_6 = \begin{bmatrix} 0 & \alpha_{11} & 0 & \alpha_{12} \\ 0 & \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{11} & 0 & \alpha_{12} & 0 \\ \alpha_{21} & 0 & 0 & \alpha_{22} \end{bmatrix}$,
$\alpha_{11} \geq 0$ e $\alpha_{12}, \alpha_{21}, \alpha_{22} \leq 0$ ou	$Z_7 = \begin{bmatrix} \alpha_{11} & 0 & \alpha_{12} & 0 \\ \alpha_{21} & 0 & 0 & \alpha_{22} \\ 0 & \alpha_{11} & 0 & \alpha_{12} \\ 0 & \alpha_{21} & \alpha_{22} & 0 \end{bmatrix}$,
$\alpha_{12} \geq 0$ e $\alpha_{11}, \alpha_{21}, \alpha_{22} \leq 0$	$Z_8 = \begin{bmatrix} \alpha_{11} & 0 & \alpha_{12} & 0 \\ 0 & \alpha_{21} & \alpha_{22} & 0 \\ 0 & \alpha_{11} & 0 & \alpha_{12} \\ \alpha_{21} & 0 & 0 & \alpha_{22} \end{bmatrix}$.

PASSO 4: verificar se os valores das triplas $(a; b; c)$ e $(d; e; f)$ cumprem a condição $a \leq b \leq c$ e $d \leq e \leq f$. Caso cumpram, são soluções do sistema linear, senão são descartadas.

2.7. Quadrados mínimos fuzzy

A partir da teoria de conjuntos fuzzy, é possível construir uma flexibilidade numérica no ajuste de curva, melhorando a interpretabilidade da análise. Como contemplado no artigo Pinto et al. (2022), essa seção adapta o método de quadrados mínimos para aceitar dados fuzzy respeitando a álgebra própria desse tipo de conjunto. Desse modo, as saídas deixam de ser números reais e passam a ser números triangulares fuzzy (Y_1, \dots, Y_m) e, portanto, os coeficientes C_1 e C_2 gerados também serão fuzzy.

$$WC = Y, \quad (2.2)$$

isto é,

$$\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}. \quad (2.3)$$

O sistema linear (2.2) pode ser resolvido pela multiplicação da transposta de W em ambos os lados à esquerda, ou seja,

$$WW^t C = YW^t, \quad (2.4)$$

isto é,

$$\begin{pmatrix} x_1^2 + \dots + x_m^2 & x_1 + \dots + x_m \\ x_1 + \dots + x_m & m \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} x_1 Y_1 +_0 \dots +_0 x_m Y_m \\ Y_1 +_0 \dots +_0 Y_m \end{pmatrix}. \quad (2.5)$$

Mesmo que $W^t W$ não admita inversa, pode-se resolver o sistema por meio de sua pseudoinversa, encontrando-se os coeficientes da função (Pinto, 2023). Neste trabalho, realizamos um ajuste linear dada pela seguinte função

$$\phi(x) = C_1 x +_0 C_2, \quad (2.6)$$

sendo C_1 e C_2 números triangulares pertencentes à $\mathbb{R}_{\mathcal{F}}$.

Caso mais de uma solução seja encontrada, deve-se realizar análises complementares de tal modo que a solução seja compatível com o que se espera para o problema. Por exemplo, a escolha pode ser feita por meio de uma análise gráfica.

Ademais, partindo dos fundamentos teóricos apresentados acima, foram desenvolvidos algoritmos computacionais por meio da linguagem de programação Python, com apoio das bibliotecas *Numpy* e *Matplotlib*, visando

transformar dados reais em dados fuzzy triangulares e automatizar o ajuste de curva linear para quaisquer conjuntos de dados. Os códigos das funções desenvolvidas estão disponíveis no repositório GitHub das autoras.

3. Discussão dos resultados

Tabela 2: Infectados por HIV no Brasil de 2015 a 2021 no formato de números triangulares fuzzy.

Ano	Solução 1	Solução 2
2015	(43283.7, 47631.79, 51979.79)	(31004.89, 47631.79, 64258.7)
2016	(41919.7, 44306.89, 46693.89)	(32710.69, 44306.89, 55903.1)
2017	(40555.7, 40982.0, 41408.0)	(34416.5, 40982.0, 47547.5)
2018	(36122.1, 37657.1, 39191.7)	(36122.29, 37657.1, 39191.9)
2019	(30836.19, 34332.2, 37827.7)	(30836.3, 34332.2, 37828.1)
2020	(25550.3, 31007.29, 36463.7)	(22480.69, 31007.29, 39533.89)
2021	(20264.4, 27682.39, 35099.7)	(14125.09, 27682.39, 41239.7)

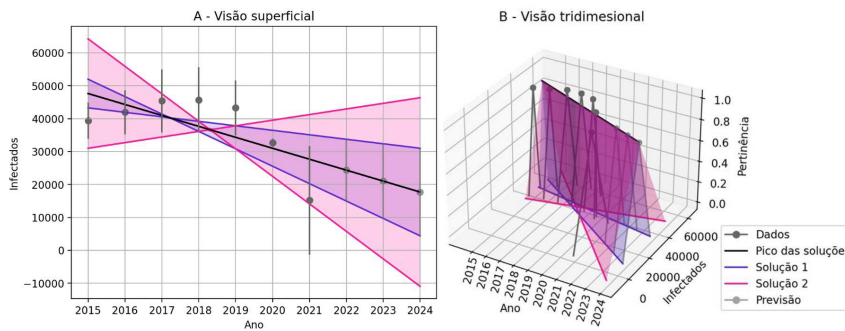


Figura 2: Ajuste de curva fuzzy para os dados de infectados por HIV do Brasil de 2015 a 2021 com a previsão até 2024. **Gráfico A:** visão superior em duas dimensões. **Gráfico B:** visão tridimensional, com o grau de pertinência expresso no eixo z (altura).

Para melhor explorar o método dos quadrados mínimo fuzzy, foi utilizada uma base de dados de infectados por HIV no Brasil de 2015 a 2021 do Ministério da Saúde (2021). Então, após a transformação em dados triangulares fuzzy, foi aplicado o ajuste de curva construído nesse projeto, obtendo duas soluções, como consta na Figura 2. Vale ressaltar que a reta que representa o α -nível com grau de pertinência $\alpha = 1$, é equivalente a um ajuste linear clássico dos dados reais.

Em consonância com a proposta da soma iterativa $+_0$, nota-se na Figura 2 que a solução 1 apresenta o melhor ajuste, uma vez que possui o menor diâmetro, tendo uma faixa de incerteza menor. No gráfico B, a visão tridimensional facilita a compreensão do ajuste, pois, a medida que os α -níveis se aproximam de $\alpha = 1$, mais aquele conjunto de dados se aproxima do “em torno de”, possibilitando a interpretabilidade na escolha da região que melhor se encaixa com o objetivo da análise dos dados. Portanto, o comportamento da solução 1 consegue descrever de uma forma mais coerente a incerteza dos dados mesmo após as manipulações matemáticas.

Tabela 3: Previsão do número de infectados por HIV no Brasil de 2022 a 2024 usando números fuzzy triangulares.

Ano	Solução 1	Solução 2
2022	(14978.5, 24357.49, 33735.7)	(5769.5, 24357.49, 42945.5)
2023	(9692.59, 21032.59, 32371.69)	(-2586.10, 21032.59, 44651.29)
2024	(4406.69, 17707.69, 31007.69)	(-10941.69, 17707.69, 46357.1)

Além disso, foi realizada a previsão do número de infectados por HIV no Brasil até 2024 a partir dos ajustes encontrados, como consta na Tabela 3. Uma vez que já é esperado que a incerteza aumente conforme se tenta extrapolar datas mais futuras, ainda que o diâmetro da solução 1 aumente progressivamente, esta ainda possui uma incerteza mais controlada, o que reafirma o seu ajuste com o mais coerente na predição dos 3 anos seguintes. Cabe destacar também a existência de valores negativos na solução 2, levantando uma interpretação incompatível com dados biológicos. No entanto, a própria construção dos dados fuzzy inseriu diâmetros que incluem valores negativos, o que tornou essa característica ainda mais expressiva.

4. Conclusões

De maneira geral, o trabalho conseguiu entregar de forma eficaz um ajuste de curva fuzzy linear, desenvolvido em código Python e aplicado em dados biológicos. O resultado foi capaz de realizar uma modelagem matemática que preserva a subjetividade tão importante no tratamento de dados de biológicos, na qual a incerteza na contagem dos infectados é compreendida. Então, essa ferramenta se mostrou importante e inovadora no processamento de dados incertos, preservando a interpretabilidade linguística mesmo após várias operações matemáticas.

Contudo, ainda há nesse projeto vértices que podem ser melhorados. Diante dos resultados expostos na discussão, as perspectivas futuras se baseiam no desenvolvimento de mecanismos que tornem menos custosa a realização de um ajuste quadrático fuzzy, haja vista que os dados utilizados apresentam um comportamento não-linear. Além disso, visando controlar melhor a amplitude da incerteza dos dados, fica como ação futura explorar outras formas de transformar dados reais em dados fuzzy.

Referências

- Barros, L. C. e Bassanezi, R. C. (2006). Tópicos de lógica fuzzy e biomatemática. In *Coleção Textos Didáticos*, volume 5. IMECC–Unicamp.
- Bede, B. (2013). *Mathematics of Fuzzy Sets and Fuzzy Logic*. 295. Springer, Berlin, 1st edition.
- Esmi, E., Sánchez, D. E., Wasques, V. F., e Barros, L. C. (2021). Solutions of higher order linear fuzzy differential equations with interactive fuzzy values. *Fuzzy Sets and Systems*, 419:122–140.
- Klir, G. J. e Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Ministério da Saúde (2021). Boletim epidemiológico hiv/aids 2021. Disponível em: <https://antigo.aids.gov.br/pt-br/pub/2021/boletim-epidemiologico-hivaids-2021>, acesso em: 30/10/2024.
- Pinto, N. J. B. (2023). *Curve fitting based on interactive fuzzy data*. Tese de Doutorado, IMECC–Unicamp, Campinas/SP.

- Pinto, N. J. B., Wasques, V. F., Esmi, E., e Barros, L. C. (2022). Sistemas lineares fuzzy e problemas de mínimos quadrados aplicados a dados longitudinais. *Biomatemática*, 32:43–60.
- Ruggiero, M. A. G. e Lopes, V. L. R. (2000). *Cálculo Numérico: Aspectos Teóricos e Computacionais*. Pearson, S.Paulo/SP.
- Sacilotto, C. (2017). Uso da teoria de conjuntos fuzzy para análise prognóstica do câncer de rim. Dissertação de Mestrado, IMECC–Unicamp, Campinas/SP.
- Santos, D. D. C. (2018). Um estudo sobre identificação de anomalias em bases de regras fuzzy aplicado a estimação do risco de endometriose. Dissertação de Mestrado, IMECC–Unicamp, Campinas/SP.
- Simões, F. S. P. (2017). *Sobre equações diferenciais para processos fuzzy linearmente correlacionados: aplicações em dinâmica de população*. Tese de Doutorado, IMECC–Unicamp, Campinas/SP.
- Wasques, V. F. (2019). *Fuzzy Differential Equations via Interactive Arithmetic: Applications in Biomathematics*. Tese de Doutorado, IMECC–Unicamp, Campinas/SP.

Estudo da epidemiologia matemática da tuberculose no Brasil

Vitor E. G. Barelli¹, Tiago M. Bigardi², Débora V. P. Chaves³, Felipe S. Minatogau⁴, Ana C. B. Loponi⁵
Ilum Escola de Ciência, CNPEM, 13.087-548, Campinas/SP

Resumo. A tuberculose é uma das doenças mais prevalentes e letais do mundo, persistindo há milhares de anos e exigindo novas estratégias e aprimoramentos para um combate mais efetivo. Ela é multifatorial e está constantemente associada a fatores socioeconômicos e a fatores de eficiência do sistema de saúde, que, no Brasil, moldam o cenário de formas diferentes para cada região do país, dadas as suas diferenças sociais e econômicas. Considerando isso, a abordagem adequada para essa doença é a união da epidemiologia matemática com a epidemiologia social. Logo, no presente trabalho, é investigada a dinâmica da tuberculose ao longo da última década, empregando o modelo matemático epidemiológico Suscetível-Infetado-Recuperado-Suscetível (SIRS) adaptado para a dinâmica da tuberculose e analisado considerando os fatores socioeconômicos e efetividade do sistema público de saúde. Esse estudo destaca os impactos desses fatores nas diferentes regiões brasileiras, mostrando como as características locais influenciam de forma única a permanência da tuberculose.

Palavras-chave: Tuberculose; epidemiologia matemática e social; modelo SIRS adaptado; macrorregiões brasileiras.

1. Introdução

A tuberculose é uma doença infectocontagiosa causada pela bactéria *Mycobacterium tuberculosis*. Sua transmissão ocorre pelas vias aéreas, afetando

¹vitorbarelli@gmail.com

²esmarques.esmarques@gmail.com

³deboravpchaves@gmail.com

⁴Felipeminatogau@gmail.com

⁵ana.loponi@gmail.com

diretamente os pulmões. No entanto, em cerca de 10 a 15% dos casos, pode também atingir outros órgãos, como os rins e as meninges. Os sintomas incluem tosse persistente com ou sem secreção, febre baixa noturna, sudorese noturna, cansaço excessivo, falta de apetite, palidez e rouquidão. Se não forem tratados, tais sintomas podem levar ao óbito (Ministério da Saúde, 2024b; Fiocruz, 2024b; Sociedade Brasileira de Pneumologia e Fisiologia, 2024).

Essa doença é reconhecida pela Organização Pan-Americana de Saúde (OPAS) como uma das doenças infecciosas mais letais do mundo. Estudos realizados em 2022 pela OPAS e pela Organização Mundial da Saúde (OMS) indicam uma estimativa de 133 novas infecções a cada 100 mil habitantes, com uma taxa de mortalidade de 14 mortes a cada 100 mil habitantes, resultando em pelo menos 11 mil mortes em decorrência da tuberculose. No Brasil, a realidade não é muito diferente: o país está na lista dos 30 países com maiores estimativas de transmissão de tuberculose, com incidência estimada entre 42 e 56 casos a cada 100 mil habitantes (OPAS, 2023; WHO, 2023).

Sobre ela, é válido ressaltar que é uma doença multifatorial, ou seja, sua incidência também sofre influência direta de outros fatores. Dentre eles, há as características socioeconômicas que influenciam a incidência da tuberculose, por exemplo, o acesso ao sistema público de saúde, a vacinação, o Índice de Desenvolvimento Humano (IDH) regional, a renda per capita familiar, a densidade populacional e a qualidade de vida. Em outras palavras, os fatores socioeconômicos são os principais agravantes do índice de transmissão da tuberculose, pois esta está intimamente ligada às condições de vida e de miséria, como desnutrição, superpopulação, moradia irregular e acesso inadequado ao sistema de saúde. Por fim, é possível estabelecer uma correlação entre as regiões brasileiras com os maiores números de casos de tuberculose e suas características socioeconômicas (Borges e Teixeira, 2020; Fiocruz, 2024a; Aguiar, 2018; Ministério da Saúde, 2022; Silva, 2011; Gonçalves, 2000; Uchimura et al., 2017).

Considerando o caráter alarmante da incidência da doença e sua complexidade, estudos, pesquisas e programas de saúde foram desenvolvidos em todo o mundo visando entender sua dinâmica e controlar sua disseminação para proteger a população (WHO, 2010; Urell, 2022; WHO, 2023; Ministério da Saúde, 2024a).

Estudos do gênero também foram realizados no Brasil, em que se destaca a dissertação de Aguiar (2018) que aborda, com enfoque principal, as caracte-

terísticas socioeconômicas da tuberculose no estado de Pernambuco e como estas afetam a incidência de casos.

Já no âmbito da modelagem matemática em relação à tuberculose, há estudos como o desenvolvido por Aparicio e Castillo-Chavez (2009), que evidencia a modelagem da tuberculose a partir de diversos modelos matemáticos e relações populacionais com a dinâmica da doença. Além deste, outro estudo, realizado por Mabud et al. (2019) analisou possíveis estratégias para a prevenção da transmissão da tuberculose e seu espalhamento entre comunidades nas prisões do estado do Mato Grosso do Sul, evidenciando métodos matemáticos como possível estratégia para o combate à transmissão da doença.

Contudo, tais estudos não apresentam uma união entre as diversas áreas desse problema complexo, a tuberculose, como a epidemiologia matemática, epidemiologia social e outras áreas. Visando se diferenciar de estudos já propostos na literatura, o presente projeto propõe uma metodologia interdisciplinar para abordar a tuberculose em cada região brasileira.

Para estudar a dinâmica de transmissão da doença considerando os fatores socioeconômicos citados anteriormente, este estudo propõe simular esse fenômeno por meio do modelo SIRS adaptado e analisá-lo correlacionando aos fatores socioeconômicos, o que permite a descrição do comportamento da doença e como esta se espalha na população de cada região de forma única (Aparicio e Castillo-Chavez, 2009; Abreu, 2020; Almeida e Silveira, 2021; Nascimento, 2020).

Para um melhor entendimento da descrição do comportamento da doença, será aplicada a abordagem numérica partindo do método de Runge-Kutta de ordem 4, o que permite calcular soluções aproximadas para as equações do modelo epidemiológico. Os resultados disso serão analisados perante a epidemiologia social.

Diante disso, o estudo busca contribuir nos seguintes pontos:

1. Produzir uma ferramenta de epidemiologia matemática que, aliada a dados socioeconômicos e de saúde, consiga compreender e prever o comportamento da tuberculose nas diferentes regiões do Brasil; e
2. Investigar a dinâmica da transmissão da tuberculose por região por uma abordagem multidisciplinar, não apenas produzindo uma ferramenta de epidemiologia matemática, mas associando seus resultados a características socioeconômicas e de saúde que influenciam a dinâmica da doença.

2. Metodologia

2.1. Modelo SIRS adaptado

Com base em conhecimentos prévios sobre epidemiologia matemática e tuberculose, o modelo SIRS (equações 2.1–2.3) foi adaptado para representar a dinâmica da tuberculose. A Figura 1 ilustra o diagrama do modelo.

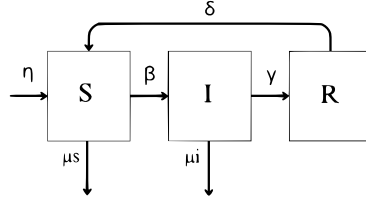


Figura 1: Diagrama do modelo SIRS adaptado.

$$\frac{dS}{dt} = -\frac{\beta SI}{100} + (\eta - \mu_s)S + \xi R \quad (2.1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{100} - \gamma I - \mu_i I \quad (2.2)$$

$$\frac{dR}{dt} = \gamma I - \xi R \quad (2.3)$$

sendo S : População suscetível (%); I : População infectada (%); R : População recuperada (%); β : Taxa de transmissão; γ : Taxa de recuperação; μ_s : Taxa de mortalidade natural dos suscetíveis; μ_i : Taxa de mortalidade dos infectados; η : Taxa de entrada na população; e ξ : Taxa de perda de imunidade.

Tal modelo foi escolhido por sua capacidade de representar a complexidade da tuberculose, incluindo características como o estado latente da doença, que resulta em transmissão prolongada, e o tratamento de longa duração necessário para a recuperação completa.

Para a implementação do modelo, foram definidos o método numérico (Runge-Kutta de 4 Ordem - RK4O), as condições iniciais e os valores dos parâmetros, conforme detalhado a seguir.

2.2. Método numérico: Runge-Kutta de 4 ordem

O método RK4O deve permitir calcular soluções aproximadas para as equações do modelo epidemiológico¹ e para aplicar ele no modelo SIRS, ou seja, nas EDOs ((2.1)–(2.3)), devem ser calculados os incrementos em cada k a cada passo de tempo t_n , ou seja, k_{1S} , k_{1I} , k_{1R} , k_{2S} , k_{2I} , k_{2R} , k_{3S} , k_{3I} , k_{3R} , k_{4S} , k_{4I} e k_{4R} , para chegar na atualização de variáveis a baixo:

$$S_{n+1} = S_n + \frac{h}{6} (k_{1S} + 2k_{2S} + 2k_{3S} + k_{4S}), \quad (2.4)$$

$$I_{n+1} = I_n + \frac{h}{6} (k_{1I} + 2k_{2I} + 2k_{3I} + k_{4I}), \quad (2.5)$$

$$R_{n+1} = R_n + \frac{h}{6} (k_{1R} + 2k_{2R} + 2k_{3R} + k_{4R}). \quad (2.6)$$

A partir da atualização das variáveis, podemos entender as equações (2.1)–(2.3) como funções que representam as derivadas em relação ao tempo.

