



# EAMC2023

XVI Encontro Acadêmico de Modelagem Computacional

Laboratório Nacional de Computação Científica - LNCC/MCTI

14 a 17 de fevereiro de 2023.

## Proceedings do evento



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÃO



## **XVI Encontro Acadêmico de Modelagem Computacional**

Laboratório Nacional de Computação Científica - LNCC/MCTI

14 a 17 de fevereiro de 2023. Petrópolis - RJ.

### **Comitê Organizador**

Ana Luiza Martins Karl - LNCC

Douglas Terra Machado - UFRJ

Lucas Menezes Silva - LNCC

Luis Alonso Mansilla Alvarez - LNCC

Mateus da Silva Melo - UFF

Matheus Müller Pereira da Silva - LNCC

Victor de Paula Dornellas Ribeiro - LNCC

### **Comitê Científico**

Caio César Graciani Rodrigues - SSM Napoli

Claudio Daniel Tenório de Barros - LNCC

Felipe Figueredo Rocha - EPFL

Gregório Kappaun Rocha - IFF

Guilherme Guilhermino Neto - IFES

Gustavo Barbosa Libotte - UFRJ

Karina Baptista dos Santos - LNCC

Lucas dos Anjos - USP

Lucas dos Santos Fernandez - USP

Luis Alonso Mansilla Alvarez - LNCC

Mariza Ferro - UFF

Otávio José Bernardes Burstolini - UFV

Pedro Carlos da Silva Lara - CEFET - RJ

Wesley da Silva Pereira - CU Denver

### **Apoio**

Laboratório Nacional de Computação Científica - LNCC

Ministério da Ciência, Tecnologia e Inovação - MCTI

Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ

## **XVI Encontro Acadêmico de Modelagem Computacional**

Laboratório Nacional de Computação Científica - LNCC/MCTI

14 a 17 de fevereiro de 2023. Petrópolis - RJ.

O Encontro Acadêmico de Modelagem Computacional do LNCC (EAMC/LNCC) é um evento dedicado à Modelagem Computacional, parte do Programa de Verão da mesma instituição, que busca promover maior integração entre os alunos, docentes, pesquisadores e demais profissionais da área de Ciência e Tecnologia. O EAMC é inteiramente organizado por discentes do LNCC e de outras instituições colaboradoras, sendo composto por minicursos, mesas redondas, palestras, além de apresentações de trabalhos em formato pôster e apresentações orais.

Criado em 2007, o EAMC tem como principal objetivo proporcionar um ambiente que promova a divulgação científica, estimule a cooperação entre os profissionais da área, além de fomentar a multidisciplinaridade na formação de novos profissionais. O EAMC abrange diferentes áreas do conhecimento, tais como Computação Científica, Controle e Filtragem de Sistemas Dinâmicos, Modelagem de Biosistemas e Bioinformática, Modelagem de Circulação e Transporte, Modelagem de Equilíbrio e Otimização e, mais recentemente, Ciência de Dados e áreas correlatas, como Inteligência Artificial.

A XVI edição do Encontro Acadêmico de Modelagem Computacional foi realizada no formato virtual desde as instalações do Laboratório Nacional de Computação Científica, em Petrópolis/RJ, durante os dias 14 a 17 de fevereiro de 2023. Em razão da pandemia do SARS-CoV-2, os minicursos e apresentações orais foram transmitidas de forma virtual visando a segurança de todos os participantes.

O comitê organizador gostaria de agradecer à comunidade do LNCC e instituições parceiras (representadas por pesquisadores, professores, discentes) pela participação e dedicação buscando melhorar o impacto e importância do EAMC a cada edição. Particularmente, o nosso agradecimento à organização do Programa de Verão do LNCC, professoras Sandra Malta e Kary Ocaña, os pesquisadores que aceitaram compor o comitê científico e os participantes dessa edição.

Comitê Organizador do XVI EAMC  
XVI Encontro Acadêmico de Modelagem Computacional  
Laboratório Nacional de Computação Científica - LNCC

## Trabalhos completos

	Página
<i>Algoritmo adaptativo para o cálculo numérico da derivada</i> Callebe R. Reis, Sandro D. P. Vitenti e Mariana Penna-Lima	6
<i>Modelagem de jornadas de pacientes como um grafo multiaspecto</i> Caroline de Oliveira Costa Souza Rosa, Márcia Ito, Alex Borges Vieira, Klaus Wehmuth e Antônio Tadeu Azevedo Gomes	15
<i>Predição in silico de compostos bioativos contra SARS-COV-2</i> Elcio Ferreira Santana e Adriano Carlos Soares	27
<i>Ataque Zumbi: a humanidade irá sobreviver?</i> Gabriel Vilela Sousa, Ulisses Ferreira Kaneko e Paulo Freitas Gomes	41
<i>Estudo e Construção de Modelo Digital de Elevação (MDE) nas Regiões de Habitat de Peixes Ameaçados de Extinção, no Noroeste Fluminense</i> Igor Martins Zanata, Vicente de Paulo Santos de Oliveira e Wagner Rambaldi Telles	53
<i>Modelos epidemiológicos baseados em redes: Uma revisão preliminar de escopo</i> Jesuliana Nascimento Ulysses e Antônio Tadeu Azevedo Gomes	66
<i>Multiscale Hybrid-Hybrid-Mixed Method</i> Franklin da Conceição de Barros, Alexandre Loureiro Madureira e Frédéric Gerard Christian Valentin	78
<i>Uso do Critério de Informação de Akaike para Estimativa das Ordens de Trajetórias de Ratos Reais e Virtuais no Labirinto em Cruz Elevado</i> Leonardo Rocha de Freitas Dias, Renato Tinós e Ariadne de Andrade Costa	88
<i>Modelagem da Dinâmica de uma Válvula Solenoide para Composição de Máquina Virtual (MV) em Funções de Blowout Preventer (BOP)</i> Lucas Constantino Mendonça, Yasmin Nunes Alves, Luiz Antônio de Oliveira Chaves, Iara Tammela, Rodolfo Cardoso e Danilo Colombo	100
<i>Discrepância de Modelo: Estudo Inicial com Regressão por Processo Gaussiano</i> Lucas de Oliveira Sá, Leonardo Tavares Stutz e Daniel Alves Castello	112
<i>Técnicas de Processamento de Linguagem Natural Para Detecção de Discurso de Ódio</i> Cássia Claudiane Silva da Rosa e Renato Porfirio Ishii	123

---

<i>Consensus docking to ensemble-docking using multiple programs: exemple with <i>Drosophila melanogaster</i> dopamine transporter.</i>	
Fabiani Fernanda Triches, Francieli Triches e Cilene Lino de Oliveira	125
<i>A molecular dynamics approach to spermine and spermidine SI-PPCs interactions with heparin and DNA</i>	
Frederico Henrique do C. Ferreira e Luiz Antônio S. Costa	126
<i>Segmentação de pragas e doenças em folhas de café utilizando redes convolucionais</i>	
Humberto da Silva Neto, Shirley Peroni Neves Cani e Mariana Rampinelli Fernandes	127
<i>A container-based environment for theparametric simulation of job schedulers</i>	
João Pedro Macleure Nunes dos Santos e Antônio Tadeu Azevedo Gomes	128
<i>Serviços de Atualização no Gateway BioinfoPortal: Suporte ao Bancos de dados de Proveniência</i>	
Marco Cabral, Marcelo Galheigo, Antônio Gomes e Kary Ocaña	130
<i>Modelagem computacional de drenagem urbana: Estudo para sub-bacia do Jandiá em Macapá-AP</i>	
Mauricio Dias Da Conceição Neto e Luana Oliveira de Sá	132
<i>Identificação de transientes astrofísicos utilizando redes neurais convolucionais</i>	
Phelipe Darc e Clécio R. Bom	134
<i>Uso de Modelagem Hidrológica visando a Mitigação dos Efeitos de Desastres Naturais</i>	
Renata Nalim Basilio Tissi, Wagner Rambaldi Telles e Antônio Silva Neto	135
<i>FLASH Radiotherapy RNA-Seq Data Mining</i>	
Samella Salles e Kary Ocaña	136
<i>Um dispositivo de Biometria para a segurança residencial</i>	
Vicente Soares e Aline Lunkes	137
<i>Human Skin Color Diversity with CUDAMCML</i>	
Victor Porto Gontijo de Lima e Lilian Tan Moriyama	138

---

## Trabalhos completos



# Algoritmo adaptativo para o cálculo numérico da derivada

Callebe R. Reis<sup>1</sup>, Sandro D. P. Vitenti<sup>2</sup> e Mariana Penna-Lima<sup>3</sup>

<sup>1</sup> *Departamento de Matemática, Universidade de Brasília, Brasília-DF 70919-900, Brazil*

<sup>2</sup> *Departamento de Física, Universidade Estadual de Londrina, Rod. Celso Garcia Cid, Km 380, 86057-970, Londrina, Paraná, Brazil*

<sup>3</sup> *Instituto de Física, Universidade de Brasília, Caixa Postal 04455, 70919-970 Brasília-DF, Brazil*

---

## Abstract

Métodos de derivação numérica são aplicados nos mais variados campos de conhecimento. Entretanto, apesar da grande versatilidade do cálculo numérico da derivada, há sempre um erro numérico associado a cada algoritmo. Muitas vezes, este erro é difícil de ser mitigado. Portanto, existe a demanda por algoritmos mais precisos, e, ao mesmo tempo, mais eficientes para a avaliação de derivadas numericamente. Neste trabalho, exploramos um método de derivação numérica cuja precisão é um parâmetro do algoritmo. Para isto, utilizamos matrizes de Vandermonde para uma aproximação da derivada com métodos de derivação numérica clássicos. Em particular, aplicamos o algoritmo para o método de derivação numérica com diferenças finitas progressivas. Com o novo algoritmo, obtemos aproximações com precisão arbitrária. Realizamos testes em uma máquina de precisão finita, onde conseguimos demonstrar a precisão do algoritmo através da comparação com resultados analíticos, a menos da precisão da máquina utilizada.

**Keywords:** Derivada numérica, Algoritmo adaptativo, Matriz de Vandermonde

---

## 1 Introdução

A derivação numérica apresenta um dos grandes pilares da computação científica moderna, visto que, muitos algoritmos dependem do cálculo da derivada, desde algoritmos para resolução de equações diferenciais ao cálculo de matrizes de Fisher e problemas de otimização. Portanto, é necessário que o cálculo da derivada seja realizado de forma eficiente e precisa.

Resultados presentes na literatura propõem a utilização da matriz de Vandermonde para o cálculo da derivada, em [9] um método é proposto utilizando a inversão da matriz

de Vandermonde, neste trabalho, apresentamos um método redescoberto de diferenciação numérica cuja ordem de precisão pode ser controlada como um parâmetro [7, 5] e propomos uma solução analítica para a matriz de Vandermonde, além de um algoritmo adaptativo para o método descrito.

A derivada de uma função real em números reais aplicada a um ponto pode ser vista como a inclinação da reta secante entre dois pontos pertencentes à função quando tomamos a distância entre eles como um limite tendendo para zero [6], i.e.,

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}. \quad (1)$$

Apesar da grande utilidade e aplicação da definição de derivada, é comum que o cálculo da derivada necessite de ser realizado várias vezes, e para diversas funções diferentes, mantendo uma precisão satisfatória para a resolução de um dado problema. Nesses casos, é muito comum que a derivada necessite ser computada através de um computador. Entretanto, no cálculo realizado por uma máquina de precisão finita, como por um computador, a avaliação deste limite se torna problemática: como definir um limite que pode levar a uma indeterminação em uma máquina de precisão finita? E como maximizar a precisão deste cálculo?

### 1.1 Métodos clássicos de diferenciação numérica

Existem muitas maneiras de aproximar a derivada de uma função sem a necessidade de calcular o limite explicitamente. Métodos clássicos de derivação sugerem a utilização da expansão em fórmula de Taylor para a derivada [1, 8]. Para isso, expandimos em séries de Taylor a função  $f(x+h)$ :

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(x)h^3 + \mathcal{O}(h^4), \quad (2)$$

onde a notação  $\mathcal{O}(h^4)$  representa a ordem da aproximação em relação ao parâmetro  $h$ , quanto maior o expoente de  $\mathcal{O}(h^n)$  melhor a aproximação.

Para obter a derivada através da expressão acima, isolamos o termo  $f'(x)$  que contém a derivada da função. Somando  $-f(x)$  à equação (2) e dividindo por  $h$  de ambos os lados da equação, obtemos a expressão abaixo:

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 + \mathcal{O}(h^3), \quad (3)$$

$$f'_h(x) := \frac{f(x+h) - f(x)}{h} + \mathcal{O}(h). \quad (4)$$

A equação (4) descreve um método clássico de aproximação para a derivada conhecido como Método de Diferenças Finitas à Direita (*FDM Forward Derivative Method*).

Note que, a expressão obtida não possui nenhum limite em sua descrição. Além disso, considerando a parte  $\mathcal{O}(h)$  como o erro na aproximação, a expressão acima torna explícito como o erro depende de uma escolha para o parâmetro  $h$ , quanto menor o  $h$ , melhor é a aproximação, conforme o que era esperado devido ao limite que aparece na equação (1).





O método FDM descrito na equação (4) é fácil de ser implementado e possui rápida execução. No entanto, devido a precisão ser de ordem  $\mathcal{O}(h)$ , muitas vezes é necessário um valor muito pequeno de  $h$  para se obter uma aproximação da derivada de forma satisfatória, e em máquinas de precisão finita, a precisão do método pode ficar bastante limitada devido ao erro de arredondamento da máquina.

### 1.2 Novo método

Observando as equações para o FDM equação (4), e a expansão de  $f(x+h)$  equação (2), vemos que a aproximação depende da escolha de um valor para  $h$ . Se considerarmos duas aproximações para a derivada utilizando o FDM, com valores diferentes de  $h$ , como  $h/g_0$  e a outra como  $h/g_1$ , obtemos as duas aproximações abaixo:

$$\begin{aligned} f'_{h/g_0}(x) &= \frac{f(x+h/g_0) - f(x)}{h/g_0} - \frac{1}{2}f''(x)h/g_0 - \frac{1}{6}f'''(x)(h/g_0)^2 - \mathcal{O}(h^3), \\ f'_{h/g_1}(x) &= \frac{f(x+h/g_1) - f(x)}{h/g_1} - \frac{1}{2}f''(x)h/g_1 - \frac{1}{6}f'''(x)(h/g_1)^2 - \mathcal{O}(h^3). \end{aligned} \quad (5)$$

Queremos escrever uma combinação linear de forma a zerar o termos das equações (5) que envolvem termos de ordem  $\mathcal{O}(h)$ , obtendo assim, uma aproximação que depende apenas de ordem  $\mathcal{O}(h^2)$ . Para isto, multiplicamos a primeira equação por uma constante  $a_0$  e a outra por uma outra constante  $a_1$  e somamos as duas:

$$f'(x) = a_0 f'_{h/g_0}(x) + a_1 f'_{h/g_1}(x) - \frac{1}{2}f''(x)h \left( \frac{a_0}{g_0} + \frac{a_1}{g_1} \right) + \mathcal{O}(h^2). \quad (6)$$

Gostaríamos que a soma das duas aproximações  $a_0 f'_{h/g_0}(x) + a_1 f'_{h/g_1}(x)$  fosse também uma aproximação para  $f'(x)$  e que essa fosse uma aproximação melhor para a derivada, ou seja, que dependa de  $\mathcal{O}(h^2)$  ou ordens superiores.

Impondo estas restrições, observamos que precisamos que as equações  $a_0 + a_1 = 1$  e  $a_0/g_0 + a_1/g_1 = 0$  sejam satisfeitas, para isto, montamos o sistema de equações (7).

$$\begin{cases} a_0 + a_1 &= 1 \\ \frac{a_0}{2g_0} + \frac{a_1}{2g_1} &= 0 \end{cases} \quad (7)$$

o que é equivalente à equação matricial:

$$\begin{bmatrix} 1 & 1 \\ \frac{1}{g_0} & \frac{1}{g_1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (8)$$

Resolvendo (8) para  $a_0$  e a  $a_1$ , obtemos  $a_0 = \frac{1}{1 - g_1/g_0}$  e  $a_1 = \frac{1}{1 - g_0/g_1}$ .

A matriz dos coeficientes em (8) é chamada matriz de Vandermonde de ordem 2 (veja o apêndice A para mais detalhes sobre a matriz de Vandermonde).

1.2.1 Caso de ordem  $n$ 

Podemos repetir o processo de maneira análoga para o caso de ordem  $n$ , ou seja, o caso em que o erro da aproximação da derivada seja da ordem de  $\mathcal{O}(h^n)$ .

Consideramos uma combinação linear de  $n$  aproximações, que dependem de  $n$  valores diferentes do parâmetro  $h$ , chamaremos esses valores de  $h/g_i$  com  $i = 0, \dots, n-1$ . Com isso,

$$a_0 f'_{h/g_0}(x) + a_1 f'_{h/g_1}(x) + \dots + a_{n-1} f'_{h/g_{n-1}}(x) = f'(x) + \mathcal{O}(h^n). \quad (9)$$

Agrupando os termos de ordem  $h, h^2, \dots, h^{n-1}$ , e impondo a restrição que os coeficientes destes  $n-1$  termos sejam nulos e que esta combinação linear ainda seja uma aproximação para a derivada, montamos o seguinte sistema linear (10)

$$\begin{cases} a_0 + a_1 + \dots + a_{n-1} = 1 \\ -\frac{a_0}{2!g_0} - \frac{a_1}{2!g_1} - \dots - \frac{a_{n-1}}{2!g_{n-1}} = 0 \\ \vdots \\ -\frac{a_0}{n!g_0^{n-1}} - \frac{a_1}{n!g_1^{n-1}} - \dots - \frac{a_{n-1}}{n!g_{n-1}^{n-1}} = 0. \end{cases} \quad (10)$$

Que é equivalente ao sistema matricial (11)

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ g_0^{-1} & g_1^{-1} & \dots & g_{n-1}^{-1} \\ g_0^{-2} & g_1^{-2} & \dots & g_{n-1}^{-2} \\ \vdots & \vdots & \ddots & \vdots \\ g_0^{-(n-1)} & g_1^{-(n-1)} & \dots & g_{n-1}^{-(n-1)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (11)$$

A matriz dos coeficientes da expressão (11) é, como já vimos, uma matriz de Vandermonde (Apêndice A). Sabendo disso, podemos expressar a solução desta matriz da seguinte forma

$$A = \begin{bmatrix} \frac{1}{\prod_{j \neq 1} (1 - g_j/g_1)} \\ 1 \\ \frac{1}{\prod_{j \neq 2} (1 - g_j/g_2)} \\ \vdots \\ 1 \\ \frac{1}{\prod_{j \neq n-1} (1 - g_j/g_{n-1})} \end{bmatrix}. \quad (12)$$

A solução (12) representa os coeficientes da combinação linear que aproxima uma derivada de uma função num ponto com um erro de ordem  $\mathcal{O}(h^n)$ :

$$f'(x) = \sum_{i=0}^{n-1} a_i f'_{h/g_i}(x) + \mathcal{O}(h^n). \quad (13)$$

Os valores para os coeficientes  $g_i$  podem ser obtidos através de uma função arbitrária, ou ainda, podem representar diferentes frequências de amostragem da função  $f$ , o que faz com que a aplicabilidade do novo método para avaliação da derivada seja estendida a casos onde não possuímos informações analíticas sobre a função  $f$ , apenas um conjunto finito de pontos onde a função foi avaliada.



## 2 Algoritmo Adaptativo

Utilizando o método descrito na Seção 1.2, construímos um algoritmo para o cálculo da derivada que leva em consideração o erro de arredondamento da máquina utilizada e o erro gerado pela aproximação por termos finitos.

Devido à forma como as soluções para as matrizes de Vandermonde são dadas, é possível obter soluções para matrizes de ordem  $n + 1$  através da solução de matrizes de ordem  $n$ , sendo necessário apenas multiplicar em cada uma das  $n$  coordenadas um termo da forma  $1/(1 - g_n/g_i)$ , onde  $i$  é a coordenada, e acrescentar uma  $(n + 1)$ -ésima coordenada com  $\prod_{i \neq n} 1/(1 - g_i/g_n)$ . Verificando a cada iteração do algoritmo, se o erro relativo está abaixo de um dado limite, e se o erro de arredondamento da máquina não está desviando a solução da convergência. O fluxograma presente na Fig. 1 ilustra o algoritmo de derivação adaptativo. Vale ressaltar que a etapa “Repete 3x” é realizada para garantir que o valor é de fato menor que o limiar imposto, ou seja, que o erro obtido no passo subsequente é de fato menor que o anterior.

Este algoritmo está implementado na biblioteca *Numerical Cosmology* (NumCosmo) [2]. Além disso, implementamos também uma técnica mais robusta para controlar o erro presente na avaliação da derivada numérica. Essa técnica está sendo descrita pelos autores em um segundo trabalho.

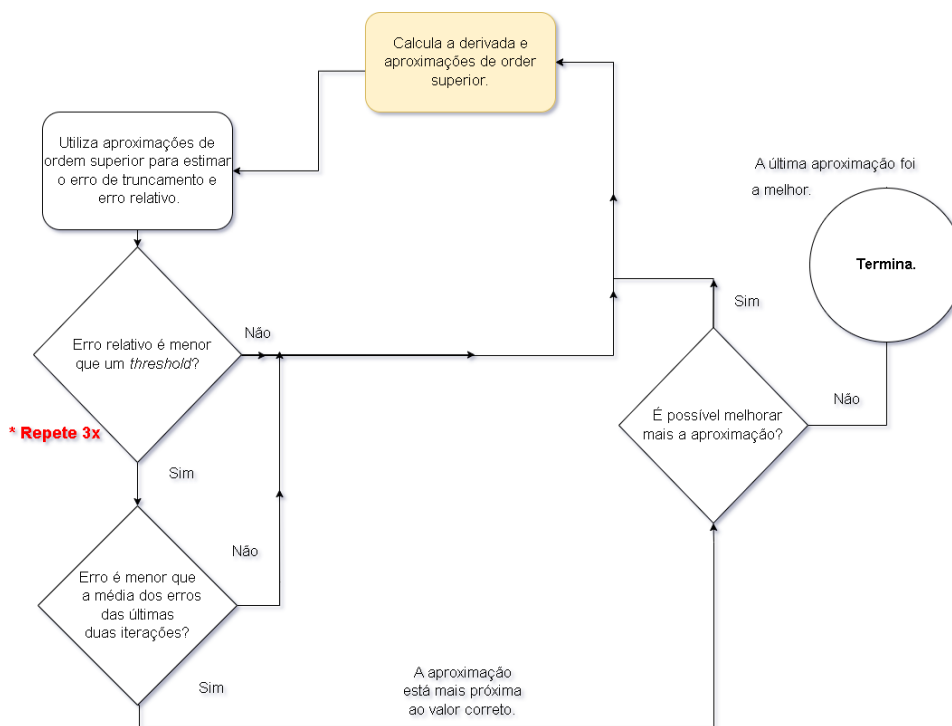


Fig. 1: Algoritmo adaptativo para o cálculo da derivada numérica.

## 3 Resultados

Implementamos o algoritmo descrito na Seção 1.2.1 na linguagem de programação Python. Utilizamos a precisão padrão da linguagem de 18 casas decimais, e como tolerância para

o algoritmo adaptativo  $10^{-8}$ . Escolhemos funções para as quais a avaliação da derivada numericamente apresentam dificuldades devido a presença de exponenciais, raízes e potências. Em especial, escolhemos as funções  $f(x) = \exp(-x^2)$  e  $g(x) = \exp(\sqrt{x})$ . Para a aplicação do algoritmo, escolhemos  $x = 3$  para função  $f(x)$ , e  $x = 0.1$  para a função  $g(x)$ . Aplicamos o algoritmo para ambas utilizando o passo inicial  $h$  como 0.1. Realizamos para as duas funções 1, 2, 3, 5 e 8 iterações do algoritmo. Utilizamos  $2^i$  como função geradora dos termos  $g_i$ .

**Tabela 1:** TABELA DE VALORES PARA A DERIVAÇÃO NUMÉRICA DA FUNÇÃO  $\exp(-x^2)$  UTILIZANDO O NOVO MÉTODO.

Iterações	valor obtido	valor analítico	erro
1	-0.000721603041819833	-0.0007404588245200774	1.8855782700244 e-05
2	-0.000739975274324733	-0.0007404588245200774	4.835501953443e-07
3	-0.0007404533668132014	-0.0007404588245200774	5.45770687595-09
5	-0.0007404588244915567	-0.0007404588245200774	2.85206e-14
8	-0.0007404588245190902	-0.0007404588245200774	9.871e-16

**Tabela 2:** DERIVADAS PARA FUNÇÃO  $f(x) = \exp(\sqrt{x})$ .

Iterações	valor obtido	valor analítico	erro
1	1.6285920044163307	2.169231878730517	0.5406398743141865
2	2.1633658043366903	2.169231878730517	0.005866074393826803
3	2.1687897319602625	2.169231878730517	0.00044214677025467
5	2.169231444806324	2.169231878730517	4.33924193021e-07
8	2.16923187873095	2.169231878730517	4.3298e-13

Observamos que utilizando o FMD com o menor passo possível  $10^{-8}$  (utilizando passos menores o erro de arredondamento da máquina faz o resultado divergir, devido a presença de exponenciais e potências na equação), obtemos um erro de ordem  $2.10^{-11}$ , para a derivada da função  $f(x) = \exp(-x^2)$  avaliada em  $x = 3$ . Para a função  $g(x) = \exp(\sqrt{x})$  avaliada em  $x = 0.1$ , obtemos um erro de ordem  $1.10^{-7}$ . Ambos os resultados são inferiores aos obtidos utilizando o novo método com 8 iterações, passo inicial de 0.1 e um passo mínimo de  $0.1 \cdot 2^{-8}$ . Vale ressaltar ainda que a cada nova iteração do novo método, são necessárias apenas uma operação de avaliação de função,  $n$  multiplicações por escalares e  $n$  somas, onde  $n$  é o número de iterações.

Observamos que a diferença entre o valor esperado, que foi calculado analiticamente, e o valor obtido através do algoritmo decresceu conforme o esperado, ou seja, decresceu como uma potência do valor do passo com o número de iterações.

#### 4 Conclusão

O método apresentado neste projeto é capaz de melhorar a precisão do algoritmo de diferenciação numérica por diferenças finitas, dentro da precisão numérica possível em um computador de memória finita, e, além disso, gerar resultados mais precisos que métodos tradicionais utilizando passos maiores.



Estes resultados se traduzem em um método mais preciso para a avaliação da derivada numérica, tanto em um regime em que se conhece a função a ser derivada quando apenas se conhece a função em um conjunto finito de pontos.

Entretanto, mais pesquisas são necessárias para a possível melhoria no tratamento do erro de arredondamento quando o algoritmo é aplicado com precisão finita. Assim, uma análise da relação entre os erros de arredondamento e o erro numérico inerente ao algoritmo, assim como uma apresentação mais detalhada da implementação presente na biblioteca *Numerical Cosmology* (NumCosmo) serão realizadas em um trabalho futuro.

## A Matriz de Vandermonde

A matriz quadrada da forma

$$A = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ a_1 & a_2 & a_3 & \dots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \dots & a_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \dots & a_n^{n-1} \end{bmatrix} \quad (14)$$

é chamada de matriz de *Vandermonde* [4, 3]. Seu determinante é dado por

$$\det(A) = \prod_{1 \leq i < j \leq n} (a_j - a_i). \quad (15)$$

A solução de  $Ax = B$ , onde  $A$  é a matriz de Vandermonde e a matriz  $B = (1, 0, 0, \dots, 0)_n$  é

$$x = \begin{bmatrix} \frac{\prod_{i \neq 1} a_i}{\prod_{j \neq 1} (a_j - a_1)} \\ \frac{\prod_{i \neq 2} a_i}{\prod_{j \neq 2} (a_j - a_2)} \\ \vdots \\ \frac{\prod_{i \neq n} a_i}{\prod_{j \neq n} (a_j - a_n)} \end{bmatrix}, \quad (16)$$

equivalentemente

$$x = \begin{bmatrix} \frac{1}{\prod_{j \neq 1} (1 - a_1/a_j)} \\ 1 \\ \frac{1}{\prod_{j \neq 2} (1 - a_2/a_j)} \\ \vdots \\ 1 \\ \frac{1}{\prod_{j \neq n} (1 - a_n/a_j)} \end{bmatrix}. \quad (17)$$

Esta solução pode ser facilmente encontrada fazendo uso da regra de Cramer, uma vez que o determinante e os determinantes das matrizes adjuntas são dados pela expressão (15).

Uma propriedade bastante útil das matrizes de Vandermonde é que a solução de uma matriz de ordem  $n$  é parcialmente a solução de uma matriz de Vandermonde de ordem  $n + 1$ .

## Referências

- [1] Gerrund Dahlquist, Ake Bjorck, and N Anderson. Numerical methods prentice-hall. *Inc., Englewood Cliffs, NJ*, 1974.
- [2] Sandro Dias Pinto Vitenti and Mariana Penna-Lima. NumCosmo: Numerical Cosmology. Astrophysics Source Code Library, record ascl:1408.013, August 2014.
- [3] Roger A Horn, Roger A Horn, and Charles R Johnson. *Topics in matrix analysis*. Cambridge university press, 1994.
- [4] Donald E Knuth. The art of computer programming. volume 1: Fundamental algorithms. volume 2: Seminumerical algorithms. *Bull. Amer. Math. Soc*, 1997.
- [5] Jianping Li. General explicit difference formulas for numerical differentiation. *Journal of Computational and Applied Mathematics*, 183(1):29–52, 2005.
- [6] Elon Lages Lima. *Análise real*, volume 1. Impa Rio de Janeiro, 2004.
- [7] JN Lyness and CB Moler. Van der monde systems and numerical differentiation. *Numerische Mathematik*, 8(5):458–464, 1966.
- [8] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes in c++. *The art of scientific computing*, 2:1002, 2007.
- [9] Yijie Zhang, Jinghuai Gao, Jigen Peng, and Weimin Han. A robust method of computing finite difference coefficients based on vandermonde matrix. *Journal of Applied Geophysics*, 152:110–117, 2018.



# Modelagem de jornadas de pacientes como um grafo multiaspecto

Caroline de Oliveira Costa Souza Rosa<sup>1</sup>, Márcia Ito<sup>2</sup>, Alex Borges Vieira<sup>3</sup>, Klaus Wehmuth<sup>1</sup> e Antônio Tadeu Azevedo Gomes<sup>1</sup>

<sup>1</sup> *Laboratório Nacional de Computação Científica, Petrópolis/RJ, Brasil*

<sup>2</sup> *Centro Paula Souza, São Paulo/SP, Brasil*

<sup>3</sup> *Universidade Federal de Juiz de Fora, Juiz de Fora/MG, Brasil*

---

## Resumo

A jornada do paciente é o registro dos atendimentos, intervenções e tratamentos a que o paciente foi submetido ao longo de sua vida. O estudo das jornadas de um conjunto de pacientes traz informações sobre a eficiência de abordagens terapêuticas, permite avaliar a aderência a elas e também revela o uso de recursos (p.ex. unidades de saúde) para prover os atendimentos necessários. No entanto, modelos de representação de jornadas são necessários para que estes tipos de análise sejam feitos. Diferentes modelos são reportados na literatura, porém poucos se adequam às características de jornadas longas ou que consideram diferentes perspectivas. Neste artigo, avalia-se a viabilidade de se utilizar um grafo multiaspecto para expressar as jornadas de pacientes. Um estudo de caso com gestantes no município de São Paulo indica que o modelo é adequado para representar visualmente as jornadas e é flexível para permitir diferentes tipos de análise da estrutura e distribuição da rede de saúde.

**Palavras-chave:** Jornadas de pacientes, Mineração de processos, Grafo multiaspecto, Serviço de saúde

---

## 1 Introdução

O estudo do histórico de encontros de um conjunto de pacientes com o sistema de saúde permite a realização de inferências sobre a evolução de doenças (Prodel et al., 2018), o cumprimento de protocolos de tratamento (Rinner et al., 2018) e a eficiência de diferentes unidades de saúde (Leonardi et al., 2018; Sato et al., 2020). Esta sequência de eventos relacionados ao tratamento de saúde do paciente é chamada de *trajetória* ou *jornada* do paciente.

Para o contexto deste estudo, uma *atividade/intervenção* é um tipo de episódio bem definido na jornada de um paciente, como uma consulta ou um exame. Um *evento* corresponde à ocorrência de uma atividade/intervenção, para um paciente específico em

determinado momento no tempo. No entanto, além do rótulo principal da atividade que caracteriza o evento, ele também pode ser visto sob diferentes *perspectivas*. Por exemplo, a ocorrência de uma consulta pode ser ainda mais especificada considerando o tipo de unidade de saúde onde ocorreu ou o especialista que a realizou.

Para avaliar as jornadas de um grupo de pacientes, é necessário adotar um modelo de representação dessas jornadas que as apresente de forma resumida, preservando os padrões mais relevantes, a ordenação temporal dos eventos e o tempo decorrido entre eles. Diferentes modelos têm sido utilizados pelos autores dessa área de pesquisa (Rosa *et al.*, 2022). O uso de uma lista de sequências (Antonelli *et al.*, 2012) é geralmente útil quando há baixa variabilidade de jornadas<sup>1</sup> ou quando o estudo tenciona usar os resultados para alimentar problemas de otimização ou predição. A adoção de grafos direcionados cíclicos (Najjar *et al.*, 2018) apoia casos com variabilidade de jornadas um pouco maior, porém sua interpretação se torna difícil quando há repetição de atividades em momentos diferentes das jornadas. Álgebras de processo se adequam bem a casos em que as jornadas são padronizadas, mas apresentam dificuldade em casos menos estruturados (Rebuge e Ferreira, 2012; Kempa-Liehr *et al.*, 2020). A elaboração de modelos relacionados com a representação de jornadas de pacientes em contextos de alta variabilidade ou repetição de atividades (De Oliveira *et al.*, 2020; Duma e Aringhieri, 2020), considerando as diferentes perspectivas de seus eventos, ainda é uma área pouco explorada.

O grafo multiaspecto (MAG – do inglês, *MultiAspect Graph*) é uma generalização do grafo tradicional (Wehmuth *et al.*, 2016) visando a representação de redes multicamada ou variantes no tempo. Essa modelagem já foi adotada, por exemplo, para representar redes de transporte aéreo (Wehmuth *et al.*, 2018) e de transições de criptomoedas (Mascarenhas *et al.*, 2020). No contexto de jornadas de pacientes, especialmente no caso de pacientes crônicos ou de jornadas com muitos eventos, é importante analisar, além da intervenção realizada, o momento e o contexto em que ela ocorreu. Um exemplo é analisar a distribuição das unidades de saúde numa região ou verificar a aderência a um protocolo clínico específico. Além disso a proposta do SUS em se organizar de acordo com uma rede de atenção à saúde demonstra que é possível pensar em jornada de pacientes como uma rede complexa com vários aspectos a serem analisados (Mendes, 2011; Jasmim *et al.*, 2017). Existe, portanto, o potencial da construção de um modelo de jornada de pacientes como uma rede complexa com múltiplas camadas (perspectivas) e variante no tempo (momento de ocorrência dos eventos).

O objetivo deste artigo é avaliar a viabilidade de representar jornadas de pacientes por meio da modelagem de MAG. Especificamente, busca-se responder as seguintes questões: (i) MAG provê uma representação visual fácil de interpretar, mesmo para jornadas de pacientes com repetição de atividades/intervenções? (ii) MAG torna possível analisar as jornadas de pacientes sob diferentes perspectivas?

Através do estudo de jornadas de gestantes no município de São Paulo, verificou-se que o MAG permite a análise visual considerando tanto a ordenação temporal dos eventos quanto os procedimentos realizados e unidades de saúde visitadas. O modelo também viabilizou a identificação de combinações de procedimento e unidade mais frequentes e a contabilização de conexões entre os diferentes estabelecimentos de saúde.

<sup>1</sup>Diz-se que a variabilidade de jornadas é baixa quando o número de jornadas distintas é pequeno quando comparado ao número total de pacientes.





O restante do artigo está estruturado como se segue: a Seção 2 apresenta diferentes modelos propostos na literatura para representar jornadas de pacientes; a Seção 3 apresenta o conjunto de dados usado neste artigo e a construção do modelo de jornadas aqui proposto; a Seção 4 apresenta os resultados obtidos; e a Seção 5 apresenta as conclusões do estudo.

## 2 Trabalhos Relacionados

A área de pesquisa que trata da modelagem e mineração de jornadas de pacientes é relativamente nova e tem crescido sobretudo nos últimos anos (Rosa et al., 2022).

A representação de cada jornada como uma sequência de eventos (Antonelli et al., 2012; Aspland et al., 2021; Egho et al., 2015) é uma alternativa frequente e direta. Contudo, ela se restringe à apresentação de sucessões consecutivas entre os eventos ocorridos e, se houver muitas jornadas distintas, a listagem delas será extensa, dificultando a interpretação dos especialistas.

A adoção de um grafo direcionado como modelo de jornadas (Najjar et al., 2018) tende a gerar uma representação mais concisa, já que ao invés de uma listagem de padrões, tem-se uma representação unificada das jornadas. Entretanto, quando as jornadas possuem repetição de atividades, o grafo resultante terá ciclos que podem dificultar a identificação da ordenação temporal dos eventos seguidos. Como alternativa, alguns autores propõem o uso de grafos direcionados acíclicos (Chiudinelli et al., 2020), também conhecidos como grafos tipo árvore, que permitem a existência de múltiplos vértices com mesmo rótulo de forma a prevenir a formação de ciclos. Se por um lado este modelo favorece a leitura da evolução temporal dos eventos das jornadas, por outro tem-se um modelo que pode se tornar muito grande e difícil de interpretar, como no caso das sequências.

Alguns autores buscaram propor novos modelos combinando características dos grafos cíclicos e acíclicos. Duma e Aringhieri (2020), por exemplo, propõem o modelo *“Hybrid Activity Tree”*, que é semelhante a um grafo acíclico, mas cujos ramos infrequentes são convertidos em grafos cíclicos. Outro exemplo é o *“Time Grid Process Model”* proposto por De Oliveira et al. (2020). Os autores utilizam camadas temporais que indicam em que posição da jornada cada evento ocorreu e cada atividade pode aparecer uma vez em cada camada. Desse modo, a ordenação dos eventos previne a formação de ciclos, enquanto que a limitação de no máximo uma ocorrência de atividade por camada ajuda a limitar o crescimento do modelo, favorecendo sua interpretação.

Outra alternativa para representar as jornadas de pacientes é o uso de álgebras de processo que buscam revelar as dependências entre atividades das jornadas (Rebuge e Ferreira, 2012; Kempa-Liehr et al., 2020), como pontos em que há escolha de sub-caminhos a serem seguidos, ocorrência de padrões paralelos e dependências em longo prazo entre atividades. A expressividade desses modelos favorece sua aplicação para modelar jornadas de pacientes com foco em identificação de protocolos clínicos ou processos organizacionais. No entanto, sua utilização em estudos de caso com jornadas muito longas ou com variabilidade alta resulta frequentemente em modelos complexos e difíceis de serem lidos (modelos “spaghetti”) (Günther e van der Aalst, 2007).

Além da ordenação e dos intervalos temporais dos eventos, a conservação e apresentação de diferentes perspectivas relacionadas a esses eventos pode ser relevante. Por exemplo, no caso de internações hospitalares (Egho et al., 2015), um procedimento é

realizado devido a um conjunto de diagnósticos, ocorre em determinada unidade ou departamento e pode envolver um ou mais medicamentos. Portanto, um modelo de jornadas que una uma boa representação temporal dos eventos e que também permita avaliar as jornadas sob diferentes perspectivas é um tópico de interesse e ainda em aberto na literatura.

### 3 Metodologia

#### 3.1 Seleção dos Dados

Os dados usados foram obtidos junto à Prefeitura Municipal de São Paulo (Jasmim et al., 2017). Eles se referem ao Sistema Integrado de Gestão de Assistência à Saúde (SIGA-SAÚDE) e contêm informações sobre agendamentos e atendimentos ambulatoriais prestados pela Secretaria Municipal da Saúde de São Paulo durante os anos de 2014 e 2015. Este intervalo correspondia, à época do acordo de cessão de dados, ao período mais recente que pôde ser fornecido. Os dados revelam, por exemplo, a data em que um paciente se submeteu a um procedimento ou recebeu um diagnóstico, em que unidade de saúde isso ocorreu e qual a categoria (tipo) desta unidade. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa (CEP) da UNIFESP e da Secretaria Municipal de Saúde de São Paulo sob CAAE 51038515.6.3001.0086.

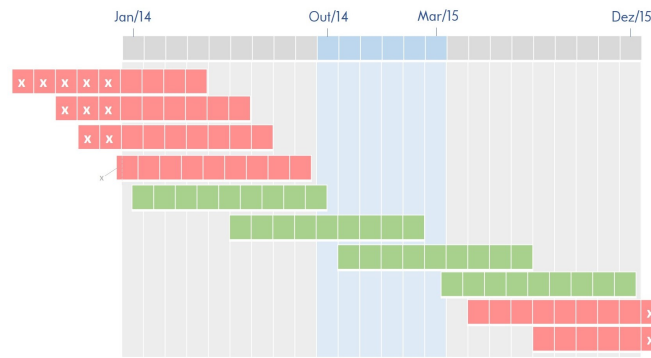
Para este estudo decidiu-se analisar as jornadas de gestantes. Assim, da base de dados original, extraiu-se apenas os atendimentos deste grupo de pacientes. Para isso, listou-se códigos de diagnósticos<sup>2</sup> e de procedimentos<sup>3</sup> associados direta e exclusivamente à gestação. Em seguida, obteve-se a listagem de pacientes com algum registro desses procedimentos ou diagnósticos. Apenas pacientes cujas jornadas envolviam ao menos três eventos foram consideradas.

No entanto, como algumas dessas pacientes podem ter iniciado sua gestação antes de janeiro de 2014 (início do período da base de dados) ou terminado após dezembro de 2015 (fim do período da base de dados), foi adotada uma estratégia de seleção visando pacientes com maior probabilidade de “gestação completa” contida no intervalo de tempo disponível. A Figura 1 apresenta um esquema visual da estratégia de filtragem adotada. Foi considerado que uma gestação de 9 meses iniciada em janeiro de 2014 faria com que a gestante tivesse atendimentos até outubro de 2014, logo, uma paciente cujo último registro relacionado à gestação ocorreu antes de outubro, possivelmente engravidou antes de janeiro de 2014, e portanto, parte da gestação não estaria incluída na base de dados. Similarmente, uma gestante cujo parto ocorreu em dezembro de 2015 teria o início de sua jornada gestacional por volta de março de 2015, portanto uma paciente cujo primeiro registro gestacional que ocorreu após março pode ter finalizado sua gestação em 2016, ou seja, fora do período coberto pela base de dados. Optou-se, então, por selecionar apenas pacientes que tinham ao menos um registro relacionado à gestação entre outubro de 2014 e março de 2015.

Destaca-se que esta estratégia está sujeita a recusar algumas pacientes com todo o período de gestação entre 2014 e 2015, como as que engravidaram antes de março de 2015

<sup>2</sup>De acordo com a Classificação Internacional de Doença (CID-10)

<sup>3</sup>De acordo com o Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP)



Cada sequência de quadrados representa a jornada de uma paciente durante a gestação, em que cada quadrado corresponde a um mês (foi considerado que cada paciente tem 9 meses de gestação). As jornadas em verde são as mantidas para análise, enquanto as em vermelho são as descartadas. Os períodos assinalados com um 'x' são aqueles que fazem parte da gestação da paciente, mas que não estão contemplados na base de dados.

**Fig. 1:** Esquema ilustrativo da estratégia de seleção de gestantes com “gestação completa” coberta pela base de dados.

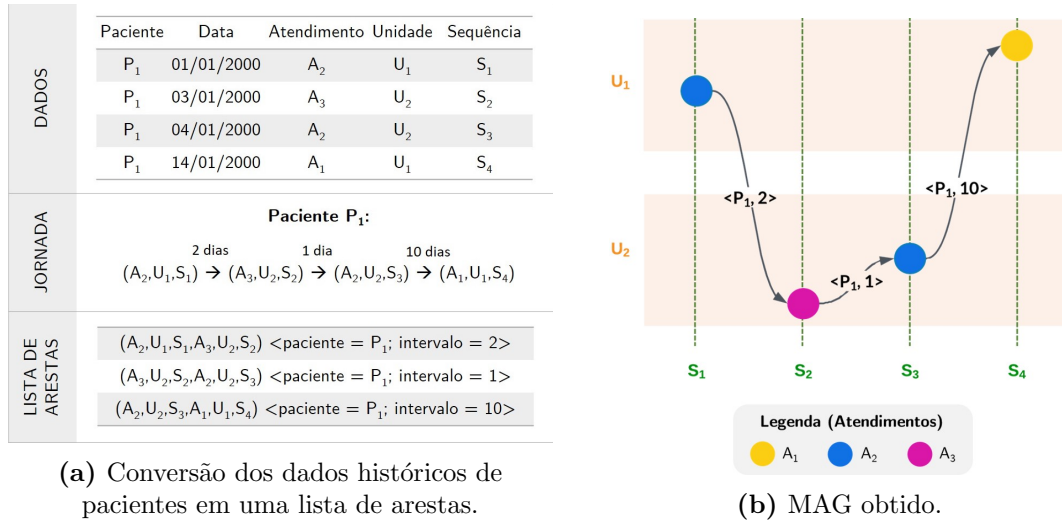
mas só iniciaram o acompanhamento pré-natal após o segundo mês de gestação, ou as que tiveram período gestacional inferior a 9 meses. Também é possível que algumas gestações não completamente cobertas no período entre 2014 e 2015 sejam consideradas, como as de duração levemente maior do que nove meses, ou as de pacientes que engravidaram duas vezes nesse período (uma gestação estaria completa na base, enquanto a outra não).

### 3.2 Construção do MAG

Um grafo é uma estrutura matemática que representa um conjunto de entidades (nós ou vértices) e o relacionamento entre elas (arestas). No contexto de jornadas de pacientes, por exemplo, os vértices podem denotar diagnósticos (doenças) que um paciente teve, enquanto uma aresta direcionada saindo de um vértice e indo até outro pode indicar que a doença do vértice de destino ocorreu imediatamente após a do vértice de origem (Prodel et al., 2018). O Grafo Multiaspecto (MAG), proposto por Wehmuth et al. (2016), generaliza o conceito tradicional de grafo e é formado por um conjunto de aspectos e um conjunto de arestas. Cada aspecto apresenta uma característica da rede sendo considerada. Uma tupla formada por um elemento de cada aspecto pode ser entendida de forma análoga ao vértice de um grafo tradicional. Por simplicidade, nos referimos a esse tipo de tupla como o “vértice” do MAG.

Para modelar jornadas de pacientes, propõe-se o MAG  $H = (A, U, S)$ , em que  $A$  é o aspecto “Atendimento”, que expressa o procedimento realizado,  $U$  é o aspecto “Unidade de Saúde”, que indica o estabelecimento onde o atendimento ocorreu, e  $S$  é o aspecto “Sequência”, que revela a posição do atendimento na jornada do paciente. Cada aresta contém um atributo categórico que a associa a um paciente, e outro quantitativo que aponta o tempo decorrido entre o vértice de origem e o de destino da aresta. A Figura 2a ilustra o histórico de encontros de um paciente, a jornada extraída a partir dele, e a lista de arestas do MAG proposto para representar essa jornada. A Figura 2b apresenta a representação visual desse MAG.

Como cada aresta contém um atributo com o intervalo de tempo e outro com o identificador do paciente que a percorreu, temos, na verdade, um multigrafo multiaspecto.



**Fig. 2:** Esquema ilustrativo do processo de conversão de dados de pacientes no MAG proposto.

Isto é, entre um mesmo par de vértices, múltiplas arestas podem existir. No entanto, o modelo pode ser facilmente convertido em um dígrafo multiaspecto.<sup>4</sup> Para isso, as arestas passariam a ter como atributo uma lista de pares (*paciente, intervalo*).

Na área de ciência de redes, métricas de centralidade são usadas para medir a importância dos vértices da rede (grafo). Uma dessas métricas é o grau, que indica quão conectado um vértice é. No caso de redes direcionadas, tem-se o grau de entrada (*in degree*) que indica o número de arestas que chegam até o vértice, e o grau de saída (*out degree*) que mede quantas arestas partem dele.

Para o MAG proposto como modelo de jornadas de pacientes, quando a sua versão multigrafo é usada, o grau do vértice indicará quantos pacientes chegam/saem dele. Já quando a versão dígrafo é considerada, o grau mede quantos vértices diferentes enviam/recebem arestas do vértice em questão. Outro ponto importante é que na versão multigrafo o grau de chegada é igual ao grau de saída para todos os vértices, com exceção do vértice de início e de fim de cada jornada. Isso ocorre pois cada aresta está associada a um único paciente, e todos os pacientes iniciam sua jornada no vértice virtual “Início” e a terminam no vértice virtual “Fim”; os pacientes “chegam” e “saem” de todos os eventos intermediários. No caso do dígrafo, os graus de chegada e de saída não são necessariamente iguais, pois indicam quantos vértices distintos se conectam ao vértice em questão.

Além da flexibilidade de poder ser convertido em multigrafo ou dígrafo dependendo do objetivo buscado, o modelo proposto também pode ser *subdeterminado* para viabilizar diferentes tipos de análise. O processo de subdeterminação (Wehmuth et al., 2016) consiste em converter o MAG original em uma versão baseada em um subconjunto de seus aspectos. No caso das jornadas de pacientes, por exemplo, seria possível desprezar os atendimentos realizados e considerar apenas a evolução temporal de unidades visitadas.

Ao acompanhar o tratamento de pacientes e suas visitas a diferentes estabelecimentos de saúde, identificar as unidades e os procedimentos mais utilizados pode amparar análises e tomadas de decisão relacionadas com o serviço de saúde prestado. Torna-se possível, por exemplo, verificar se há predomínio de uma unidade específica para algum procedimento.

<sup>4</sup>Um dígrafo é um grafo direcionado no qual cada par de vértice é ligado por no máximo uma aresta.



Se houver, isso pode implicar que o não funcionamento daquela unidade para aquele procedimento (p.ex. falha em um equipamento de exame) atingirá uma parcela maior de pacientes, resultando no cancelamento da execução do procedimento, pois a realocação não é imediata. O paciente pode ser ou não comunicado com antecedência e neste caso terá que providenciar novo agendamento, dependendo do município que este paciente se encontra. No SUS nem sempre a realocação de atendimento é feita de modo automático por conta da complexidade que é fazer isto. Compreender a estrutura da rede nesse sentido pode auxiliar a tomada de decisão na alocação de recursos, atuação preventiva e até em ter unidades de saúde de retaguarda para procedimentos críticos.

## 4 Resultados

A avaliação do modelo proposto se deu em três etapas. Na primeira, foram considerados os três aspectos em conjunto; em seguida, o modelo foi subdeterminado e os aspectos Atendimento e Unidade foram usados; por fim, avaliou-se o modelo no aspecto Unidade. As análises foram feitas em *python* utilizando o módulo MAG desenvolvido por Mascarenhas (2019) e o pacote *NetworkX* (Hagberg et al., 2008).

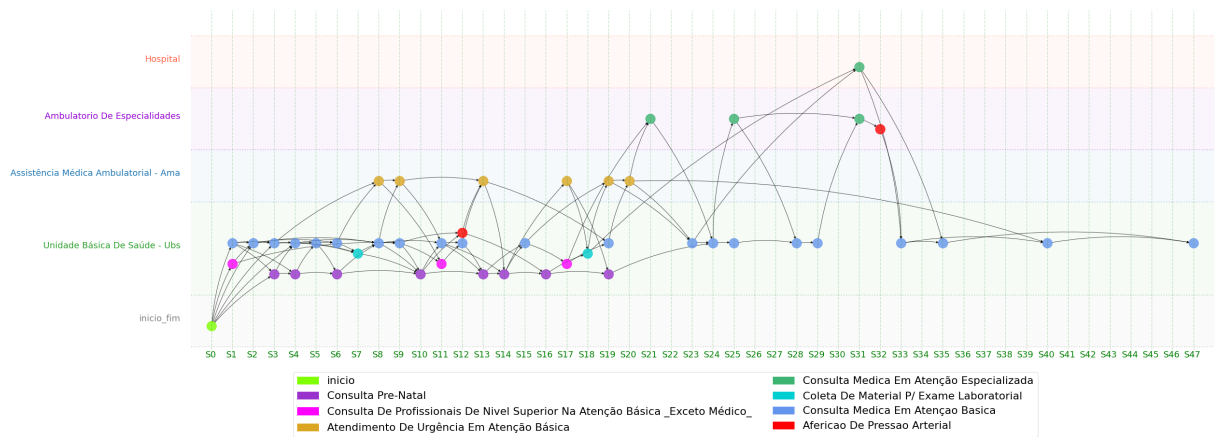
### 4.1 Mantendo os três aspectos do MAG

Inicialmente verificou-se a viabilidade do modelo de jornadas com base no MAG para representar visualmente a jornada de pacientes. A Figura 3 apresenta o multigrafo multiaspecto envolvendo as jornadas de um subconjunto de pacientes classificadas como gestantes de risco por registrarem diagnósticos associados a diabetes e a hipertensão.<sup>5</sup> Foram dez as pacientes com as duas doenças, e apenas vértices que ocorreram para ao menos duas delas são apresentados. É destacado que para esta análise foram usados os tipos de unidade de saúde ao invés dos estabelecimentos exatos<sup>6</sup>, pois o objetivo é verificar o uso dos diferentes níveis de cuidado providos pelo SUS.

Podemos observar que os primeiros eventos estão associados a visitas a Unidades Básicas de Saúde (UBS). A UBS corresponde ao nível de atenção primária, sendo a porta de entrada pela qual o paciente recebe um atendimento com complexidade mais baixa, o que se espera de uma gestação sem complicações (Mendes, 2011). Este padrão é percebido no gráfico, mesmo considerando as gestantes de risco. A partir do oitavo evento das jornadas, destacam-se visitas a unidades de Assistência Médica Ambulatorial (AMA) devido a ocorrências de urgência. Sendo estas gestantes de risco é esperado que algo assim aconteça, porém ter a noção em que momento do tempo isso ocorre com mais frequência permite que medidas preventivas sejam tomadas para diminuir a necessidade destas intervenções, principalmente por se tratar de atendimentos de urgências. Saber quando adotar medidas para diminuir a necessidade de ida da gestante a consultas de

<sup>5</sup>Para a seleção dessas pacientes, recorreremos à base de dados para identificar quais pacientes registraram ao menos um diagnóstico de diabetes e um de hipertensão. Em seguida, filtramos o multigrafo multiaspecto de modo que apenas arestas e, conseqüentemente, vértices referentes a essas pacientes fossem mantidos.

<sup>6</sup>O MAG foi criado de modo que o aspecto Unidade contivesse os nomes das unidades de saúde. A informação sobre o tipo delas foi adicionada como atributo dos vértices. Para trabalhar com os tipos de unidade, os vértices referentes ao mesmo tipo e com os mesmos elementos do aspecto Atendimento e Sequência foram contraídos.



**Fig. 3:** Jornadas de gestantes diabéticas e hipertensas, considerando apenas eventos que ocorreram para ao menos duas pacientes.

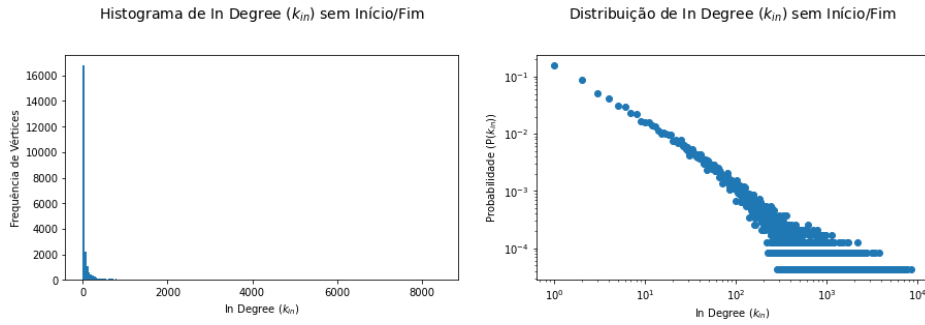
urgência é o desejável. A AMA e o Ambulatório de Especialidades fazem parte dos níveis secundários e o registro nesses tipos de unidades permite identificar quando estas gestantes tiveram algum tipo de complicação ou necessidade de maior atenção por conta da condição delas de hipertensas e diabéticas.