2.3. Condições iniciais do modelo

As condições iniciais dos compartimentos Suscetíveis (S), Infectados (I) e Recuperados (R) foram definidas com base nos dados oficiais e ajustes metodológicos descritos a seguir.

2.3.1 Definição da população total média (P_{tm})

Primeiro, define-se a população total média que representa 100% dos indivíduos da região e foi estimada a partir dos dados de 2013 e 2023. Para aprimorar a acurácia dos cálculos, considerou-se a taxa anual média de crescimento de 0,52% (Cabral, 2023). A Tabela 1 apresenta os valores de P_{2013} , P_{2023} e a população total média (P_{tm}) para cada região.

¹O método numérico RK4O funciona através do cálculo de soluções aproximadas e isso pode apresentar certo erro e de forma adicional, utilizamos quatro métricas de erro para analisar isso: erro médio absoluto (MAE), erro quadrático médio (MSE), raiz do erro quadrático médio (RMSE) e erro médio percentual absoluto (MAPE), que seus cálculos e códigos estão disponíveis no GitHub: https://github.com/Leprechias/TCC_Tuberculose.

Tabela 1: População total média (P_{tm}) de 2013 e 2023 para cada região.

Região	P_{2013}	P_{2023}	P_{tm}
Norte	16.113.228	17.445.129	16.779.179
Nordeste	53.914.341	54.942.739	54.428.540
Centro-Oeste	14.278.542	16.374.243	15.326.393
Sudeste	81.624.625	86.286.481	83.955.553
Sul	27.816.351	30.093.382	28.954.867

2.3.2 População infectada inicial (I_0)

A determinação de I_0 baseou-se na média anual dos casos de tuberculose registrados entre 2013 e 2023. Seja X a média dos casos anuais, a proporção inicial de infectados foi calculada por:

$$I_0 = \left(\frac{X}{P_{tm}} \right) \times 100\% \quad (2.7)$$

Para representar adequadamente a realidade epidemiológica, aplicou-se uma correção com base na variação percentual dos casos ao longo do período.

2.3.3 População recuperada inicial (R_0)

De forma semelhante, R_0 foi definida a partir da média anual dos casos recuperados (denotada por Y), normalizada em relação à população total média:

$$R_0 = \left(\frac{Y}{P_{tm}} \right) \times 100\% \quad (2.8)$$

2.3.4 População suscetível inicial (S_0)

Para que a soma dos compartimentos totalizasse 100%, o valor inicial de indivíduos suscetíveis foi determinado por:

$$S_0 = 100\% - (I_0 + R_0) \quad (2.9)$$

2.3.5 Valores iniciais para cada região

A aplicação dos procedimentos acima permitiu estabelecer os valores iniciais dos compartimentos para cada região, conforme resumido na Tabela 2.

Tabela 2: Proporções iniciais dos compartimentos S , I e R para cada região.

Região	S_0 (%)	I_0 (%)	R_0 (%)
Norte	99,99137	0,00519	0,00343
Nordeste	99,99436	0,00354	0,00209
Sudeste	99,99352	0,00399	0,00248
Sul	99,99496	0,00314	0,00189
Centro-Oeste	99,99621	0,00234	0,00144

Os valores estabelecidos serviram como condições iniciais para a simulação numérica do modelo SIRS, possibilitando a análise da dinâmica de transmissão da tuberculose e a avaliação da eficácia das medidas de controle implementadas.

2.4. Valores dos parâmetros do modelo

A partir das condições iniciais, as variáveis do modelo SIRS foram estabelecidas com base em dados oficiais e procedimentos específicos para cada parâmetro, conforme descrito a seguir.

2.4.1 Taxa de mortalidade natural dos suscetíveis (μ_s)

Utilizando dados de DATASUS (2024), determinou-se o total de mortes naturais (M) de 2013 a 2022 para cada região. A taxa anual média é calculada por:

$$\mu_{\text{anual}} = \frac{M}{P_{tm} \times 10}, \quad (2.10)$$

em que P_{tm} é a população total média, e o fator 10 refere-se ao período de 10 anos. A conversão para base diária foi feita dividindo-se por 365:

$$\mu_s = \frac{\mu_{\text{anual}}}{365}. \quad (2.11)$$

Os valores calculados estão apresentados na Tabela 3.

Tabela 3: Taxa de mortalidade natural dos suscetíveis por região.

Região	M (2013-22)	P_{tm}	μ_s Anual	μ_s Diário
Norte	884160	16.779.179	0,52694	0,00144
Nordeste	3632709	54.428.540	0,66743	0,00183
Sudeste	6336787	83.955.553	0,75478	0,00207
Sul	2136378	28.954.867	0,73783	0,00202
Centro-Oeste	934350	15.326.393	0,60963	0,00167

2.4.2 Taxa de entrada na população (η)

A taxa de entrada, predominantemente decorrente dos nascimentos, foi estimada a partir dos dados de Depto de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis (2024). Para cada região, o total de nascimentos (N) de 2013 a 2022 foi dividido por $P_{tm} \times 10$, obtendo-se a taxa anual:

$$\eta_{\text{anual}} = \frac{N}{P_{tm} \times 10}. \quad (2.12)$$

A taxa diária é dada por:

$$\eta = \frac{\eta_{\text{anual}}}{365}. \quad (2.13)$$

Os valores estão resumidos na Tabela 4.

Tabela 4: Taxa de entrada (natalidade) na população por região.

Região	Nascimentos totais (N)	η_{anual}	η Diário
Norte	3.100.165	0,01847	0,00506
Nordeste	8.002.214	0,01470	0,00403
Centro-Oeste	2.377.227	0,01551	0,00425
Sudeste	11.097.956	0,01321	0,00362
Sul	3.858.973	0,01333	0,00365

2.4.3 Taxa de mortalidade dos infectados (μ_i)

A partir dos óbitos associados à tuberculose (DATASUS, 2024), obteve-se a média normalizada de mortes entre 0 e 1 para cada região. Devido a possíveis subnotificações (Melo et al., 2018), aplicou-se um ajuste (incremento variando entre $-0,0003$ e $+0,0003$). Os valores finais de μ_i estão na Tabela 5.

Tabela 5: Taxa de mortalidade dos infectados por região.

Região	Média normalizada	Incremento	μ_i
Norte	0,0025	+0,0002	0,0027
Nordeste	0,0031	+0,0003	0,0034
Centro-Oeste	0,0033	-0,0003	0,0030
Sudeste	0,0027	+0,0003	0,0030
Sul	0,0034	+0,0003	0,0037

2.4.4 Taxa de recuperação (γ) e a taxa de perda de imunidade (ξ)

A taxa de recuperação foi obtida a partir da média de recuperados no período 2013–2022 (DATASUS, 2024), normalizada e ajustada com incrementos (de 0 a +0,0007) para considerar subnotificações e fatores do tratamento (Melo et al., 2018; U.S. Centers for Disease Control and Prevention, 2024). Os resultados estão na Tabela 6.

Já a taxa de perda de imunidade foi definida considerando um tempo mínimo de nove meses para reinfecção ou reativação. Define-se a taxa de perda de imunidade como metade de γ :

$$\xi = \frac{\gamma}{2}. \quad (2.14)$$

A Tabela 6 apresenta os valores obtidos.

Tabela 6: Taxas de recuperação dos infectados por região e de perda de imunidade por região.

Região	Média normalizada	Incremento	γ	ξ
Norte	0,0066	+0,0007	0,0073	0,00365
Nordeste	0,0058	+0,0006	0,0064	0,00320
Centro-Oeste	0,0061	+0,0006	0,0067	0,00335
Sudeste	0,0062	+0,0006	0,0068	0,00340
Sul	0,0060	0,0	0,0060	0,00300

2.4.5 Taxa de transmissão (β)

A determinação de β utilizou uma abordagem híbrida que combina otimização global e local. Inicialmente, o algoritmo *Differential Evolution* (DE) foi empregado para explorar o intervalo $[0,1]$ utilizando as médias percentuais

mensais dos casos de tuberculose (2013–2023), excluindo outliers (vide Tabelas 7 e 8).

Tabela 7: Média de casos percentuais mensais (Jan–Jun) por região.

Região	Jan	Fev	Mar	Abr	Mai	Jun
Norte	0,00587	0,00542	0,00631	0,00614	0,00629	0,00617
Nordeste	0,00374	0,00341	0,00385	0,00360	0,00372	0,00344
Sudeste	0,00419	0,00379	0,00446	0,00391	0,00398	0,00390
Sul	0,00328	0,00291	0,00328	0,00313	0,00312	0,00303
Centro-Oeste	0,00240	0,00224	0,00272	0,00241	0,00244	0,00234

Tabela 8: Média de casos percentuais mensais (Jul–Dez) por região.

Região	Jul	Ago	Set	Out	Nov	Dez
Norte	0,00612	0,00629	0,00601	0,00584	0,00570	0,00460
Nordeste	0,00373	0,00387	0,00358	0,00363	0,00333	0,00266
Sudeste	0,00410	0,00431	0,00412	0,00416	0,00382	0,00318
Sul	0,00324	0,00350	0,00326	0,00338	0,00306	0,00255
Centro-Oeste	0,00240	0,00251	0,00239	0,00239	0,00217	0,00173

Em seguida, o método *L-BFGS-B* refinou os valores encontrados com base em estimativas numéricas do gradiente. Esses valores estão dispostos na Tabela 9. Além disso, foi necessário realizar ajustes empíricos subsequentes nesses valores, para que o modelo reproduzisse satisfatoriamente a dinâmica dos dados, conforme mostra a Tabela 9.

Tabela 9: Comparação entre os valores de β obtidos por otimização e ajustes empíricos.

Região	β (Otimização)	β (Empírico)
Norte	0,0107	0,01397
Nordeste	0,0099	0,01230
Centro-Oeste	0,0098	0,01261
Sudeste	0,0098	0,01174
Sul	0,0097	0,01159

3. Resultados

Além de todas as tabelas já apresentadas essa metodologia forneceu gráficos que comparam as projeções do modelo com os dados reais em cada região.

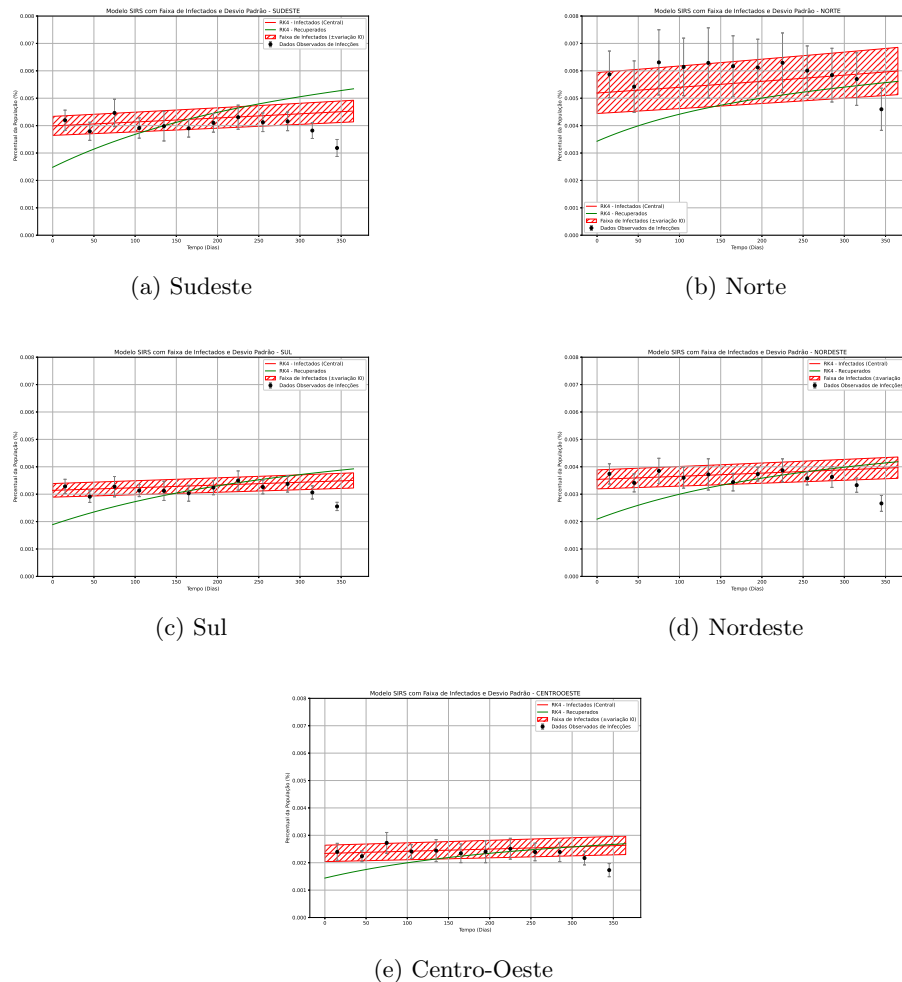


Figura 2: Modelo proposto para projeção de casos de tuberculose nas regiões brasileiras. A faixa de infecção indica a variação estimada nos casos infectados, considerando o desvio padrão dos dados observados.

4. Discussão

A partir dos gráficos, é possível realizar interpretações multifatoriais para explicar a realidade de cada região.

4.1. Sudeste

Para a região Sudeste (Fig. 2a), o modelo SIRS prevê um leve aumento percentual no número de infectados, passando de aproximadamente 0,0035% para 0,0040%, o que numericamente representa uma variação de cerca de 1959 para 2177 casos no período analisado. Apesar desse aumento, o gráfico evidencia que a taxa de recuperação ultrapassa a taxa de infectados em torno de 160 dias, indicando que o sistema público de saúde da região consegue controlar gradualmente a epidemia, especialmente no que tange à redução do número de óbitos. Porém, evidencia também que a região não consegue reduzir significativamente o número absoluto de casos.

Esse cenário epidemiológico pode ser explicado por fatores socioeconômicos conflitantes presentes na região. Por um lado, o elevado Índice de Desenvolvimento Humano (IDH) exerce um efeito protetor, ao facilitar maior acesso à educação e serviços básicos de saúde. Por outro lado, questões estruturais relacionadas à urbanização acelerada, como alta densidade populacional em áreas metropolitanas, moradias inadequadas, baixa disponibilidade de água potável e infraestrutura deficiente em regiões periféricas proporcionam condições favoráveis à disseminação da tuberculose e dificultam a adesão ao tratamento. Essas vulnerabilidades urbanas, sem o contrapeso positivo do alto IDH regional, poderiam resultar em taxas significativamente maiores de infecção (Dvorak, 2011; Agência O Globo, 2024; Weise, 2023).

Portanto, embora o desempenho do modelo seja aceitável e suas métricas de erro estejam dentro de limites razoáveis, pequenas variações nas previsões podem ter implicações relevantes na saúde pública. Sugere-se que futuros estudos aprofundem a investigação dessas vulnerabilidades específicas, com abordagens que permitam melhor compreender como fatores socioeconômicos interagem e influenciam a dinâmica epidemiológica da tuberculose na região Sudeste.

4.2. Norte

A Figura 2b ilustra a evolução do compartimento de infectados na região Norte. Observa-se que, embora o crescimento seja discreto, passando de apro-

ximadamente 0,0052 para 0,006, essa variação representa um aumento de cerca de 872 casos para 1006 casos durante o período analisado, sendo a maior entre as regiões estudadas. Por outro lado, a curva dos recuperados apresenta um aumento marginal, permanecendo sempre abaixo da dos infectados. Esse comportamento indica um ritmo de recuperação lento, evidenciando que o sistema público de saúde na região Norte enfrenta dificuldades para controlar a epidemia de tuberculose.

A situação observada pode ser atribuída a diversos fatores interconectados, tais como o isolamento geográfico, manifestado pela vasta extensão territorial e pelo acesso restrito a comunidades remotas, o que dificulta a distribuição de recursos médicos e a presença de profissionais de saúde (Silva e Bacha, 2014; Guimarães et al., 2020). Outro aspecto relevante é a concentração de renda aliada à escassez de oportunidades econômicas, agravando a exclusão social e comprometendo o acesso aos serviços de saúde (Agência O Globo, 2024). Além disso, a infraestrutura insuficiente dificulta a implementação de medidas eficazes de controle, enquanto o subfinanciamento do setor afeta negativamente a qualidade dos serviços prestados (Weise, 2023). Da mesma forma, dados incompletos ou desatualizados nos registros de saúde prejudicam o planejamento e a execução das políticas públicas (Ribeiro et al., 2023).

Esses fatores podem contribuir para a subnotificação dos casos de tuberculose, comprometendo a precisão dos dados e, conseqüentemente, do modelo. Mesmo assim, os resultados obtidos demonstram que o modelo é capaz de capturar robustamente a dinâmica da doença na região, evidenciando a necessidade de políticas públicas mais efetivas para o controle epidemiológico na área.

4.3. Sul

Para a região Sul, o gráfico da Figura 2c demonstra que a projeção de novos casos apresenta um leve aumento, passando de aproximadamente 0,0032 para 0,0035 ao longo de um ano. Paralelamente, a taxa de recuperação ultrapassa a dos infectados por volta dos 200 dias, o que indica que a região Sul tem uma maior capacidade de lidar com a tuberculose.

Este comportamento, entretanto, suscita questionamentos quanto à persistência da doença. Considerando os elevados indicadores socioeconômicos da região - como o alto IDH, exemplificado pelas capitais estaduais - aliados a uma densidade populacional historicamente mediana - resultado de um processo de colonização que privilegiou o espaçamento entre os assentamentos - e

ao desempenho eficiente do sistema público de saúde - com infraestrutura de qualidade e efetividade nas campanhas de combate à tuberculose -, o resultado ideal sugeriria um cenário praticamente livre da doença (Sousa, 2020; Dvorak, 2011; Amorim, 2016; Secretaria da Saúde, 2024).

Contudo, a permanência da tuberculose pode estar relacionada à importação de casos, uma vez que a região Sul, por apresentar uma alta densidade de cidades em áreas de fronteira, registra um fluxo de casos provenientes de estados como São Paulo e Mato Grosso do Sul, além de países vizinhos (Argentina, Paraguai e Uruguai). Adicionalmente, fatores como a tendência de aglomeração em ambientes fechados e aquecidos em períodos de temperaturas mais baixas também podem favorecer a disseminação da doença (Kossakovski, 2020).

Assim, embora o modelo SIRS projete um cenário estável e um sistema de saúde robusto, os níveis atuais da doença ainda estão distantes dos padrões recomendados pela OMS (WHO, 2024), evidenciando a necessidade de políticas públicas que adotem uma abordagem multifatorial, focando nos desafios específicos das áreas de fronteira.

4.4. Nordeste

Na região Nordeste, representada no gráfico da Figura 2d, o comportamento da curva de infectados apresentou um aumento moderado ao longo de um ano, subindo de cerca de 0,00354 para 0,004, correspondendo numericamente ao aumento de aproximadamente 1959 para 2177 casos. Apesar desse crescimento, a taxa de recuperação alcança a taxa de infectados próximo aos 220 dias, indicando que o sistema de saúde consegue controlar a epidemia, embora com maior dificuldade do que outras regiões.

Este cenário epidemiológico pode ser explicado por fatores estruturais e socioeconômicos interligados. A região possui um legado colonial que deixou uma infraestrutura urbana insuficiente para atender as demandas atuais da população, dificultando a implementação eficaz de medidas de saúde pública (Silva e Bacha, 2014; Guimarães et al., 2020). Além disso, as condições socioeconômicas precárias, como moradias inadequadas, superlotadas e mal ventiladas, favorecem a transmissão da tuberculose (César, 2020). Soma-se a isso o subfinanciamento histórico da saúde pública, resultando em escassez de medicamentos, profissionais qualificados e campanhas de prevenção eficazes (Zanon, 2024). Outro fator relevante é a alta mobilidade populacional interna, especi-

almente migrações sazonais para regiões Centro-Oeste e Sudeste, que podem interromper tratamentos em andamento, afetando negativamente o controle da doença.

Portanto, o controle efetivo da tuberculose no Nordeste requer estratégias integradas entre governos federal, estadual e municipal, com investimentos na infraestrutura da rede de saúde e fortalecimento das políticas públicas de saúde coletiva.

4.5. Centro-Oeste

A região Centro-Oeste (gráfico da Figura 2e) apresentou uma leve elevação no número de infectados no período estudado, indo de aproximadamente 0,00234 (~ 359 casos) para cerca de 0,0025 (~ 420 casos). Apesar do pequeno crescimento, observa-se que a curva de recuperados supera a de infectados somente próximo ao dia 300, demonstrando um ritmo relativamente lento de recuperação. O modelo mostrou resultados satisfatórios na captura dessa dinâmica epidemiológica, conforme evidenciado pelas métricas de erro apresentadas.

Esse comportamento pode estar associado a fatores específicos da região, como a significativa presença de população indígena, especialmente no Mato Grosso do Sul e em áreas da Amazônia Legal, o que dificulta o acesso aos serviços de saúde. Soma-se a isso a intensa mobilidade populacional resultante da migração sazonal ligada à atividade agropecuária, o que torna mais difícil o rastreamento de casos e o acompanhamento contínuo dos tratamentos. Adicionalmente, o acesso limitado aos serviços públicos de saúde, especialmente em municípios menores e em áreas rurais, a fragilidade dos registros epidemiológicos e a infraestrutura precária de transporte contribuem para potenciais subnotificações e tratamentos interrompidos (Silva Neto, 2023; Famato, 2018; Correa e Ramos, 2010).

Diante desse cenário, reforça-se a necessidade de políticas públicas voltadas ao fortalecimento da infraestrutura básica dos municípios menores, melhoria dos sistemas de notificação, maior investimento na rede de atenção básica e na capacitação dos profissionais de saúde para garantir uma resposta eficaz ao desafio epidemiológico representado pela tuberculose na região.

5. Considerações finais

Os resultados obtidos nesse estudo demonstraram que o modelo SIRS foi capaz de representar, com uma precisão razoável e robusta, a dinâmica da doença em cada região. Este estudo evidencia a complexidade da tuberculose no Brasil e a necessidade de abordagens multidisciplinares para seu controle e considerando as disparidades regionais, é necessária a implementação de estratégias específicas, considerando as particularidades socioeconômicas, culturais e de infraestrutura de cada área. Além de ter revelado que políticas públicas integradas, investimentos em saúde, melhoria das condições de vida e educação em saúde são fundamentais para a redução dos casos e eventual erradicação da tuberculose no país.

Por fim, reconhecemos as limitações deste trabalho, especialmente em relação à qualidade e disponibilidade dos dados, possíveis subnotificações e simplificações inerentes aos modelos matemáticos. Recomendamos que em estudos futuros seja considerado a incorporação de dados mais atualizados, inclusão de outras variáveis relevantes e o uso de modelos mais complexos e equilibrados que possam capturar de forma ainda mais precisa a dinâmica da doença em diferentes contextos.

Agradecimentos

Gostaríamos de agradecer aos pesquisadores Ivya Minelli, Paola Ferrari e Vinícius Wasques pelas eventuais colaborações no decorrer desse projeto e a instituição financiadora Ilum - Escola de Ciência por viabilizar esse projeto.

Referências

- Abreu, M. O. R. (2020). Modelagem matemática da Covid-19 usando SIR. *Jornal Eletrônico de Ensino e Pesquisa de Matemática*, 4(1):27–48.
- Aguiar, R. S. C. (2018). Características socioeconômicas da incidência de tuberculose no Brasil e em Pernambuco. Dissertação de Mestrado, CCSA–UFPE, Recife/PE.
- Agência O Globo (2024). Região norte é a mais afetada pela fome. no Pará, quase 10% das famílias sofrem privação. Disponível

em: <https://exame.com/brasil/regiao-norte-e-a-mais-afetada-pela-fome-no-para-quase-10-das-familias-sofrem-privacao/>, acesso em 25/04/2024.

Almeida, L. A. F. e Silveira, G. P. (2021). Análise do comportamento da Covid-19 no Brasil via modelo SIR. *C.Q.D. - Revista Eletrônica Paulista de Matemática*. Publicado em 21/12/2021.

Amorim, W. (2016). Região Sul recebe prêmios por combate à tuberculose. Relatório técnico, Secretaria Municipal da Saúde.

Aparicio, J. P. e Castillo-Chavez, C. (2009). Mathematical modelling of tuberculosis epidemics. *Mathematical Biosciences and Engineering: MBE*, 6(2):209–237.

Borges, V. F. e Teixeira, E. C. (2020). Efeito da tuberculose sobre os rendimentos do trabalho no Brasil. *RDE - Revista de Desenvolvimento Econômico*, 2(46).

Cabral, U. (2023). De 2010 a 2022, população brasileira cresce 6,5% e chega a 203,1 milhões. <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37237-de-2010-a-2022-populacao-brasileira-cresce-6-5-e-chega-a-203-1-milhoes>, acesso em 20/04/2025.

Correa, V. H. C. e Ramos, P. (2010). A precariedade do transporte rodoviário brasileiro para o escoamento da produção de soja do Centro-Oeste: situação e perspectivas. *Revista de Economia e Sociologia Rural*, 48(2).

César, D. (2020). Região Nordeste possui quase metade de toda a pobreza no Brasil, segundo IBGE. <https://www.fecop.seplag.ce.gov.br/2020/11/20/regiao-nordeste-possui-quase-metade-de-toda-a-pobreza-no-brasil-segundo-ibge/>.

DATASUS (2024). Departamento de Informática do SUS - Informações de Saúde (TABNET). url=<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>, acesso em: 24/04/2024.

Depto de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis (2024). Painel de monitoramento de nascidos vivos.

Dvorak, A. (2011). A geografia histórica da colonização e da imigração europeia no Sul do Brasil. *Revista de reflexão Brasil-Polônia*, 4:106–122.

- Famato (2018). Infraestrutura em transporte agrícola: como melhorar no Centro-Oeste? Disponível em: <https://cargox.com.br/blog/infraestrutura-em-transporte-agricola-comomelhorar-no-centro-oeste>, acesso em 25/04/2024.
- Fiocruz (2024a). Crise econômica e política fiscal ameaçam combate à tuberculose no Brasil. Disponível em: <https://abrir.link/NsraI>, acesso em: 22/04/2024.
- Fiocruz (2024b). Tuberculose. Disponível em: <https://portal.fiocruz.br/doenca/tuberculose>, acesso em: 22/04/2024.
- Gonçalves, H. (2000). A tuberculose ao longo dos tempos. *História, Ciências, Saúde-Manguinhos*, 7(2):305–327.
- Guimarães, A. F., Barbosa, V. L. M., Silva, M. P., Portugal, K. A., e Reis, H. S. (2020). Acesso a serviços de saúde por ribeirinhos de um município no interior do estado do Amazonas, Brasil. *Revista Pan-Amazônica de Saúde*, 11.
- Kossakovski, F. (2020). Por que nosso corpo tem mais dificuldade para combater infecções respiratórias no frio? Disponível em: <https://www.nationalgeographicbrasil.com/ciencia/2020/12/por-que-nosso-corpo-tem-mais-dificuldade-para-combater-infeccoes-respiratorias-frio>, acesso em 24/04/2024.
- Mabud, T. S., Alves, M. L. D., Ko, A. I., Basu, S., Walter, K. S., Cohen, T., Mathema, B., Colijn, C., Lemos, E., Croda, J., e Andrews, J. R. (2019). Evaluating strategies for control of tuberculosis in prisons and prevention of spillover into communities: An observational and modeling study from Brazil. *PLOS Medicine*, 16(1):e1002737.
- Melo, M. A. S., Silva, L. L. M., Melo, A. L. S., e Castro, A. M. (2018). Subnotificação no Sinan e fatores gerenciais e operacionais associados: revisão sistemática da literatura. *Revista de Administração da UEG*, 9 (1):25–43.
- Ministério da Saúde (2022). Boletim epidemiológico: Tuberculose (2022). Publicação especial, Brasil – Ministério da Saúde, Brasília, DF.
- Ministério da Saúde (2024a). Relatório da tuberculose. Disponível em: <https://bvsmms.saude.gov.br/bvs/publicacoes/ProgramaTB.pdf>, acesso em: 24/04/2024.

- Ministério da Saúde (2024b). Tuberculose. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/t/tuberculose/tuberculose>, acesso em: 22/04/2024.
- Nascimento, M. (2020). Dinâmica da Covid-19 descrita pelo modelo SIR: Estudo de caso para três cidades brasileiras. Dissertação de Mestrado, ICSA–Unifal-MG, Varginha/MG.
- OPAS (2023). OPAS pede aos países que acelerem as ações para acabar com a tuberculose. Disponível em: <https://abrir.link/gKE0k>, acesso em: 22/04/2024.
- Ribeiro, L. F. T., Sousa, D. S., Rangel, M. E. M., Curzio, R. L., e Silva, F. J. A. (2023). Dinâmica da coinfeção de tb-hiv na região norte antes (2017-2019) e durante a pandemia(2020-2022). *Brazilian Journal of Implantology and Health*, 5(5):1960–1976.
- Secretaria da Saúde (2024). 28 municípios são reconhecidos por atingir metas de prevenção à tuberculose em Santa Catarina.
- Silva, F. M. (2011). Um estudo econômico da tuberculose no Brasil. Dissertação de Mestrado, FACE–PUCRS, Porto Alegre/RS.
- Silva, R. R. e Bacha, C. J. C. (2014). Acessibilidade e aglomerações na Região Norte do Brasil sob o enfoque da Nova Geografia Econômica. *Estante de Economia e Sociedades Brasileiras*, 24(1).
- Silva Neto, A. L. (2023). Efetividade do tempo de resposta da vigilância epidemiológica em um município da região Centro Oeste do Brasil. Trabalho de conclusão de curso, ECSS–PUCGO, Goiânia/GO.
- Sociedade Brasileira de Pneumologia e Fisiologia (2024). Tuberculose. Disponível em: <https://sbpt.org.br/portal/publico-geral/doencas/tuberculose>, acesso em: 22/04/2024.
- Sousa, M. F. (2020). Região Sul. Relatório técnico, UnB. Disponível em: https://arbocontrol.unb.br/?page_id=1231#:~:text=Possui%20o%20maior%20%C3%8Dndice%20de,a%20menor%20incid%C3%Aancia%20de%20pobreza, acesso em 20/04/2024.

- Uchimura, L. Y. T., Felisberto, E., and M. P. Ferreira, E. R. F., e Viana, A. L. A. (2017). Evaluation performance in health regions in Brazil. *Revista Brasileira de Saúde Materno Infantil*, 17:S259–S270.
- Urell, A. (2022). Remembering Dr. Paul Farmer. <https://eji.org/news/remembering-dr-paul-farmer/>, acesso em: 22/04/2024.
- U.S. Centers for Disease Control and Prevention (2024). Treating active tuberculosis disease.
- Weise, A. (2023). Saúde no norte do Brasil: novos investimentos não dão conta de gargalos de infraestrutura, logística e falta de médicos especialistas. <https://futurodasaude.com.br/saude-regiao-norte/>.
- WHO (2010). Health equity. https://www.who.int/health-topics/health-equity#tab=tab_1, acesso em: 22/04/2024.
- WHO (2023). Global tuberculosis report 2023. Disponível em: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023>, acesso em: 22/04/2024.
- WHO (2024). The end tb strategy. Internet.
- Zanon, M. (2024). Sistema de saúde da região Nordeste é o mais frágil do Brasil, aponta estudo. <https://encurtador.com.br/dCegn>.

Estudo de dinâmica populacional de *Brachionus plicatilis* em função da toxicidade da bituca de cigarro em diferentes concentrações

André A. Caetano,¹ Carlos E. G. Santos,² José D. A. Sales,³ Vinicius F. Wasques⁴

Ilum Escola de Ciencia, CNPEM – 13.087-548, Campinas/SP.

Resumo. Neste estudo é proposta uma modelagem de dinâmica populacional via adaptação do modelo de Malthus e de Verhulst que descrevem o crescimento populacional e a taxa de crescimento, respectivamente, para populações de *Brachionus plicatilis* submetidos a diferentes concentrações de um agente tóxico, no caso a bituca de cigarro. Com isso, focamos em adaptar esses modelos para adicionar um termo em função da concentração da bituca (C) e modelar como diferentes concentrações afetam o crescimento e a taxa de crescimento das populações. Além dessa modelagem, buscamos discorrer sobre os riscos da poluição das bitucas bem como seus efeitos negativos no organismo modelo utilizado.

Palavras-chave: Modelagem; Ecotoxicidade; Malthus; Verhulst

1. Introdução

As bitucas de cigarro (BCs) são os resíduos mais frequentemente descartados em ambientes aquáticos ao redor do mundo, havendo ainda um volume expressivo desse tipo de resíduos em áreas urbanas, sobretudo nas praias (Novotny e Slaughter, 2014; Yousefi Nasab et al., 2022). Estima-se que entre 19% e 38% dos resíduos encontrados em zonas costeiras sejam compostos por resíduos

¹andre.araujoaetano.profissional@gmail.com

²kaduedugsantos@gmail.com

³jdalvessales@gmail.com

⁴vwasques@outlook.com

de BCs (Lian et al., 2024). Esses resíduos por parte são transportadas para corpos d'água por diversos fatores, como pela ação do vento e da chuva ou até por correntes de água, eventualmente alcançando o ambiente marinho. As BCs contêm diferentes materiais, incluindo papel, cinzas, restos de tabaco e, sobretudo, fibras de acetato de celulose, as quais representam um polímero semissintético que apresenta alta resistência à degradação natural (Novotny e Slaughter, 2014). As BCs podem conter mais de 7.000 substâncias químicas, como nicotina, metais pesados e hidrocarbonetos aromáticos, que representam um risco significativo não só para os seres humanos, como também para todo um ecossistema (Lima et al., 2021).

Com todos esses fatores, e os potenciais riscos a vida marinha, é necessário a utilização de métodos experimentais e teóricos a fim de gerar *insights* sobre a atuação e os efeitos deste resíduo em organismos aquáticos. Uma das alternativas teóricas é o estudo da dinâmica populacional de um sistema de interesse, que utiliza dados experimentais do comportamento desse sistema que fundamentam modelos matemáticos capazes de descrever a evolução de um sistema ao longo do tempo, considerando diferentes condições iniciais. Esse campo teve início em 1798, com a publicação do ensaio “*An Essay on the Principle of Population*”, de Thomas Robert Malthus, que, de forma pioneira, sugeriu que a população cresce em progressão geométrica, enquanto os recursos de subsistência aumentam em progressão aritmética. Contudo, ele não considerou que vivemos em um ecossistema fechado, onde limitações como alimento, água, ar ou espaço físico inevitavelmente impõem um teto ao crescimento populacional, levando a uma estabilização.

Com o intuito de testar os efeitos das BCs em organismos a diferentes concentrações (Lian et al., 2024) utilizou o rotífero eurialino *Brachionus plicatilis*, um microzooplâncton herbívoro. Com isso, gerando diversos dados acerca dos efeitos gerados pelas BCs nas populações de *B. plicatilis*. Dentre esses efeitos, destaca-se a redução na taxa de reprodução e a redução na quantidade de ovos, tópicos que serão o foco do estudo.

Com base nesses dados, propõe-se o desenvolvimento de um modelo matemático baseado em dinâmica populacional para descrever o comportamento populacional de *B. plicatilis* em função da concentração de bitucas de cigarro. Tal abordagem permitirá compreender melhor a dinâmica dessas populações sob diferentes níveis de contaminação.

2. Objetivos

Com base nos dados experimentais sintetizados a partir do estudo de Lian et al. (2024), este trabalho propõe o desenvolvimento de duas equações fundamentadas em modelos de dinâmica populacional para descrever os efeitos adversos da exposição à bituca de cigarro sobre populações de *Brachionus plicatilis*. Os principais impactos a serem modelados incluem a variação da densidade populacional e a taxa de crescimento, ambos observados em declínio conforme o aumento da concentração de BC nos experimentos. Logo, nossos objetivos principais para estudo são:

1. Desenvolver um modelo baseado na equação de Malthus para descrever a dinâmica da densidade populacional, incorporando um termo que represente a influência da concentração de BC ao longo do tempo;
2. Adaptar o modelo logístico de Verhulst para representar a taxa de crescimento populacional, incluindo um termo que considere os efeitos da concentração de BC ao longo do tempo.

3. Metodologia

O atual estudo foi realizado a partir da modelagem matemática de dados experimentais descritos anteriormente, em que pretende-se atribuir a concentração da bituca de cigarro como termos adicionais em cada modelo estudado. Ademais, o experimento realizado utilizou tanto as fibras de acetato de celulose que compõem a parte física da bituca, quanto o lixiviado filtrado da água que interagiu com a bituca. Ao analisar os gráficos experimentais, é perceptível que a densidade populacional ao longo do tempo para cada concentração tem um caráter de crescimento Malthusiano. Já para a taxa de crescimento o comportamento aparenta ser um crescimento descrito pelo modelo de Verhulst.

Os dados foram retirados do artigo da referência utilizando o aplicativo *automeris-io* (Automeris.io, 2024) que tem funcionalidade de obter os pontos (x,y) de um determinado gráfico. Dessa forma, foi possível obter com a fidelidade semelhante ao dos gráficos do artigo os dados construído pelos dados experimentais.

3.1. Modelo de Malthus

O modelo clássico de crescimento populacional de Malthus é descrito pela equação diferencial (3.1):

$$\frac{dP}{dt} = rP, \quad (3.1)$$

em que:

- $P(t)$ = representa a densidade populacional no tempo t ;
- t = representa o tempo, dado em dias;
- r = taxa de crescimento intrínseca da população.

A solução desta equação assume a forma exponencial descrita pela equação (3.2) abaixo:

$$P(t) = P_0 e^{rt}, \quad (3.2)$$

sendo P_0 a densidade populacional inicial.

No entanto, este modelo não considera efeitos externos que possam influenciar a taxa de crescimento, como a presença de substâncias tóxicas no ambiente, no caso a bituca. Para capturar esses efeitos, foi necessário modificar a equação (3.2) para incluir um termo de toxicidade que afeta a taxa de crescimento r .

A presença de um composto tóxico na população reduz a taxa de crescimento, um fenômeno frequentemente modelado como uma função decrescente. Uma das alternativas adotada para simular o efeito tóxico é utilizar a relação de exponencial negativa como o da equação (3.3):

$$r_{\text{eff}} = r e^{-kC}, \quad (3.3)$$

nesse caso, $k > 0$ é um coeficiente que representa a intensidade do efeito tóxico sobre r . No entanto, tentativas de ajuste da função mostraram que essa formulação simples pode não representar corretamente situações em que há um limite de concentração acima do qual a população não sobrevive. Para evitar que r_{eff} assumia valores pequenos, mas ainda positivos para concentrações elevadas, foi introduzida uma função logística de transição demonstrada na equação (3.4) abaixo:

$$r_{\text{eff}} = r e^{-kC} \cdot \frac{1}{1 + e^{10(C-2.5)}}. \quad (3.4)$$

Esse termo adicional faz com que r_{eff} tenda rapidamente a zero para $C > 2.5$ (valor próximo a concentração que elimina totalmente a população), garantindo que, em concentrações elevadas do composto, a população efetivamente se extinga, refletindo a realidade experimental.

Realizando outro ajuste de função para os dados experimentais, notou-se que a intensidade da toxicidade pode não ser uma função linear da concentração do composto. Para capturar esse efeito, foi adotado um modelo logístico para k , presente na equação (3.5), garantindo que a relação entre C e a inibição do crescimento seja suavizada.

$$k = \frac{a}{1 + e^{-b(C-c)}}. \quad (3.5)$$

Aqui:

- a = controla a intensidade máxima do efeito tóxico;
- b = define a inclinação da transição;
- c = determina a concentração crítica em que a toxicidade se torna significativa;

Esse ajuste permite uma transição mais realista entre condições de baixa e alta toxicidade, garantindo que o efeito do composto seja minimamente perceptível para pequenas concentrações e se intensifique progressivamente conforme C aumenta.

Substituindo os ajustes acima na equação original de Malthus, obtém-se a função final para a densidade populacional afetada pela toxicidade das BC:

$$P(t) = P_0 \cdot e^{\left(\frac{r \cdot e^{-\left(\frac{a}{1+e^{-b(C-c)}} \right) C}}{1+e^{10(C-2.5)}} \right) t}. \quad (3.6)$$

Então,

- O parâmetro k , dado na equação (3.5), é calculado pela função logística;
- A taxa de crescimento efetiva r_{eff} , equação (3.4), é ajustada por um termo exponencial e uma transição logística;
- O crescimento populacional segue a solução clássica de Malthus, mas agora modificado pelo impacto da toxicidade;

Dessa forma, o modelo proposto mantém a estrutura fundamental do modelo de Malthus, ao mesmo tempo que incorpora mecanismos realistas para descrever o impacto da toxicidade sobre o crescimento populacional. Essa abordagem permite representar adequadamente fenômenos biológicos como a

inibição progressiva do crescimento e a extinção da população em concentrações críticas do composto.