Destacamos que há arestas na Figura 3 que ligam pontos não consecutivos do aspecto Sequência. Elas ocorrem porque, como os vértices do grafo foram filtrados, optou-se por criar arestas que conectam os eventos frequentes da jornada de cada paciente. Por exemplo, se uma gestante possuía duas arestas –  $(A_1, U_1, S_1, A_2, U_2, S_2)$  e  $(A_2, U_2, S_2, A_3, U_3, S_3)$  – mas apenas os vértices  $(A_1, U_1, S_1)$  e  $(A_3, U_3, S_3)$  possuíam frequência suficiente para serem mantidos na rede, então a jornada da paciente passaria a ser representada apenas com a aresta  $(A_1, U_1, S_1, A_3, U_3, S_3)$ .

#### 4.2 Subdeterminando o MAG para manter apenas os aspectos Atendimento e Unidade

A distribuição de grau de chegada ( $k_{in}$ ) da versão subdeterminada do MAG de jornadas de gestantes que mantém apenas os aspectos Atendimento e Unidade é apresentada na Figura 4. Nesta subseção e na próxima, tratamos das unidades enquanto estabelecimentos específicos, ao invés de categorizá-las por seus tipos. O grau foi calculado a partir da versão multigrafo do MAG e, portanto, indica o número de ocorrências de cada vértice. Como o vértice virtual que indica o fim das jornadas possui grau de chegada elevado, igual ao total de gestantes, ele não é apresentado nos gráficos da Figura 4. O vértice virtual que marca o início das jornadas (grau de chegada nulo) também não é apresentado.

Pode-se observar que quase metade (47.56%) dos vértices (combinações de atendimento e unidade) possuem grau relativamente baixo ( $k_{in} \leq 10$ ), mas há poucos vértices (3.32%) com grau elevado ( $k_{in} > 1000$ ). Ao identificar os vértices de maior grau de chegada (Figura 5), foi verificado que o vértice de maior grau é o vértice virtual que indica o final das jornadas, o que era esperado já que todas as pacientes chegam nele uma vez. Nota-se que o procedimento “Atendimento de Urgência em Atenção Básica” associado a unidades AMA ocorre três vezes na lista, e o atendimento “Consulta Médica em Atenção Especializada” ocorre duas vezes, associado a dois hospitais diferentes. Também apresen-



**Fig. 4:** Histograma de frequência e Distribuição de probabilidade de grau de chegada (*in degree*) dos vértices do multigrafo multispecto, subdeterminado para manter os aspectos *A* e *U*.

taram grau alto duas unidades provendo consultas pré-natais, o atendimento domiciliar de uma UBS, e uma unidade que realiza ultrassonografias obstétricas.

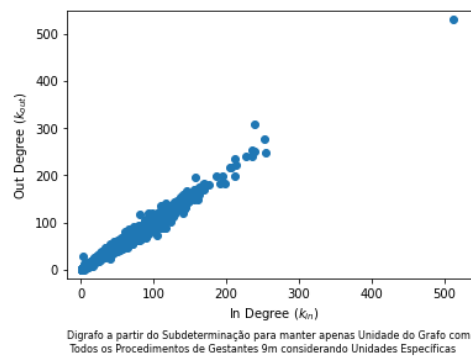
Atendimento	Unidade	Grau de Chegada
fim	inicio_fim	137503
Atendimento De Urgência Em Atenção Básica	Ama Parelheiros	8434
Atendimento De Urgência Em Atenção Básica	Ama Paraisópolis	7612
Consulta Medica Em Atenção Especializada	Hosp Dia Da Rhc M Boi Mirim I	7446
Consulta Medica Em Atenção Especializada	Hosp Mun Cachoeirinha Mario De Moraes Altenfelder Silva	7387
Visita Domiciliar Por Profissional De Nível Médio	Ubs Parque Residencial Cocaia Independente	7012
Consulta Pre-Natal	Ubs / Nir Jardim Soares	6873
Atendimento De Urgência Em Atenção Básica	Ama Capao Redondo	6625
Ultra-Sonografia Obstetrica	Casa Ser Maria Auxiliadora Lara Barcelos	6408
Consulta Pre-Natal	Ama/Ubs Integrada Jd Das Oliveiras	6261

**Fig. 5:** Vértices com maiores graus de chegada (*in degree*) no multigrafo multispecto, subdeterminado para manter os aspectos Atendimento e Unidade.

#### 4.3 Subdeterminando o MAG para manter apenas o aspecto Unidade

Fazendo uma subdeterminação para manter apenas o aspecto Unidade, é possível verificar como é a movimentação de pacientes entre unidades. Pode-se identificar, por exemplo, qual é a relação entre o grau de chegada e o de saída de cada unidade. Com isso, é possível verificar se existem unidades que recebem pacientes de muitas unidades diferentes e os que concentram em poucas unidades de destino, ou se os graus de chegada e de saída são similares. A Figura 6 apresenta um diagrama de dispersão do grau de chegada comparado ao grau de saída de cada vértice da rede subdeterminada para manter apenas o aspecto Unidade, convertida na versão de dígrafo. Os vértices virtuais de início e fim das jornadas não estão incluídos no gráfico.

Observou-se que para a maior parte das unidades o grau de chegada é similar ao de saída. No entanto, neste trabalho não investigamos se isto é uma coincidência ou se as unidades das quais um estabelecimento recebe pacientes são as mesmas para as quais ele envia pacientes. Este segundo cenário permite levantar a hipótese de que há grupos de unidades que frequentemente partilham pacientes; neste caso, o gestor pode direcionar estudos específicos para determinar o que acontece nestas unidades e tomar as medidas necessárias para administrar a situação. Outra característica que se destaca é que uma unidade (no canto superior direito) possui grau de chegada/saída muito superior ao das demais: o *Instituto Cema De Oftalmologia e Otorrinolaringologia*, com grau de chegada de 512 e grau de saída de 529. Este tipo de constatação chama a atenção e permite que



**Fig. 6:** Diagrama de dispersão com o grau de chegada (*in degree*) e de saída (*out degree*) dos vértices do dígrafo multiaspecto obtido pela subdeterminação que mantém apenas o aspecto *U*.

o gestor possa investigar melhor estas ocorrências. Com este tipo de análise é possível encontrar as discrepâncias que necessitam de uma análise mais aprofundada.

## 5 Conclusões

O uso do MAG com os aspectos Atendimento, Unidade e Sequência para modelar jornadas de gestantes foi avaliado tanto como modelo visual de representação das jornadas, quanto como ferramenta para revelar a estrutura da rede de saúde que dá suporte a essas pacientes.

Visualmente, o MAG foi capaz de apresentar jornadas com repetição de atividades sem prejuízo na interpretação sequencial dos eventos ocorridos. Além disso, possibilitou avaliar as jornadas não considerando apenas os procedimentos realizados, mas também os tipos de unidade visitados. No entanto, a visualização se restringiu a um conjunto pequeno de gestantes.

Ao adotar o recurso das subdeterminações, foi possível identificar que uma pequena parcela dos pares de atendimento e unidade de saúde são significativamente mais frequentes que os demais. Dentre os pares de maior frequência, destacam-se procedimentos e unidades relacionados à atenção especializada, aos atendimentos de urgência e à realização de exames.

Considerando outro tipo de subdeterminação, para manter apenas o aspecto Unidade, foi possível comparar os graus de chegada e de saída de diferentes unidades de saúde para avaliar de/para quantos estabelecimentos distintos cada unidade recebe/envia pacientes. Observou-se que, em geral, as unidades possuem graus de chegada e saída parecidos. Nas próximas etapas da pesquisa, pretende-se investigar a causa da semelhança dos graus; uma hipótese é que o alto compartilhamento de pacientes entre as unidades ocorre porque há especialidades (profissionais de saúde) ou procedimentos que se complementam entre as unidades (uma tem e a outra não tem) o que faz com que seja necessário esta troca. Em se comprovando isto, o estudo de distribuição de recursos pode ser necessário para verificar se é saudável este tipo de compartilhamento ou não.

O modelo de jornadas proposto neste trabalho mostrou a flexibilidade que o modelo traz para analisar os dados de jornadas de pacientes sob diferentes perspectivas, trabalhando-os de diferentes formas. Além disso, este estudo foi feito sem a necessidade de utilizar registros clínicos de pacientes, que são dados de maior sensibilidade.





Futuramente, além da investigação do relacionamento de estabelecimentos de saúde, pretende-se preparar um método de filtragem de dados baseado não exclusivamente na frequência para viabilizar a representação visual das jornadas de um conjunto maior de pacientes, e avaliar outras métricas de centralidade de rede que possam revelar outras características do modelo de jornadas de pacientes e de suas versões subdeterminadas. Há também interesse em identificar grupos de pacientes com jornadas semelhantes levando em conta não apenas a ordenação dos atendimentos (aspecto “Sequência”), mas também os intervalos de tempo decorridos entre eles (atributo “intervalo” das arestas do MAG proposto).

## 6 *Agradecimentos*

Agradecemos à CAPES, FAPERJ, CNPq, e ao INCT-MACC2 pelo apoio financeiro.

## Referências

- Antonelli, D., Baralis, E., Bruno, G., Chiusano, S., Mahoto, N. A., e Petrigni, C. (2012). “Analysis of diagnostic pathways for colon cancer”. *Flexible Services and Manufacturing Journal*, 24(4):379–399.
- Aspland, E., Harper, P. R., Gartner, D., Webb, P., e Barrett-Lee, P. (2021). “Modified Needleman–Wunsch algorithm for clinical pathway clustering”. *Journal of Biomedical Informatics*, 115.
- Chiudinelli, L., Dagliati, A., Tibollo, V., Albasini, S., Geifman, N., Peek, N., Holmes, J. H., Corsi, F., Bellazzi, R., e Sacchi, L. (2020). “Mining post-surgical care processes in breast cancer patients”. *Artificial Intelligence in Medicine*, 105.
- De Oliveira, H., Augusto, V., Jouaneton, B., Lamarsalle, L., Prodel, M., e Xie, X. (2020). “Optimal process mining of timed event logs”. *Information Sciences*, 528:58–78.
- Duma, D. e Aringhieri, R. (2020). “An ad hoc process mining approach to discover patient paths of an Emergency Department”. *Flexible Services and Manufacturing Journal*, 32(1):6–34.
- Egho, E., Raïssi, C., Calders, T., Jay, N., e Napoli, A. (2015). “On measuring similarity for sequences of itemsets”. *Data Mining and Knowledge Discovery*, 29(3):732–764.
- Günther, C. W. e van der Aalst, W. M. P. (2007). “Fuzzy mining – adaptive process simplification based on multi-perspective metrics”. In: Alonso, G., Dadam, P., e Rosemann, M., editors, *Business Process Management*, Springer Berlin Heidelberg, págs. 328–343.
- Hagberg, A. A., Schult, D. A., e Swart, P. J. (2008). “Exploring network structure, dynamics, and function using networkx”. In: Varoquaux, G., Vaught, T., e Millman, J., editors, *Proceedings of the 7th Python in Science Conference*, págs. 11–15.
- Jasmim, L., Ziviani, A., Ito, M., e Paiva, P. (2017). “Caracterização de atendimentos em uma rede de atenção à saúde”. In: *Anais do XVII Workshop de Informática Médica*, SBC, págs. 2044–2047.

- Kempa-Liehr, A. W., Lin, C. Y. C., Britten, R., Armstrong, D., Wallace, J., Mordaunt, D., e O’Sullivan, M. (2020). “Healthcare pathway discovery and probabilistic machine learning”. *International Journal of Medical Informatics*, 137.
- Leonardi, G., Striani, M., Quaglini, S., Cavallini, A., e Montani, S. (2018). “Leveraging semantic labels for multi-level abstraction in medical process mining and trace comparison”. *Journal of Biomedical Informatics*, 83:10–24.
- Mascarenhas, J. Z. G. (2019). *Modelagem e Análise Temporal da Rede de Transações de uma Plataforma de Consenso Distribuído*. LNCC, Petrópolis, RJ.
- Mascarenhas, J. Z. G., Ziviani, A., Wehmuth, K., e Vieira, A. B. (2020). “On the transaction dynamics of the Ethereum-based cryptocurrency”. *Journal of Complex Networks*, 8(4). cnaa042.
- Mendes, E. V. (2011). *As Redes de Atenção a Saúde*. 2 ed.
- Najjar, A., Reinharz, D., Girouard, C., e Gagné, C. (2018). “A two-step approach for mining patient treatment pathways in administrative healthcare databases”. *Artificial Intelligence in Medicine*, 87:34–48.
- Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L., e Xie, X. (2018). “Optimal Process Mining for Large and Complex Event Logs”. *IEEE Transactions on Automation Science and Engineering*, 15(3):1309–1325.
- Rebuge, Á. e Ferreira, D. R. (2012). “Business process analysis in healthcare environments: A methodology based on process mining”. *Information Systems*, 37(2):99–116.
- Rinner, C., Helm, E., Dunkl, R., Kittler, H., e Rinderle-Ma, S. (2018). “Process mining and conformance checking of long running processes in the context of melanoma surveillance”. *International Journal of Environmental Research and Public Health*, 15(12).
- Rosa, C. d. O. C. S., Ito, M., Vieira, A. B., e Gomes, A. T. A. (2022). “Modelling and mining of patient pathways: A scoping review”. *arXiv*.
- Sato, D. M., Mantovani, L. K., Safanelli, J., Guessier, V., Nagel, V., Moro, C. H., Cabral, N. L., Scalabrin, E. E., Moro, C., e Santos, E. A. (2020). “Ischemic stroke: Process perspective, clinical and profile characteristics, and external factors”. *Journal of Biomedical Informatics*, 111(1):103582.
- Wehmuth, K., Costa, B., Bechara, J. V., e Ziviani, A. (2018). “A multilayer and time-varying structural analysis of the brazilian air transportation network”. In: Ziviani, A., Hara, C. S., Ogasawara, E. S., de Macêdo, J. A. F., e Valduriez, P., editors, *Proceedings of the Latin America Data Science Workshop co-located with 44th International Conference on Very Large Data Bases (VLDB 2018), Rio de Janeiro, Brazil, Aug 27, 2018*, CEUR-WS.org, págs. 57–64.
- Wehmuth, K., Fleury, É., e Ziviani, A. (2016). “On MultiAspect graphs”. *Theoretical Computer Science*, 651:50–61.



# Predição *in silico* de compostos bioativos contra SARS-COV-2

Elcio Ferreira Santana<sup>1</sup>, Adriano Carlos Soares<sup>2</sup>

<sup>1</sup>Centro Universitário Vértice -

Univértix, Matipó/MG, Brasil <sup>2</sup>

---

## Resumo

A pandemia do novo coronavírus se mostrou contagiosa e letal na população humana nos últimos anos, a variante ômicron ou B.1.1.529 demonstrou algumas mutações na proteína *spike* (pico ou S), em seu domínio de ligação ao receptor na unidade S1 que apresentou maior afinidade a Enzima Conversora de Angiotensina 2 (ECA2). Este estudo demonstrou a análise de proteínas e ligantes presentes em seus domínios para serem extraídos e analisados a partir do *docking* molecular para possível terapêutica, dentre as 22 proteínas, que apresentaram em seu domínio ligantes semelhantes ao da estrutura 7ZXU (obtida no *Protein Data Bank*). Na análise do índice de instabilidade, ponto isoelétrico e meia vida e *Grand Average of Hydropathicity (GRAVY)*, no software suíço *ProtParam* foram refinados 5 compostos e seus ligantes; pela acetilação refinados 2 compostos e depois 1 deles apresentou ligante e sua estrutura foi validada e preparada. Seus ligantes foram extraídos no banco de dados *Protein Data Bank*, os resultados do *docking* por *AutoDock* demonstraram energia livre favorável aos aminoácidos e o *DockThor* mostrou uma energia diferente justificada pelo grupo químico ligado. Os resultados mostraram ligações aos aminoácidos inibidores o que sustentou a hipótese que um composto molecular obtido a partir das proteínas pesquisadas (alinhadas a partir do composto “2-acetamido-2-deoxy-beta-D-glucopyranose” da proteína 7JMP) (PDB DOI: 10.2210/pdb7JMP/pdb) pode ser um potencial inibidor da proteína *spike*.

Palavras-chave: Variante ômicron. Docking molecular. Bioinformática.

---

# 1 INTRODUÇÃO

O novo coronavírus é um vírus causador da pandemia de SARS-COV-2 iniciada no ano de 2020, sua afinidade pela Enzima Conversora de Angiotensina 2 (ECA2) permitiu que infectasse os seres humanos e causasse sinais e sintomas respiratórios, inclusive a síndrome respiratória aguda grave que é um dos fatores de risco para a saúde, principalmente idosos; Khan *et al* [2020]; Wu *et al* [2020]. Em sua região codificadora há duas regiões principais: pp1a (capaz de codificar proteínas NSP1 ao NSP11), pp1ab (codifica do NSP1 ao NSP16) e a região que codifica as proteínas estruturais M, S (*spike* ou pico), E, N; Robson *et al*, [2020]; Haq *et al* [2021].

Com isso, a variante ômicron surgiu em 2021 como uma das mais preocupantes devido a sua alta taxa de contaminação. Esse fator levou a ondas de casos confirmados no mundo; em seu domínio de ligação ao receptor foi observado 15 mutações fundamentais para infecção, essas mutações dificultam o tratamento e a produção de vacinas para o combate; Alkhatib *et al* [2022]. Seu domínio de ligação ao receptor da proteína de membrana *spike* (pico) pode ser representado por uma sequência de 202 aminoácidos que em uma de suas pontas apresenta 25 aminoácidos responsáveis pela sua interação com a ECA2 para realizar a infecção e os resíduos 452 e 486 também participam da interação como potentes inibidores; Tuekprakhon *et al* [2022]; He *et al* [2021]. A ômicron infelizmente ainda não possui uma vacina ou medicamento que combata o vírus diretamente, os avanços da tecnologia demonstram que a alta taxa de mutações na variante dificulta esse processo que ainda pode levar tempo o que preocupa muito os cientistas; Khan *et al* [2020].

Observando esses dados, na bioinformática os softwares permitem a partir de cálculos de dinâmica molecular, mecânica quântica, estatísticos e muitos outros, fazer a predição *in silico* de compostos a partir de experimentação comprovada e aplicada nesses *softwares*. Essa abordagem pode ser muito bem utilizada por diversas instituições de ensino, empresas e laboratórios, economizando muito tempo e dinheiro nas pesquisas; Kaur, Madgulkar, Bhalekar e Asgaonkar [2019].

Observando a urgência que a pandemia do novo coronavírus causou, se faz necessária a busca por compostos terapêuticos que auxiliem na luta contra esse vírus. Como mencionado a técnica da bioinformática utiliza softwares para predição *in silico* de compostos bioativos; o *docking* molecular (encaixe molecular) é uma técnica que utiliza a mecânica molecular, cálculos termodinâmicos e mecânica quântica a partir de softwares desenvolvidos por grandes instituições. Para isso se utiliza uma molécula alvo e um ligante que irá fazer a interação com o alvo, essa *docking* permite obter as poses e as interações que ocorrem entre os compostos estudados e fornecendo assim dados para pesquisa de novos fármacos; Sanos; Ferreira e Caffarena [2019]. É demonstrado que os estudos *in silico* são bastante eficientes, economizam tempo e dinheiro, o que é de suma importância visto a urgência da pandemia, ao contrário de estudos *in vitro* que muitas das vezes são demorados e demandam enormes custos financeiros; Soares [2015]. Portanto a abordagem de *docking* molecular permite a predição e triagem de compostos bioativos moleculares para combate a patógenos, visando a validação por cálculos muito bem

estruturados e propriedades químicas dos compostos estudados; Sanos; Ferreira e Caffarena [2019]. Logo, o objetivo deste estudo é a realização da predição *in silico* de compostos bioativos candidatos a novos compostos bioativos contra a proteína *spike* codificada pela variante ômicron do SARS-CoV-2 através do *docking* molecular.

## 2 METODOLOGIA

Se trata de um estudo descritivo com abordagem quantitativa, procura delinear softwares que realizam cálculos de *docking* molecular para obtenção de afinidade da proteína *spike* com compostos bioativos. Sendo também uma pesquisa de caráter qualitativa; Cathala e Moorley [2018].

Para que seja feita uma análise de compostos contra o novo coronavírus é necessário a busca pelo *Receptor Binding Domain* (que fornece o local exato da interação da *spike*) da proteína *spike* da variante ômicron; Lan *et al* [2020]. No banco de dados do *Protein Data Bank* (<https://www.rcsb.org/>) obtendo a proteína complexada com algum composto neutralizante que será utilizado para estudo de novos ligantes, fornecendo informações sobre o ligante da *spike* para a predição de novos compostos; Soares [2015].

Para obtenção de compostos semelhantes ao ligante foi feita a busca no banco de dados *Uniprot (Universal Protein)*, a busca por proteínas que apresentem domínios funcionais com identidade favorável ao ligante, Morgat *et al* [2020].

Os cálculos de alinhamento foram feitos a partir do *software BLAST (Basic Local Alignment Search Tools)* (Ferramenta Básica de Pesquisa de Alinhamento Local) (<https://www.uniprot.org/blast/>) que fornece o alinhamento de proteínas semelhantes a obtida anteriormente e que estejam depositadas no banco de dados do *Swiss Prot* (Proteínas Suíças) (<https://www.sib.swiss/swiss-prot>) com estruturas validadas Soares [2015]. Posteriormente será feito o alinhamento dos dados obtidos no *software MEGA11 (Molecular Evolutionary Genetics Analysis)* (Análise de Genética evolutiva molecular) (<https://www.megasoftware.net/>) que possibilitará a retirada de peptídeo de sinal, caso haja; Tamura, Stecher, Kumar, [2021].

Na predição de possíveis ancestrais comuns e relações de ancestralidade, o *software MEGA11* é capaz de fazer esse papel, com a sequência *FASTA* (sequência de aminoácidos de uma proteína) fazendo diferentes análises gerando uma *Topology Only* Soares [2015].

O estudo de parâmetros físico-químicos é imprescindível para realização de posteriores testes *in vitro*, servindo como um funil para economia de tempo e recursos. Os melhores parâmetros para predição são o índice de estabilidade, Ponto isoelétrico, meia vida estimada e *Grand average of hydropathicity* (Grande média de hidropatia) (*GRAVY*), utilizando o *software ProtParam (Protein Parameters)* Parâmetros Proteicos, *Expasy (Expert Protein Analysis System)* (Sistema Especialista de análises de Proteínas) (<https://web.expasy.org/protparam/>); Dutta, Banerjee, Chakraborty [2018].

O coronavírus necessita de uma acetilação importante na sua interação. Portanto, a predição destes locais é fundamental para o método de *docking* de melhor estado espacial

entre as moléculas. Este visa prever interações em sítios de ligação e ativação das moléculas para melhorar a interação proteína-ligante. Foram utilizados os servidores *NetAcet* (N-terminal acetylation in eukaryotic proteins) (Acetilação N-terminal em proteínas eucarióticas) e *Phosphorylation Sites Prediction* (Previsão de Locais de Fosforilação) (*NetPhos*, <https://www.cbs.dtu.dk/service.php?NetPhos>); Foroutan *et al* [2021].

Esta etapa contribuirá no entendimento da percepção e recebimento celular, esse mecanismo ajuda a analisar como o possível composto bioativo irá “navegar” nos meios extracelulares e intracelulares. Para essa análise, será utilizado os servidores *Prediction of transmembrane helices in protein* (TMHMM versão 2.0, <https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) e a análise de seus depósitos no *Uniprot* para localização celular; Foroutan *et al* [2021].

Esta etapa será importante para conseguir a predição das conformações de enovelamento das moléculas com potencial bioativo e também da proteína alvo, então será possível obter uma noção de sua estrutura nas etapas seguintes. Usando o servidor e o *Secondary Structure Prediction – NPS* (SOPMA, [https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)) para prever e validar a estrutura secundária. Servidores como o *Protein Data Bank* também oferecem uma estrutura secundária da variante ômicron e do coronavírus, o que facilitará a análise de docking molecular; Ghaffari *et al* [2021].

Essa etapa tem como objetivo identificar a polaridade de cadeias laterais e hidrofobicidade, conseguir melhor alinhamento possível e determinar o modelo 3D. Sendo assim, foi utilizado o banco de dados *SWISS-MODEL* onde essas proteínas foram validadas e refinadas a partir de seus depósitos e do gráfico de Ramachandran; Motamedpour, Dalimi, Pirestani, Ghaffarifar [2020].

Após a predição de possíveis ligantes obtidos das proteínas até então preditas foi realizado, o método de docagem (*docking* molecular) que visa a análise entre o ligante e a molécula alvo, estimando a afinidade ligante-molécula alvo (pontuação, score). Esta etapa utiliza a conformação e orientação do ligante (pose), analisando a complementariedade molecular entre o ligante e o alvo. O *software* utilizado foi o *AutoDock* e o servidor *DockThor*, ambos com ligante rígido, conseguindo prever a energia de ligação e afinidade, respectivamente; Goodsell, Sanner, Olson, Forli, [2020]; Guedes *et al* [2021].

Para esse processo são necessários alguns parâmetros fundamentais. Os parâmetros de cálculos considerados nessa função de busca serão: para o *DockThor* baseados no campo de força MMFF94S que realiza os cálculos a partir de sua função de avaliação, utiliza o sistema *grid* (grade) para representação do atrancamento mapeando as interações de van der Waals e eletrostáticas, gerando uma grade no local onde será direcionado o *Docking* a partir das coordenadas inseridas. O ligante depositado no servidor é adequado ao campo de força MMFF94S utilizando o *OpenBabel* analisando ligação e grupos para realização de cálculos e para o domínio de ligação ao receptor (RBD) da *spike*, o *PdbThorBoxO* define átomos de proteínas, ajusta ao campo de força do servidor e ajusta cadeias laterais. Já para o *AutoDock* é utilizado para os cálculos uma função de avaliação híbrida (Empírico e

Campo de Força), sendo gerada uma *grid box* (caixa de grade) formada na região tridimensional fornecida a partir das coordenadas inseridas e abrangendo o sítio alvo desejado, é possível fazer todas as ligações não rotacionáveis a partir da verificação dos ângulos de torção, os hidrogênios não polares são adicionados no próprio software e para realização dos cálculos os comandos são efetuados no *prompt* de comando onde os pacotes instalados “AD4.1\_bound”, “Autodock4.exe” e “Autogrid4.exe” são fundamentais para execução das tarefas de *docking*, tendo, logo após, os resultados avaliados no melhor *cluster* na interface do *software*. Avaliando também a análise de campo de força (visando todas as interações entre o complexo). Para validação final, foi utilizado as conformações e orientação do ligante (pose) no complexo, observado as energias de afinidade, Entropia e energia livre de ligação fornecidas no *AutoDock* e *DockThor*; Goodsell, Sanner, Olson, Forli, [2020]; Guedes *et al* [2021].

A análise de resultados é o processo final no qual se busca obter a real validação do complexo ligante-molécula alvo predito no processo de *docking* molecular. Foram analisados os valores de pontuação no menor nível de energia calculada do complexo, maior a estabilidade e melhor pose. É necessário ainda, analisar as interações ocorridas entre o ligante e o sítio de ligação, principalmente as ligações de hidrogênio que são extremamente importantes e outras. Essas análises demonstram como está sendo o reconhecimento molecular e permite a estruturação de pose. Por se tratar de um *docking* consenso, será buscada a energia de afinidade fornecida pelo *DockThor* e a energia livre do *AutoDock*. Caso os parâmetros sejam satisfatórios a molécula então apresentará um potencial bioativo contra a proteína analisada contra o SARS-CoV-2; Heinzelmann e Gilson [2021].

### 3 RESULTADOS E DISCUSSÕES

#### 3.1 Busca pelo alvo com seu ligante

O banco de dados utilizado para a busca do ligante foi o *Protein Data Bank* (<https://www.rcsb.org/>), nele está depositado várias proteínas e assim foi procurado a proteína *spike* da variante ômicron. Foi encontrado a proteína *spike* complexada com anticorpo neutralizante com resolução de 1,78Å (PDB DOI: 10.2210/pdb7JMP/pdb) de onde havia o ligante com a nomenclatura: “2-acetamido-2-deoxy-beta-D-glucopyranose”, com fórmula molecular C<sub>8</sub> H<sub>15</sub> O<sub>16</sub>; Tuekprakhon *et al* [2022].