3.2. Modelo de Verhulst

O modelo de Verhulst, descrito pela equação (3.7), sugere que o crescimento populacional é limitado por fatores ambientais que restringem a expansão indefinida da população, de modo que uma capacidade suporte K atribuída ao ambiente seja considerada na equação.

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right), \quad (3.7)$$

em que:

- $P(t)$ = representa a população no instante de tempo t .
- $\frac{dP}{dt}$ = é a taxa de variação da população ao longo do tempo.
- r = corresponde à taxa de crescimento populacional.
- K = define a capacidade de suporte do ambiente.

A solução dessa equação diferencial pode ser expressa como:

$$P(t) = \frac{K}{1 + P_0(e^{rt} - 1)}, \quad (3.8)$$

em que:

- P_0 representa a população inicial no instante $t = 0$.
- O termo e^{rt} descreve o crescimento exponencial desconsiderando limitações ambientais.
- A fração expressa o efeito da limitação dos recursos, garantindo que $P(t)$ se aproxime de K ao longo do tempo.
- O crescimento inicial é aproximadamente exponencial, mas desacelera à medida que a população se aproxima da capacidade de suporte.

Embora o modelo de Verhulst seja tradicionalmente utilizado para descrever o crescimento de uma população ao longo do tempo, neste trabalho ele foi adaptado para modelar a taxa de crescimento da densidade populacional de *Brachionus plicatilis*. Dessa forma, a equação assume uma nova interpretação, na qual:

- $P(t)$ passa a ser representado como $P_{tt}(t)$, descrevendo a taxa de crescimento da densidade populacional de rotíferos ao longo do tempo.

- K não representa mais a capacidade suporte da população, e sim o valor máximo que a taxa de crescimento pode atingir, ou seja, a capacidade suporte da taxa de crescimento.

- r deixa de ser a taxa de crescimento da população e passa a representar a taxa de variação da própria taxa de crescimento, caracterizando a rapidez com que $P_{rt}(t)$ se aproxima de K .

Essa reformulação permite que o modelo logístico seja utilizado para descrever a dinâmica da taxa de crescimento dos rotíferos. Além da mudança do significado dos coeficientes e variáveis da equação, é também necessário acrescentar nesta um modo de tornar a função dependente da concentração de substâncias tóxicas.

A abordagem aplicada foi incorporar a dependência da capacidade suporte da taxa de crescimento em relação à concentração de substâncias tóxicas. Essa dependência foi modelada por meio da seguinte expressão:

$$K(C) = K_0 e^{-\beta C} + K_\infty (1 - e^{-\beta C}). \quad (3.9)$$

Na equação (3.9), $K(C)$ representa a capacidade suporte da taxa de crescimento em função da concentração C da substância tóxica; K_0 é a capacidade suporte inicial, correspondente ao ambiente sem contaminação; K_∞ é a capacidade suporte em altas concentrações da substância tóxica, refletindo o efeito limite da toxicidade; β é um coeficiente que controla a sensibilidade da capacidade suporte à presença da substância tóxica, modulando a transição entre K_0 e K_∞ e C é a concentração da substância tóxica no meio.

A dinâmica da taxa de crescimento foi então modelada por uma função logística modificada, dada pela equação (3.10)

$$P_{rt}(t) = \frac{K(C)}{1 + e^{-r(t-t_0)}}, \quad (3.10)$$

em que t_0 representa o tempo no qual ocorre a inflexão da curva de crescimento.

Essa reformulação permite descrever a dinâmica da taxa de crescimento dos rotíferos considerando o impacto de substâncias tóxicas no ambiente. Assim, o modelo não apenas mantém a estrutura fundamental do modelo logístico, mas também incorpora um mecanismo para capturar os efeitos da toxicidade sobre a capacidade suporte da taxa de crescimento.

3.3. Otimização por evolução diferencial

O objetivo da otimização é ajustar a equação adaptada do modelo de Malthus (3.6) aos dados sintéticos. Para isso, utilizamos o método de Evolução Diferencial (DE, do inglês "Differential Evolution"), uma abordagem estocástica para otimização global que busca minimizar o erro entre os valores experimentais e os valores previstos pelo modelo. Os parâmetros ajustáveis a serem otimizados são r , a , b e c .

A Evolução Diferencial é um algoritmo baseado em população que busca iterativamente um conjunto de parâmetros ótimos para minimizar uma função objetivo (Price, 2013). No nosso caso, queremos encontrar os valores de r , a , b , c que minimizam a soma dos quadrados dos resíduos, onde um resíduo é a diferença entre os valores observados $P_{\text{exp},i}$ e os valores previstos pelo modelo $P_{\text{mod},i}$ para cada ponto experimental i :

$$\text{Erro} = \sum_{i=1}^N (P_{\text{exp},i} - P_{\text{mod},i})^2, \quad (3.11)$$

ou seja, queremos minimizar:

$$S(r, a, b, c) = \sum_{i=1}^N (P_{\text{exp},i} - P_0 e^{r_{\text{eff},i} \cdot t_i})^2. \quad (3.12)$$

O método de Evolução Diferencial funciona através das seguintes etapas:

1. Inicialização da População: O algoritmo começa com uma população inicial de vetores de parâmetros, gerados aleatoriamente dentro de um intervalo predefinido.
2. Mutação: A cada iteração, novos vetores candidatos são gerados ao combinar aleatoriamente três indivíduos da população original. Essa combinação é ponderada por um fator de mutação que controla a variação dos novos candidatos.
3. Recombinação: Para cada vetor da população, um novo vetor candidato é gerado combinando suas características com o vetor mutado, segundo uma taxa de recombinação.
4. Seleção: O novo vetor candidato é avaliado e comparado ao indivíduo original. Se apresentar um menor erro, ele substitui o indivíduo na próxima geração.

5. Convergência: O processo continua até que a variação nos parâmetros seja suficientemente pequena ou o número máximo de iterações seja atingido.

Para a implementação numérica desse conceito, utilizamos a função `differential_evolution` da biblioteca `scipy.optimize`, que executa automaticamente os passos do algoritmo. Esse método apresenta vantagens sobre os quadrados mínimos, pois não depende de derivadas e é mais robusto para funções altamente não lineares ou com múltiplos mínimos locais (Mishra, 2007).

Após a otimização, o modelo ajustado é comparado com os dados experimentais visualmente através de gráficos gerados pela biblioteca `matplotlib`. Um bom ajuste é caracterizado por uma curva que segue bem a tendência dos pontos experimentais, com resíduos pequenos e distribuídos aleatoriamente.

3.4. Otimização por quadrados mínimos

Devido ao modelo de Verhulst apresentar um comportamento mais bem condicionado, a otimização do ajuste da função aos dados foi realizada por meio do método dos Mínimos Quadrados. Esse método busca minimizar a soma dos quadrados dos resíduos, ou seja, as diferenças entre os valores experimentais e os valores previstos pelo modelo. A função objetivo geral a ser minimizada pode ser expressa pela equação (3.13)

$$S = \sum_{i=1}^n (y_i - f(x_i, \theta))^2. \quad (3.13)$$

sendo y_i os valores observados, $f(x_i, \theta)$ a função do modelo parametrizado pelos coeficientes θ , e n o número total de observações.

Esse método é amplamente utilizado para encontrar os parâmetros ótimos que melhor descrevem os dados, garantindo um ajuste adequado ao modelo. Para a modelagem de Verhulst, os parâmetros a serem ajustados são b , k_0 , k_∞ e r da função $P(t)$ da equação (3.10). O ajuste dos parâmetros é realizado minimizando a soma dos quadrados dos resíduos entre os valores observados $P_{rt,i}$ e os valores previstos pelo modelo $P_{mod,i}$:

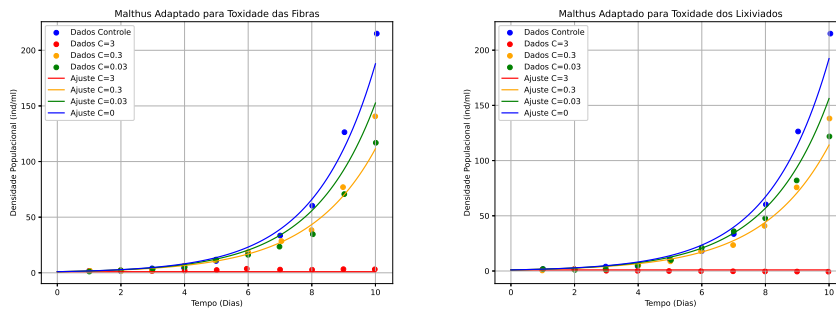
$$S(b, k_0, k_\infty, r) = \sum_{i=1}^N (P_{rt,i} - P_{mod,i})^2. \quad (3.14)$$

A otimização foi implementada utilizando a função `curve_fit` da biblioteca `scipy.optimize`, que ajusta os parâmetros minimizando a função objetivo.

4. Resultados

4.1. Malthus adaptado

Com a equação final do modelo adaptado (3.6), utilizamos o método de otimização por evolução diferencial para encontrar os parâmetros que melhor ajustam a função levando em conta todos os dados reais. Também removendo os dados de $C = 0,03$ que estavam causando sobreajuste na função pois alguns pontos demonstravam que essa concentração era mais tóxica que a de $C = 0,3$. É possível que esse resultado decorra de alguma inconsistência experimental na determinação da densidade populacional. Com isso, foi possível obter os gráficos presentes na Figura 1.



(a) Ajuste otimizado das fibras da bituca. (b) Ajuste otimizado do lixiviado da bituca.

Figura 1: Modelo de Malthus adaptado com ajuste otimizado da evolução diferencial.

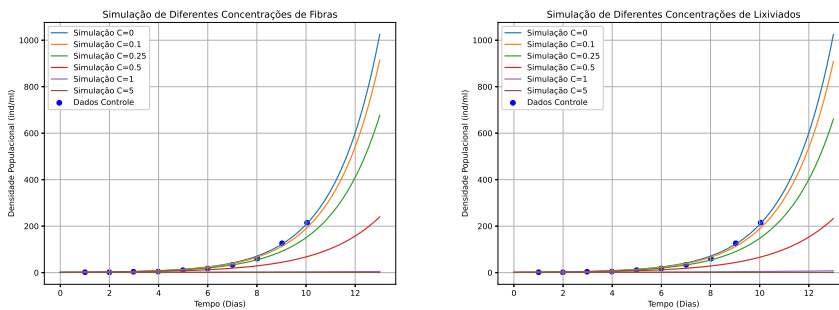
Ao analisar os gráficos da Figura 1, observa-se que tanto as fibras das bitucas de cigarro quanto o lixiviado apresentam efeitos semelhantes na densidade populacional. Podemos notar que a função obteve bons resultados no ajuste dos dados reais, apresentando baixa disparidade nas curvas. Além disso, o ajuste segue a lógica de que ao extremo ($C = 3$) a população não pode crescer novamente, visto que todos os indivíduos morreram. Abaixo os parâmetros otimizados por evolução diferencial:

Tabela 1: Otimização dos parâmetros do modelo de Malthus adaptado.

Teste	r	a	b	c
fibras	0,533	4,887	2,766	1,310
lixiviado	0,533	2,266	2,897	0,952

Na Tabela 1, podemos perceber que apesar dos gráficos serem semelhante os parâmetros a e c demonstram uma certa diferença entre as duas otimizações.

Com base nessa otimização, extrapolamos o modelo para simular valores sem dados experimentais de concentração, propondo assim um possível cenário a partir das concentrações simuladas.



(a) Simulação das fibras da bituca.

(b) Simulação dos lixiviados da bituca.

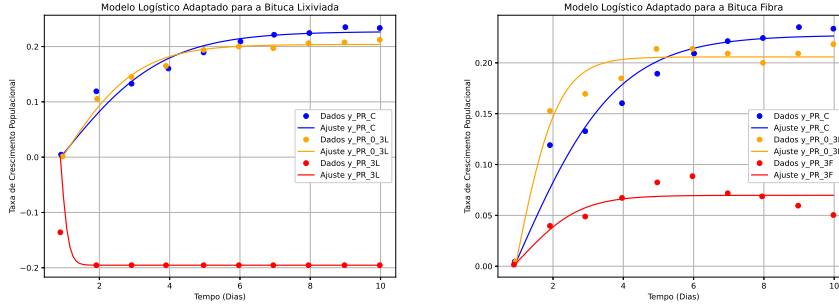
Figura 2: Simulações feitas com o modelo otimizado para os dados experimentais.

Com base na Figura 2, é possível ter uma simulação que extrapola até 12 dias (T) de experimento, com as concentrações C sendo: 0,1, 0,25, 0,5, 1 e 5. Porém, pela simplicidade da equação de Malthus o fator exponencial de um mundo sem limitação de espaço e recursos torna simulações para T maior que 10 sem sentido.

4.2. Verhulst adaptado

Através da aplicação do método dos quadrados mínimos na equação adaptada final do modelo de Verhulst (3.10), foi possível encontrar os melhores coeficientes para ajustar a função aos dados. É válido ressaltar que os dados da

concentração $C = 0,03$ também foram removidos desse processo de otimização devido ao sobreajuste. Assim, os gráficos da Figura 3 foram obtidos.



(a) Ajuste otimizado das fibras da bituca. (b) Ajuste otimizado do lixiviado da bituca.

Figura 3: Modelo de Verhulst adaptado com ajuste otimizado.

Comparando os gráficos da Figura 3, é possível observar que o comportamento da taxa de crescimento da densidade populacional de rotíferos é bastante semelhante para a bituca lixiviada e para as fibras nos dados do grupo controle e na concentração de 0,3. No entanto, nota-se uma diferença significativa no crescimento ao longo do tempo, sendo que a fibra da bituca apresenta um aumento mais acentuado em comparação ao lixiviado. Devido ao fato do processo de otimização por quadrados mínimos ocorrer de forma individual para cada concentração e caso, a Tabela 2 apresenta um aproximação dos coeficientes por meio de uma média.

Tabela 2: Média dos parâmetros do modelo de Verhulst adaptado.

Teste	r	K_0	K_∞	b	t_0
Fibras	3,284	0,60	0,30	0,38	0,92
Lixiviado	3,284	0,60	0,30	0,38	0,92

5. Conclusões

Com base na função adaptada de Malthus, foi possível criar um modelo que se ajustou bem aos dados experimentais utilizados. A partir desse modelo foi possível realizar simulações computacionais, a fim de inferir possíveis

cenários com outras concentrações do agente tóxico. Para a função adaptada de Verhulst, foi possível fazer um ajuste geral para modelar a taxa de crescimento da densidade populacional no tempo com base nas otimizações feitas para cada caso e concentração.

Com isso, entende-se que ambos os modelos podem servir como um material de apoio para outros estudos que busquem entender como o crescimento populacional de um conjunto de organismos pode variar, ou desaparecer, dado a concentração de um agente tóxico.

Agradecimentos

Os autores agradecem a Ilum Escola de Ciência e ao Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) pela disponibilidade dos recursos computacionais utilizados nesse projeto.

Referências

- Automeris.io (2024). WebPlotDigitizer. Disponível em: <https://github.com/automeris-io/WebPlotDigitizer>, acesso em: 20/11/2024.
- Lian, H., Zhu, L., Li, M., Feng, S., Gao, F., Zhang, X., Zhang, F., Xi, Y., e Xiang, X. (2024). Emerging threat of marine microplastics: Cigarette butt contamination on yellow sea beaches and the potential toxicity risks to rotifer growth and reproduction. *Journal of Hazardous Materials*, 478:(art.nb. 135534).
- Lima, C. F., Pinto, M. A. S., Choueri, R. B., Moreira, L. B., e Castro, I. C. B. (2021). Occurrence, characterization, partition, and toxicity of cigarette butts in a highly urbanized coastal area. *Waste Management*, 131:10–19.
- Mishra, S. K. (2007). Performance of differential evolution method in least squares fitting of some typical nonlinear curves. Disponível em <https://ssrn.com/abstract=1010508>, DOI: <http://dx.doi.org/10.2139/ssrn.1010508>.
- Novotny, T. E. e Slaughter, E. (2014). Tobacco product waste: An environmental approach to reduce tobacco consumption. *Current Environmental Health Reports*, 1(3):208–216.

- Price, K. V. (2013). Differential evolution. In Zelinka, I., Snášel, V., e Abraham, A., editors, *Handbook of Optimization*, volume 38 de *Intelligent Systems Reference Library*, páginas 153–171. Springer, Berlin.
- Yousefi Nasab, A., Oskoei, V., Rezasab, M., Alinejad, N., Hosseinzadeh, A., e Kashi, G. (2022). Cigarette butt littering consequences: a study of pollution rate on beaches and urban environments. *Environmental Science and Pollution Research*, 29(30):45396–45403.

Avaliação de parâmetros em modelos de carga hidráulica para aquíferos confinados via equação de difusão

Ana C. J. Coutinho¹, Elaine C. C. Poletti,²

Divisão de Ambiental, Faculdade de Tecnologia, Unicamp – 13.484-332,
Limeira/SP.

Resumo. O objetivo deste artigo é apresentar avaliação numérica e comportamental de sistemas de águas subterrâneas, a partir de um modelo de difusão derivado da combinação entre a lei de Darcy, que rege o escoamento em meios porosos, e a equação da continuidade, de estudos sobre a conservação de massa em um sistema, resultando em uma equação diferencial parcial de segunda ordem. Considerando como condição inicial uma função gaussiana centrada no domínio físico e como condições de contorno as de Dirichlet homogêneas, o modelo segue resolvido analiticamente e numericamente. Para a solução numérica, adota-se o método das diferenças finitas, com implementação computacional em ambiente Matlab^R. As simulações realizadas permitem analisar a influência de três parâmetros físicos fundamentais dos aquíferos confinados: a condutividade hidráulica, o armazenamento específico e a espessura saturada, fornecendo subsídios para a compreensão do comportamento de sistemas aquíferos mediante perturbações no sistema.

Palavras-chave: Modelagem matemática; poluição de recursos hídricos; águas subterrâneas.

1. Introdução

A crescente degradação dos corpos hídricos do mundo moderno, impulsionada por práticas inadequadas de uso e ocupação do solo, evidencia a necessidade de abordagens quantitativas que integrem aspectos físicos, químicos e

¹a258852@dac.unicamp.br

²elainec@unicamp.br

biológicos, capazes de descrever e prever o comportamento desses sistemas complexos. Neste sentido, a modelagem matemática tem se consolidado como uma ferramenta de suporte técnico-científico de auxílio e subsídio de planejamento e gestão de sistemas (Bonganha et al., 2005; Corseuil et al., 2006).

No tocante aos recursos hídricos, as águas subterrâneas merecem atenção especial, dado seu papel estratégico de desenvolvimento e de abastecimento humano, na produção agrícola e na manutenção da estabilidade ecológica. Cerca de 50% das cidades brasileiras dependem ao menos parcialmente das águas subterrâneas para abastecimento público (Hirata et al., 2019a). Embora relativamente protegidas por camadas geológicas, essas reservas não estão imunes à contaminação (Hirata et al., 2019b; Zhang, 2019). Vazamentos em tanques de combustíveis, falhas em sistemas de esgotamento sanitário, uso excessivo de agrotóxicos e disposição inadequada de resíduos sólidos estão entre as principais fontes de poluição, muitas vezes, silenciosa e de difícil detecção devido à natureza lenta e subterrânea dos processos de infiltração (Corseuil e Marins, 1997; Ribeiro et al., 2007).

As águas subterrâneas compõem uma fração significativa das reservas de água doce acessíveis para consumo humano, desempenhando papel vital de abastecimento de comunidades urbanas e rurais, bem como na sustentação dos fluxos de rios e lagos. Sua contaminação, no entanto, acarreta riscos à saúde e impacta economicamente sistemas com o aumento dos custos de tratamento e a redução da disponibilidade hídrica (Dao et al., 2024; Otálvaro, 1999).

Entre os contaminantes mais preocupantes estão pesticidas, metais pesados e hidrocarbonetos, cuja infiltração pode comprometer a qualidade dos aquíferos de forma duradoura (Pinheiro et al., 2019; Ribeiro et al., 2007). A percolação de líquidos contaminados, sejam eles oriundos de aterros sanitários mal gerenciados, depósitos clandestinos de resíduos sólidos, e ainda outros, constitui uma das vias mais comuns de poluição desses sistemas (Alves e Santos, 2023).

Adicionalmente, falhas estruturais em sistemas de esgoto e deficiências de infraestruturas, inclusive de armazenamento de efluentes e compostos, agravam o cenário de vulnerabilidade dos aquíferos, especialmente em áreas urbanas (Ferracini et al., 2001; Reynolds, 2007).

Nesse contexto, a construção de modelos hidrogeológicos, baseados em leis físicas como a lei de Darcy e a equação da continuidade, torna-se fundamental para representar o fluxo de água em meios porosos e simular a dispersão

de poluentes ao longo do tempo, bem como o comportamento do sistema, mediante outras formas de perturbações pontuais (Wendland, 2003). Tais modelos permitem identificar zonas de maior ou menor capacidade de transmissão de água, fornecendo subsídios para o planejamento de captações, estratégias de remediação e políticas públicas de proteção dos mananciais subterrâneos.

Diversos estudos evidenciam a aplicabilidade e a relevância desses modelos. Mustafa e Mawlood (2024), por exemplo, analisaram o esgotamento de águas subterrâneas em uma bacia hidrográfica do Oriente Médio por meio de modelagem tridimensional baseada em dados de poços, enquanto Bonganha et al. (2005) desenvolveu um modelo hidrológico para compreender o comportamento do sistema aquífero de Araraquara (SP) (Mustafa e Mawlood, 2024; Bonganha et al., 2005). No campo da contaminação, destacam-se os trabalhos de Martínez (2023), que modelou vazamentos de hidrocarbonetos em Valência na Espanha, e o desenvolvimento do modelo solução corretiva baseada no risco (SCBR) no Brasil, voltado à simulação de cenários de contaminação por combustíveis (Corseuil et al., 2006). Firmiano e Wendland (2009), por sua vez, propuseram soluções numéricas para a determinação do campo de velocidades em aquíferos confinados, contribuindo para o avanço da modelagem de fluxos subterrâneos.

Compreender a influência dos parâmetros que modelam os sistemas hídricos subterrâneos é fundamental para entender os fatores que controlam seu comportamento e identificar parâmetros relevantes na dinâmica do sistema. Esse conhecimento pode orientar discussões voltadas à gestão sustentável desses recursos, especialmente em cenários de crescente demanda e vulnerabilidade ambiental. Nesse contexto, os modelos matemáticos que representam com precisão os fenômenos físicos do escoamento subterrâneo tornam-se ferramentas estratégicas para embasar a tomada de decisões em políticas públicas, no planejamento territorial e no manejo de captações.

Este trabalho se justifica pela necessidade de aprofundar o estudo da taxa de variação da carga hidráulica em aquíferos confinados, por meio de uma abordagem que contempla a formulação do modelo matemático, sua solução analítica e numérica, bem como a realização de simulações computacionais. A metodologia empregada é baseada na equação de difusão e na discretização pelo método das diferenças finitas, com implementação em ambiente Matlab^R e permite analisar a influência de três parâmetros fundamentais desses sistemas: de condutividade hidráulica, de armazenamento específico e de espessura

saturada.

Do ponto de vista aplicado, a condutividade hidráulica é especialmente relevante, por expressar a facilidade com que a água se movimenta no meio poroso e refletir alterações geológicas, processos naturais de compactação e impactos antrópicos, como a extração intensiva de água. Ao investigar esses fatores, o estudo contribui para a compreensão dos mecanismos de recarga, transporte de poluentes e resposta dos aquíferos a diferentes cenários de uso e perturbações. Assim, reforça-se o papel da modelagem matemática como instrumento valioso na proteção dos recursos hídricos subterrâneos e na promoção da sustentabilidade ambiental (Foster e Chilton, 2003).

2. Objetivos

Estudar o desenvolvimento de um modelo matemático para análise da carga hidráulica de aquíferos confinados, a partir da dedução que combina a lei de Darcy com a equação da continuidade, visando representar adequadamente os fenômenos físicos envolvidos. O trabalho apresenta a resolução do modelo de forma analítica e numérica, via método de diferenças finitas, implementa computacionalmente as soluções em ambiente Matlab^R e explora cenários para análise da influência de diferentes parâmetros.

3. Descrição do problema

A água é essencial para a manutenção da vida, circulando entre a atmosfera, a crosta terrestre e os oceanos através do ciclo hidrológico. Após a precipitação, uma parte da água é utilizada como subsistência de plantas e animais, uma parte escoia superficialmente até alimentar lagos e rios e, outra parte, por fim, se infiltra no solo, podendo alcançar formações geológicas subterrâneas que são capazes de armazená-la e de permitir seu escoamento, formando os aquíferos (Feitosa et al., 2008).

Os aquíferos, portanto, são unidades hidrogeológicas compostas por materiais porosos e permeáveis, cuja estrutura permite tanto o armazenamento quanto a circulação da água, podendo ser livres ou confinados. De forma simplificada, os aquíferos confinados podem ser considerados uma reserva natural de água no subsolo, contida entre materiais geológicos que permitem que a água se acumule e circule (Woessner e Poeter, 2020).

O comportamento da água subterrânea nos aquíferos confinados depende de propriedades físicas do solo, entre as quais se destacam a porosidade, o teor de umidade e, principalmente, a condutividade hidráulica (Freeze e Cherry, 1979). A porosidade corresponde à razão entre o volume de vazios e o volume total de uma amostra de solo; o teor de umidade representa a fração do volume total ocupado pela água e, por fim, a condutividade hidráulica é uma propriedade que expressa a capacidade do solo ou rocha de transmitir água sob gradiente hidráulico.

A quantificação e análise espacial da carga hidráulica de aquíferos confinados, delimitados por camadas que impedem/restringem a movimentação vertical da água, é importante para a modelagem do escoamento subterrâneo, do transporte de contaminantes e dos reflexos de perturbações pontuais. Nesse contexto, modelos matemáticos baseados na lei de Darcy e na equação da continuidade são formulados, resultando em equações diferenciais parciais que descrevem o fluxo de água no meio poroso, como a equação da difusão aplicada a águas subterrâneas. A solução dessas equações pode ser obtida analiticamente ou de forma numérica, através de métodos como das diferenças finitas, dos elementos finitos, entre outros, e favorece a realização de simulações computacionais que dão suporte à predição de comportamentos do aquífero e à gestão sustentável dos recursos hídricos subterrâneos.

3.1. A lei de Darcy e a equação da continuidade

A lei de Darcy foi descoberta por Henry Darcy em 1856 e trata-se de uma equação fundamental, usada para descrever o fluxo de fluido através de meios porosos, incluindo águas subterrâneas, e é dada pela equação (3.1) (Wendland, 2003; Woessner e Poeter, 2020):

$$Q = -K \frac{dh}{dl} A \quad (3.1)$$

onde:

Q representa a vazão volumétrica (m^3/s);

K indica condutividade hidráulica, é a constante de proporcionalidade que reflete a facilidade com que a água flui através de um material (m/s);

$\frac{dh}{dl}$ é a taxa de variação de carga hidráulica em função da distância l ;

A , a área da seção transversal do fluxo perpendicular à direção do fluxo (m^2).

Tomando-se a vazão específica: $q = \frac{Q}{A}$, a equação (3.1) pode ser reescrita na forma equação (3.2):

$$q = -K \frac{dh}{dl} \quad (3.2)$$

que descreve como a altura manométrica, os gradientes hidráulicos e a condutividade hidráulica estão ligados para quantificar e descrever o fluxo das águas subterrâneas.

A respeito da equação da continuidade, que descreve como a massa de um fluido se conserva durante seu movimento, é utilizada para garantir que as vazões de entrada e saída do aquífero sejam iguais (Freeze e Cherry, 1979). Assim, dado um volume de controle $V = \Delta x \Delta y \Delta z$ unidades de volume, o balanço de massa é uma equação que expressa a conservação da matéria em um sistema e estabelece que a variação da massa dentro de um volume de controle, ao longo do tempo, seja igual à diferença entre as entradas e saídas de massa, mais (ou menos) qualquer geração ou consumo (como fontes, reações químicas, etc). A essa variação chamamos acúmulo:

$$\text{acumulação} = \text{fluxo} \pm \text{fonte}$$

Matematicamente, o acúmulo é indicado por $\frac{\partial m}{\partial t}$, com o tempo t e a massa de fluido no volume de controle m . Considerando a massa de fluido armazenada em um meio poroso m , definida como o produto da massa específica ρ pelo volume de fluido V_f , temos a equação (3.3):

$$m = \rho V_f = \rho \phi \Delta V, \quad (3.3)$$

em que,

ρ é a massa específica do fluido (kg/m³);

ϕ é a porosidade do meio;

ΔV é o volume total do elemento de controle;

$V_f = \phi \Delta V$ é o volume de fluido armazenado no meio poroso.

Daí, a taxa de variação da massa m no tempo t é dada pela equação (3.4):

$$\frac{\partial m}{\partial t} = \frac{\partial(\rho \phi \Delta V)}{\partial t} = \frac{\partial(\rho \phi)}{\partial t} \Delta V \quad (3.4)$$

e representa o acúmulo de massa no interior do volume de controle devido à variação da densidade ou da porosidade.

Quanto ao fluxo de massa, ele pode sofrer variação em qualquer direção espacial x , y e z . Considerando que a quantidade de massa que entra ou sai do sistema, por unidade de volume, é descrita pelas derivadas espaciais dos fluxos de massa em cada uma dessas direções, temos que, na direção x , o fluxo de massa que passa através de uma face do volume de controle é dado pelo produto da massa específica ρ pelo fluxo volumétrico q_x e pela área da seção transversal $A_x = \Delta y \Delta z$, assim: $m_x = \rho q_x A_x = \rho q_x \Delta y \Delta z$.

Considerando tal variação líquida nas três direções espaciais, a quantidade de massa que entra ou sai do sistema, por unidade de volume, é descrita pela equação (3.5):

$$\text{fluxo} = -\frac{\partial m_x}{\partial x} \Delta x - \frac{\partial m_y}{\partial y} \Delta y - \frac{\partial m_z}{\partial z} \Delta z \quad (3.5)$$

sendo que $m_i = \rho q_i$ representa o fluxo de massa na direção i (com $i = x, y, z$).

Portanto, substituindo m_i , temos a equação (3.6):

$$\text{fluxo} = -\frac{\partial(\rho q_x)}{\partial x} \Delta x - \frac{\partial(\rho q_y)}{\partial y} \Delta y - \frac{\partial(\rho q_z)}{\partial z} \Delta z. \quad (3.6)$$

que corresponde à perda líquida de massa por transporte convectivo nas três direções espaciais, sendo consistente com a forma diferencial da equação de continuidade em meio poroso.

Em sua forma vetorial, para fins de generalização, podemos reescrever a equação (3.6) em termos vetoriais, resultando na equação (3.7):

$$\text{fluxo} = -\nabla \cdot (\rho \vec{q}) \Delta V \quad (3.7)$$

sendo $\vec{q} = (q_x, q_y, q_z)$ o vetor fluxo específico de Darcy, $\rho \vec{q}$ é o vetor fluxo de massa e $\nabla \cdot$ representa a divergência.

Dando seqüência aos desenvolvimentos, considerando todos os termos que compõem o balanço de massa em um meio poroso tridimensional, podemos agora escrever a equação geral incorporando, ao acúmulo de massa no volume de controle, a variação líquida do fluxo de massa nas direções x , y e z e uma fonte ou sumidouro de massa.

Conforme indicado, se Q representa a taxa de ganho ou perda de volume de fluido por unidade de volume de meio poroso, então a taxa de variação de massa devido à fonte ou sumidouro é dada pela equação (3.8):

$$\text{fonte} = \rho Q \Delta V \quad (3.8)$$

Portanto, o balanço de massa completo relaciona o acúmulo de massa ao fluxo líquido de entrada e à fonte, resultando na equação (3.9):

$$\underbrace{\frac{\partial(\rho\phi)}{\partial t}\Delta V}_{\text{acúmulo}} = \underbrace{-\nabla \cdot (\rho\vec{q})\Delta V}_{\text{fluxo}} \pm \underbrace{\rho Q\Delta V}_{\text{fonte/Sumidouro}} \quad (3.9)$$

Manipulando-se algebricamente a equação (3.9), com substituições favorecidas pela equação (3.2), seguindo com simplificações para fluidos incompressíveis em meio homogêneo e isotrópico e considerando $Q = 0$, ausência de fonte ou sumidouro no volume de controle, tem-se a equação (3.10), que é a equação da continuidade:

$$\frac{\partial\phi}{\partial t} = K\nabla^2 h \quad (3.10)$$

Ressalta-se que, em regime permanente de escoamento, $\frac{\partial}{\partial t} = 0$, resultando na equação de Laplace, uma equação diferencial parcial muito importante na matemática e na física, especialmente no contexto da análise de fenômenos estacionários (Boyce et al., 2020).

3.2. O modelo de estudo

Considerando que ϕ pode representar uma função escalar dependente do tempo e do espaço, tal como a carga hidráulica h , o modelo matemático que estuda a taxa de variação da carga hidráulica de aquíferos confinados é governado pela equação da difusão da carga hidráulica, dada pela equação (3.11) (Wendland, 2003):

$$S\frac{\partial h}{\partial t} = T\nabla^2 h \quad (3.11)$$

em que,

$h(x, y, t)$ representa a carga hidráulica (altura piezométrica);

S indica o armazenamento específico (sem unidade);

$T = Kb$ é transmissividade do aquífero (em m^2/s);

K representa a condutividade hidráulica do meio (em m/s);

b indica a espessura saturada do aquífero (em m);

$\nabla^2 h$ é o operador laplaciano em coordenadas espaciais.

Para abordagem do modelo em uma variável espacial x , substituindo $T = Kb$, a equação (3.11) resulta na equação (3.12) de interesse do nosso estudo:

$$S \frac{\partial h}{\partial t} = Kb \left[\frac{\partial^2 h}{\partial x^2} \right] \quad (3.12)$$

que descreve como a carga hidráulica se propaga no aquífero confinado ao longo da direção x devido à difusão hidráulica.

3.3. Resolução analítica e numérica do modelo

Manipulando-se a equação (3.12) e considerando, por simplificação, a constante de difusão $D = \frac{Kb}{S} > 0$, tem-se a equação (3.13):

$$\frac{\partial h}{\partial t} = D \frac{\partial^2 h}{\partial x^2}, \quad (3.13)$$

uma equação diferencial parcial de segunda ordem, linear, cuja resolução analítica pode ser obtida pelo método de separação de variáveis, em que $h(x, t) = X(x)T(t)$ (Boyce et al., 2020)

Substituindo e reorganizando a equação (3.13), tem-se:

$$X(x) \frac{dT}{dt} = DT(t) \frac{d^2 X(x)}{dx^2}$$

que, separando-se as variáveis, resulta em duas equações diferenciais ordinárias independentes: uma relacionada à variável temporal t e a outra relacionada à variável espacial x , com λ constante de separação, indicadas na equação (3.14):

$$\begin{cases} \frac{d^2 X(x)}{dx^2} = -\lambda X(x) \implies X(x) = A \cos(\sqrt{\lambda}x) + B \sin(\sqrt{\lambda}x) \\ \frac{dT(t)}{dt} = -\lambda DT(t) \implies T(t) = C e^{-\lambda Dt} \end{cases} \quad (3.14)$$

com $\lambda > 0$; A , B e C constantes, dependentes de condições iniciais e de contorno do problema.

Considerando uma perturbação inicial localizada, suave, no campo de carga hidráulica $h(x, 0) = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-x_0)^2/2\sigma^2}$, com centro $x_0 = L/2$ e largura σ , e as condições de contorno ditas de Dirichlet homogêneas: $h(0, t) = h(L, t) = 0 \implies X(0) = 0, X(L) = 0$, a solução geral da equação de estudo, após manipulação algébrica, resulta na série de Fourier dada pela equação (3.15):

$$h(x, t) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) e^{-\left(\frac{n\pi}{L}\right)^2 Dt} \quad (3.15)$$

tal que $B_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx$ e $\lambda_n = \left(\frac{n\pi}{L}\right)^2$.

Por fim, manipulando-se, a solução analítica da equação (3.13), conforme exposto, é dada pela equação (3.16):

$$h(x, t) = \sum_{n=1}^{\infty} \left(\frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \right) \sin\left(\frac{n\pi x}{L}\right) e^{-\left(\frac{n\pi}{L}\right)^2 Dt} \quad (3.16)$$

Com vistas à resolução numérica, aproximada, da equação (3.13), aplicou-se o método de diferenças finitas, no esquema explícito, avançado no tempo e centrado no espaço, com $x \in [0, L]$, $t \in [0, T]$, $h(x, 0) = f(x)$ e as condições de Dirichlet homogêneas.

Assumindo-se o espaçamento Δx e o passo Δt para a aproximação: $h(x_i, t^n) \approx h_i^n$, com pontos $x_i = i\Delta x$, para $i = 0, 1, \dots, N$ e $t^n = n\Delta t$, para $n = 0, 1, 2, \dots$, adotou-se a aproximação temporal, dada pela equação (3.17) (Cunha, 2000):

$$\left. \frac{\partial h}{\partial t} \right|_{x_i, t^n} \approx \frac{h_i^{n+1} - h_i^n}{\Delta t} \quad (3.17)$$

de modo que,

h_i^n indica o valor aproximado da função h no ponto x_i no tempo n ;

h_i^{n+1} indica o valor aproximado da função h no ponto x_i no tempo $n + 1$;

t representa o incremento no tempo entre os instantes n e $n + 1$.

e a aproximação espacial centrada, dada pela equação (3.18) (Cunha, 2000):

$$\left. \frac{\partial^2 h}{\partial x^2} \right|_{x_i, t^n} \approx \frac{h_{i+1}^n - 2h_i^n + h_{i-1}^n}{\Delta x^2} \quad (3.18)$$

sendo $h_{i-1}^n, h_i^n, h_{i+1}^n$ os valores nos pontos espaciais vizinhos x_{i-1}, x_i, x_{i+1} , no instante n .

Assim, manipulando-se, tem-se a equação (3.19), referente à solução numérica do modelo:

$$h_i^{n+1} = h_i^n + \alpha (h_{i+1}^n - 2h_i^n + h_{i-1}^n) \quad (3.19)$$

com $\alpha = \frac{D\Delta t}{\Delta x^2} \leq \frac{1}{2}$, critério de estabilidade do método explícito.

Quanto à ordem do erro da aproximação, tem-se que o erro de truncamento local é de primeira ordem na variável temporal, ou seja, $\mathcal{O}(\Delta t)$ e de segunda ordem na variável espacial x , ou seja, $\mathcal{O}(\Delta x^2)$.

Na equação (3.19), adotando-se as condições de contorno de Dirichlet homogêneas: $h(0, t) = h(L, t) = 0 \Rightarrow h_0^n = h_N^n = 0, \forall n$; a cada passo no tempo, temos, para $i = 1, \dots, N - 1$, a atualização da equação (3.19), com: $h_0^{n+1} = h_N^{n+1} = 0$.

De modo a avaliar a relação entre as soluções analítica e numérica do problema proposto, apresenta-se a Figura 1:

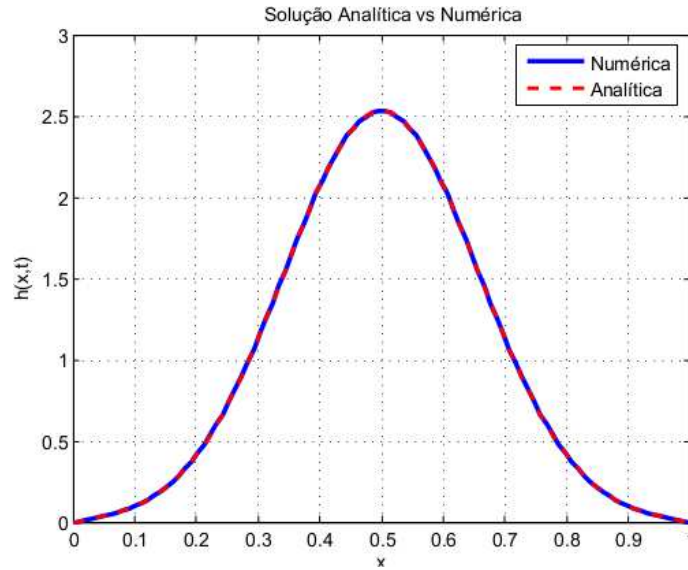


Figura 1: Apresentação de solução analítica (vermelho) e a numérica (azul).