#### 3.2 Alinhamento do ligante

No banco de dados *UniprotKB* (<https://www.uniprot.org/help/uniprotkb>) o ligante foi pesquisado e resultou em 1 proteína com o ligante complexado com o código: Q96LB9, apresenta 340 resíduos de aminoácidos e os 17 primeiros sendo peptídeo de sinal. O alinhamento foi feito pelo servidor americano *BLASTP* que consegue alinhar os resíduos de aminoácidos e entregar proteínas com identidade desejada. Nessa busca foi utilizado como parâmetro serem proteínas depositadas no *Swiss Protein (SwissProt)* o que resultou num total de 36 proteínas. Como essas proteínas precisam ser refinadas foram considerados valores acima de 40% de identidade, resultou assim em 22 sequências de proteínas com os seguintes identificadores: Q96LB9, A1A547, Q96LB8, Q0VB07, O75594, B5T255, Q8SPP7, O88593, Q9GK12, Q9JLN4, Q9VYX7, O76537, Q70PY2, Q765P3, Q70PR8, Q9VS97, Q9VXN9, Q765P4, Q8VCS0, Q866Y3, Q96PD5 e Q8SXQ7; Liberman [2004]. Após essa análise foi consultado no *Uniprot* a localização do peptídeo sinal de cada

proteína para que fossem removidos, logo após o software *PyMOL* (<https://www.schrodinger.com/products/pymol>) permitiu a edição de cada proteína para retirada dos mesmos. Posteriormente suas sequências de aminoácidos foram alinhadas novamente, para isso o software *MEGA11* alinhou os códons de todas elas. Em teste estatístico de significância feita pelo software *bioEstat 5.0* (<https://bioestat.software.informer.com/>) foi utilizado para o cálculo do *p-value* obtendo o resultado *p-value* < 0,0001 o que demonstra confiabilidade na pesquisa; Soares [2015].

### 3.3 Análise filogenética

Nessa fase o software *MEGA11* a partir do alinhamento de códons feito foi feita a análise e construção da árvore; utilizou o método *Construct/Test Neighbor-Joining Tree* demonstrando que mesmo a diferença entre as espécies a linhagem evolutiva está conectada, apresentando ancestrais comuns (TAMURA, STECHER, KUMAR, 2021).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
[ 1]																							
[ 2]	30.00																						
[ 3]	39.00	43.00																					
[ 4]	40.00	45.00	27.00																				
[ 5]	77.00	77.00	81.00	78.00																			
[ 6]	75.00	76.00	76.00	77.00	34.00																		
[ 7]	75.00	76.00	76.00	77.00	35.00	1.00																	
[ 8]	80.00	77.00	79.00	76.00	39.00	40.00	41.00																
[ 9]	79.00	76.00	81.00	78.00	31.00	27.00	27.00	33.00															
[10]	79.00	81.00	81.00	76.00	42.00	45.00	45.00	17.00	35.00														
[11]	77.00	79.00	86.00	83.00	78.00	81.00	81.00	77.00	80.00	82.00													
[12]	79.00	79.00	83.00	82.00	78.00	81.00	81.00	78.00	76.00	79.00	75.00												
[13]	83.00	87.00	87.00	86.00	88.00	91.00	91.00	87.00	87.00	89.00	75.00	80.00											
[14]	84.00	86.00	88.00	85.00	83.00	89.00	89.00	84.00	87.00	85.00	85.00	89.00	87.00										
[15]	82.00	82.00	83.00	83.00	81.00	81.00	81.00	81.00	80.00	83.00	77.00	79.00	85.00	93.00									
[16]	75.00	82.00	79.00	76.00	87.00	80.00	80.00	84.00	85.00	89.00	80.00	81.00	86.00	83.00	87.00								
[17]	84.00	86.00	89.00	86.00	83.00	87.00	87.00	84.00	85.00	85.00	84.00	90.00	87.00	4.00	95.00	83.00							
[18]	83.00	85.00	86.00	88.00	84.00	84.00	84.00	82.00	83.00	84.00	73.00	78.00	57.00	79.00	91.00	83.00	78.00						
[19]	81.00	85.00	83.00	85.00	88.00	90.00	90.00	95.00	91.00	96.00	88.00	87.00	90.00	92.00	87.00	87.00	93.00	85.00					
[20]	79.00	83.00	85.00	85.00	87.00	89.00	89.00	92.00	89.00	93.00	85.00	85.00	88.00	90.00	86.00	87.00	91.00	82.00	32.00				
[21]	81.00	86.00	88.00	89.00	83.00	88.00	88.00	91.00	86.00	92.00	86.00	86.00	90.00	93.00	87.00	90.00	94.00	83.00	25.00	21.00			
[22]	85.00	91.00	88.00	85.00	91.00	94.00	94.00	92.00	89.00	91.00	89.00	88.00	85.00	93.00	82.00	91.00	93.00	85.00	90.00	88.00	91.00		

**Figura 1:** Comparação das distâncias para par – número de diferenças.

No gráfico da figura 1 cada número corresponde as respectivas proteínas: Q96LB9, A1A547, Q96LB8, QOVB07, O75594, B5T255, Q8SPP7, O88593, Q9GK12, Q9JLN4, Q9VYX7, O76537, Q70PY2, Q765P3, Q70PR8, Q9VS97, Q9VXN9, Q765P4, Q8VCS0, Q866Y3, Q96PD5 e Q8SXQ7. Podemos observar o número de posições de cada proteína em relação à proteína superior, estas são vistas as mutações não sinônimas pois apresentaram troca do resíduo de aminoácido. Os resultados obtidos não influenciam na execução do *docking* molecular pois são seus compostos moleculares complexados que serão utilizados no *docking* Soares [2015].

### 3.4 Análise físico-química

A análise físico-química foi adicionado as sequências FASTA de todas as 22 proteínas encontradas no servidor suíço *Swiss ProtParam ExPasy* na análise todas possuíam valor de *GRAVY* negativo (o que demonstrava caráter hidrofílico), ressalta que todas as proteínas são solúveis em água e que não levará a precipitação; Kyte e Doolittle [1982]. O índice de instabilidade também foi feito para previsão de proteínas que são estáveis em sistemas biológicos, os valores considerados foram os menores que 40. Após a análise foram obtidos 8 compostos que apresentaram estabilidade e índice de instabilidade menores que 40, sendo elas: Q96LB8, Q9VYX7, O76537, Q765P3, Q70PR8, Q9VS97,



Q765P4 e Q8VCS0; Hoda, Tafaj, Sallaku [2021]. Logo após foram analisados os pontos isoelétricos de cada proteína, visto que a solubilidade e repulsão elétrica são menores nesse ponto, portanto apresentarão precipitação e agregação maior. Para isso foram consideradas proteínas que não tivessem seu ponto isoelétrico entre 7 e 8 para evitar assim problemas no pH fisiológico e os resultados demonstraram que as proteínas consideráveis são: Q9VYX7 (pI= 6,58), O76537 (pI= 6,49), Q70PR8 (pI= 5,79), Q9VS97 (pI= 5,56) e Q8VCS0 (pI= 6,49). A meia-vida estimada destes compostos considerando sua cadeia N-terminal a partir da metionina por testes *in vivo* parametrizados, os valores encontrados foram respectivamente para as 5 proteínas descritas anteriormente: >20 horas (fermento, *in vivo*), > 10 horas (*Escherichia coli*, *in vivo*); >20 horas (fermento, *in vivo*), > 10 horas (*Escherichia coli*, *in vivo*); >20 horas (fermento, *in vivo*), > 10 horas (*Escherichia coli*, *in vivo*); >20 horas (fermento, *in vivo*), > 10 horas (*Escherichia coli*, *in vivo*) e >20 horas (fermento, *in vivo*), > 10 horas (*Escherichia coli*, *in vivo*); Dutta, Banerjee, Chakraborty [2018].

### 3.5 Predição de locais de acilação e fosforilação

Na predição de locais de acetilação foi relatado em estudos recentes que a proteína spike utiliza de uma maquinaria de acetilação para facilitar sua entrada na célula hospedeira. Com isso o *software* dinamarquês *NetAcet* 1.0 foi utilizado para predição devido seu alto grau de confiança, os resultados foram: proteínas Q9VS97 apresentou acetilação no contexto MTWIG e um score de 0,487 e a Q8VCS0 apresentou acetilação no contexto MKAWGA e um score de 0,479; Foroutan *et al.*, 2021; Rouka, Gourgoulianis, Zarogiannis [2021].

**Tabela 1:** Resultado da predição da fosforilação

Sequência	Resíduo	Contexto	Score	Kinase
Q9VS97	12T	IVGLTAAIV	0.672	PKC
Q9VS97	42T	DSMETPLPR	0.807	unsp
Q9VS97	42T	DSMETPLPR	0.533	p38MAPK
Q9VS97	52T	VIAHTAGGA	0,511	cdc2
Q9VS97	64S	DVTCSQHMQ	0,573	DNAPK
Q9VS97	64S	DVTCSQHMQ	0,516	ATM
Q9VS97	76S	NFQMSKQKF	0,566	PKC
Q9VS97	81S	KQKFSDIGY	0,714	unsp
Q9VS97	81S	KQKFSDIGY	0,542	PKG
Q9VS97	100S	YEGRSPSQR	0,733	unsp
Q9VS97	100S	YEGRSPSQR	0,663	cdk5
Q9VS97	100S	YEGRSPSQR	0,508	GSK3
Q9VS97	102S	GRSPSQRGA	0,994	unsp
Q9VS97	102S	GRSPSQRGA	0,565	DNAPK
Q9VS97	115S	NNDGSLGIA	0,572	DNAPK
Q9VS97	115S	NNDGSLGIA	0,519	PKA
Q9VS97	163S	HRQVSATKS	0,988	unsp
Q9VS97	163S	HRQVSATKS	0,559	PKA

Q9VS97	163S	HRQVSATKS	0,549	PKG
Q9VS97	163S	HRQVSATKS	0,531	PKC
Q9VS97	165T	QVSATKSPG	0,548	cdc2
Q9VS97	167S	SATKSPGEA	0,993	unsp
Q9VS97	167S	SATKSPGEA	0,580	cdk5
Q9VS97	167S	SATKSPGEA	0,504	GSK3
Q9VS97	173Y	GEALYALIQ	0,587	unsp
Q8VCS0	24S	GAASSLPLL	0,626	PKA
Q8VCS0	58S	AWILSAKNS	0,632	unsp
Q8VCS0	58S	AWILSAKNS	0,579	PKC
Q8VCS0	63S	AKNSSTHNS	0,649	unsp
Q8VCS0	67S	STHNSLHQR	0,557	PKA
Q8VCS0	81T	PSHNTTEPD	0,867	unsp
Q8VCS0	88S	PDPHLSPE	0,813	unsp
Q8VCS0	90S	PHLSPELQ	0,986	unsp
Q8VCS0	98S	QALISEVAQ	0,504	cdc2
Q8VCS0	121T	PDGSTVAVK	0,884	PKC
Q8VCS0	139S	LQAHSVANL	0,690	unsp
Q8VCS0	164T	NIRATWPGL	0,523	PKC
Q8VCS0	177S	PNASSPDVG	0,689	unsp
Q8VCS0	177S	PNASSPDVG	0,528	cdc2
Q8VCS0	191S	DKAKTPTTV	0,693	unsp
Q8VCS0	191S	DKAKTPTTV	0,521	p38MAPK
Q8VCS0	194T	KTPTTVDRL	0,770	unsp
Q8VCS0	194T	KTPTTVDRL	0,558	PKC
Q8VCS0	215S	FLHRSQTWS	0,639	unsp
Q8VCS0	215S	FLHRSQTWS	0,555	DNAPK
Q8VCS0	215S	FLHRSQTWS	0,551	ATM
Q8VCS0	217T	HRSQTWSPP	0,926	unsp
Q8VCS0	217T	HRSQTWSPP	0,710	PKC
Q8VCS0	219S	SQTWSPPGL	0,658	cdk5
Q8VCS0	219S	SQTWSPPGL	0,518	GSK3
Q8VCS0	239T	PRVFTLLDP	0,771	unsp
Q8VCS0	239T	PRVFTLLDP	0,563	cdc2
Q8VCS0	246S	DPQASRLTM	0,501	CaM-II
Q8VCS0	267S	GNHLSQIPR	0,660	ATM
Q8VCS0	307T	GAALTSAPT	0,864	unsp
Q8VCS0	307T	GAALTSAPT	0,605	PKC
Q8VCS0	337S	LQNISQEQL	0,616	DNAPK
Q8VCS0	337S	LQNISQEQL	0,547	ATM

Q8VCS0	378T	RGHPTPLRL	0,769	unsp
Q8VCS0	378T	RGHPTPLRL	0,565	p38MAPK
Q8VCS0	378T	RGHPTPLRL	0,525	cdk5
Q8VCS0	378T	RGHPTPLRL	0,508	GSK3
Q8VCS0	400T	APPCTTFQS	0,508	GSK3
Q8VCS0	427Y	DDIGYSFVV	0,573	unsp
Q8VCS0	427Y	DDIGYSFVV	0,529	INSR
Q8VCS0	428S	DIGYSFVVG	0,520	cdc2
Q8VCS0	433S	FVVGSDGYL	0,617	PKA
Q8VCS0	436Y	GSDGYLYQG	0,758	unsp
Q8VCS0	436Y	GSDGYLYQG	0,536	EGFR
Q8VCS0	436Y	GSDGYLYQG	0,504	INSR
Q8VCS0	450T	VGAHTRGYN	0,647	PKC
Q8VCS0	455S	RGYNSRGFG	0,937	unsp
Q8VCS0	469S	NYTGSLPNE	0,636	PKA
Q8VCS0	469S	NYTGSLPNE	0,589	DNAPK
Q8VCS0	469S	NYTGSLPNE	0,519	cdc2
Q8VCS0	478T	AALNtvrda	0,932	unsp
Q8VCS0	478T	AALNtvrda	0,723	PKC
Q8VCS0	508T	QLVLTHCPG	0,518	cdc2
Q8VCS0	521T	NLLRTWPHF	0,652	PKC
Q8VCS0	526T	WPHFTEVEN	0,896	unsp
Q8VCS0	526T	WPHFTEVEN	0,618	CKII

Para a fosforilação, essa predição é importante devido ao seu fator fosforilativo no reconhecimento celular, os dados obtidos foram através de também um *software* dinamarquês *Netphos* 3.1 ele prevê locais de fosforilação de serina, treonina ou tirosina nas proteínas eucariontes com a previsão para as quinases ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 e p38MAPK, observando scores maiores que 0,5. Como demonstrado na Tabela 1 os dados demonstraram muitos locais de fosforilação o que levaram a crer que a aceitação dessas proteínas é muito grande pelas quinases e que a células as reconhecem com facilidade; Foroutan *et al* [2021].

### 3.6 Busca de domínios transmembrana e predição da localização celular

Com a previsão de domínios transmembrana para prever se essas proteínas são proteínas de transmembrana ou se estão em livre circulação, o *servidor* dinamarquês *TMHMM* 2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) que ao acessar o servidor oferece um plug-in para o servidor *deepTMHMM* (<https://dtu.biolib.com/DeepTMHMM>) que fornece a previsão da localização de sequências de aminoácidos de uma proteína que podem ser hélices transmembrana utilizando o algoritmo *N-best*. Na pesquisa foram adicionadas a sequência *FASTA* das proteínas Q9VS97 e Q8VCS0 (refinadas na etapa anterior) que foram obtidas anteriormente no alinhamento e utilizada como input no software, ele então previu que

as proteínas Q9VS97 e Q8VCS0 não são proteínas transmembrana. A localização celular foi fornecida pelo banco de dados *UniprotKB* que ao digitar os códigos Q9VS97 e Q8VCS0 fornece informações sobre as proteínas, sendo então que na análise as proteínas se demonstraram excretadas e se localizam fora da célula (extracelular); Foroutan *et al* [2021].

### 3.7 Análise da estrutura secundária

A avaliação da estrutura secundária demonstra os diversos tipos de estruturas que a proteína pode ter de acordo com a sua composição de resíduos de aminoácidos, eles de acordo com sua composição química podem ter alfa hélices, folhas beta, bobinas e giro. Para essa análise foi utilizado o servidor *PRABI-GERLAND RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE (SOPMA)* onde através da sequência *FASTA* obtida anteriormente no alinhamento foi utilizada como input para a análise. A proteína com o código Q9VS97 que tem um total de 186 resíduos de aminoácidos apresentou: *Alpha helix* (Hh) : 45 que é 24,19%, *310 helix* (Gg) : 0%, *Pi Helix* (Ii): 0%, *Beta bridge* (Bb): 0%, *Extended strand* (Ee): 35 que é 18,82%, *Beta turn* (Tt): 15 que é 8,06%, *Bend region* (Ss): 0%, *Random coil* (Cc): 91 que é 48,92%, *Ambiguous states* (?): 0%, *Other states*: 0%; a proteína com o código Q8VCS0 que possui 530 resíduos de aminoácidos apresentou *Alpha helix* (Hh) : 180 que é 34,34%, *310 helix* (Gg) : 0%, *Pi Helix* (Ii): 0%, *Beta bridge* (Bb): 0%, *Extended strand* (Ee): 66 que é 12,45%, *Beta turn* (Tt): 32 que é 6,04%, *Bend region* (Ss): 0%, *Random coil* (Cc): 250 que é 44,17%, *Ambiguous states* (?): 0%, *Other states*: 0%; Ghaffari *et al* [2021].

### 3.8 Construção, refinamento e validação da estrutura 3D

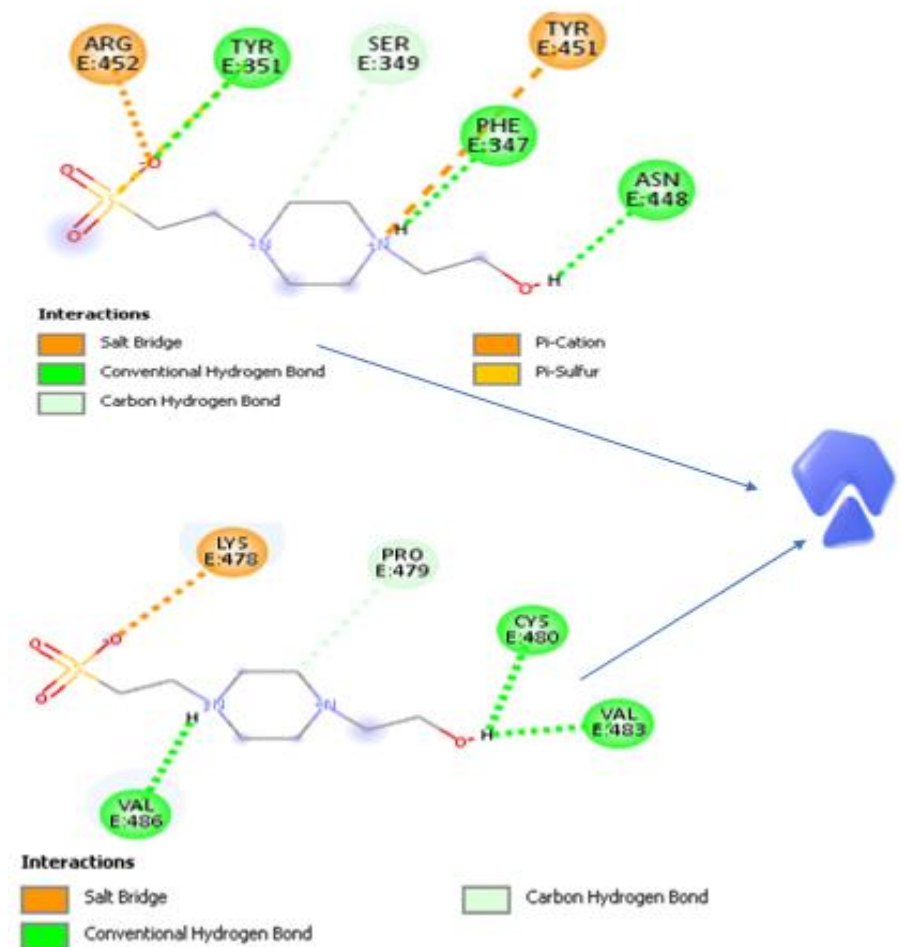
No modelo 3D todas estão depositadas no banco de dados suíço *Swiss Protein* (*SwissProt*) que fornece estruturas validadas das proteínas com parâmetros de confiabilidade. A estrutura 3D da proteína com o código Q9VS97 está depositada no *SwissProt* que está ligado a estrutura dessa proteína construída pelo maior banco de dados do mundo o *AlphaFold* (<https://alphafold.com/>) pelo código AF-Q9VS97-F1-model-v3 com confiança do modelo de 93,94 com a sua análise confirmada pela análise do gráfico de Ramachandran com 0,85 pontos (podendo ser maior com a retirada do peptídeo de sinal pois esta proteína está depositada com peptídeo de sinal onde suas estruturas secundárias não estão bem validadas e fora da área de aceitação, porém como o peptídeo de sinal foi retirado ele não irá interferir na validação) e também um modelo depositado com o código 2rkq.1.A com confiança de 92 com a sua qualidade validada com a análise do gráfico de Ramachandran que obteve 0,92 pontos. Já a proteína com o código Q8VCS0 também se encontra depositada no *SwissProt* com a sua estrutura ligada pelo *AlfaFold* pelo código AF-Q8VCS0-F1 com confiança de 81,14 com a sua qualidade foi confirmada através da análise do gráfico de Ramachandran que obteve 0,71 pontos e também há uma região de domínio depositada com o código 7nsz.1. A com confiança de 69 com a qualidade de 0,69 pontos confirmadas pela análise do gráfico de Ramachandran; Motamedpour, Dalimi, Pirestani, Ghaffarifar [2020].

### 3.9 Docking Molecular

Sendo assim, um ligante foi fornecido pela proteína 7nsz com o nome: 4-(2-hidroxiilo)-1-piperazina ácido elfnônico, o *docking* molecular foi feito utilizando o *software AutoDock* onde foram fornecidas a proteína *spike* (localizada na membrana do coronavírus) e o ligante foi otimizado e fornecido ao *software*, utilizando o protocolo para *docking* fornecido por ([AutoDock \(scripps.edu\)](http://Autodock.scripps.edu)), que foi interpretado e executado com êxito nas

posições dos aminoácidos 452 (Arginina) e 456 (Valina). O *DockThor* também foi utilizado para *docking* e o ligante otimizado também foi fornecido, a proteína alvo foi preparada no próprio servidor onde oferece recursos para o mesmo; foi obtido com êxito o *docking* nas posições 452 e 486; Goodsell, Sanner, Olson, Forli [2020]; Guedes *et al* [2021].

### 3.10 Análise dos resultados



**Figura 2:** (*Docking* molecular com *Autodock*).

Os resultados do docking demonstraram que pelo *AutoDock* nas posições 452 e 486 apresentaram energia livre de ligação: -5.50 kcal/mol e -3.99 kcal/mol, respectivamente (FIGURA 2).

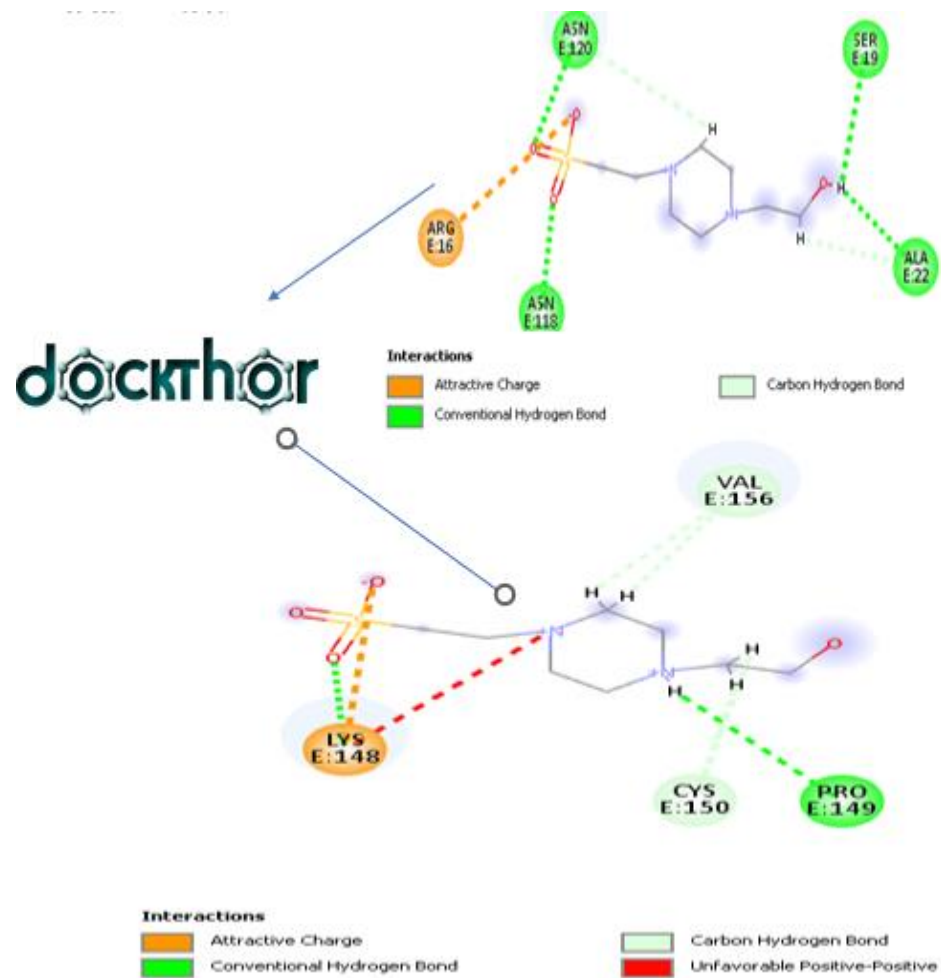


Figura 3: (Docking molecular com DockThor).

Já o *DockThor* foi apresentado para as mesmas posições (posição 452 se refere a posição 16 e 486 se refere a posição 156) uma energia livre de ligação: -6.221 Kcal/mol e -6.331kcal/mol (FIGURA 3). Logo, a partir da validação por energia livre calculada pelo *AutoDock* o complexo proteína-ligante (*spike* com o ligante) apresentou afinidade na ligação devido a um processo espontâneo e seus valores negativos; Guedes *et al* [2021]. Os dados fornecidos pelo *DockThor* apresentando a energia de afinidade demonstraram a energia de ligação favorável em relação valor de referência citado, uma das observações foi que neste *software* houve a rotação de um dos carbonos da cadeia e na posição 452 houve uma ligação *attractive charge* (carga atrativa). Sendo os do *AutoDock* apesar de a posição 452 ligar ao grupo com enxofre a sua ligação foi pi-enxofre e da posição 486 foi ligação de hidrogênio; Ahmad *et al* [2021].

## 4 CONCLUSÕES

As proteínas são grandes fontes de compostos complexados e podem fornecer bastantes informações. Dentro do contexto do SARS-COV-2, a proteína *spike* da variante ômicron se mostrou um importante alvo terapêutico, sendo ela bastante estudada em todo mundo. Os dados de *docking* molecular forneceram pelo *AutoDock* energias de ligação favoráveis que permite observar um potencial ligação nas posições 452 e 486 do domínio de ligação ao receptor da proteína *spike*. O *DockThor* chegou a um valor maior devido a sua ligação ser em uma região do ligante diferente pois a ligação do aminoácido Valina ao ligante se dá em um carbono e é uma ligação *carbon hydrogen bond* (ligação carbono hidrogênio), ao contrário do *Autodock* que se ligou a uma região nitrogenada no ligante formando uma *Conventional hydrogen bond* (ligação de hidrogênio convencional). Portanto, é possível concluir que esta pesquisa forneceu dados de extrema importância para um possível inibidor da proteína *spike* da variante ômicron do SARS-COV-2 através do *docking* molecular. Portanto, esses dados são muito importantes e estudos posteriores são fundamentais para maiores confirmações.