Ressalta-se que a solução analítica refere-se à solução exata do problema matemático para as condições impostas, servindo como referência para validar métodos numéricos. Com relação à solução numérica, ela aproxima a solução analítica, e sua precisão é bastante dependente da malha espacial e temporal.

Na Figura 1, vemos as soluções analítica e numérica da carga hidráulica $h(x, t)$ no instante final do tempo simulado. A boa proximidade entre os gráficos mostra concordância entre os métodos, especialmente considerando o número reduzido de termos da série analítica.

Ressalta-se que, para valores de α maiores que o limite de estabilidade ($1/2$ para o esquema explícito), podem surgir oscilações ou instabilidades.

4. Resultados

Consideradas as soluções analítica e numérica apresentadas, simulações computacionais foram realizadas utilizando a aproximação numérica em ambiente Matlab^R, adotando-se valores hipotéticos para fins ilustrativos, com discussões sobre parâmetros importantes do modelo, para aquíferos confinados.

Importante ressaltar que são três as principais características dos aquíferos: condutividade hidráulica, chamada de K (m/s), armazenamento específico, chamado de S (adimensional) e, por fim, espessura saturada do aquífero b (m). Desta forma, considerando um aquífero de 100 m de comprimento, com um passo $\Delta x = 2$, condição inicial dada por uma função gaussiana centrada no domínio com $\sigma = 5$, com representação de uma curva mais estreita, e condições de contorno de Dirichlet homogêneas, têm-se as simulações representadas nas Figuras 2, 3 e 4, conforme indicado.

No tocante aos parâmetros, a condutividade hidráulica K é uma propriedade do meio poroso que expressa a facilidade com que um fluido se move através dos poros do solo ou da rocha, sendo influenciada tanto pelas características do fluido (como viscosidade e densidade), quanto pelas propriedades do meio (como tamanho e conectividade dos poros) (Fetter, 2001). Isso significa que, para cada tipo de solo ou rocha que compõe o aquífero, há um valor específico de K .

Via de regra, aquíferos confinados apresentam menores valores de K quando comparados aos não confinados, uma vez que são delimitados por camadas de menor permeabilidade, que restringem o fluxo de água subterrânea (Kozhevnikov et al., 2021).

Desta forma, para discutir a influência dos valores de K no modelo, foram simulados três cenários com diferentes materiais (meios porosos), apresentados na Figura 2, gráficos (a), (b) e (c), considerando-se diferentes condutividades hidráulicas, descritas na Tabela 1:

Tabela 1: Valores do parâmetro K (m/s) adotados.

Material	Condutividade Hidráulica
Areia fina	10^{-4}
Calcário fraturado	10^{-5}
Formações argilosas	10^{-8}

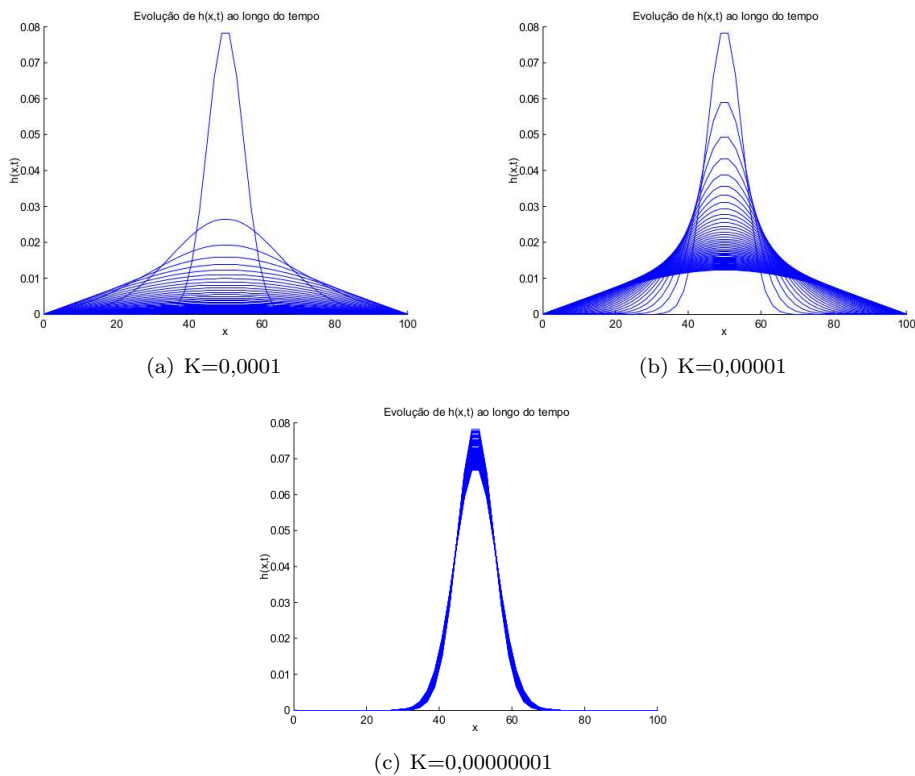


Figura 2: Variação de condutividade hidráulica K (meios porosos): areia fina, calcário fraturado e formações argilosas (praticamente impermeáveis), respectivamente, com $S = 0,0001$ e $b = 50$.

Os gráficos apresentados na Figura 2 consideram aquíferos confinados com espessura de $b = 50m$ e armazenamento específico $S = 0,0001$. O parâmetro b é definido como a extensão saturada entre o topo e o fundo do aquífero (para aquíferos livres, b representa a altura saturada do nível freático, acima do topo do aquífero que delimita o aquífero) (Freeze e Cherry, 1979).

Os valores de b , de modo geral, podem variar significativamente e a literatura recomenda uma faixa de espessura entre 5 e 100 m (Freeze e Cherry, 1979) para aquíferos confinados.

Um exemplo de aquífero dentro dessa faixa é o Mahomet Aquifer, localizado em Illinois, Estados Unidos: aquífero, predominantemente confinado, com espessura entre 30 e 60 m .

No Brasil, mais especificamente no estado de São Paulo, encontramos mais de cinco aquíferos. Dentre eles, destaca-se o aquífero Guarani, que abrange vários países da América do Sul, cobrindo bacias hidrográficas importantes como a do Paraná e do Paraguai. Confinado por basaltos, possui uma extensão confinada de aproximadamente 180 mil Km^2 no Brasil, e profundidades que variam entre dezenas a mais de mil metros. Sua espessura varia entre 100 e 400m (Rocha et al., 2006).

Outro importante aquífero do estado de São Paulo é o Aquífero Bauru, que serve de grande fonte de abastecimento para a porção oeste do Estado. Sua espessura saturada média contempla cerca de 75m, porém pode chegar até 300m em sua extensão (Rocha et al., 2006).

No tocante ao parâmetro S , refere-se ao armazenamento específico do aquífero e é definido como a propriedade que mede a capacidade de um aquífero de liberar (ou armazenar) água, em resposta a variações na hidráulica por unidade de área (Freeze e Cherry, 1979).

Essa propriedade varia conforme o tipo de confinamento, sendo distinta em aquíferos confinados e livres. Para os aquíferos livres, a propriedade de armazenamento é conhecida como rendimento específico (Freeze e Cherry, 1979).

Valores elevados de armazenamento específico S indicam que o aquífero possui maior capacidade de liberar água em resposta a variações na carga hidráulica.

De acordo com a literatura, valores típicos de S para aquíferos confinados variam, aproximadamente, entre valores da ordem de 10^{-3} e 10^{-5} , enquanto para aquíferos livres situam-se entre 0,01 e 0,30 (Freeze e Cherry, 1979).

Desta forma, ao se analisar os gráficos apresentados na Figura 2, observa-se que um maior valor de K , ou seja, uma condutividade hidráulica mais elevada, resulta em um espalhamento mais rápido da curva. Representando, assim, uma diluição mais rápida do pulso, que é o que acontece em meio composto por areia fina, por exemplo, em contraponto às formações argilosas, que são praticamente impermeáveis. Quanto menor a condutividade hidráulica, mais concentrada fica a curva ao redor do pulso.

Portanto, valores maiores de condutividade hidráulica K indicam maior facilidade de movimento da água no meio poroso.

Alterando o valor de armazenamento específico, $S = 0,001$, mantendo as condições de K e b , conforme simulado anteriormente, resulta que $K = 0,0001$ gera instabilidade no sistema e $K = 0,00001$ apresenta simulação equivalente

ao gráfico da Figura 2 (a). Por fim, para $K = 0,00000001$ temos a Figura 3:

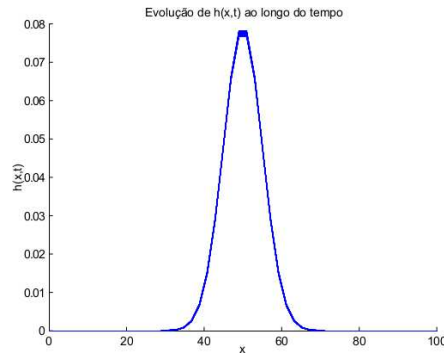


Figura 3: Armazenamento específico $S = 0,001$, $b = 50$ e $K = 0,00000001$.
Fonte: autoral.

Analisando o gráfico gerado, Figura 3, em comparação ao gráfico da Figura 2 (c), podemos entender que um valor maior de S (Figura 2 (c)) concentra mais os valores ao redor do pulso.

Por fim, considerando valores tal qual abordados na Figura 2, com $b = 10$, temos os gráficos 4, 4 e 4 da Figura 4.

A análise da Figura 4 possibilita inferir a influência da espessura b do aquífero exercida no espalhamento da curva ao redor do pulso: quanto maior a espessura b , mais rápido o espalhamento do gráfico ao redor do pulso (Figura 4).

Ou seja, quanto maior a espessura b , maior a transmissividade, a quantidade de água que flui horizontalmente através de uma faixa unitária, determinando a capacidade do aquífero transmitir água.

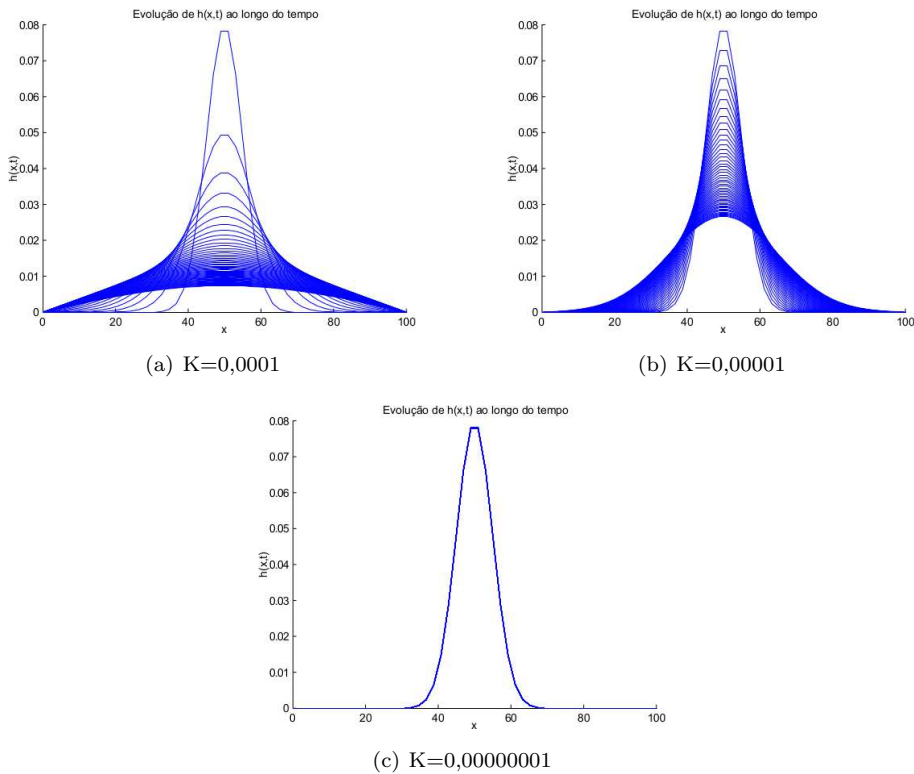


Figura 4: Espessura $b = 10$, com $S = 0,001$ e variação de K .

Sobre a função pulso utilizada, a gaussiana localizada na região central do domínio, ela representa uma perturbação inicial na carga hidráulica, com distribuição espacial suave e simétrica.

Verifica-se nas simulações apresentadas nas Figuras 2, 3 e 4, que a curva se alarga e se achata, representando o efeito da difusão em cada um dos casos abordados. Também se verifica a simetria do gráfico em torno da região central, como esperado em virtude das condições de contorno homogêneas e da perturbação no centro.

Importante ressaltar que a equação de difusão, na ausência de termos de fonte ou sumidouro, assim como tomado neste estudo, conserva a massa total do sistema. Dessa forma, espera-se que a área sob a curva permaneça constante ao longo do tempo.

A cada instante de tempo analisado, observa-se que o perfil da curva

torna-se progressivamente mais largo e menos elevado, evidenciando o espalhamento da massa no meio, considerando que os gráficos exibem a evolução temporal da variável $h(x, t)$, representada para cada múltiplo de 100 passos de tempo.

5. Conclusão

O estudo mostrou boa concordância entre as soluções analítica e numérica para a equação de difusão da carga hidráulica $h(x, t)$, com simetria e conservação de massa garantidas pelas condições de contorno homogêneas e pela condição inicial gaussiana centralizada. O modelo numérico mostrou-se estável dentro do limite teórico de $\alpha \leq 1/2$, sendo observadas instabilidades para valores superiores.

Em relação aos parâmetros físicos do aquífero, constatou-se que maiores valores de condutividade hidráulica K promovem maior espalhamento da carga, enquanto menores valores limitam sua propagação. Por outro lado, o armazenamento específico S apresenta efeito inverso: valores mais altos concentram a carga ao redor do pulso. A espessura saturada b também influencia o espalhamento, aumentando-o conforme o valor de b cresce.

De modo geral, o modelo reproduziu adequadamente o comportamento esperado da difusão em aquíferos confinados, evidenciando o papel dos parâmetros K , S e b na propagação da carga hidráulica e respeitando os princípios físicos e matemáticos envolvidos.

Agradecimentos

Agradecemos ao CNPq - Conselho Nacional de Desenvolvimento Científico pelo fomento à pesquisa e concessão de bolsa.

Referências

- Alves, J. B. e Santos, M. M. (2023). Relações entre contaminação de aquíferos subterrâneos no semiárido e disposição inadequada de resíduos sólidos. *Brazilian Journal of Development*, 9:14810–14827.
- Bonganha, C. A., Ribeiro, M. L., e Guiguer, N. (2005). A utilização da

- modelagem matemática como ferramenta para a gestão integrada e sustentável de recursos hídricos subterrâneos: uma proposta para o município de Araraquara-SP. In *Anais dos XIV Encontro Nacional de Perfuradores de Poços e II Simpósio de Hidrogeologia do Sudeste*, páginas 1–14, Ribeirão Preto/SP. ABAS.
- Boyce, W. E., Diprima, R. C., e B.Meade, D. (2020). *Equações diferenciais elementares e problemas de valores de contorno*. LTC, R.Janeiro/RJ.
- Corseuil, H. X. e Marins, M. D. M. (1997). Contaminação de águas subterrâneas por derramamentos de gasolina: o problema é grave? *Revista Engenharia Sanitária e Ambiental*, 2:50–54.
- Corseuil, H. X., Schneider, M. R., Rosário, M., e Júnior, C. M. (2006). Solução corretiva baseada no risco (SCBR): Modelo matemático de tomada de decisão para o gerenciamento ambiental de Águas subterrâneas contaminadas. In *XIV Congresso Brasileiro de Águas Subterrâneas*, páginas 1–14, Curitiba/PR. ABAS.
- Cunha, M. C. C. (2000). *Métodos numéricos*. Unicamp, Campinas/SP.
- Dao, P. U., Heuzard, A. G., Le, T. X. H., Yin, R., Shang, C., e Fan, C. (2024). The impacts of climate change on groundwater quality: A review. *Sci Total Environ.*, 912:(art.nb.169241).
- Feitosa, F. A. C., M., F. J., Feitosa, E. C., e Demetrio, J. G. A. (2008). *Hidrogeologia: conceitos e aplicações*. CPRM.
- Ferracini, V. L., Pessoa, M. C. Y. P., Silva, A. S., e Spadotto, C. A. (2001). Análise de risco de contaminação das águas subterrâneas e superficiais da região de Petrolina (PE) e Juazeiro (BA). *Revista de Ecotoxicologia e Meio Ambiente*, 11:1–106.
- Fetter, C. W. (2001). *Applied Hydrogeology*. Prentice Hall, N.York – USA.
- Firmiano, A. e Wendland, E. (2009). Solução da equação de fluxo subterrâneo a partir de estimador de erro a posteriori. *Trends in Computational and Applied Mathematics*, 10(1):11–20.
- Foster, S. S. D. e Chilton, P. J. (2003). Groundwater: the processes and global significance of aquifer degradation. *Philosophical Transactions of the Royal Society B*, 358:1957–1972.

- Freeze, A. e Cherry, J. (1979). *Água Subterrânea*. Prentice-Hall, N.York – USA.
- Hirata, R., Suhogusoff, A. V., Marcellini, S. S., Villar, P. C., e Marcellini, L. (2019a). *As águas subterrâneas e sua importância ambiental e socioeconômica para o Brasil*. IGc/USP, S.Paulo/SP.
- Hirata, R., Suhogusoff, A. V., Marcellini, S. S., Villar, P. C., e Marcellini, L. (2019b). *A revolução silenciosa das águas subterrâneas no Brasil: uma análise da importância do recurso e os riscos pela falta de saneamento*. Instituto Trata Brasil, S.Paulo/SP.
- Kozhevnikov, E. V., Turbakov, M. S., Riabokon, E. P., e Poplygin, V. V. (2021). Effect of effective pressure on the permeability of rocks based on well testing results. *Energies*, 14(8):1–20.
- Martínez, D. M. (2023). *Modelación matemática de la contaminación de agua subterránea por hidrocarburos: aplicación a una estación de servicio en Al-daia (Valencia)*. Universitat Politècnica de València, València – Espanha.
- Mustafa, J. S. e Mawlood, D. K. (2024). Mathematical modelling for groundwater management for multilayers aquifers (Erbil basin). *Ain Shams Engineering Journal*, 15:(art.nb.102782).
- Otálvaro, M. V. V. (1999). *Hidráulica de aguas subterráneas*. Universidad Nacional de Colombia Sede Medellín. Facultad de Minas. Escuela de Geociencias y Medio Ambiente, Medellín – Colombia.
- Pinheiro, D., Monte, C., e Saldanha, E. C. (2019). Índice de estado trófico e a proveniência do fósforo e clorofila-a em diferentes estações do ano em uma microbacia amazônica. *Revista Ibero Americana de Ciências Ambientais*, 10:89–100.
- Reynalds, J. H. (2007). A review of the effects of sewer leakage on groundwater quality. *Water and Environment Journal*, 17(1):34–39.
- Ribeiro, M. L., Lourencetti, C., Pereira, S. Y., e Marchi, M. R. R. (2007). Contaminação de águas subterrâneas por pesticidas: avaliação preliminar. *Quim. Nova*, 30:688–694.

- Rocha, G., Fernandes, A. J., Mancuso, M. A., Campos, J. E., Takahashi, A. T., Oda, G. H., Perrotta, M. M., Salvador, E. D., Albuquerque F , J. L., Gimenez F^o, A., Azevedo, S. G., Denser, M., Pizzatto, L. G., e Paulon, N. (2006). Mapa de Águas subterrâneas do estado de São Paulo. *Águas Subterrâneas*, 14(8):1–20.
- Wendland, E. (2003). *Modelos Matemáticos e Métodos Numéricos em Águas Subterrâneas*. EESC-USP, S.Carlos/SP.
- Woessner, W. W. e Poeter, E. P. (2020). *Hydrogeologic Properties of Earth Materials and Principles of Groundwater Flow*. Guelph, Ontario, Canadá: The groundwater project.
- Zhang, M. (2019). Challenges of solving the problem of soil and groundwater contamination. *Synthesiology English edition*, 12(1):41–50.

Influência dos diferentes métodos de defuzzificação nos resultados do estudo da qualidade da água usando Python

Jônathas D. S. Oliveira¹, Vinícius S. Oliveira²,
DEE, CEFET-MG – 30.510-000, Belo Horizonte/MG.