## 5 AGRADECIMENTOS

*Agradecimentos ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) fornecido pelo Centro Universitário Vértice – Univértix pelo financiamento desta pesquisa e ao Professor. Dr Adriano Carlos Soares pelas orientações durante o trabalho.*

## Referências

- A. D. Ghaffari *et al.* Immunoinformatic analysis of immunogenic B-and T-cell epitopes of MIC4 protein to designing a vaccine candidate against *Toxoplasma gondii* through an in-silico approach. *Clinical and Experimental Vaccine Research*. 10(1):59-71, 2021.
- A. Hoda, M. Tafaj, E. Sallaku. *In silico* Structural, Functional and Phylogenetic Analyses of cellulase from *Ruminococcus albus*. *Journal of Genetic Engineering & Biotechnology*. 19(1):58-73, 2021.
- A. Morgat *et al.* Bridge; Consórcio UniProt. Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*. 36(6):1896–1901, 2020.
- A. Soares. (2015). Identificação de lipocalinas com domínio de ligação a histamina e serotonina e validação de genes de referência em transcriptomas de glândula salivar, intestino e ovário do carrapato *Amblyomma sculptum*. *Tese de doutorado (tese em Bioquímica Aplicada) - UFV, Departamento de Bioquímica e Biologia Molecular. Viçosa. 113p.*
- A. Tuekprakhon *et al.* Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell*. 185(14):2422-2433, 2022.
- A. U. Haq *et al.* Annotation of Potential Vaccine Targets and Design of a Multi-Epitope Subunit Vaccine against *Yersinia pestis* through Reverse Vaccinology and Validation through an Agent-Based Modeling Approach. *Vaccines*. 9(11):1327-1346, 2021.
- B. Dutta, A. Banerjee, P. Chakraborty, R. Bandopadhyay. In silico studies on bacterial xylanase enzyme: Structural and functional insight. *Journal of Genetic Engineering & Biotechnology*. 16(2):749–756, 2018.
- D. Goodsell, M. Sanner, A. Olson, S. Forli. A suite AutoDock aos 30. *Protein Science*. 30(1):31-43, 2020.

- E. Rouka, K. Gourgoulianis, S. Zarogiannis. *In silico* investigation of the viroporin E as a vaccine target against SARS-CoV-2. *American Journal of Physiology Lung Cellular and Molecular Physiology*. 320(6):1057-1063, 2021.
- F. Liberman. Análise dos fatores determinantes para a qualidade da anotação genômica automática. *Tese de Mestrado (Biotecnologia e Ciências genômicas) – Universidade Católica de Brasil*. Brasília, 2004.
- F. Robson *et al.* Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Molecular Cell*. 79(5):710-727, 2020.
- G. Heinzelmann e M. Gilson. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Science Republic*. 11(1):1116-1134, 2021.
- I. A. Guedes *et al.* Novas funções de aprendizagem de máquina e pontuação baseada em física para descoberta de drogas. *Science Republic*. 11(1):3198-3217, 2021.
- I. A. Guedes *et al.* Design de drogas e redefinição com o servidor web DockThor-VS com foco em alvos terapêuticos SARS-CoV-2 e suas variantes não-sinônimos. *Science Republic*. 11(1):5543-5563, 2021.
- J. Kyte e F. Doolittle. A simple method for displaying the hydrophobic character of a protein. *Molecular Biology*. 157(1):105-132, 1982.
- J. Lan *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 581(7807):215-220, 2020.
- K. Tamura, G. Stecher, S. Kumar. MEGA11: Molecular Evolutionary Genetics Analysis Versão 11. *Molecular Biology and Evolution*. 38(7):3022-3027, 2021.
- L. Motamedpour, A. Dalimi, M. Pirestani, Ghaffarifar. *In silico* analysis and expression of a new chimeric antigen as a vaccine candidate against cutaneous leishmaniasis. *Iranian Journal of Basic Medical Sciences*. 23(11):1409-1418, 2020.
- L. Sanos, R. Ferreira, E. Caffarena. Integrating Molecular docking and Molecular Dynamics Simulations. *Methods Molecular Biology*. 2053(1):13-34, 2019.
- M. Alkhatib *et al.* Update on SARS-CoV-2 Omicron Variant of Concern and Its Peculiar Mutational Profile. *Microbiology Spectrum*. 10(2):2732-2749, 2022.
- M. Khan *et al.* COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far. *Molecules*. 26(1):39-64, 2020.
- M. Foroutan *et al.* Sabaghan. Bioinformatics analysis of calcium-dependent protein kinase 4 (CDPK4) as *Toxoplasma gondii* vaccine target. *BMC research notes*. 14(1):50-59, 2021.
- N. C. Wu *et al.* Um modo de ligação alternativo de anticorpos IGHV3-53 ao domínio de ligação do receptor SARS-CoV-2. *Cell Reports*. 33(3):108274-108307, 2020.
- S. Ahmad, Y. Waheed, A. Abro, S. W. Abbasi, Ismail. Molecular screening of glycyrrhizin-based inhibitors against ACE2 host receptor of SARS-CoV-2. *Journal of Molecular Modeling*. 27(7):206-219, 2021.
- T. Kaur, A. Madgulkar, M. Bhalekar, K. Asgaonkar. Molecular Docking in Formulation and Development. *Drug Discovery Technologies*. 19(1):30-39, 2019.
- X. Cathala e C. Moorley. How to appraise quantitative research. *Evidence-Based Nursing*. 21(4):99-101, 2018.
- X. He *et al.* SARS-CoV-2 Omicron variant: Characteristics and prevention. *MedComm*. 2(4):838-845, 2021





# Ataque Zumbi: a humanidade irá sobreviver?

Gabriel Vilela Sousa<sup>1</sup>, Ulisses Ferreira Kaneko<sup>2</sup> e Paulo Freitas Gomes<sup>1</sup>

<sup>1</sup> Grupo de Redes Complexas Aplicadas de Jataí, Universidade Federal de Jataí, Jataí/GO, Brasil

<sup>2</sup> Laboratório Nacional de Luz Síncrotron (LNLS), Centro Brasileiro para Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo 13083-970, Brazil

---

## Abstract

Neste trabalho estudamos a dinâmica de um apocalipse no qual parte da população se torna zumbis temporariamente, podendo se tornar vivo novamente ou morrer. Descrevemos essa dinâmica utilizando uma versão modificada do modelo epidêmico SIS (do inglês: *susceptible-infected-susceptible*). Os zumbis podem morrer quando em contato com um indivíduo vivo do tipo exterminador. Para definir quem interage com quem utilizamos a rede aleatória Erdős-Rényi e calculamos como as fases absorvente e ativa do modelo SIS original se altera em função do grau médio da rede. Além disso a densidade de mortos atua como uma força dissipativa diminuindo a população efetiva.

**Keywords:** Modelo epidêmico SIS, Método de Monte Carlo, rede Erdős-Rényi, fase absorvente, transição de fase

---

## 1 INTRODUÇÃO

Os chamados modelos epidêmicos referem-se a um conjunto de dinâmicas onde um indivíduo saudável pode se tornar infectado quando em contato com outros indivíduos infectados. Além desses dois estados, há vários outros que podem ser incluídos no modelo de forma a deixar a dinâmica mais sofisticada como: morto, recuperado, latente, exposto, etc [Hethcote, 2000]. Modelos epidêmicos tem sido amplamente estudados em diversas aplicações nas suas diversas variações, apresentando uma fase absorvente a partir da qual o sistema não consegue sair, como também uma fase ativa onde as densidades dos parâmetros de ordem não variam com o tempo [Barthélemy et al., 2022, da Silva and Fernandes, 2015, Miranda et al., 2020, Reia and Fontanari, 2022].

O contexto de apocalipse zumbi é um tema atual tanto na indústria do entretenimento quanto no meio acadêmico. Algumas abordagens encontradas envolvem a utilização de

equações diferenciais [Miranda et al., 2020], método de Monte Carlo [Miranda et al., 2020], rede quadrada [Alemi et al., 2015], equações de difusão [Smith?, 2014], questões de ensino [Lofgren ET, 2016] dentre outras.

Neste trabalho utilizamos um modelo epidêmico SIS modificado para descrever a dinâmica entre indivíduos vivos e zumbis em um apocalipse zumbi. Além disso, incrementamos o estado de morto e uma variante do estado vivo chamado exterminador. No fim buscamos observar o comportamento durante a fase absorvente e ativa.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Modelo SIS

O modelo epidêmico SIS considera apenas dois estados para os indivíduos: saudável (susceptible) e infectado (infected). Um indivíduo saudável pode se tornar infectado quando em contato com outros indivíduos infectados. Já um indivíduo infectado pode se tornar saudável com uma chance constante ao longo do tempo, sem a necessidade de ter contato com indivíduos vivos. Esse modelo descreve bem por exemplo uma gripe comum.

No nosso trabalho usaremos os estados vivo e zumbi no lugar de saudável e infectado, tendo uma variação de exterminadores para indivíduos vivos e incremento do estado morto. Ao invés de colocarmos a clássica dinâmica ao qual vivos matam zumbis, optamos em colocar exterminadores [Smith?, 2014]. A existência de exterminadores é concebível em um apocalipse zumbi, assim como somente uma pequena parcela da população está apta a matar zumbis devido a questões de idade, habilidade e rigor físico. Os exterminadores podem interferir na dinâmica para que a fase absorvente chegue mais rapidamente, exterminando a população de zumbis. Logo, conforme a porcentagem de exterminadores aumenta, mais tempo demora para a fase ativa começar.

Explorando o modelo SIS tradicional, seja  $N$  o número total de indivíduos do sistema dos quais  $V$  refere-se ao número de indivíduos vivos e  $Z$  o número de zumbis. As respectivas densidades serão  $v = V/N$  e  $z = Z/N$ . Neste trabalho iremos analisar o sistema no chamado estado estacionário, no qual as densidades não se alteram mais com o tempo. Este estado é obtido após um certo intervalo de tempo. A equação diferencial que descreve o modelo SIS no estado estacionário é [Newman, 2010]:

$$\frac{dv}{dt} = cz - bvz, \quad \frac{dz}{dt} = bvz - cz, \quad (1)$$

onde  $b$  e  $c$  são as chances de conversões de vivo para zumbi e vice-versa. Estas equações referem-se a aproximação de campo médio na qual cada indivíduo interage com todos os outros. Veja que se  $z = 0$  os vivos continuam vivos, já os zumbis sempre se tornam zumbis após algum tempo, independente da existência de indivíduos vivos. Como temos que  $v + z = 1.0$ , podemos combinar essas duas equações em uma:

$$\frac{dz}{dt} = (b - cb - bz)z. \quad (2)$$

A solução é:

$$z(t) = (1 - c/b) \frac{De^{(b-c)t}}{1 + De^{(b-c)t}}, \quad D = \frac{bz_0}{b - c - bz_0}. \quad (3)$$



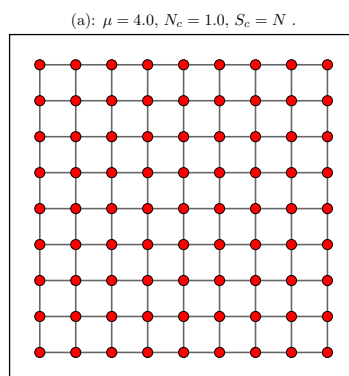
$z_0$  é a concentração inicial de zumbis.

Se  $b > c$  temos a chamada fase ativa na qual há uma conversão de vivo para zumbi e vice-versa. Além disso a taxa dessas duas conversões é idêntica de forma que as densidades não mais se alteram com o tempo (havendo apenas a flutuação estatística). Apesar de que o conjunto de indivíduos vivos e zumbi estão se alterando continuamente.

Por outro lado se  $b < c$ , a taxa de conversão de zumbi para vivo é maior de forma que o número de zumbis cai para zero e todos os indivíduos se mantêm vivos. Esta é a chamada fase absorvente pois uma vez havendo apenas vivos não há chance de alguém se tornar zumbi (é necessário contato com algum zumbi para um vivo se tornar zumbi). Uma vez nessa fase, o sistema não mais sai dela, daí o nome de fase absorvente. A condição  $b = c$  é exatamente a transição entre essas duas fases. Para o nosso modelo será considerado  $c = 0.1$  enquanto  $b$  varia na escala de 0.0 a 1.0, com o finalidade de observar de forma mais rápida a fase ativa no sistema.

## 2.2 Rede Erdős-Rényi

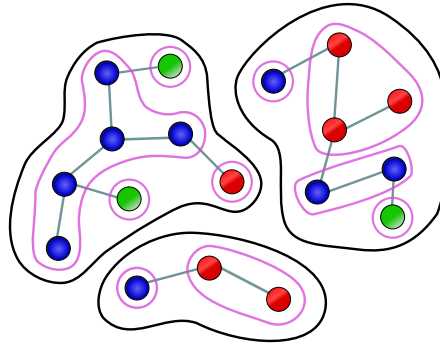
A dinâmica é definida pelo modelo em questão: modelo epidêmico SIS neste trabalho. Ela define o que acontece quando um indivíduo interage com outro. Também é necessário definir o padrão de comunicação entre os indivíduos: quem interage com quem. Tradicionalmente a rede quadrada (veja figura 1) tem sido utilizada em simulações numéricas de Monte Carlo devido a sua fácil implementação computacional. Neste trabalho escolhemos utilizar uma rede aleatória que permite alterar sua topologia: a chamada rede aleatória do tipo Erdős-Rényi (RER) [Solomonoff and Rapoport, 1951, Erdős and Rényi, 1959, 1960, 1961]. Sua definição é: cada dupla de vértices tem uma chance  $p$  de possuir uma aresta (ou seja, de ser conectada). Os parâmetros  $N$  e  $p$  são os chamados parâmetros de controle da rede. Enquanto  $N$  é o número total de indivíduos da rede durante as simulações, o grau  $k_i$  de um vértice  $i$  é o número de conexões (arestas) que ele tem com os outros vértices. Dessa maneira, os vizinhos de  $i$  são todos os vértice que possuem conexão direta com  $i$  por uma aresta. O grau médio da rede é a média de  $k_i$  entre todos os vértices da rede:  $\mu = (1/N) \sum_i k_i$ . No caso da RER temos que  $\mu = p(N - 1)$  [Barabási, 2016], de forma que podemos utilizar  $\mu$  ao invés de  $p$  como o parâmetro de controle, por ter um significado mais intuitivo. Uma vez definido  $\mu$ , basta fazer  $p = \mu/(N - 1)$  no algoritmo.



**Fig. 1:** (a) Topologia da rede quadrada:  $\mu = 4.0, N_c = 1$  e  $S_c = N$ .

Uma segunda característica importante de redes é a sua distribuição de componentes. Uma componente é um aglomerado conectado de vértices, e o seu tamanho é o número de vértices contidos [Newman, 2010]. Um vértice isolado é uma componente de tamanho 1. Seja  $N_c$  e  $S_c$  o número de componentes e o tamanho (número de vértices) da maior componente de uma rede. Uma rede conectada tem apenas uma componente, logo:  $N_c = 1$  e  $S_c = N$ . Além das componentes, podemos definir também os domínios: grupo de vértices conectados dentro de uma componente tendo o mesmo estado [Gomes et al., 2022]. Seja  $N_d$  e  $S_d$  o número de domínios e o tamanho do maior domínio de uma dada rede. Este conceito é apenas aplicado em redes nas quais os vértices tem estados diferentes, como em nosso caso: vivo, zumbi e morto. A figura 2 mostra uma ilustração desses conceitos em uma rede fictícia. Nas análise utilizamos as densidades desses parâmetros:

$$\begin{aligned} n_c &= \frac{N_c}{N}, & s_c &= \frac{S_c}{N}, \\ n_d &= \frac{N_d}{N}, & s_d &= \frac{S_d}{N}. \end{aligned}$$



**Fig. 2:** Ilustração de componentes e domínio. Os contornos pretos indica componentes e os contornos rosas indicam domínios. A cor dos vértices (azul, vermelho ou verde) indicam o estado deles. Nesta rede temos:  $N = 18$ ,  $N_c = 3$ ,  $S_c = 8$ ,  $N_d = 10$  e  $S_d = 5$ .

### 3 METODOLOGIA

#### 3.1 Modelo SIS com mortos e exterminadores

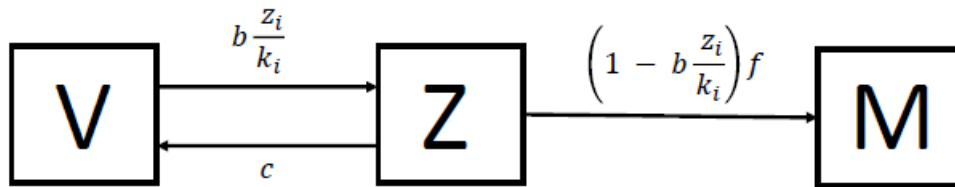
Nosso modelo é uma adaptação do modelo SIS e utilizamos os estados vivo e zumbi no lugar dos estados saudável e infectado, tendo uma variação de exterminadores para indivíduos vivos e incrementamos do estado morto. A dinâmica entre esses estados é exatamente a do modelo SIS descrita na seção 2.1. A seguir descrevemos os passos de uma análise, na qual  $\Gamma$  é um número aleatório entre 0.0 e 1.0 gerado por distribuição uniforme.

- Um indivíduo  $i$  aleatório é selecionado:  $i = 1 + I(\Gamma N)$ , onde  $I(a)$  é a função inteiro, que retorna a parte inteira de  $a$ .
- Se o indivíduo selecionado é zumbi, ele pode se tornar vivo com uma chance  $c$ . Se  $\Gamma < c$ , esse indivíduo passa de zumbi para vivo.



- Se o indivíduo  $i$  selecionado é vivo, ele pode se tornar um zumbi com chance  $bz_i/k_i$ , onde  $b$  é a taxa de infecção,  $z_i$  é o número de vizinhos zumbi e  $k_i$  é o número total de vizinhos do indivíduo  $i$ . Se  $\Gamma < bz_i/k_i$ , o indivíduo passa de vivo para zumbi.
- Se  $\Gamma > bz_i/k_i$  e esse indivíduo vivo também for exterminador, um vizinho zumbi aleatório passa para o estado morto.

Os 3 passos anteriores constituem uma análise. Uma ilustração dessas conversões está na figura 3. Se não houver zumbis no sistema, ou se o indivíduo vivo selecionado para a análise não tiver vizinhos zumbis, nada acontece na análise (mas ela é contabilizada). Um passo de Monte Carlo é definido como  $N$  análises. Na média todos os indivíduos são analisados, mas em um dado passo pode haver indivíduos não analisados e outros analisados mais de uma vez. A termalização constitui de  $Q$  passos de Monte Carlo (uma calibração é feita para se obter o valor adequado desse parâmetro).



**Fig. 3:** Ilustração das mudanças de estados e suas probabilidades.

Os exterminadores perfazem uma fração aleatória  $f$  da população de vivos definidos no instante inicial. Quando um vivo exterminador se torna zumbi e eventualmente se torna vivo novamente, ele não volta como exterminador. Assim a população de exterminadores diminui com o tempo até desaparecer. Após esse instante a densidade de mortos fica constante no tempo.

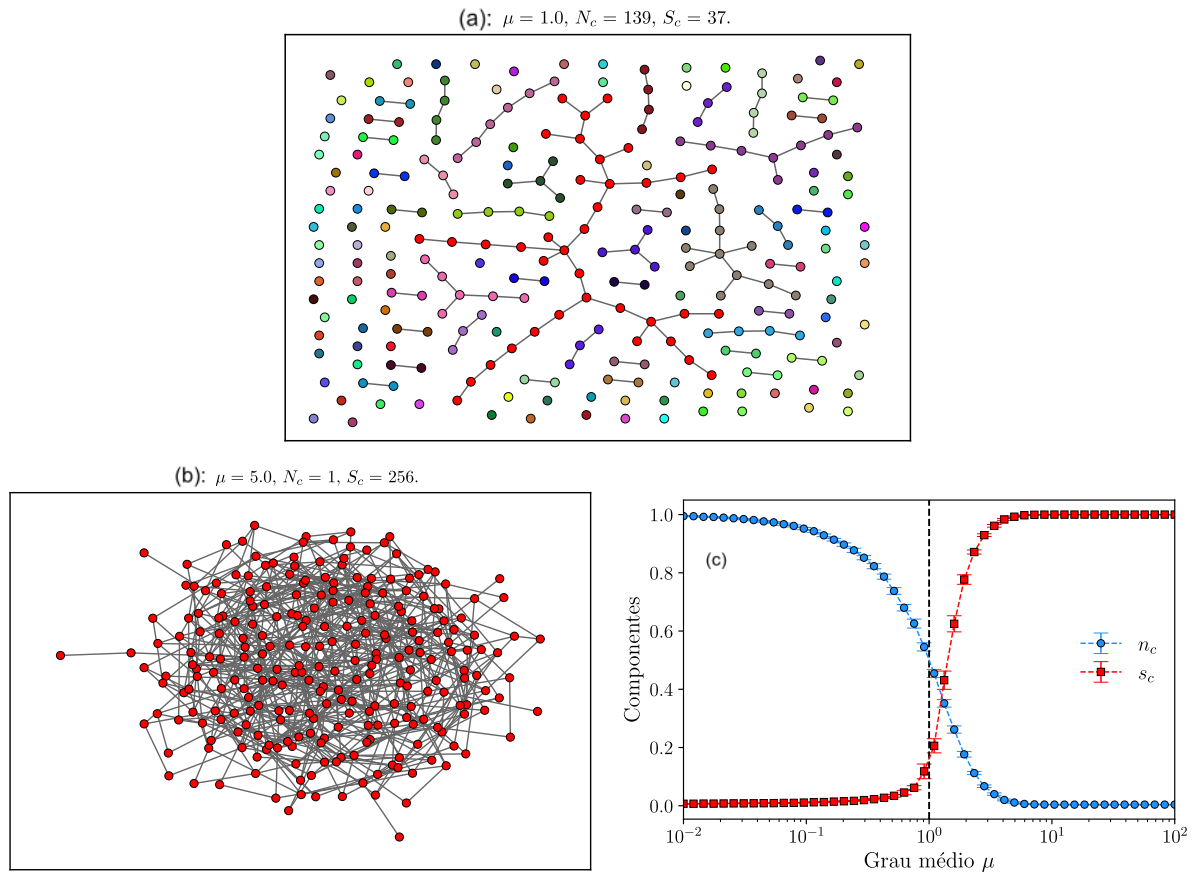
### 3.2 Ferramentas computacionais

O modelo foi implementado em linguagem Fortran para a geração dos dados enquanto sua análise e todos os gráficos foram feitos em Python. Os seguintes pacotes em Python foram utilizados: Numpy [Harris and *et al.*, 2020], Matplotlib [Hunter, 2007], NetworkX [Hagberg et al., 2008] e Pandas [McKinney, 2010]. Todos os dados e códigos estão disponíveis mediante solicitação.

## 4 RESULTADOS

Na figura 1 está uma ilustração de uma rede quadrada (RQ) na qual pode-se ver que cada vértice está ligado com os seus 4 primeiros vizinhos. A RQ é uma rede dita conectada pois há somente uma componente:  $N_c = 1$ . Assim o tamanho da maior (única) componente é exatamente  $S_c = N$ . Já na figura 4(a) está uma instância aleatória da RER com  $\mu = 1.0$ . Repare que a maior componente tem  $S_c = 37$  vértices (em vermelho) enquanto que há  $N_c = 139$  componentes no total (cada vértice isolado é uma componente de tamanho 1.0). A medida que  $\mu$  aumenta, a rede se torna mais conectada:  $N_c$  diminui e  $S_c$  aumenta. Na figura 4(b) está uma instância da RER gerada com  $\mu = 5.0$ , na qual a rede já está

conectada:  $N_c = 1$  e  $S_c = N$ . A figura 4(c) mostra a variação de  $n_c = N_c/N$  e  $s_c = S_c/N$  em função de  $\mu$  ao longo de 5 ordens de grandeza. No limite  $\mu$  muito pequeno a rede está fragmentada contendo apenas vértices isolados:  $n_c = 1.0$  e  $s_c = 1/N$ . Já no limite de  $\mu$  muito grande a rede se torna conectada:  $n_c = 1/N$  e  $s_c = 1$ . Assim a rede apresenta as fases não percolada e percolada sendo que a transição é em torno de  $\mu = 1.0$  Newman [2010], Vilela et al. [2020]. Todos os resultados mostrados a seguir são na rede percolada:  $\mu \geq 1.0$

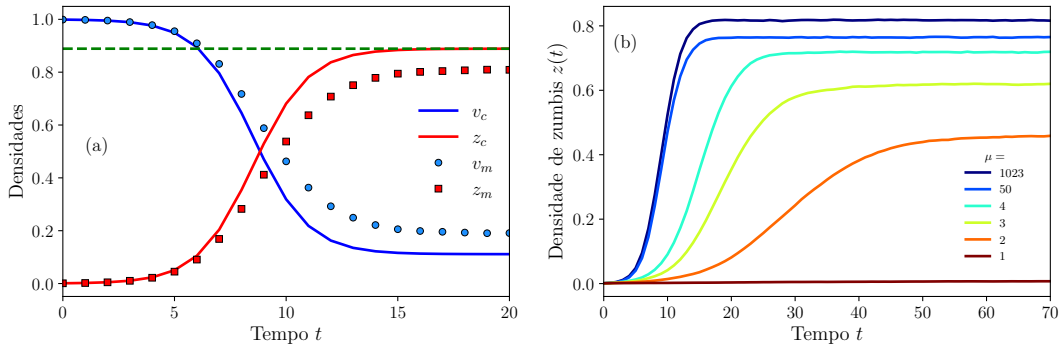


**Fig. 4:** Exemplos aleatórios (a) e (b) da topologia da rede Erdos-Renyi com  $N = 16^2$ . As maiores componentes ficam no centro enquanto os vértices isolados ficam na parte externa. Os vértices da maior componente estão em vermelho para facilitar a visualização. (a)  $\mu = 1.0$ ,  $N_c = 139$  e  $S_c = 37$ . (b)  $\mu = 5.0$ ,  $N_c = 1$  e  $S_c = N$ . (c) Número de componentes  $n_c$  e tamanho da maior componente  $s_c$  (normalizados) em função do grau médio  $\mu$  para a RER. O eixo horizontal está em escala log. A linha vertical tracejada indica a transição de fase em  $\mu = \mu_c$ . Fase não percolada:  $\mu < \mu_c$ . Fase percolada:  $\mu > \mu_c$ .

Para evidenciar o efeito do uso de uma rede arbitrária, apresentamos uma comparação dos resultados utilizando aproximação de campo médio e Método de Monte Carlo na figura 5(a). Mesmo utilizando  $\mu = N - 1$  (rede totalmente conectada), o resultado da evolução temporal não é idêntico ao do campo médio. Esta variação temporal também serve para calibrar o tempo mínimo  $Q$  para se obter o estado estacionário. Já na figura 5(b) podemos ver como a topologia da rede influencia na evolução temporal da densidade de zumbis.



O caso totalmente conectado refere-se a curva  $\mu = 1023$ , que está bem próxima da curva com  $\mu = 50$ . Ou seja, para redes densamente conectadas há pouca variação na densidade. Já para  $\mu = 4$  e valores menores há uma variação drástica até que em  $\mu = 1$  a densidade não apresenta uma evolução significativa. Ou seja, neste caso, como cada indivíduo só tem um vizinho, o zumbi inicial único não foi suficiente para infectar outros zumbis, não houve a formação de um apocalipse.

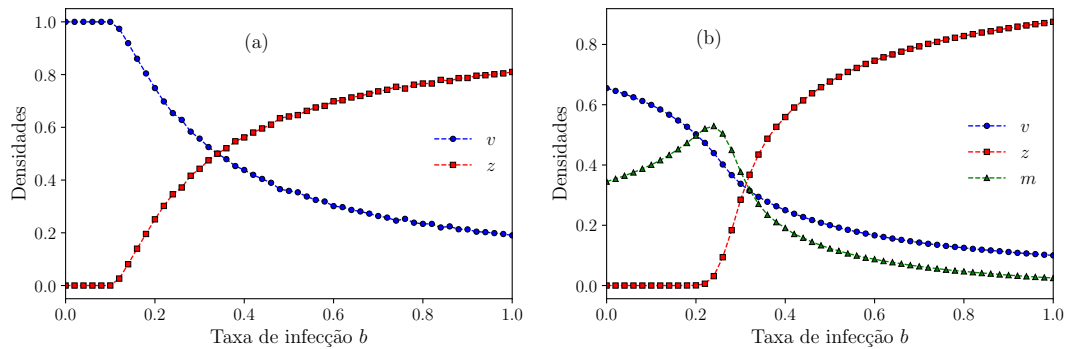


**Fig. 5:** Evolução temporal das densidades. Parâmetros:  $N = 32^2 = 1024$ ,  $b = 0.9$ ,  $c = 0.1$ ,  $S = 100$ . A condição inicial para esses resultados é  $z_0 = z(0) = 1/N$ , ou seja, no instante inicial havia apenas um zumbi no sistema. (a) Densidades de vivos  $v$  e zumbis  $z$  vs. tempo de evolução. O sub-índice  $c$  indica aproximação de campo médio (Eq. 3) e o sub-índice  $m$  Método de Monte Carlo com  $\mu = N - 1$ , o que configura uma rede totalmente conectada. O estado estacionário começa quando as densidades não variam mais com o tempo, neste gráfico, após  $t = 15$  aproximadamente. As densidades no estado estacionário dependem das taxas  $b$  e  $c$ , sendo a linha tracejada o valor  $z_c = (b - c)/b$  na aproximação de campo médio. (b) Densidade de zumbis vs. tempo para diferentes graus médio  $\mu$  na RER obtido com Método de Monte Carlo.

#### 4.1 Diagrama de fases

Vamos agora analisar as fases em função do parâmetro de controle  $b$ . A figura 6(a) apresenta o diagrama de fases do modelo SIS original. A fase absorvente é quando não há zumbis ( $z = 0$ ), pois nesse casos os vivos continuam vivos e nada acontece. Essa fase ocorre para valores baixos de  $b$ . Já para valores maiores a densidade de zumbis é maior que zero e a taxa de conversão de zumbis para vivos é igual a taxa de vivos para zumbi, de forma que ambas as densidades permanecem constantes (a menos da flutuação estatística) em função do tempo. Esta é a fase ativa. De modo geral a transição será em torno de  $b = c$ .

A figura 6(b) apresenta o diagrama de fases modificado com a inclusão dos exterminadores (um grupo dos indivíduos vivos). No instante inicial uma fração  $f$  aleatória dos  $v(0)$  indivíduos vivos é escolhido como exterminadores também. Agora há um terceiro estado possível: mortos (triângulos em verde). A fase absorvente agora é maior sendo que a transição de fase ocorre para  $b$  um pouco maior que 0.2. Neste mesmo valor há um máximo da densidade de mortos. A densidade de zumbis tem aproximadamente a mesma forma que no caso SIS original, enquanto a densidade de vivos é bastante alterada



**Fig. 6:** Densidades em função da taxa de infecção  $b$  no estado estacionário considerando uma rede totalmente conectada:  $\mu = N - 1$ . Parâmetros:  $N = 32^2 = 1024$ ,  $c = 0.1$ ,  $S_1 = 100$  e  $Q = 500$ . Condição inicial:  $z_0 = 1/2$ , ou seja, metade da população é zumbi no instante inicial  $t = 0$ . (a) Modelo SIS original contendo apenas vivos e zumbis como estados possíveis.  $S_2 = 100$ . (b) Modelo SIS modificado contendo os exterminadores e mortos como estados adicionais.  $S_2 = 20$  e  $f = 0.1$ .

#### 4.2 Influência da topologia da rede

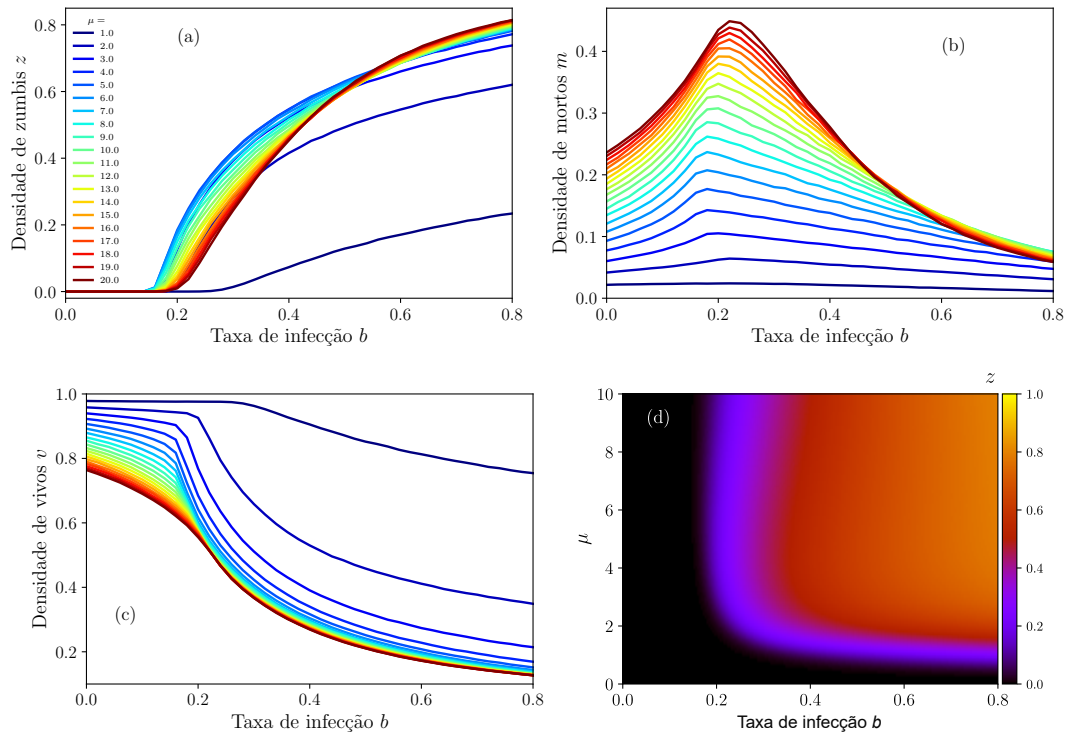
Vamos analisar em mais detalhes a influência da topologia da rede, na figura do seu grau médio  $\mu$ . Como foi observado na figura 5(b) que valores altos de  $\mu$  não criam grandes alterações, vamos focar agora em valores mais baixos. Da figura 7(a) podemos observar que o grau médio  $\mu$  não altera significativamente a transição de fase. Com exceção de  $\mu = 1.0$ , a transição ocorre para valores de  $b$  um pouco antes de 0.2. Lembrando que a fase absorvente ocorre para  $z = 0$ . Já a densidade de mortos na figura 7(b) apresenta um máximo para um  $\mu$  maior e um pouco acima de  $b = 0.2$ . Na figura 7(c) podemos observar que a densidade de vivos cai rapidamente na fase ativa a medida que aumenta a taxa  $b$ . Na figura 7(d) está o diagrama de fases 2D no plano  $(b, \mu)$ . Observa-se que a fase absorvente ocorre para  $b$  ou  $\mu$  pequenos.

#### 4.3 Domínios

Os domínios também trazem informações que contribuem para o entendimento das fases do sistema. Os domínios são limitados pela topologia da rede e são definidos em última instância pela dinâmica do modelo. Como as unidades são as mesmas, podemos fazer uma comparação numérica entre os domínios e as densidades. Por exemplo: se o maior domínio  $s_d$  for igual a densidade de vivos (ou mortos ou zumbis), podemos concluir que todos os vivos estão no maior domínio. Da mesma forma se por exemplo  $z > s_d$ , podemos concluir que esse maior domínio só contém zumbis.

A figura 8(a) apresenta as variações das densidades de indivíduos e de domínios em função de  $b$  para  $\mu = 2.0$ . De imediato é possível notar a grande variação dos domínios, especialmente no intervalo  $0.2 < b < 0.4$ . Na fase absorvente ( $b < 0.2$ ) o maior domínio é composto por vivos ( $v > s_d$ ) mas sofre uma queda brusca na transição de fase  $b \sim 0.2$ . Até  $b = 0.6$  não é possível identificar o tipo de indivíduo que compõe o maior domínio, já que  $v$  e  $z$  são maiores que  $s_d$ . A medida que o maior domínio cai o número de domínios  $n_d$  aumenta. Para  $b$  grande, maior que 0.6, o maior domínio fica ocupado todo por zumbis

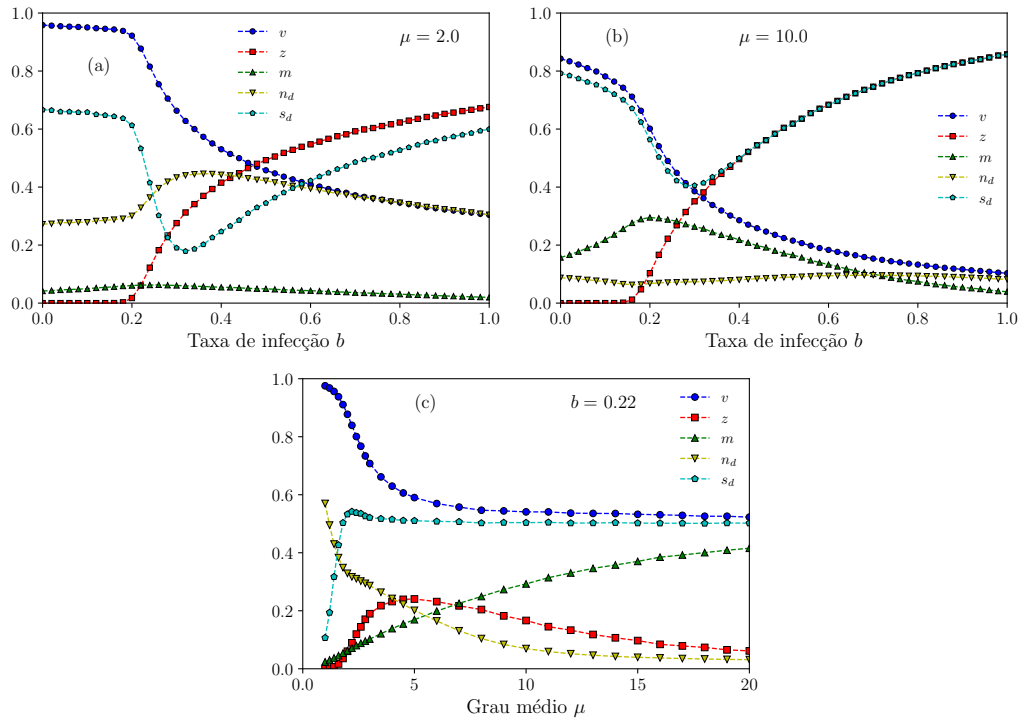




**Fig. 7:** Densidades em função da taxa de infecção  $b$  para diferentes valores do grau médio  $\mu$ . Parâmetros:  $N = 32^2 = 1024$ ,  $c = 0.1$ ,  $f = 0.1$ ,  $S_1 = 100$  e  $S_2 = 20$ . (a) Densidade de zumbis  $z$ . (b) Densidade de mortos  $m$ . (c) Densidade de vivos  $v$ . A legenda de cores da parte (a) vale para as partes (b) e (c). (d) Gráfico 2D da densidade de zumbi  $z$  no plano  $b, \mu$ . A fase absorvente ( $z = 0.0$ ) é a região preta. Parâmetros:  $N = 16^2 = 1024$ ,  $c = 0.1$ ,  $f = 0.1$ ,  $S_1 = 100$  e  $S_2 = 280$ . Os espaçamentos são  $\Delta b = 0.004$  no eixo horizontal e  $\Delta \mu = 0.1$  no eixo vertical. Q quantidade total de pontos é  $201 \times 101 = 20301$ .

pois  $z > s_d$ . Já na figura 8(b) temos as mesmas densidades mas para  $\mu = 10.0$ . A primeira diferença é que devido ao grau médio maior há um menor número de domínios  $n_d$ . De fato, aumentando a conectividade as componentes aumentam de tamanho possibilitando os domínios aumentarem também. Na fase absorvente quase a totalidade dos vivos estão no maior domínio:  $v$  um pouco maior que  $s_d$ . Já na fase ativa a partir de  $b \sim 0.4$  todos os zumbis estão na maior domínio:  $z = s_d$ .

Na figura 8(c) está mostrado os mesmos parâmetros porém em função do grau médio  $\mu$  para  $b = 0.22$ . O interessante neste resultado é que o tamanho do maior domínio  $s_d$  é aproximadamente constante no intervalo  $3 < \mu < 20$  apesar do número de domínios  $n_d$  diminuir consideravelmente. Isso significa que os domínios menores estão se juntando entre si enquanto o maior domínio se mantém inalterado. Como  $v > s_d$  podemos afirmar que o maior domínio é formado inteiramente por indivíduos vivos. Além disso a densidade de mortos aumenta enquanto a de zumbis diminui. Isso nos permite concluir que os domínios pequenos estão se juntando pois indivíduos zumbis que separam esses domínios de mortos estão se convertendo em mortos, o que agrega os domínios.



**Fig. 8:** Densidades vivos ( $v$ ), zumbis ( $z$ ), mortos ( $m$ ), número de domínios ( $n_d$ ) e maior domínio ( $s_d$ ). Parâmetros:  $N = 16^2 = 256$ ,  $z_0 = 1/2$ ,  $Q = 500$ ,  $c = 0.1$ ,  $f = 0.1$ ,  $S_1 = 100$  e  $S_2 = 50$ . (a) Em função de  $b$  com  $\mu = 2.0$ . (b) Em função de  $b$  com  $\mu = 10.0$ . (c) Em função de  $\mu$  com  $b = 0.22$ .

## 5 CONCLUSÕES

Neste trabalho aplicamos uma variação do modelo epidêmico SIS e observamos que um baixo valor do grau médio dificulta o sistema a atingir a fase ativa, favorecendo o aumento da população dos indivíduos vivos. Além disso há um máximo na densidade



de mortos para uma baixa taxa de infecção e alto valor de grau médio, já que assim a chance de um exterminador matar um zumbi aumenta. Utilizamos também a distribuição de domínios e componentes da rede para descrever em detalhes a população no estado estacionário.

Apesar do trabalho ser um estudo sobre as propriedades de um modelo, é possível a adaptação para algumas aplicações. Em jogos é possível usar o mesmo modelo variando os parâmetros para assim diversificar cenários e dificuldades em jogos. No contexto de ensino, pode ocorrer a adaptação e inserção de mais parâmetros, semelhante ao trabalho de Lofgren ET [2016] para estudos de conceitos epidemiológicos, principalmente ligados a topologia.

## 6 Agradecimentos

G. V. Sousa agradece o suporte financeiro de Rafael Grisotto e Souza e da Universidade Federal de Jataí. P. F. Gomes agradece o suporte financeiro da Fapeg (<http://www.fapeg.go.gov.br/>). As simulações numéricas foram realizadas no LAMCAD / UFG (<https://lamcad.ufg.br/>).

## Referências

- Alexander A. Alemi, Matthew Bierbaum, Christopher R. Myers, and James P. Sethna. You can run, you can hide: The epidemiology and statistical mechanics of zombies. *Phys. Rev. E*, 92:052801, Nov 2015. doi: 10.1103/PhysRevE.92.052801. URL <https://link.aps.org/doi/10.1103/PhysRevE.92.052801>.
- A.-L. Barabási. *Network Science*. Cambridge University Press, Glasgow, 2016. URL <http://networksciencebook.com/>.
- M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Jour. of Theoretical Biology*, 21:275–88, 2022. doi: 10.1016/j.jtbi.2005.01.011. URL <https://doi.org/10.1016/j.jtbi.2005.01.011>.
- Roberto da Silva and Henrique A Fernandes. A study of the influence of the mobility on the phase transitions of the synchronous sir model. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(6):P06011, jun 2015. doi: 10.1088/1742-5468/2015/06/P06011. URL <https://dx.doi.org/10.1088/1742-5468/2015/06/P06011>.
- P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959. URL <https://doi.org/10.1109/MCSE.2007.55>.
- P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.
- P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, 12:261–267, 1961.
- P. F. Gomes, H. A. Fernandes, and A. A. Costa. Topological transition in a coupled dynamics in random networks. *Physica A*, 597:127269, 2022. URL <https://doi.org/10.1016/j.physa.2022.127269>.

- A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics and function using networkx. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, volume 445, page 11–15, 2008. doi: [http://conference.scipy.org/proceedings/SciPy2008/paper\\_2/](http://conference.scipy.org/proceedings/SciPy2008/paper_2/).
- C. R. Harris and *et al.* Array programming with numpy. *Nature*, 585:357–362, 2020. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4): 599–653, 2000. doi: 10.1137/S0036144500371907. URL <https://doi.org/10.1137/S0036144500371907>.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9:90–95, 2007. URL <https://doi.org/10.1109/MCSE.2007.55>.
- Smith TC Cartwright RA Lofgren ET, Collins KM. Equations of the end: Teaching mathematical modeling using the zombie apocalypse. *National Center for Biotechnology Information*, 17, março 2016. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4798798/#b1-jmbe-17-134>.
- W. McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56, 2010. doi: <https://doi.org/10.25080/Majora-92bf1922-00a>.
- L. H. F. Miranda, B. V. Ribeiro, P. M. M. Rocha, D. D. A. Santos, and N. C. de Sena. Surviving the zombie apocalypse: A population dynamics based approach. *Revista Brasileira de Ensino de Física*, 42:1806–9126, 2020. doi: 10.1590/1806-9126-RBEF-2020-0071. URL <https://doi.org/10.1590/1806-9126-RBEF-2020-0071>.
- M. E. J. Newman. *Networks: an introduction*. Oxford University Press, Oxford, 2010. URL <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>.
- S. M. Reia and J. F. Fontanari. Long-term scientific impact revisited. *Eur. Phys. J. Plus*, 137:161, 2022. doi: 10.1140/epjp/s13360-022-02376-5. URL <https://doi.org/10.1140/epjp/s13360-022-02376-5>.
- R. Smith? *Mathematical Modelling of Zombies*. EBSCO ebook academic collection. University of Ottawa Press, 2014. ISBN 9780776621685. URL <https://books.google.com.br/books?id=B7aVBQAAQBAJ>.
- R. Solomonoff and A. Rapoport. Connectivity of random nets. *B. Math. Biophys.*, 13: 107–117, 1951.
- E. B. Vilela, H. A. Fernandes, F. L. P. Costa, and P. F. Gomes. Phase diagrams of the ziff–gulari–barshad model on random networks. *Journal of Computational Chemistry*, 41:1964–1972, 2020. URL <http://dx.doi.org/10.1002/jcc.26366>.



# Estudo e Construção de Modelo Digital de Elevação (MDE) nas Regiões de Habitat de Peixes Ameaçados de Extinção no Noroeste Fluminense

Igor Martins Zanata<sup>1</sup>, Vicente de Paulo Santos de Oliveira<sup>1</sup> e Wagner Rambaldi Telles<sup>2</sup>

<sup>1</sup> Instituto Federal Fluminense, Campos dos Goytacazes/RJ, Brasil

<sup>2</sup> Universidade Federal Fluminense, Santo Antônio de Pádua/RJ, Brasil

---

## Resumo

O trabalho proposto tem como abordagem o estudo do Modelo Digital de Elevação (MDE) em um trecho do Rio Pomba e do Rio Paraíba do Sul, localizados no noroeste do estado do Rio de Janeiro. Buscas dos MDE's da região foram realizadas em sites governamentais e gratuitos e observou-se a necessidade da construção manual de um novo MDE da região, visto que os Modelos Digitais de Elevação disponibilizados não estão coerentes com a região de estudo. O *software* utilizado para o auxílio no desenvolvimento do MDE da região foi o QGIS, sendo gratuito e de código-fonte livre. Para as simulações hidráulicas foi utilizado o *software* de modelagem computacional chamado Iber. Desta forma, após o desenvolvimento manual do MDE, foram estudadas as características topográficas e hidráulicas da região para implementação da simulação.

**Palavras-chave:** Digital, Elevação, MDE, Iber, Modelagem.

---

## 1 INTRODUÇÃO

Os Objetivos de Desenvolvimento Sustentável (ODS) são um apelo global à ação para acabar com a pobreza, proteger o meio ambiente e o clima e garantir que as pessoas, em todos os lugares, possam desfrutar de paz e de prosperidade. Os ODS são divididos em 17 tópicos, (figura 1), contendo ainda seus subtópicos, onde tem os objetivos para os quais as Nações Unidas estão contribuindo a fim de que se possa atingir a Agenda 2030 no Brasil.

Este trabalho tem como referências os ODS números 6 (Água Potável e Saneamento) e 14 (Vida na Água) que contribuem para melhorar a qualidade da água, reduzir a poluição, proteger e restaurar ecossistemas contribuindo assim para a vida aquática.



**Fig. 1:** Ícones representativos das ODS.

**Fonte:** <https://odsbrasil.gov.br/>

Os ODS buscam implementar planos de gestão com base científica, para restaurar populações de peixes no menor tempo possível, pelo menos a níveis que possam produzir rendimento máximo sustentável e legal para a conservação das espécies e de seus recursos nos rios e nos oceanos, conforme registrado no parágrafo 158 do “Futuro Que Queremos” assinalado no manual do ODS, tópico 14c.

No entanto, a distribuição espacial dos índices de adequação de habitat requer dados hidráulicos mais detalhados para determinar a distribuição espacial de profundidade e velocidade ao longo dos rios [Bovee et al., 2007]. Além disso, a geomorfologia do rio é um fator muito importante para a produção de modelos hidrodinâmicos.

Além de estudar espécies de peixes nativas da região de atuação do Projeto Piabanha, que é uma Organização Não Governamental que faz a reprodução de algumas espécies ameaçadas da região, pode-se combinar índices de adequação de habitat com o resultado do modelo hidráulico para avaliar a relação entre o fluxo do rio e a habitabilidade dos peixes que serão estudados.

Estas informações são pilares recomendados para um rio com uma boa diversidade de vida aquática, através de simulações, obtendo índice quali-quantitativo para cada espécie que relaciona a vazão e a área de habitat utilizável.

O trabalho tem como objetivo localizar, simular e comparar os MDE’s disponíveis em um trecho do Rio Pomba. Esta busca pelo MDE ideal é importante para realizar estudos futuros sobre as espécies de peixes ameaçados de extinção que vivem na região e também realizar estudos mais precisos sobre áreas de alagamentos em zona urbana em épocas de enchentes.

E para iniciar o estudo mencionado no título deste trabalho, necessita-se de estudos dos MDE’s da área em questão, onde são informadas as elevações sobre o terreno, incluindo até mesmo a batimetria (profundidade) da calha dos rios da região de estudo, para posteriormente realizar as simulações hidrológicas dos rios Pomba e Paraíba do Sul.

## 2 REVISÃO DA LITERATURA

Os MDEs são representações matemáticas da topografia, servindo de suporte para o desenvolvimento de pesquisas e políticas, sobretudo os modelos que são gratuitos, como o SRTM, ASTER GDEM, TOPODATA e ALOS [Farr *et al.*, 2007; Tachikawa *et al.*, 2011; Valeriano & Rosseti, 2012]. Considerando a utilização de MDE’s, é possível reduzir custos, entregar informações relevantes e aplicar decisões de melhoria para uma região, por exemplo.

O sensor da missão SRTM (*Shuttle Radar Topography Mission*) da NASA (*National Aeronautics and Space Administration*) é um dos mais conhecidos e confiáveis na disponibilização e utilização de suas imagens para análises digitais. Foi o primeiro sensor, lançado em 2000, a produzir o mais completo mapeamento da topografia terrestre [Farr

et al, 2007].

Ao assumir que a topografia é um fenômeno dependente da escala [Florinsky, 2012], as imagens da missão SRTM disponíveis podem não ser suficientes para caracterização de determinadas feições devido a resolução espacial máxima alcançada por este tipo de imagem.

O MDE SRTM é distribuído na resolução de 30m para o território dos EUA e de 90m para as demais áreas continentais. O Brasil possui estes dados na resolução de 30m derivados de interpolação dos modelos de 90m [Valeriano & Rossetti, 2011], os quais podem ser adquiridos na plataforma TOPODATA de dados geomorfológicos disponibilizada pelo INPE (Instituto Nacional de Pesquisas Espaciais), pelo site <http://www.dpi.inpe.br/topodata/data/grd/>.

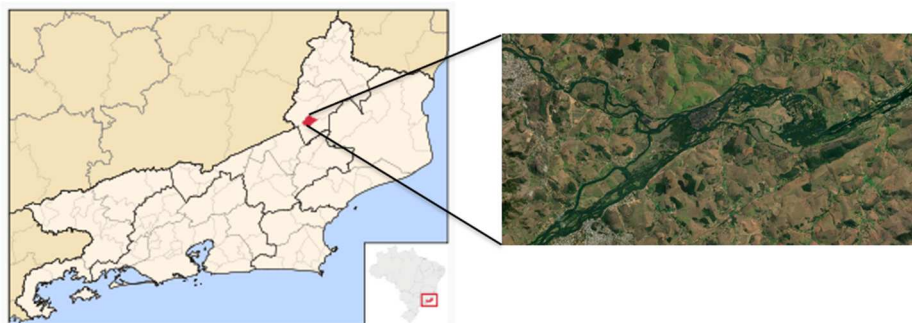
De acordo com Macêdo & Surya [2019] o sensor PALSAR do satélite ALOS (*Advanced Land Observing Satellite*) foi lançado em 2006 pela missão da agência de exploração aeroespacial japonesa (*Japan Aerospace Exploration Agency - JAXA*), desenvolvido para contribuir nas áreas de mapeamento, observação precisa da cobertura regional da terra, monitoramento de desastres e levantamento de recursos, utilizando-se de um radar de abertura sintética que opera na Banda L, capaz de obter imagens diurnas ou noturnas e em quaisquer condições atmosféricas.

Macêdo & Surya [2019] ressaltam que a coleta de dados do sensor PALSAR durou até 2011 e foi projetado com mais dois instrumentos para contribuir no mapeamento da topografia terrestre, precisão no monitoramento da cobertura vegetal e de desastres e no levantamento de recursos naturais [ASF DAAC, 2018]. Os MDE's ALOS PALSAR podem ser adquiridos gratuitamente, já corrigidos e projetados no sistema UTM WGS84, nas resoluções baixa e alta, de 30m e 12,5m, respectivamente, através do site <https://www.asf.alaska.edu/sar-data/palsar/about-palsar/>.

Todas estas informações citadas anteriormente foram analisadas para a região de estudo em questão, coletando informações sobre elevações e batimetrias na proximidade do rio Pomba e do rio Paraíba do Sul, considerando a sobrevivência e reprodução de algumas espécies de peixes da região que estão sob ameaça de extinção, como por exemplo a Piabanha (*Brycon insignis*), que é uma espécie nativa e endêmica do rio Paraíba do Sul, de acordo com Fowler [1950].

### 3 METODOLOGIA

A região escolhida para o desenvolvimento deste trabalho foi o trecho na proximidade da foz do rio Pomba com o rio Paraíba do Sul, entre os municípios de Aperibé e Itaocara, no noroeste do estado do Rio de Janeiro, conforme figura 2.



**Fig. 2:** Localização da área de estudo

**Fonte:** Autores

Esta localização é fundamental, pois fica próximo ao chamado Domínio das Ilhas Fluviais (DIF), identificada também como área de reprodução das espécies nativas da região, que tem proximidade da sede do Projeto Piabanha. Apesar da grande importância ecológica, não há registros deste tipo de estudo na região.

Modelos matemáticos e computacionais têm sido amplamente utilizados em estudos sobre os rios e mares, por serem instrumentos tecnológicos capazes de avaliar hidrológicamente os impactos que a água tem sobre os peixes viventes em determinada região, principalmente os nativos ameaçados de extinção, podendo ser mais ou menos precisos, dependendo das hipóteses adotadas na formulação.

Toda análise, simulação e estudo para o desenvolvimento deste trabalho foi realizado com o auxílio do *software* Iber, que é um *software* gratuito e de modelo hidráulico bidimensional para simulação de escoamento superficial em rios e estuários, além de fazer análise de hidrodinâmica, turbulência, transporte de sedimentos, processos de qualidade da água e habitat dos peixes.

Para a calibração do modelo, foram realizadas buscas nas bases de dados disponibilizadas nas plataformas da ANA (Agência Nacional de Águas) e do INEA (Instituto Estadual do Ambiente) sobre o rio Pomba e o rio Paraíba do Sul, além das características físico-químicas da água, medições de profundidade dos rios e velocidade de fluxo.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Coleta de informações no Topodata

Após selecionada a região de estudo, foram coletadas informações sobre MDE da região. No primeiro momento foi verificado o MDE disponibilizado pelo Banco de Dados Geomorfométricos do Brasil, também conhecido por Topodata.

A região de estudo possui uma vasta área territorial, que para fins didáticos e para fins de validação observou-se um pequeno trecho do Rio Pomba, conforme figura 3.