Resumo. Este artigo investiga a influência dos diferentes métodos de defuzzificação nos resultados do estudo da qualidade da água usando sistemas baseados em regras fuzzy. O objetivo principal é comparar o impacto dos métodos de centroide, média dos máximos, menor dos máximos, bissetriz e maior dos máximos na determinação da qualidade da água, utilizando como variável de saída a potabilidade da mesma. Para isso, será implementado um sistema fuzzy em Python (biblioteca *scikit-fuzzy*), que utiliza cor aparente, turbidez e pH como variáveis de entrada para avaliar a potabilidade. Os resultados obtidos com os diferentes métodos de defuzzificação serão comparados com dados de um artigo de referência que empregou apenas o método do centroide via MATLAB^R, permitindo uma análise aprofundada das variações e implicações para a precisão e robustez desses sistemas na avaliação da qualidade hídrica.

Palavras-chave: Potabilidade; *scikit-fuzzy*; base de regras.

1. Introdução

A lógica fuzzy tem se consolidado como uma ferramenta essencial para o tratamento de incertezas e imprecisões na gestão e avaliação de recursos hídricos, com diversas aplicações no Brasil. No campo da avaliação da qualidade da água, pesquisadores têm proposto índices que superam as limitações dos métodos tradicionais. O estudo “Nova abordagem do índice de qualidade de água bruta utilizando a lógica fuzzy” (Oliveira et al., 2014) propõe um índice de qualidade de água bruta fuzzy (IQABF) mais restritivo e consistente. De forma similar, a pesquisa “Análise qualitativa de água subterrânea utilizando sistema de inferência fuzzy” (Teixeira, 2022) desenvolveu o índice de qualidade de água subterrânea fuzzy (IQASF) para interpretar a qualidade da água de poços a partir de parâmetros físicos, químicos e bacteriológicos. Para ambientes lóticos, o trabalho “Índice fuzzy de qualidade de água para ambiente lótico - IQAFAL” (Pessoa et al., 2020) introduz um índice que demonstra maior sensibilidade na detecção de alterações, mitigando o “efeito eclipse” (um fenômeno em que parâmetros favoráveis podem mascarar o impacto de um indicador crítico).

A aplicabilidade da lógica fuzzy, contudo, transcende a avaliação da qualidade, estendendo-se à análise de riscos hidrológicos. O artigo “Abordagem fuzzy para caracterização de incertezas estruturais e quantificação de riscos hidrológicos: estudo de caso do reservatório Orós no semiárido brasileiro” (Raulino et al., 2022) ilustra como números fuzzy triangulares (NFTs) podem ser utilizados para caracterizar incertezas e quantificar os riscos associados à gestão de sistemas hídricos complexos. A versatilidade da metodologia é ainda evidenciada em setores como a aquicultura, conforme demonstrado no estudo “Sistema baseado na lógica fuzzy para diagnóstico da qualidade da água para o cultivo de tilápia-do-Nilo” (Castro Jr. et al., 2022), que desenvolve um sistema de apoio à decisão para correlacionar parâmetros da água à produtividade do cultivo, capturando pequenas variações sem alterações bruscas na classificação final.

¹jonathasoliveira@cefetmg.br

²viniciussilas85@gmail.com

Visando a esse contexto, o presente trabalho tem como objetivo analisar a influência dos diferentes métodos de defuzzificação (centroide, média dos máximos, menor dos máximos, bissetriz e maior dos máximos) na avaliação da qualidade da água em sistemas baseados em regras fuzzy. Para tanto, será implementado um sistema em Python, com a biblioteca *scikit-fuzzy*, empregando como variáveis de entrada a cor aparente, a turbidez e o pH, a partir dos dados utilizados por Neves et al. (2019). Dessa forma, busca-se verificar como a escolha do método impacta os resultados obtidos e, conseqüentemente, a determinação da potabilidade.

Enquanto o trabalho de Neves et al. (2019) concentrou-se na utilização de um sistema baseado em regras fuzzy para avaliação da qualidade da água, adotando exclusivamente o método do centroide na defuzzificação e implementado em MATLAB^R, o presente estudo amplia essa abordagem ao investigar comparativamente diferentes métodos em ambiente Python. Essa opção não apenas favorece a análise das variações decorrentes da escolha do método de defuzzificação, mas também se justifica por tornar a metodologia mais acessível e reproduzível, uma vez que Python é uma linguagem livre e amplamente utilizada na comunidade científica, ao contrário do MATLAB^R, que requer licença proprietária. Assim, o trabalho contribui tanto para a compreensão da robustez e precisão dos sistemas fuzzy quanto para a difusão de ferramentas abertas no estudo da qualidade hídrica.

Para a implementação do sistema fuzzy, será utilizada a biblioteca *scikit-fuzzy*, desenvolvida para a linguagem Python e disponibilizada como software livre. Essa biblioteca fornece recursos para a modelagem completa de sistemas fuzzy, abrangendo desde a definição de variáveis linguísticas e funções de pertinência até a aplicação de métodos de inferência e defuzzificação. Nesse sentido, o *scikit-fuzzy* desempenha papel análogo ao da Fuzzy Logic Toolbox do MATLAB^R, com a vantagem de estar acessível de forma gratuita e aberta, o que facilita a reprodutibilidade dos experimentos e amplia o alcance das aplicações acadêmicas e profissionais. Mais informações sobre as funcionalidades da biblioteca podem ser encontradas em Zavaleta (2020).

2. Sistemas baseados em regras fuzzy

Um sistema baseado em regras fuzzy, ou SBRF, pode ser definido como um sistema que emprega a teoria dos conjuntos fuzzy para transformar entradas numéricas em saídas correspondentes. O desenvolvimento de tais sistemas foi motivado pela tentativa de simular o raciocínio humano que, frequentemente, se baseia em um conjunto de regras linguísticas para a tomada de decisões (Jafelice et al., 2023). A estrutura de um SBRF é composta por quatro elementos principais, como ilustrado na figura 1.

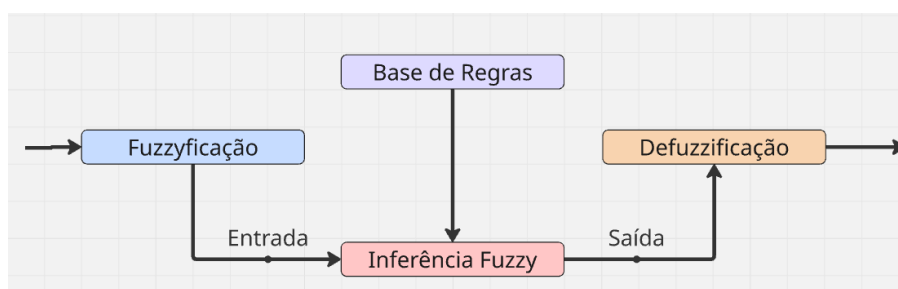


Figura 1: Diagrama da estrutura de um sistema baseado em regras fuzzy (SBRF)

O primeiro componente é o módulo de fuzzificação, que realiza a conversão dos dados de entrada numéricos (crisp) em conjuntos fuzzy, utilizando funções de pertinência para atribuir graus de associação a termos linguísticos. A base de regras, por sua vez, constitui o núcleo do sistema, armazenando o conhecimento especialista na forma de um conjunto de proposições “Se-Então” que relacionam as variáveis de entrada e saída.

O motor de inferência fuzzy é o componente que processa as regras. Ele combina os dados fuzzificados da entrada com as regras da base de conhecimento para derivar uma conclusão, que é expressa como um conjunto fuzzy de saída. Finalmente, o módulo de defuzzificação converte este conjunto fuzzy de saída em um valor numérico (crisp) e acionável. O método de inferência de Mamdani é um dos mais utilizados para implementar essa arquitetura (Jafelice et al., 2023, p. 36). A figura 2 ilustra as etapas centrais deste processo. Para cada regra, os graus de pertinência das entradas são combinados através de operadores fuzzy (tipicamente o mínimo para o operador E e o máximo para o OU) para determinar um único valor de ativação. Este valor é então utilizado na etapa de implicação para truncar o conjunto fuzzy consequente da regra. Por fim, na etapa de agregação, todos os conjuntos fuzzy de saída resultantes da implicação são combinados através do operador de união (máximo), formando o resultado agregado, como mostrado na figura 3.

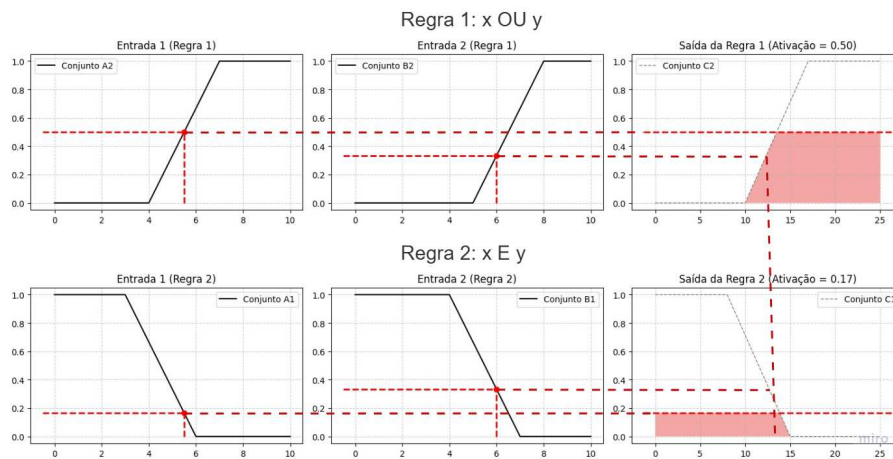


Figura 2: Processo de inferência para a regra 1 (superior) e regra 2 (inferior), mostrando a fuzzificação das entradas e a implicação de Mamdani sobre os conjuntos de saída.



Figura 3: Conjunto fuzzy de saída agregado, formado pela união (operador máximo) das saídas individuais das regras.

3. Métodos de defuzzificação

A etapa de defuzzificação é um componente crucial em um sistema baseado em regras fuzzy (SBRF), pois é responsável por converter o conjunto fuzzy de saída, que é o resultado da etapa de

agregação, em um valor numérico (crisp) que pode ser utilizado para a tomada de decisão ou ação de controle. Trata-se do processo inverso à fuzzificação, traduzindo os valores fuzzy de saída em um único valor nítido e acionável (Mazraeh, 2025). A escolha do método de defuzzificação pode influenciar significativamente o resultado final do sistema. Nesse estudo, foram analisados cinco métodos distintos, cujas definições teóricas e formulações são detalhadas a seguir e podem ser encontradas em Ross (2010).

3.1. Método do centroide

O método do centroide, também conhecido como centro de gravidade, é definido como o cálculo do ponto de equilíbrio da área sob a curva da função de pertinência agregada. De forma análoga ao centro de massa de um objeto físico, ele representa a média de todos os pontos do universo de discurso, ponderada por seu respectivo grau de pertinência. Por levar em consideração a forma completa e a distribuição do conjunto fuzzy de saída, é considerado o método mais preciso e robusto da literatura. Sua principal contrapartida, no entanto, é a maior intensidade computacional em comparação com métodos mais simples, especialmente para conjuntos de saída com formas complexas. As equações que descrevem suas formas contínua e discreta são apresentadas em Ross (2010).

Equação contínua:

$$x_{centroide}^* = \frac{\int_X x \cdot \mu(x) dx}{\int_X \mu(x) dx} \quad (3.1)$$

A equação (3.1) é apresentada em (Ross, 2010, p. 99, Eq. 4.5). Nesta equação, $x_{centroide}^*$ é o valor nítido (crisp) de saída, X representa todo o universo de discurso, x é a variável de saída, $\mu(x)$ é a função de pertinência do conjunto fuzzy agregado e \int_X denota a integral sobre todo o universo de discurso.

Equação discreta:

$$x_{centroide}^* = \frac{\sum_{i=1}^n x_i \cdot \mu(x_i)}{\sum_{i=1}^n \mu(x_i)} \quad (3.2)$$

Nesta equação, que é a aproximação discreta da forma contínua, x_i é o i -ésimo elemento do universo de discurso, $\mu(x_i)$ é o seu grau de pertinência e \sum denota o somatório sobre todos os elementos. A figura 4 ilustra o método.

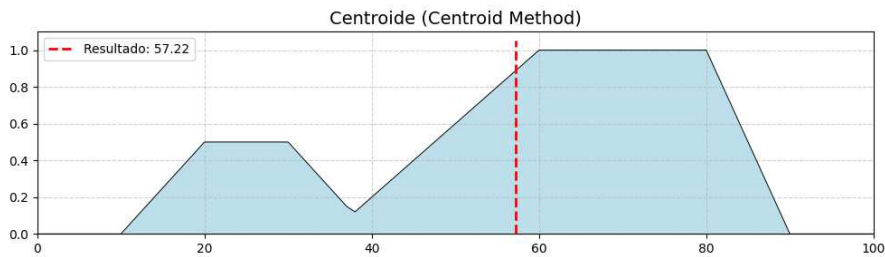


Figura 4: Ilustração da operação do método do centróide

3.2. Método da bissetriz da área

O método da bissetriz, por sua vez, adota uma abordagem puramente geométrica, sendo definido como o procedimento que encontra o ponto no universo de discurso que divide a área do conjunto fuzzy agregado em duas metades de tamanho exatamente igual. Diferentemente do centroide, seu foco é exclusivo na distribuição da área, o que o torna uma alternativa robusta que também considera todo o conjunto de saída. Contudo, de forma similar ao centroide, pode ser computacionalmente mais oneroso que os métodos baseados em máximos, e seu resultado pode ser menos intuitivo para formas muito assimétricas.

Equação contínua: Busca-se $x_{bisetor}^*$ tal que:

$$\int_a^{x_{bisetor}^*} \mu(x) dx = \int_{x_{bisetor}^*}^b \mu(x) dx \quad (3.3)$$

Nesta equação, $[a, b]$ são os limites inferior e superior do universo de discurso X . A figura 5 ilustra o método.

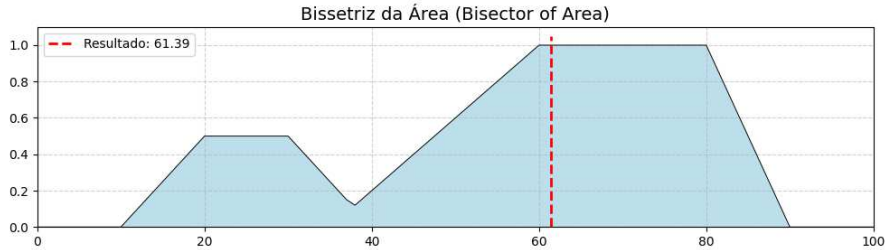


Figura 5: Ilustração da operação do método da bisetriz.

Equação discreta: Encontra-se o índice k que satisfaz a equação (3.4).

$$\sum_{i=1}^k \mu(x_i) \approx \sum_{i=k+1}^n \mu(x_i) \quad (3.4)$$

3.3. Método da média dos máximos

Visando a uma maior eficiência computacional, o método da média dos máximos (MOM) é definido por um processo que simplifica a análise, focando apenas nos resultados de máxima pertinência. O cálculo padrão consiste em identificar o conjunto de pontos com o maior grau de pertinência e calcular sua média aritmética. Embora essa abordagem garanta grande simplicidade e rapidez, ela o faz ao custo de ignorar a forma geral do conjunto fuzzy, o que pode levar a uma perda de informação e a resultados menos precisos.

A complexidade do método surge em um caso particular: quando o grau de pertinência máximo ocorre em múltiplos intervalos não contíguos (separados), resultando em mais de um platô (segmento horizontal de pertinência máxima). Nessa situação, uma simples média entre os pontos centrais de cada platô seria matematicamente incorreta se estes tivessem comprimentos diferentes, pois o cálculo não ponderaria a influência de cada um. A figura 7 ilustra esse cenário, onde o valor defuzzificado é visivelmente deslocado em direção ao platô de maior comprimento.

Para garantir essa precisão, deve-se implementar a equação contínua do método, que funciona como uma média ponderada de todos os pontos pertencentes a esses múltiplos intervalos, representando o verdadeiro centroide do domínio de máxima pertinência. As equações que definem este método são apresentadas em Ross (2010).

Equação contínua: Seja $M = \{x \in X \mid \mu(x) = \max(\mu)\}$.

$$x_{mom}^* = \frac{\int_M x dx}{\int_M dx} \quad (3.5)$$

Nesta equação, M é o conjunto de todos os pontos x com o grau de pertinência máximo e $\max(\mu)$ é o supremo (o valor máximo) da função de pertinência.

Equação discreta: Para o caso de um platô entre os pontos a e b :

$$x_{mom}^* = \frac{a + b}{2} \quad (3.6)$$

A equação (3.6) é apresentada em (Ross, 2010, p. 100, Eq. 4.7).

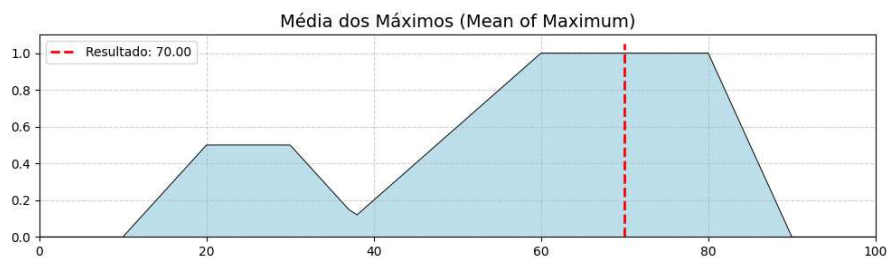


Figura 6: Ilustração da operação do método da média dos máximos

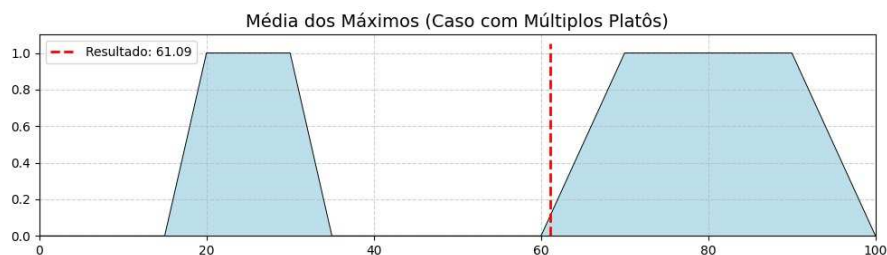


Figura 7: Caso especial do método MOM com múltiplos platôs

3.4. Método do menor dos máximos

Como uma variação do MOM, o método do menor dos máximos (SOM) é definido pela seleção do menor valor (o mais à esquerda no universo de discurso) pertencente ao conjunto de pontos com o maior grau de pertinência. Sua utilidade se destaca em cenários que exigem uma decisão conservadora ou de “pior caso”, pois representa o limite inferior dos resultados mais plausíveis.

Essa simplicidade, no entanto, o torna um método muito sensível, que baseia sua decisão em um único ponto extremo, ignorando quase toda a informação do conjunto fuzzy. A equação para este método pode ser encontrada em Ross (2010) e a figura 8 traz uma interpretação geométrica para esse método.

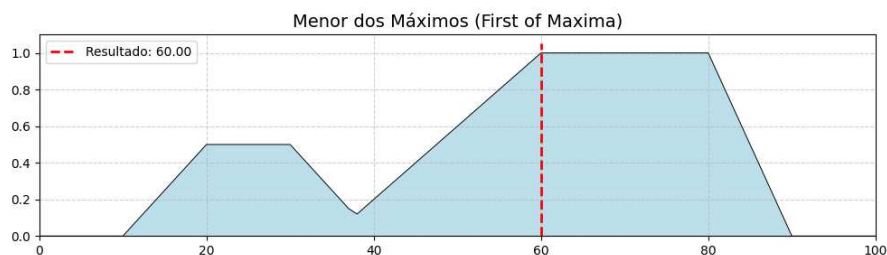


Figura 8: Ilustração da operação do método do menor dos máximos

Equação contínua:

$$x_{som}^* = \min(M) \quad (3.7)$$

Nesta equação, $\min(M)$ é o menor valor do conjunto M de máximos.

A equação (3.7) é apresentada em (Ross, 2010, p. 107, Eq. 4.11).

3.5. Método do maior dos máximos

De forma análoga ao SOM, o método do maior dos máximos (LOM) é definido pela seleção do maior valor (o mais à direita no universo de discurso) do conjunto de pontos com pertinência máxima. Ele

é igualmente rápido e útil para cenários otimistas, onde se deseja obter o limite superior dos resultados mais prováveis. Contudo, compartilha da mesma desvantagem de ser altamente sensível e desconsiderar a forma geral do conjunto de saída. A equação para este método também é apresentada em Ross (2010) e a figura 9 ilustra o método.