**Fig. 3:** Trecho de estudo do Rio Pomba.

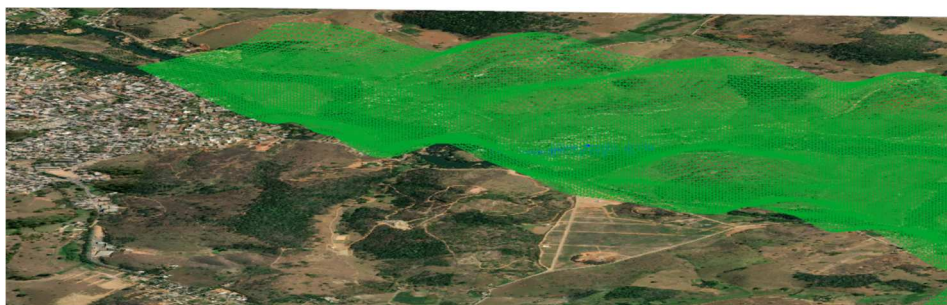
**Fonte:** Autores

Para fins de referência foi utilizada a estação meteorológica da ANA, localizada no mapa como “S02” (em vermelho), ou também conhecida por “Seção Transversal-2”, conforme estudo realizado pela ANA. Esta estação coleta informações sobre o rio Pomba, localizada em Aperibé-RJ, que de acordo com a ANA, o rio possui neste ponto cota de 59,66m, onde utilizaremos desta informação para a realizar as simulações hidráulicas. Como segundo ponto de referência utilizaremos a “Seção Transversal 1” - S01 (também em



vermelho no mapa), com cota de 57,62m.

Para a simulação utilizando o *software* Iber, foi criada uma malha de tamanho de 20m, pois entende-se como um tamanho aceitável para o estudo. Em relação à malha, sabe-se que quanto maior o tamanho da malha, mais rápido é o processamento via computador e menor a precisão dos resultados. E quanto menor a malha, maior é o curso computacional (maior tempo de processamento) e melhores são as definições na simulação, demonstrando a situação mais próxima à realidade. Após criada a malha, foi inserido o MDE do Topodata da região de estudo, observando-se notoriamente a inserção de elevações no mapa, conforme figura 4.

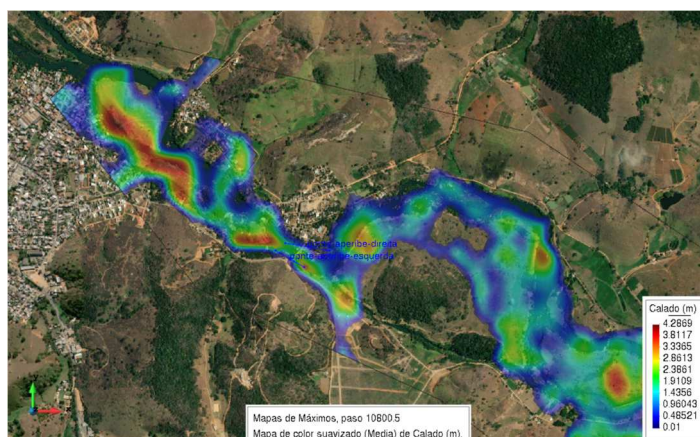


**Fig. 4:** Trecho de estudo do Rio Pomba com o MDE do Topodata.

**Fonte:** Autores

Além das informações sobre o MDE, foram inseridas as condições de contorno sobre a cota e vazão do rio Pomba e rio Paraíba do Sul, coletadas juntamente com a ANA, em período de condições normais, ou seja, sem transbordos ou outras adversidades.

Pode-se observar na figura 5, que após a conclusão da simulação hidráulica do *software* Iber (tempo aproximado de 2h e 16min de processamento), algumas regiões ficaram inundadas, em especial a área urbana.

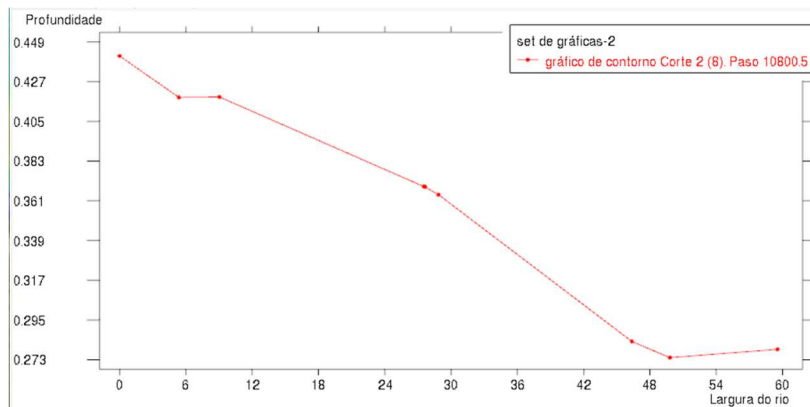


**Fig. 5:** Simulação utilizando MDE do Topodata.

**Fonte:** Autores

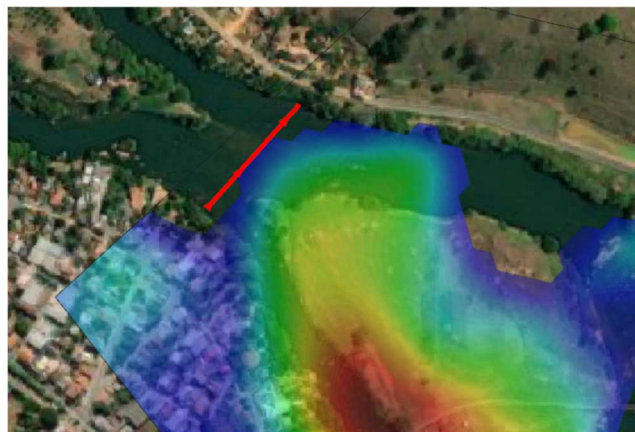
Analisando a seção transversal S01 e S02 no rio Pomba, com o *software* Iber, pode-se coletar informações gráficas sobre a cota naquele ponto em questão e verificar a causa do transbordo.

Na seção transversal S01 observa-se a largura do rio em aproximadamente 60m com uma fina lâmina d'água, variando de 27cm à 45cm de profundidade, conforme figura 6.



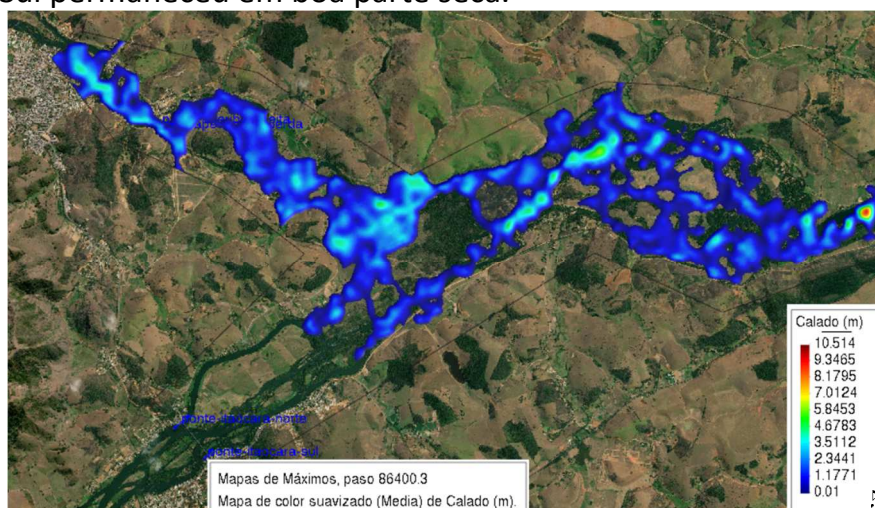
**Fig. 6:** Profundidade do rio na seção S01.  
**Fonte:** Autores

No momento da extração do gráfico para a seção transversal S02 ocorreu um erro, informando não haver rio naquele ponto, pois há uma indicação de possível problema no MDE, visto que dentro da calha do rio existe uma área seca, conforme figura 7.



**Fig. 7:** Seção transversal do rio Pomba.  
**Fonte:** Autores

Para uma percepção mais ampla da região de estudo, pode-se observar na figura 8 a região total simulada, com uma profundidade máxima de 10.51m em alguns pontos. Um ponto importante para observação foi a região próxima à cidade de Itaocara, onde a calha do rio Paraíba do Sul permaneceu em boa parte seca.



**Fig. 8:** Região simulada com MDE do Topodata.

**Fonte:** Autores

Como observado anteriormente nas simulações, o MDE coletado no Topodata não foi coerente com a realidade da região de estudo, visto que após a simulação pode-se observar pontos de alagamento em locais urbanizados, e até mesmo áreas secas dentro da própria calha do rio. Havendo também uma profundidade de até 4,29m em áreas urbanas. Por tanto, pode-se afirmar que o MDE em questão não está coerente com a realidade. Estas situações podem ocorrer por problemas de resolução espacial do próprio arquivo (MDE) no momento em que ele é coletado pelo satélite ou até mesmo devido ao tamanho da malha que utilizamos no momento da configuração do Iber.

#### **4.2 Coleta de informações no Alos Palsar**

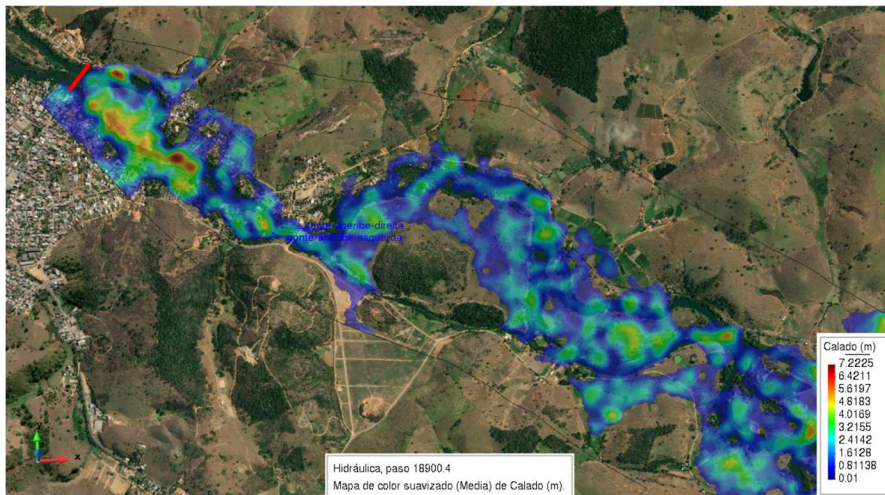
Após testes realizados com o Topodata, o próximo passo foi realizar estudos e testes com o MDE fornecido pelo Alos Palsar. E para manter as simulações de forma igualitária para posterior comparação, sendo utilizado novamente o *software* Iber com malha de 20m. Observe na figura 9, o trecho de estudo em questão já se encontra com as elevações do MDE do Alos Palsar.



**Fig. 9:** Trecho de estudo do Rio Pomba com o MDE do Alos Palsar.

**Fonte:** Autores

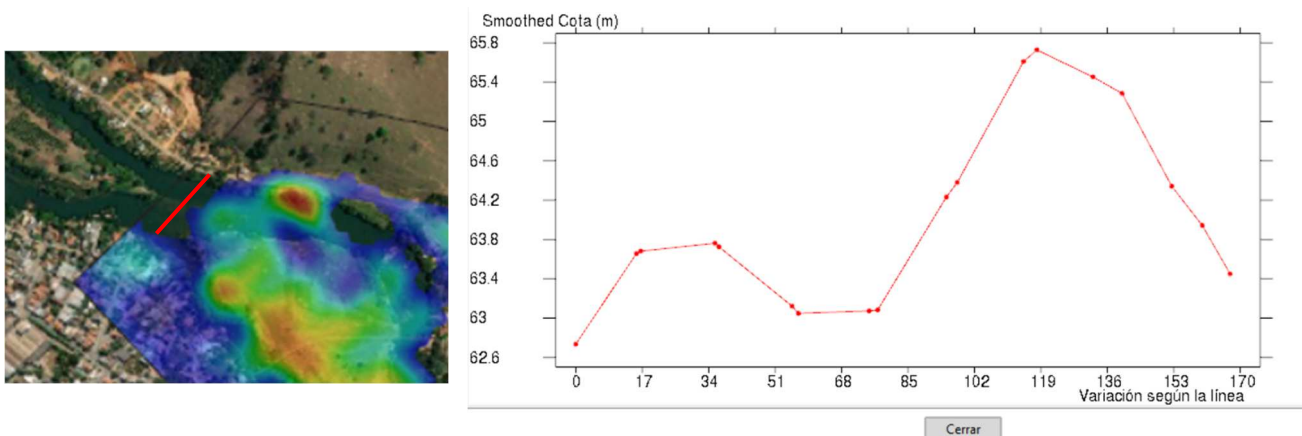
Para esta simulação, foram utilizadas as mesmas condições de contorno da simulação anterior, utilizando-se ainda a seção transversal S02 como referência. Observa-se que após concluída a simulação (com tempo aproximado de 1h e 12min de processamento), na figura 10, houve transbordo do rio Pomba para a área urbana e também para regiões adjacentes. Com profundidade máxima de 7,22m (nas regiões destacadas em vermelho). Um detalhe que se pode observar também foram alguns pontos de regiões secas dentro da própria calha do rio Pomba, mesmo problema apresentado pelo Topodata.



**Fig. 10:** Simulação utilizando MDE do Alos Palsar  
**Fonte:** Autores

Pode-se observar que a figura 11, ilustra o tamanho do rio naquela posição de leitura (seção transversal S02), contendo os mesmos 170m anteriormente medidos, de uma borda à outra e com elevações dentro do rio, ou seja, com o uso deste MDE observa-se que há elevações dentro da calha do rio, ultrapassando até mesmo a cota do rio, fato que não é coerente com a realidade.

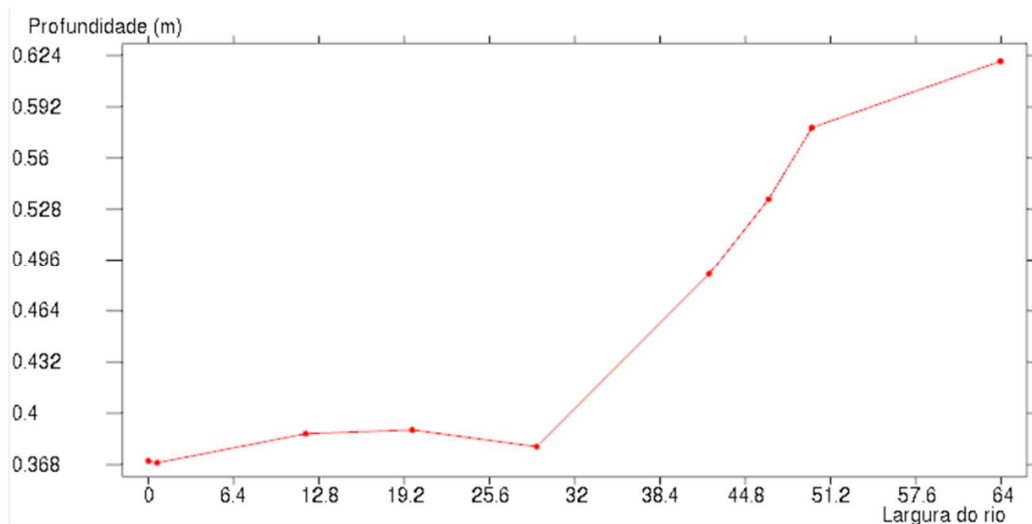
Observa-se que de um lado da borda do rio, a elevação está em aproximadamente 62,6m e vai aumentando de 63,8m até 65,8m. Chegando na outra borda do rio com aproximadamente 63,4m, ou seja, acima dos 59,66m informados pela estação da ANA. Para fins de comparação, é como se tivesse um morro/elevação dentro da calha do rio, e, portanto, justificando o motivo de estar sem água naquela localidade.



**Fig. 11:** Plotagem do gráfico da topografia – MDE Alos Palsar – Cota x Largura do rio  
**Fonte:** Autores

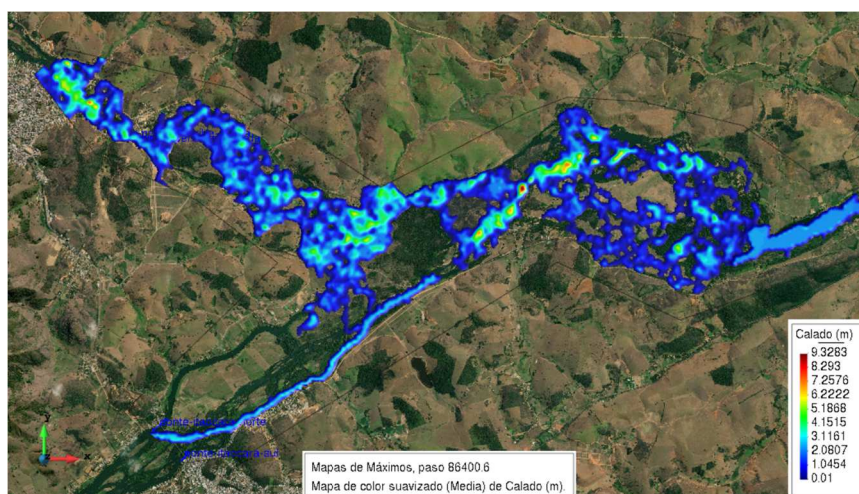
Além de detectada a elevação excessiva em alguns pontos na calha do rio, foi observada também transbordo inesperado em algumas partes da região de estudo.

Na seção transversal S01, conforme figura 12, obteve uma medida aproximada entre 36cm à 62cm de profundidade, um pouco mais profunda em relação aos dados simulados pelo Topodata. Todas estas informações obtidas após a simulação são incompatíveis com a realidade.



**Fig. 12:** Ponto de leitura para o gráfico em S01– Profundidade x Largura do rio  
**Fonte:** Autores

Nota-se na figura 13, uma visão mais ampla de toda a região simulada, a profundidade em alguns pontos do rio chega a até 9,32m. Observa-se também que tanto a simulação utilizando o MDE do Topodata quanto a simulação do Alos Palsar, ocorreram erros similares. Porém observa-se que o MDE do Alos Palsar já se mostra mais definida, ressaltando a calha do rio Paraíba do sul, mas ainda com os mesmos erros de incoerência em relação ao MDE do Topodata, como alagamentos e transbordos em áreas urbanas.

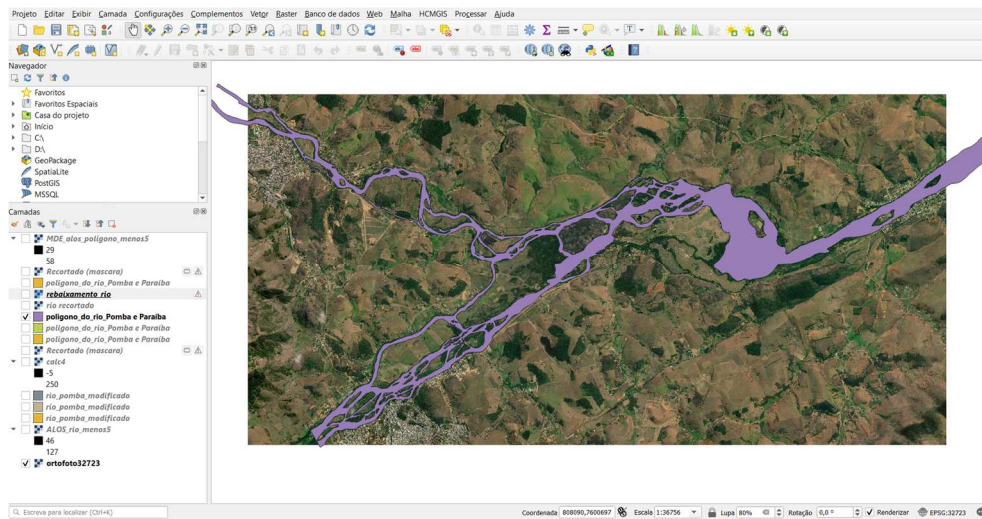


**Fig. 13:** Ponto de leitura para o gráfico  
**Fonte:** Autores

### 4.3 Construção manual de um novo MDE

Para o desenvolvimento de um novo MDE, foi necessário a utilização do *software* QGis, juntamente com o MDE do Alos Palsar que utilizamos como referência, por ter uma maior resolução espacial e por ter se mostrado mais eficaz para a simulação.

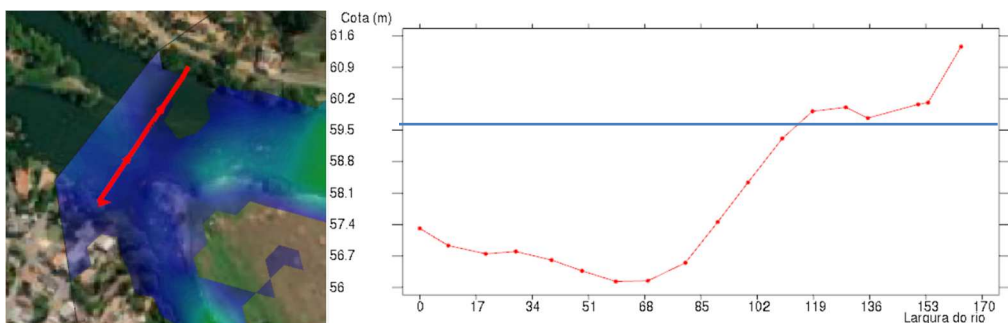
Como primeiro passo, foi adicionado uma ortofoto da região de estudo e criado um polígono contornando toda a calha do rio Pomba e do rio Paraíba do Sul, conforme figura 14, fazendo o recorte de ilhas e de pequenos córregos.



**Fig. 14:** Criação do polígono no Qgis  
**Fonte:** Autores

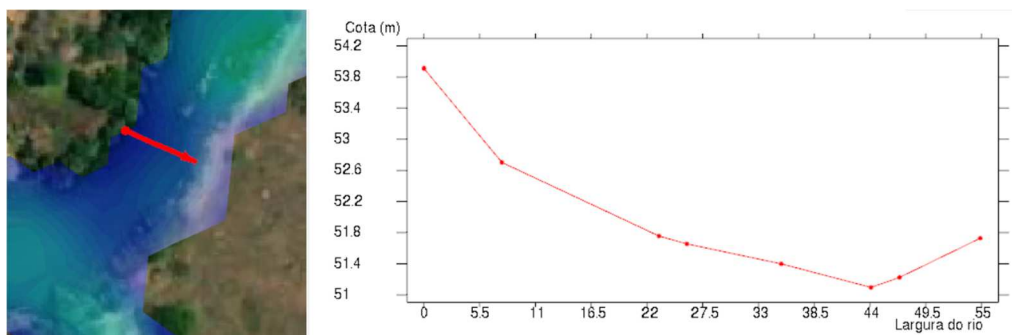
Posteriormente foi adicionado o MDE do Alos Palsar para realizar a marcação de toda a calha do rio e basicamente foram utilizados recursos do *software* Qgis para o rebaixamento da calha do rio em 7m, seguindo a profundidade média observada na tabela de topobatimetria da ANA, dos rios em questão.

Após realizada a simulação no *software* Iber, com tempo aproximado de 1h e 17min de processamento, obteve-se a figura 15 como resultado da topografia da seção transversal S01. Observa-se que o gráfico mostra o nível da água do rio (linha em azul) juntamente com a elevação do fundo do rio. É possível notar que este MDE ainda carece de ajustes, pois continua com algumas regiões com áreas secas dentro da calha do rio.



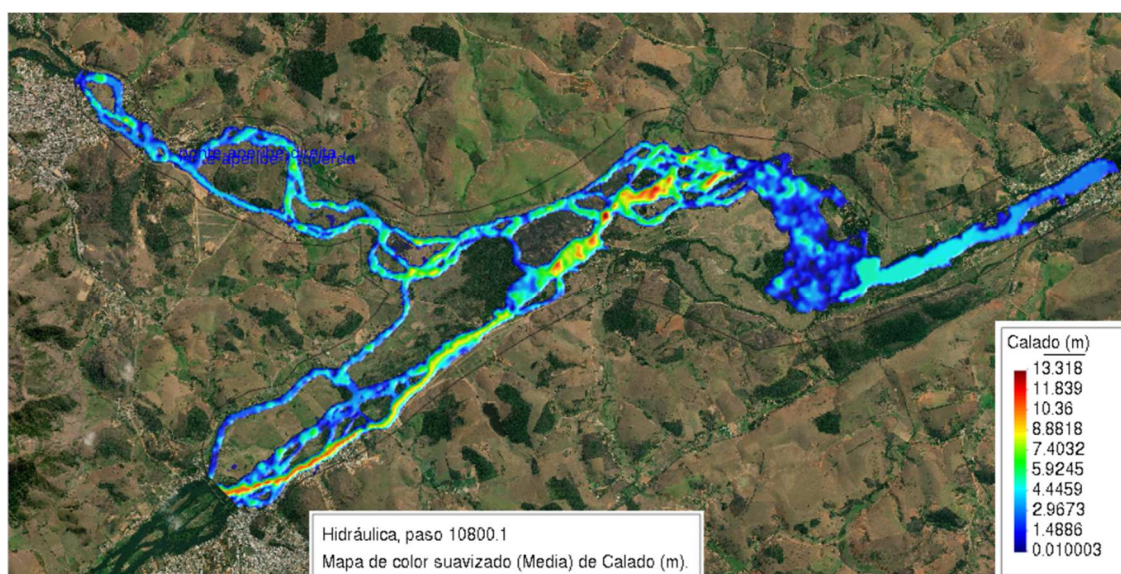
**Fig. 15:** Plotagem do gráfico da topografia – Novo MDE  
**Fonte:** Autores

Na seção transversal S01, figura 16, pode-se observar a cota em relação à largura do rio. Note também que ocorre o transbordo do rio Pomba nesta região. Observa-se ainda que a linha em vermelho representa elevação do fundo do rio em relação ao nível do mar (cota), onde segundo a ANA é de 57,62m, justificando o alagamento na região, pois trata-se da própria calha do rio.



**Fig. 16:** Plotagem do gráfico da topografia – Novo MDE  
**Fonte:** Autores

Porém, analisando o novo MDE adaptado com o rebaixamento da calha do rio, pode-se observar a simulação completa, figura 17, o encontro dos Rios Pomba e Paraíba do Sul, com sua profundidade em todo o curso da área de estudo.



**Fig. 17:** Simulação com a profundidade do novo MDE.  
**Fonte:** Autores

## 5 CONCLUSÕES

Após analisar os MDE's disponibilizados originalmente pelo Topodata e pelo Alos Palsar pode-se observar que ambos não condizem com a realidade da região de estudo do rio Pomba e do rio Paraíba do Sul. Devido a isto, foi necessário o desenvolvimento de um novo MDE, adaptado do Alos Palsar. Este novo MDE foi criado de forma manual e com o auxílio dos recursos do *software* QGis.

As informações sobre a batimetria foram baseadas nas seções transversais informadas pela ANA, com isso obteve-se o novo MDE com as informações mais próximas da realidade sobre a região de estudo em questão.

Após analisadas as opções estudadas, conclui-se que o melhor MDE foi o MDE adaptado do Alos Palsar (construído manualmente) incluindo o rebaixamento de 7m. Este MDE trouxe maior coerência com a realização da simulação.

Para estudos futuros serão utilizados o módulo de Qualidade da Água (IberWQ), que calcula a evolução espacial e temporal de várias informações, incluindo: Oxigênio dissolvido, demanda bioquímica de oxigênio (DBO), nitrogênio, nitrito, nitrato,

temperatura e salinidade, além do Iber Habitat, que analisa parâmetros de habitabilidade das espécies de peixes na região de estudo.

## **6 *Agradecimentos***

Aos orientadores Vicente e Wagner pelo apoio incondicional na estruturação deste trabalho, além de toda equipe do Doutorado em Modelagem e Tecnologia para Meio Ambiente Aplicadas em Recursos Hídricos – Ambhidro do Instituto Federal Fluminense.