Equação contínua:

$$x_{lom}^* = \max(M) \quad (3.8)$$

Nesta equação, $\max(M)$ é o maior valor do conjunto M de máximos.

A equação (3.8) é apresentada em (Ross, 2010, p. 107, Eq. 4.12).



Figura 9: Ilustração da operação do método do maior dos máximos

4. O problema da qualidade da água

Para analisar a qualidade da água, este estudo considera três aspectos principais de potabilidade, seguindo metodologias apresentadas por Jafelice et al. (2023) e Neves et al. (2019). As informações de Jafelice et al. (2023) foram obtidas a partir de dados fornecidos pela SABESP (Companhia de Saneamento Básico do Estado de São Paulo), que regulamenta e fiscaliza a qualidade da água para consumo humano no estado de São Paulo. Já Neves et al. (2019) utilizou dados disponibilizados pela empresa Águas Guariroba, com base em políticas de qualidade que garantem a potabilidade da água, acessíveis em <http://www.aguasguariroba.com.br>.

As variáveis de entrada escolhidas para o sistema baseado em regras fuzzy (SBRF) são: cor aparente, medida em UH (Unidade Hazen); pH, representando a concentração de íons de hidrogênio, com valores variando de 0 a 14; e turbidez, causada pela presença de substâncias suspensas e coloidais, determinada pela quantidade de luz dispersada ao atravessar a amostra e medida em UT (unidade de turbidez). Além dessas variáveis, outras poderiam ser incluídas na análise, como odor, sabor, nível de flúor, cloro residual e quantidade de coliformes fecais e totais.

A variável de saída do sistema é a potabilidade da água, definida linguisticamente com os termos boa, adequada e inadequada para consumo. Essa estrutura permite avaliar a qualidade da água de forma flexível, considerando diferentes níveis de variação das variáveis de entrada e apoiando decisões relacionadas ao consumo seguro e à gestão hídrica.

Baseado nos dados apresentados em Neves et al. (2019), as variáveis de entrada do sistema fuzzy são **cor aparente**, **pH** e **turbidez**, cujas funções de pertinência são do tipo trapezoidal (números fuzzy trapezoidais (Barros et al., 2017)). As funções trapezoidais são definidas como:

$$\text{Trap}(x; a, m, n, b) = \max \left\{ 0, \min \left(\frac{x - a}{m - a}, 1, \frac{b - x}{b - n} \right) \right\}.$$

Para a **cor aparente**, os conjuntos fuzzy são:

$$\text{boa} : \text{Trap}(x; -1, 5, 0, 5)$$

$$\text{adequada} : \text{Trap}(x; 5, 5, 10, 15)$$

inadequada : $\text{Trap}(x; 15, 15, 20, 25)$.

Tabela 1: Funções de pertinência para a variável cor aparente.

Rótulo linguístico	Parâmetros trapezoidais [a, m, n, b]
Boa	[-6, 3, 3, 6]
Adequada	[3, 5, 15, 17]
Inadequada	[13, 17, 30, 34]

Para o **pH**, os conjuntos são:

inadequado baixo : $\text{Trap}(x; -6, -5, 5, 6)$;

adequado : $\text{Trap}(x; 5, 6, 8, 9, 5)$;

bom : $\text{Trap}(x; 5, 5, 6, 5, 8, 8, 5)$;

inadequado alto : $\text{Trap}(x; 9, 9, 5, 14, 14, 5)$.

Tabela 2: Funções de pertinência para a variável pH.

Rótulo linguístico	Parâmetros trapezoidais [a, m, n, b]
Inadequado baixo	[-6, -5, 5, 6]
Adequado	[5, 6, 8.5, 9.5]
Bom	[5.5, 6.5, 8, 8.5]
Inadequado alto	[9, 9.5, 14, 14.5]

Para a **turbidez**, os conjuntos são:

boa : $\text{Trap}(x; -1, 5, -0, 5, 0, 5, 1, 5)$;

adequada : $\text{Trap}(x; 0, 5, 1, 5, 4, 5, 5, 5)$;

inadequada : $\text{Trap}(x; 4, 5, 5, 5, 10, 11)$.

Tabela 3: Funções de pertinência para a variável turbidez.

Rótulo linguístico	Parâmetros trapezoidais [a, m, n, b]
Boa	[-1.5, -0.5, 0.5, 1.5]
Adequada	[0.5, 1.5, 4.5, 5.5]
Inadequada	[4.5, 5.5, 10, 11]

A variável de saída, **potabilidade da água**, também é representada por conjuntos trapezoidais:

inadequada: $\text{Trap}(x; -0,5, -0,4, 0,4, 0,5)$;

adequada: $\text{Trap}(x; 0,4, 0,5, 0,7, 0,8)$;

boa: $\text{Trap}(x; 0,7, 0,8, 1,2, 1,3)$.

Tabela 4: Funções de pertinência para a variável de saída potabilidade.

Rótulo linguístico	Parâmetros trapezoidais [a, m, n, b]
Inadequada	[-0.5, -0.4, 0.4, 0.5]
Adequada	[0.4, 0.5, 0.7, 0.8]
Boa	[0.7, 0.8, 1.2, 1.3]

4.1. Construção das regras fuzzy

A etapa de definição das regras fuzzy é fundamental para o funcionamento do sistema, pois determina como as variáveis de entrada se relacionam para produzir a avaliação da potabilidade da água.

Neste estudo, as regras foram estabelecidas com base em duas variáveis de entrada: **pH (H)** e **turbidez (T)**, considerando os diferentes níveis da **cor aparente**. Dessa forma, para cada valor linguístico da cor aparente (boa, adequada ou inadequada), foi construída uma matriz de decisão relacionando as categorias linguísticas de pH e turbidez à variável de saída (potabilidade da água).

As regras, retiradas de Neves et al. (2019) foram elaboradas a partir das informações fornecidas pela empresa Águas Guariroba, da consulta a um especialista e adaptadas de Jafelice et al. (2023).

A seguir, são apresentadas as tabelas que sintetizam esse conjunto de regras:

Tabela 5: Regras fuzzy quando a cor aparente é **boa**.

pH (H)	Turbidez (T)		
	Boa	Adequada	Inadequada
Inadequado Baixo	Inadequada	Inadequada	Inadequada
Adequado	Adequada	Adequada	Inadequada
Bom	Boa	Boa	Inadequada
Inadequado Alto	Inadequada	Inadequada	Inadequada

Tabela 6: Regras fuzzy quando a cor aparente é **adequada**.

pH (H)	Turbidez (T)		
	Boa	Adequada	Inadequada
Inadequado Baixo	Inadequada	Inadequada	Inadequada
Adequado	Adequada	Adequada	Inadequada
Bom	Boa	Adequada	Inadequada
Inadequado Alto	Inadequada	Inadequada	Inadequada

Tabela 7: Regras fuzzy quando a cor aparente é **inadequada**.

pH (H)	Turbidez (T)		
	Boa	Adequada	Inadequada
Inadequado Baixo	Inadequada	Inadequada	Inadequada
Adequado	Inadequada	Inadequada	Inadequada
Bom	Adequada	Adequada	Inadequada
Inadequado Alto	Inadequada	Inadequada	Inadequada

Assim, as tabelas acima representam a base de regras fuzzy do sistema, permitindo que, a partir de diferentes combinações de valores linguísticos das variáveis de entrada, seja inferido o grau de potabilidade

da água.

4.2. Processo de inferência fuzzy

Uma vez estabelecida a base de regras, o sistema fuzzy utiliza um processo de inferência para mapear os valores de entrada (cor, pH e turbidez) para um valor de saída (potabilidade). Esse processo, baseado no método de Mamdani, ocorre em duas etapas principais: a ativação das regras e a agregação dos resultados (Neves et al., 2019).

4.2.1. Ativação das regras

Para um conjunto de medições de entrada, o sistema primeiro determina o grau de pertinência de cada medição aos conjuntos fuzzy definidos (ex: o quanto o valor do pH pertence ao conjunto “bom”). Em seguida, para cada uma das 36 regras da base de conhecimento, o grau de ativação (R_i) é calculado.

Esse valor, formalmente conhecido como o grau de ativação da regra (que determina a influência que sua conclusão terá no resultado final), é determinado pelo operador **mínimo** (\wedge) entre os graus de pertinência das três condições de entrada daquela regra, conforme a equação (4.9) (Neves et al., 2019):

$$R_i = \mu_{Cor}(x) \wedge \mu_{pH}(y) \wedge \mu_{Turbidez}(z) \quad (4.9)$$

sendo que i varia de 1 a 36, representando cada regra do sistema.

4.2.2. Agregação dos resultados

Após o cálculo da força de cada regra, os resultados de todas as regras ativadas são combinados para formar um único conjunto fuzzy de saída para a “potabilidade”. Este processo é chamado de agregação. O conjunto fuzzy consequente de cada regra (ex: “potabilidade é inadequada”) é “truncado” ou “cortado” na altura correspondente ao grau de ativação (R_i) daquela regra. Finalmente, todos esses conjuntos fuzzy truncados são unificados através do operador **máximo** (\cup), formando a figura geométrica final do conjunto de saída. A representação matemática desse processo é abordada no artigo Neves et al. (2019):

$$\mu_{Potabilidade} = \bigcup_{i=1}^{36} (R_i \wedge \mu_{Q_i}) \quad (4.10)$$

em que μ_{Q_i} é a função de pertinência do conjunto consequente da regra i (boa, adequada ou inadequada).

4.2.3. Exemplo de cálculo

Para ilustrar o processo, podemos utilizar o primeiro conjunto de dados da tabela 8: Cor aparente = 2,16, Turbidez = 8,25 e pH = 8,73. O sistema executa os seguintes passos:

1. **Fuzzificação:** O sistema calcula o grau de pertinência de cada entrada aos seus respectivos conjuntos fuzzy. Por exemplo, um pH de 8,73 terá um alto grau de pertinência ao conjunto “adequado” e um grau menor ao conjunto “bom”.
2. **Ativação das Regras:** Apenas as regras cujas condições de entrada correspondem a esses graus de pertinência serão ativadas. Por exemplo, uma regra que exige “pH é inadequado baixo” terá uma ativação de zero, enquanto uma regra que exige “Cor é boa”, “pH é adequado” e “Turbidez é inadequada” será ativada com uma força R_i igual ao valor **mínimo** entre os três graus de pertinência.

3. **Agregação:** As saídas das regras ativadas são agregadas usando o operador **máximo**. Isso resulta em um conjunto fuzzy final para a variável “Potabilidade”, cuja forma geométrica representa a conclusão combinada de todas as regras relevantes.
4. **Defuzzificação:** A este conjunto fuzzy agregado, aplica-se um dos métodos de defuzzificação. Por exemplo, ao aplicar o método do **centroide**, conforme a equação (3.2), o sistema calcula o centro de gravidade dessa figura, resultando no valor numérico final de **0,228**, que classifica a água como “Inadequada”, conforme apresentado na tabela 8.

4.3. Discussão e resultados

A base de regras fuzzy foi construída utilizando a biblioteca *scikit-fuzzy* em Python. Para a análise da água, o sistema de inferência fuzzy desenvolvido processou dez conjuntos de medições fornecidas pela empresa Águas Gabiroba, extraídas da referência Neves et al. (2019), calculando o nível de potabilidade. O objetivo principal deste estudo é realizar uma análise comparativa dos valores nítidos gerados por cinco métodos de defuzzificação distintos: centroide, bissetriz, média dos máximos (MOM), menor dos máximos (SOM) e maior dos máximos (LOM). Os dados consolidados na tabela a seguir servem como base para avaliar o impacto de cada método no resultado final da classificação da água, evidenciando as diferentes abordagens para converter o conjunto fuzzy agregado em um valor final interpretável para a tomada de decisão no mundo real.

Tabela 8: Tabela comparativa dos resultados de potabilidade.

Entradas			Saída - Potabilidade											
Cor apar.	Turb.	pH	Método de defuzzificação										Ref.do artigo	
			Centroide		Bissetriz		MOM		SOM		LOM		Valor	Class.
			Valor	Class.	Valor	Class.	Valor	Class.	Valor	Class.	Valor	Class.		
2,16	8,25	8,73	0,228	Inad.	0,227	Inad.	0,200	Inad.	0,000	Inad.	0,400	Inad.	0,223	Inad.
15	7	5	0,387	Inad.	0,388	Inad.	0,375	Inad.	0,000	Inad.	0,750	Boa	0,385	Inad.
3,96	7	5	0,500	Adq.	0,500	Adq.	0,500	Adq.	0,000	Inad.	1,000	Boa	0,500	Adq.
3,96	5,68	5	0,432	Inad.	0,425	Inad.	0,379	Inad.	0,000	Inad.	0,750	Boa	0,430	Inad.
4,4	5,89	5,95	0,233	Inad.	0,233	Inad.	0,215	Inad.	0,000	Inad.	0,430	Inad.	0,223	Inad.
12,5	6,44	5,15	0,336	Inad.	0,315	Inad.	0,215	Inad.	0,000	Inad.	0,430	Inad.	0,334	Inad.
5,3	6,72	2,06	0,642	Adq.	0,625	Adq.	0,595	Adq.	0,500	Adq.	0,690	Adq.	0,643	Adq.
2,46	7,91	1,87	0,722	Boa	0,714	Boa	0,744	Boa	0,494	Inad.	1,000	Boa	0,726	Boa
2,91	10,7	9,08	0,226	Inad.	0,225	Inad.	0,205	Inad.	0,000	Inad.	0,401	Inad.	0,223	Inad.
15	7	0	0,576	Adq.	0,606	Adq.	0,751	Boa	0,500	Adq.	1,000	Boa	0,500	Adq.

Cada linha da tabela acima representa uma amostra distinta com valores de entrada específicos para cor aparente, turbidez e pH. As colunas exibem o valor de saída para a potabilidade, calculado por cada um dos métodos analisados, além do valor de referência extraído do artigo base deste estudo. Uma primeira análise dos resultados agregados na tabela permite observar padrões de comportamento distintos para cada método de defuzzificação, que se alinham às suas respectivas definições teóricas.

Os métodos centróide e bissetriz apresentaram resultados notavelmente estáveis e consistentes entre si ao longo de todas as 10 aferições. O método centróide, que calcula o centro de gravidade do conjunto fuzzy de saída, gerou valores praticamente idênticos ao da referência que utilizou o mesmo método por meio do software MATLAB^R. A bissetriz, que encontra o ponto que divide a área do conjunto em duas metades iguais, também produziu resultados bem próximos, indicando que a forma do conjunto de saída agregado foi, na maioria dos casos, relativamente simétrica.

Diferentemente dos métodos de área mencionados anteriormente, MOM, SOM e LOM focam exclusivamente nos pontos do universo de discurso que possuem o maior grau de pertinência. Essa

abordagem em grande parte (com exceção do método MOM) resultou em uma variação de saída mais ampla e revelou informações importantes sobre a certeza do sistema. O menor dos máximos (SOM), que retorna o menor valor com pertinência máxima, estabelece o limite inferior do intervalo de resultados mais plausíveis. Em sete das dez aferições, seu valor foi 0 ou próximo de 0.5, alinhando-se à classificação de potabilidade “inadequada”. Em contrapartida, o maior dos máximos (LOM), que retorna o maior valor, representa o limite superior desse mesmo intervalo. Em três dos casos, o valor foi 1, correspondendo a uma potabilidade “boa”. A média dos máximos (MOM), por sua vez, calculou o valor médio desses pontos de ativação máxima, obtendo resultados próximos aos métodos de área mencionados anteriormente.

A análise se torna ainda mais crítica quando se observa o impacto desses valores na classificação qualitativa final da água (inadequada, adequada ou boa). Enquanto os métodos centróide, bissetriz e MOM se mostraram consistentes, mantendo a mesma classificação para a maioria das amostras, a verdadeira divergência ocorre com os métodos SOM e LOM. Em 5 dos 10 resultados, a escolha entre o limite inferior (SOM) e o limite superior (LOM) do intervalo de máxima pertinência resultou em uma mudança direta na classificação da potabilidade. Por exemplo, para a amostra 3, o método SOM produziu um valor que se enquadra na faixa “inadequada”, enquanto o LOM, para as mesmas entradas, gerou um resultado classificado como “adequado”.

Esta observação é fundamental, pois demonstra que quando múltiplas regras com conclusões divergentes são ativadas, a seleção do método de defuzzificação não é apenas um detalhe técnico, mas uma decisão que pode alterar drasticamente a ação a ser tomada, como emitir um alerta de contaminação ou aprovar a qualidade da água.

A análise comparativa dos métodos de defuzzificação confirma a robustez e a estabilidade dos métodos baseados em área (centroide e bissetriz) e da média dos máximos (MOM), que se mostraram os mais adequados para a classificação geral da potabilidade da água neste sistema. Dentre estes, a adequação do método centroide é particularmente reforçada por sua semelhança aos valores de referência do artigo.

5. Considerações finais

Este estudo permitiu avaliar de forma comparativa a influência de diferentes métodos de defuzzificação – centroide, bissetriz, média dos máximos, menor dos máximos e maior dos máximos – na classificação da potabilidade da água em sistemas baseados em regras fuzzy.

Observou-se que, quando comparados ao artigo de referência, os métodos baseados em área, como o centroide e a bissetriz, apresentaram maior estabilidade e consistência nos resultados, enquanto os métodos baseados em máximos evidenciaram variações mais acentuadas, podendo alterar significativamente a decisão final sobre a qualidade hídrica. Nesse contexto, a implementação em Python por meio da biblioteca *scikit-fuzzy* mostrou-se não apenas adequada para a análise, mas também uma alternativa acessível e reprodutível ao MATLAB^R, favorecendo a disseminação de estudos semelhantes e a continuidade das pesquisas na área.

Como trabalhos futuros, sugere-se a inclusão de novas variáveis de entrada, como cloro residual, coliformes ou metais pesados, de modo a enriquecer a análise da potabilidade. Além disso, a aplicação da metodologia a diferentes bases de dados e ambientes hídricos pode contribuir para validar os modelos.

Referências

- Barros, L. C., Bassanezi, R. C., e Lodwick, W. A. (2017). *A first course in fuzzy logic, fuzzy dynamical systems, and biomathematics*. Springer, N.York.

- Castro Jr., S. L., Lamarca, D. S. F., Kraetzer, T. L., Balthazar, G. R., e Caneppele, F. L. (2022). Sistema baseado na lógica fuzzy para diagnóstico da qualidade da água para o cultivo de tilápia-donilo. *Research, Society and Development*, 11(4):(art.nb.e3211426933).
- Jafelice, R. S. M., Barros, L. C., e Bassanezi, R. C. (2023). *Teoria dos Conjuntos Fuzzy com Aplicações*, volume 17 de *Notas em Matemática Aplicada*. SBMAC, S.Carlos/SP, 3 edição.
- Mazraeh, A. (2025). Week 3: Fuzzification and defuzzification. Medium. Disponível em: <https://medium.com/@adnan.mazraeh1993/fuzzification-and-defuzzification-6fbfe059047e>.
- Neves, E. P., Duarte, M. A. Q., e Alvarado, F. V. (2019). Sistema baseado em regras fuzzy para avaliação da qualidade da água. *CQD-Revista Eletrônica Paulista de Matemática*, 14:95–109.
- Oliveira, M. D., Rezende, O. L. T., Oliveira, S. M. A. C., e Libânio, M. (2014). Nova abordagem do índice de qualidade de água bruta utilizando a lógica fuzzy. *Engenharia Sanitária e Ambiental*, 19(4):361–372.
- Pessoa, M. A. R., Souza, F. J., Domingos, P., e Azevedo, J. P. S. (2020). Índice fuzzy de qualidade de água para ambiente lótico-iqa fal. *Engenharia Sanitária e Ambiental*, 25:21–30.
- Raulino, J. B. S., Malveira, V. T. C., e Lima Neto, I. E. (2022). Abordagem fuzzy para caracterização de incertezas estruturais e quantificação de riscos hidrológicos: estudo de caso do reservatório orós no semiárido brasileiro. *Engenharia Sanitária e Ambiental*, 27(3):507–517.
- Ross, T. J. (2010). *Fuzzy Logic with Engineering Applications*. John Wiley & Sons, Ltd., 3rd edição.
- Teixeira, L. M. (2022). Análise qualitativa de água subterrânea utilizando sistema de inferência fuzzy. Dissertação de Mestrado, CTRN–UFCCG, Campina Grande/PB.
- Zavaleta, J. (2020). Modelagem e simulação de sistemas usando lógica fuzzy com Python. Disponível em: <https://controlautomaticoeducacion.com/python-para-ingenieros/logica-fuzzy-con-python/>.