## Referências

- ASF DAAC. Alaska Satellite Facility Distributed Active Archive Center. Disponível em: <https://www.asf.alaska.edu/sar-data/palsar/about-palsar/>. Acesso em: 08/08/2022.
- Bovee, K.D. et al., Stream habitat analysis using the instream flow incremental
- Buckup, P.A., Menezes, N.A., Ghazzi, M.S. (eds.) (2007) *Catálogo das espécies de peixes de água doce do Brasil*. Rio de Janeiro, Museu Nacional. 195p. (Série Livros, 23) ISBN: 978-85-7427-018-0.
- Fernandes, A.F.D. *Instrumentação e Controle para Regulação de pH em Sistemas de Cultivo de Organismos Aquáticos* (Dissertação). Universidade do Algarve – Faculdade de Ciência e Tecnologia. 2012. 101p.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S. Kobrick, M., Paller, M., Rodriguez, E., Roth, L. 2007. The shuttle radar topography mission. *Rev. Geophys*, v. 45
- Florinsky, I. V. Digital terrain analysis in soil science and geology. 1ª ed. Russia: Elsevier, 2012.
- Fowler, H.W.(1950). *Os peixes de água doce do Brasil*. *Arq. Zool.*, 6:333-340.
- Laranjeira, R.B. *Controlo remoto e autónomo de um sistema de aquacultura*. (Dissertação). Universidade de Aveiro. 2014. 99p.
- Macêdo, R. J. A., & Surya, L. (2019). *Comparação entre modelos digitais de elevação dos sensores srtm e alos palsar para análise digital de terreno*. *Revista Contexto Geográfico*, 3(6), 47–55. <https://doi.org/10.28998/contegeo.v3i6.6968>
- Nomura, H.(1984) *Dicionário de Peixes do Brasil*. Brasília: Editerra, 482p.
- Pereira, R. (1986) *Peixes de Nossa Terra*. 2. ed: NOBEL, 129p.
- Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D.B., Oimoen, M.J., Zhang, Z., Danielson, J.J., Krieger, T., Curtis, B., Haase, J. 2017. *Aster global digital elevation model version 2- summary of validation results*. Tech. rep. NASA. 2011.
- Topodata. Banco de Dados Geomorfométricos do Brasil - TOPODATA. 2019. Disponível em: < <http://www.dsr.inpe.br/topodata/dados.php> >. Acesso em: 08/08/2022.
- Valeriano, M. M.; Rossetti, D. F. *Topodata: Brazilian full coverage refinement of SRTM data*. *Applied Geography* (Sevenoaks), 32: 300-309. 2011



# Modelos epidemiológicos baseados em redes: Uma revisão preliminar de escopo

Jesuliana Nascimento Ulysses<sup>1</sup> and Antônio Tadeu Azevedo Gomes<sup>1</sup>

<sup>1</sup> LNCC, Petrópolis/RJ, Brasil

---

## Abstract

A COVID-19 afetou bilhões de pessoas e interferiu na economia mundial, forçando controles sociais sem precedentes para mitigar a rápida propagação da doença. Esse cenário estimulou a realização de pesquisas sobre modelos epidemiológicos visando entender e combater tal propagação. Inicialmente, os principais modelos explorados foram os compartimentais, os quais não consideraram a evolução da população nos compartimentos ao longo do espaço, apresentando limitações, uma delas é não levar em conta a mobilidade humana. Nesse sentido, os modelos baseados em redes fornecem informações importantes para entender o papel dos padrões de mobilidade na dinâmica da transmissão de doenças. Apresentamos neste artigo uma revisão preliminar de escopo sobre modelos baseados em redes combinados com modelos compartimentais. Tal revisão permitiu-nos identificar diferentes estratégias para a representação da mobilidade nesses modelos combinados. Nossa contribuição com este artigo está na categorização inicial da bibliografia encontrada segundo as estratégias identificadas, permitindo-nos observar que: (i) a inclusão da mobilidade nos modelos epidemiológicos empregados no estudo da COVID-19 é de fato importante; mas (ii) os modelos combinados consideram somente a origem e o destino de indivíduos da população no estabelecimento dos padrões de mobilidade, sem considerarem a interferência do percurso desses indivíduos na propagação da doença. Entendemos que este levantamento preliminar abre uma perspectiva para o desenvolvimento futuro de modelos com melhor capacidade analítica e preditiva.

**Keywords:** Modelos baseados em rede, Mobilidade, COVID-19, Modelos epidemiológicos

---

## 1 Introdução

Em dezembro de 2019 a COVID-19, doença causada pelo coronavírus SARS-CoV-2, teve seu primeiro caso registrado na China e se espalhou rapidamente por vários países. Em 11 de março de 2020, o surto de COVID-19 foi declarado como pandemia global pela Organização Mundial da Saúde (OMS), tendo nesse dia afetado 126.702 pessoas em todo

o mundo (Kuhl, 2021). Comparado com outras epidemias o coronavírus SARS-CoV-2 espalhou-se pelo mundo em um ritmo muito mais rápido (Zhang e Li, 2021). Essa propagação deveu-se não somente à característica epidêmica do vírus, mas também à influência da globalização e de uma sociedade marcada por uma intensa mobilidade. Como consequência, os sistemas públicos de saúde ficaram sobrecarregados e o bem-estar social e econômico de todos os países foi perturbado. A vida diária de bilhões foi impactada pelos controles sociais sem precedentes impostos pelos governos para inibir a propagação do vírus (Kuzdeuov et al., 2020).

A rápida disseminação da COVID-19 estimulou o desenvolvimento de pesquisas visando compreender e combater essa nova ameaça (Costa et al., 2020). O interesse deveu-se aos questionamentos e desafios relacionados à natureza do coronavírus, tais como (Cao e Liu, 2022): (i) as características epidêmicas do vírus (processos de transmissão e influência); (ii) as características médicas e genômicas da doença (dinâmica e evolução); e (iii) os prós e contras das estratégias existentes de contenção, diagnóstico, tratamento e prevenção. Neste contexto, muitos pesquisadores desenvolveram modelos epidêmicos cada vez mais específicos ao estudo do SARS-CoV-2 e assim buscaram contribuir na definição das melhores estratégias para o controle da sua disseminação (Bertaglia e Pareschi, 2021).

Inicialmente, a maioria dos pesquisadores adotou modelos compartimentais, adaptando-os às características epidemiológicas da COVID-19. No entanto, como colocam Kuzdeuov et al. (2020),

“os modelos compartimentais têm várias limitações baseadas em suposições simplificadoras que não se correlacionam bem com a propagação viral real. Por exemplo, os modelos assumem que as taxas de recuperação e infecção são as mesmas para todos os indivíduos da população. Além disso, os modelos não levam em conta a mobilidade populacional.” (Kuzdeuov et al., 2020, tradução nossa)<sup>1</sup>

Em outras palavras, o constante movimento das pessoas pode atuar como um mecanismo de potencialização da contaminação. De fato, muitos países durante o período crítico da pandemia impuseram *lockdown*, forçando um distanciamento social e diminuindo a circulação de pessoas, não somente a nível local, mas também entre regiões. A adoção de características espaciais/geográficas que consigam capturar a dinâmica de mobilidade humana nos modelos epidemiológicos que estudam a COVID-19 permite avaliar a eficácia dessas medidas.

Neste contexto, os modelos baseados em rede fornecem informações importantes para entender o papel dos padrões de mobilidade na dinâmica da transmissão de doenças infecciosas. Esses modelos foram construídos considerando uma rede formada por nós representando indivíduos, cidades, países ou estados, e arestas responsáveis pela representação da conexão entre esses nós.

Este artigo tem como objetivo apresentar uma **revisão preliminar de escopo** de modelos desenvolvidos para o estudo da COVID-19 que incluem mobilidade, os quais utili-

<sup>1</sup>The compartmental models, however, have several limitations based on simplifying assumptions that do not correlate well to actual viral propagation. For instance, models assume that recovery and infection rates are the same for all individuals in the population. Also, the models do not take population mobility into account.



zam modelos baseados em rede combinados com modelos compartimentais. Até onde é de nosso conhecimento, não há trabalhos na literatura que se propusessem a fazer esse tipo de revisão, embora duas revisões sistemáticas existentes abordem aspectos semelhantes de nosso trabalho. A revisão sistemática realizada por Zhang et al. (2022) buscou por pesquisas que associassem mobilidade humana e COVID-19, organizando os resultados em termos de dados e fonte de dados utilizados, modelos adotados, propostas de estudo e principais conclusões. A partir dos resultados obtidos os autores propuseram pesquisas futuras, observando as lacunas existentes relacionadas ao tema. Por sua vez, o levantamento feito por Benita (2021) examinou publicações associadas às pesquisas relacionadas à COVID-19 e ao comportamento da mobilidade humana, principalmente no tocante ao uso do transporte urbano. Além disso, propôs um *framework* baseado em aprendizado de máquina que pudesse fornecer um mapeamento dessas pesquisas. As revisões citadas visam pesquisas que tratem a interferência da mobilidade humana na propagação da COVID-19, selecionando artigos que trouxessem esses dois assuntos associados, diferentemente do escopo de pesquisa do presente artigo, que objetiva estudos que envolvam o desenvolvimento de um modelo baseado em rede combinado com modelos compartimentais.

Como resultado dessa revisão, observamos o fato dos trabalhos estudados concluírem que a diminuição da mobilidade, através de medidas restritivas impostas pelos governos, está diretamente relacionada com a diminuição da disseminação da COVID-19. Contudo, identificamos algumas potenciais lacunas de investigação que abrem uma perspectiva para o desenvolvimento futuro de modelos com melhor capacidade analítica e preditiva. Em particular, notamos que a influência da movimentação da população na disseminação da COVID-19 é analisada nos trabalhos estudados considerando somente a origem e o destino de indivíduos da população no estabelecimento dos padrões de mobilidade, sem levar em conta a interferência do percurso desses indivíduos na propagação da doença. Também observamos que, até o momento, a maioria dos modelos ainda considera a população homogênea com relação às taxas de recuperação e infecção, sem explorar o potencial dos modelos baseados em rede na representação da heterogeneidade.

O restante deste artigo está estruturado como se segue. A Seção 2 explica a metodologia adotada na pesquisa para a seleção das publicações dentro do escopo pretendido. A Seção 3 apresenta como os trabalhos na literatura analisada até o momento utilizam redes para modelar e estudar a COVID-19. Por fim, a Seção 4 apresenta nossas considerações finais.

## 2 Metodologia

Este estudo realizou uma pesquisa em 9 bases de dados – *ACM Digital Library*, *IEEE Digital Library*, *ISI Web of Science*, *PubMed*, *SIAM*, *Science@Direct*, *Scopus*, *Springer Link* e *Wiley* – visando coletar a produção científica existente que adotou modelos baseados em rede no estudo da propagação e prevenção da COVID-19. Na realização da busca foi utilizada a *string*:

```
("COVID-19" OR "SARS-CoV" OR "coronavirus") AND  
("Mobility" OR "complex network" OR  
"graphs" OR "network-based epidemiological") AND
```

("compartmental model" OR "epidemiological model")

Essa busca retornou um total de 511 artigos.

Inicialmente foi lido o *abstract* dos artigos, a fim de selecionar aqueles que atendessem ao escopo do estudo, separando no total 90 artigos. Após essa seleção prévia, um estudo mais aprofundado foi realizado com 30 artigos, escolhidos aleatoriamente desse conjunto de 90 artigos. Ao final, 17 artigos foram selecionados dentre esses 30 artigos para inclusão no estudo realizado neste trabalho, por apresentarem modelos baseados em rede que consideram padrões de mobilidade e modelos compartimentais. A partir da análise desses 17 artigos foi possível realizar um levantamento preliminar incluindo as características das redes empregadas e modelos compartimentais adotados.

### 3 Resultados e Discussão

Modelos de rede têm sido usados para investigar o impacto da estrutura espacial na transmissão de infecções. Uma **rede** (ou **grafo**) é formada por um conjunto de **nós** (ou **vértices**) que são interligados através de **arestas**. Os nós em uma rede podem representar qualquer objeto e conter informações relevantes ao problema representado. Neste estudo, no conjunto de artigos selecionados, foi possível identificar os nós representando principalmente indivíduos, grupos de indivíduos, ou regiões geográficas. Por sua vez, as arestas podem representar um provável caminho de transmissão da infecção. Logo, as conexões entre a estrutura da rede, a dinâmica da doença e as técnicas matemáticas nos permitem compreender e prever os resultados das epidemias. Portanto, visando analisar como as diferentes formas de representação de nós e arestas influenciam nos modelos baseados em rede, na Subseção 3.1 categorizamos os trabalhos estudados de acordo com o tipo de objeto sendo representado por cada nó da rede.

Outra característica que julgamos importante é o modelo compartimental utilizado, uma vez que a COVID-19 tem características epidemiológicas importantes tais como infeciosidade de indivíduos pré-sintomáticos e assintomáticos (Liu e Yamamoto, 2022; Wang et al., 2021). Por isso, na Subseção 3.2 discorreremos sobre os diferentes modelos compartimentais usados em combinação com modelos baseados em redes nos trabalhos estudados.

#### 3.1 Identificação dos nós nos modelos baseados em rede

**Nós como indivíduos.** A dinâmica de transmissão da rede nos modelos nos quais os nós são indivíduos dá-se segundo uma rede de contatos que depende da estrutura de interação entre os indivíduos. A estrutura de interação normalmente é representada por uma rede complexa de pequeno mundo (*small world*), como em (Batlle et al., 2022; Thurner et al., 2020), ou livre de escala (*scale-free*), isto é, com a distribuição do grau<sup>2</sup> dos nós obedecendo uma lei de potência, como em (Singh e Arquam, 2022). Nessas redes a dinâmica de transmissão é governada por modelos compartimentais, assim cada nó em um instante  $t$  de tempo pode assumir um estado referente ao modelo adotado.

Atualmente, as pesquisas relacionadas com a COVID-19, diferentemente do início da pandemia, não estão centradas somente em prever a trajetória da doença, mas também entender as características peculiares que a difere dos outros surtos de SARS (síndrome respiratória aguda grave) e o impacto das políticas públicas adotadas na contenção da

<sup>2</sup>O **grau** de um nó em uma rede é o número de arestas que ligam este nó a outros nós da rede.



pandemia (Caetano et al., 2021; Gozzi et al., 2021; Singh e Arquam, 2022; Tagliazucchi et al., 2020; Thurner et al., 2020).

Thurner et al. (2020) propuseram um estudo para entender o crescimento quase linear apresentado nas curvas de infecção da COVID-19 por um longo período nos casos confirmados. Os autores mostraram que para as taxas de transmissão existe um grau crítico na rede de contatos abaixo do qual ocorrem curvas lineares de infecção e acima aparecem as curvas clássicas em formato de “S” que são padrão nos modelos epidemiológicos.

Singh e Arquam (2022) desenvolveram um modelo simulando a rede social de contato humana através da inclusão da informação do grau de um nó em cada estado do modelo compartimental. Além disso, esses autores buscaram entender como a taxa de disseminação de indivíduos assintomáticos e a taxa de indivíduos testados afetam a população total em um determinado compartimento no tempo  $t$  referente aos nós (indivíduos) com grau  $k$ .

Batlle et al. (2022) propuseram a construção de um modelo que possibilitasse a detecção de um novo surto epidêmico precoce e a recuperação do máximo de informações sobre a evolução da epidemia. Uma rede complexa de mundo pequeno foi construída para simular a propagação da doença, sendo utilizado um modelo compartimental estocástico para o estudo da transmissão da COVID-19.

**Nós como regiões geográficas.** No estudo da COVID-19, diversas pesquisas consideraram a mobilidade humana, a fim de investigar a influência do fluxo de pessoas nas medidas de restrição adotadas pelos governos na prevenção e na propagação da COVID-19, analisando seus impactos socio-econômicos, a interferência na dinâmica de transmissão de uma região em outras regiões e a eficácia das medidas de restrição (Goel et al., 2021; Gösgens et al., 2021; Liu e Yamamoto, 2022; Wang et al., 2021). Nesse contexto, modelos baseados em rede tomam os nós como regiões, na maioria das vezes coincidindo com as divisões geográficas adotadas pelos países. Ding et al. (2021) foram a única exceção encontrada até momento, pois consideram aeroportos como nós da rede. Por sua vez, as arestas conectam essas regiões fazendo com que fluxos de indivíduos de uma região para outra possam ser representados no modelo. Informações sobre esses fluxos são obtidas, principalmente, através de dados de viagens entre regiões, utilizando informações de um ou mais sistemas de transporte, e de dados de mobilidade obtidos a partir de operadoras de telefonia móvel.

Liu e Yamamoto (2022) adicionaram mobilidade em seu modelo visando verificar como as medidas de restrição e os pedidos de permanência em casa interferem na propagação da doença. No modelo proposto os autores consideraram a movimentação de pessoas durante o dia e a noite. Nas equações utilizadas para modelar a dinâmica de transmissão durante o dia foi adicionada uma matriz contendo os dados de mobilidade intermunicipal.

Scala et al. (2020) pretenderam estudar como as intervenções de restrição de mobilidade e o momento da suspensão do *lockdown* afetam, conjuntamente, a fração total de pessoas infectadas, o pico de prevalência da doença e o atraso na dinâmica da epidemia. No desenvolvimento do modelo os autores assumiram ser conhecida a fração de indivíduos que comutam de uma região para outra, adicionando este termo no modelo compartimental SIOR (Suscetível, Infectado, Observado e Recuperado) executado em um determinado nó.

Gösgens et al. (2021) usaram os dados de mobilidade para analisar se a divisão das regiões determinada por decisões governamentais reflete o movimento real de pessoas em todo país, examinando se essa divisão é a mais eficaz na mitigação da COVID-19. Assim, o objetivo principal do artigo é criar regiões que permitam o máximo de mobilidade, mas consigam restringir ao máximo as infecções. Para isto, um conjunto de áreas atômicas foi utilizado para representar um país e os indivíduos movimentam-se entre essas áreas. Um método de detecção da comunidade foi empregado para obter as divisões com base na mobilidade e uma variação do modelo compartimental SEIR (Suscetível, Exposto, Infectado e Recuperado) é executado em cada região.

Wang et al. (2021) incorporaram mobilidade para verificar o impacto do fluxo da população na disseminação da COVID-19. No modelo proposto a dinâmica de transmissão da doença em um determinado nó origina-se de duas fontes: local e migração. A dinâmica de transmissão local é dada pelo modelo compartimental SaucIR (Suscetível, Assintomático, Infectado mas não confirmado, Infectado confirmado e Recuperado). Já a migração descreve a dinâmica de mobilidade humana na rede.

Zhang e Li (2021) investigaram a relação da mobilidade com a disseminação da doença, procurando prever as regiões que seriam as fontes principais de transmissão mais prováveis no futuro. Esse conhecimento é importante para formular estratégias para mitigar a disseminação da doença e alocar recursos de saúde pública eficientemente. O modelo desenvolvido nesse estudo integrou um modelo de aprendizado de máquina baseado em *deep embedding*<sup>3</sup> e um modelo compartimental.

A fim de observar a necessidade de medidas de mitigação da doença Kuzdeuov et al. (2020) desenvolveram um simulador epidêmico estocástico baseado em rede. O simulador pode ser usado para estimar a extensão e duração da epidemia ao longo do tempo e modelar o impacto das medidas não farmacológicas. No modelo desenvolvido cada nó da rede executa um modelo SEIR consistindo de 4 superestados,  $S^S$ ,  $E^S$ ,  $I^S$  e  $R^S$ , sendo que esses contêm estados intermediários, e os indivíduos transitam entre os estados baseados em probabilidades de transição. Na rede, a transição de uma região para outra é dada por uma matriz que combina dados de viagens diárias de diferentes meios de transporte (aéreo, ferroviário e rodoviário).

Costa et al. (2020) adotaram a abordagem de meta-população para estudar a variabilidade geográfica e o impacto de uma doença altamente contagiosa em escala continental, quantificando o quão diversificada pode ser a pandemia considerando a extensão territorial de um país como o Brasil. Esse trabalho sugere que medidas uniformes de mitigação não são eficientes para reduzir o impacto da doença. Também no contexto nacional, Grave et al. (2022) apresentaram uma metodologia que incorpora uma estrutura de rede espacial em um sistema de equações diferenciais parciais (vide Subseção 3.2) descrevendo a propagação espaço-temporal da epidemia por meio de um modelo de reação e difusão.

**Nós como grupos de indivíduos.** Nonato et al. (2022) desenvolveram uma plataforma integrada de otimização computacional chamada *Robot Dance*. A plataforma avalia e estima as consequências de intervenções a nível regional quando há um surto e a influência

<sup>3</sup>Um *embedding* em aprendizado de máquina é um mapeamento de uma variável categórica em um espaço discreto de dimensionalidade alta em um vetor em um espaço contínuo de dimensionalidade reduzida.



da movimentação de pessoas residentes em uma região na propagação da doença. O *Robot Dance* divide a população de uma região em grupos. O estado epidemiológico da região é caracterizado pelo modelo compartimental SEIR, sendo que a cada compartimento é associado um vetor  $[0, 1]^I$  (sendo  $I$  o conjunto de nós da rede) com as correspondentes porcentagens referentes ao estado dos grupos de população em relação à transmissão da doença e o estado da região no tempo  $t$  é fornecido por uma função vetorial em  $[0, 1]^{4 \times |I|}$ . Além disso, o modelo considera que não há mobilidade durante a noite.

Shi et al. (2022) adotaram a teoria do lugar central (Hall e Pain, 2006) e a teoria do fluxo central (Castells, 2010; Meijers, 2007) para entender a influência da organização espacial na propagação da COVID-19 no fluxo de viagens, bem como o planejamento e as políticas públicas adotadas em uma região. No tocante ao modelo, cada nó da rede é composto por um grupo de indivíduos dentro de uma região (cidade), sendo que os nós conectam regiões diferentes e a transmissão pode ocorrer de forma intra-nós ou entre-nós. Observando as teorias adotadas, o modelo primeiro calcula a relação da cidade na rede, estimado pelo grau do nó refletindo a teoria do fluxo central e a sua posição estrutural relacionada à teoria do lugar central. Em seguida, o modelo compartimental SI (Suscetível, Infectado) é empregado para representar a disseminação da transmissão da COVID-19 e, por último, as ligações entre as regiões são reorganizadas para diferentes estruturas de rede, a fim de testar a associação entre estruturas de rede e contágio viral.

**Modelos hierárquicos.** Por fim, alguns trabalhos (Chang et al., 2021; Goel et al., 2021) adotaram modelos considerando a interação entre indivíduos, grupos e regiões, sendo cada um deles nós em grafos distintos dispostos em uma estrutura hierárquica.

No modelo desenvolvido por Goel et al. (2021) as regiões são distribuídas em uma rede em forma de malha quadriculada. Cada célula da rede é considerada uma região e os indivíduos dentro de uma região se conectam de acordo com uma entre duas estruturas de rede possíveis: *fully mixed* e rede complexa. Na versão *fully mixed*, os indivíduos em cada região estão todos conectados; já na versão rede complexa, os indivíduos que pertencem a uma dada região se conectam segundo uma rede complexa livre de escala gerada sinteticamente.

Chang et al. (2021) apresentam um modelo que inclui três submodelos organizados em níveis. O nível I representa o modelo de rede de viagens entre cidades. Já o nível II representa o modelo de viagens entre comunidades e o movimento entre uma comunidade e outra segue a característica de redes livres de escala do tipo BA (Barabási e Albert, 1999). Por fim, o nível III representa o modelo de rede de contato entre pessoas, construído como uma rede complexa completa. Através dessa divisão os autores propõem um modelo de transmissão adaptativo baseado em redes complexas com mobilidade populacional, estrutura social e transmissão de doenças infecciosas. Os autores realizaram simulações com esse modelo analisando o impacto das políticas de restrição adotadas pelas cidades na disseminação da COVID-19.

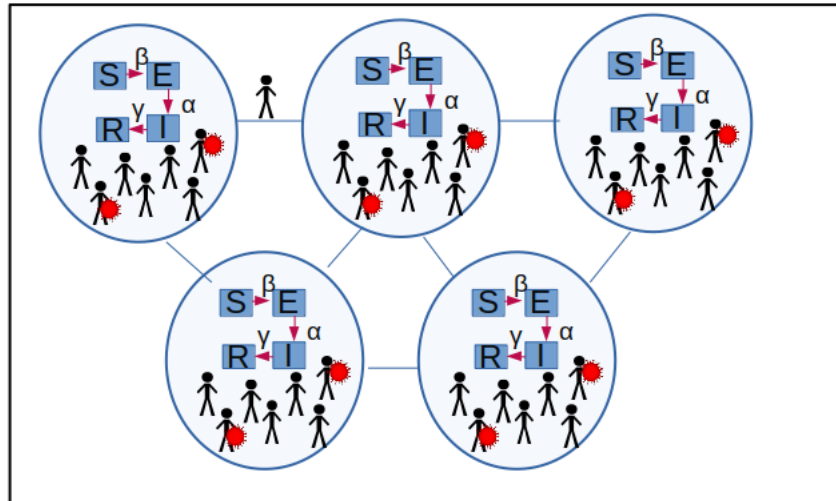
### 3.2 Modelos baseados em rede e seus modelos compartimentais

A dinâmica da COVID-19 nos modelos baseados em rede estudados até o momento é dada pelos modelos compartimentais. Nos modelos compartimentais a população é dividida em compartimentos. Cada compartimento simula o comportamento do seu subgrupo e os



indivíduos se movem entre esses compartimentos. Por exemplo, no modelo SEIR a ordem indica o movimento sucessivo entre os compartimentos Suscetível, Exposto, Infectado e Recuperado (Kuhl, 2021).

A maioria dos trabalhos analisados executa o modelo compartimental no nó da rede, conforme exemplificado na Figura 1 para o caso do nó representar uma região ou um grupo de indivíduos. Desta forma cada nó tem seu estado epidêmico e adiciona ao mesmo a contribuição de um modelo de mobilidade, tipicamente por meio de expressões algébricas que descrevem a influência mútua desse nó e nós vizinhos a ele.



**Fig. 1:** Representação esquemática de um modelo baseado em rede combinado com o modelo SEIR.

Segundo Kuzdeuov et al. (2020)

“Em geral, os compartimentos e as transições de estado são selecionados com base nas características de uma doença específica e no propósito do modelo.”  
(Kuzdeuov et al., 2020, tradução nossa)<sup>4</sup>

Assim, para refletir as características da COVID-19, a maioria dos trabalhos analisados incluíram novos compartimentos aos modelos compartimentais clássicos, como o SEIR. Um dos compartimentos mais presentes nesses modelos modificados é o compartimento assintomático (Batlle et al., 2022; Bertaglia e Pareschi, 2021; Chang et al., 2021; Costa et al., 2020; Ding et al., 2021; Grave et al., 2022; Kuzdeuov et al., 2020; Liu e Yamamoto, 2022; Singh e Arquam, 2022; Wang et al., 2021), uma vez que pessoas assintomáticas da COVID-19 podem transitar e iniciar a disseminação da COVID-19 em outras localidades. Identificamos também, em alguns modelos, a presença de estados representando grupos de indivíduos testados e não testados. Cabe observar a existência somente no trabalho de Kuzdeuov et al. (2020) de um compartimento referindo-se a grupos de indivíduos vacinados. No entanto, esse não foi explorado no artigo, uma vez que o desenvolvimento de vacinas ainda estava em andamento.

<sup>4</sup>In general, the compartments and related state transitions are selected based on the characteristics of a specific disease and the purpose of the model.



Nos modelos compartimentais o fluxo entre os compartimentos é descrito tipicamente através de um sistema de equações diferenciais ordinárias (EDOs). No entanto, há exceções (Bertaglia e Pareschi, 2021; Grave et al., 2022). A motivação para se utilizar sistemas de equações diferentes de EDOs no estudo de epidemias está relacionado ao fato que EDOs descrevem somente a evolução temporal da propagação da epidemia, não considerando a componente espacial, sendo essa importante, principalmente, quando é necessário examinar intervenções heterogêneas espacialmente.

Bertaglia e Pareschi (2021) propõem um modelo de transporte hiperbólico na rede que utiliza o sistema de velocidade discreta associado ao modelo epidêmico SEIAR (Suscetível, Exposto, Infectado, Assintomático, Recuperado). Além de considerar a movimentação espacial da população, o modelo possibilita a incorporação de incertezas para descrever a propagação espacial da epidemia.

Grave et al. (2022) propuseram a combinação do sistema de equações diferenciais parciais (EDPs) de reação e difusão apresentado por Viguerie et al. (2020) combinado com um modelo baseado em rede. O sistema de EDPs foi utilizado presumindo que a difusão ocorre localmente em cada nó (isto é, dentro das cidades) e a dinâmica não local foi modelada através de operadores de transporte entre nós, considerando somente origem e destino da movimentação populacional.

#### 4 Conclusão

A Organização Mundial da Saúde continua considerando a COVID-19 uma emergência global de saúde, sendo relatado até a presente data mais de 629 milhões de casos confirmados e mais de 6 milhões de mortes no mundo. Essa situação justifica o interesse dos cientistas em continuar pesquisando sobre a doença e seus impactos socioeconômicos no mundo.

Este artigo realizou uma revisão preliminar de estudos que combinaram modelos epidemiológicos compartimentais com modelos baseados em rede, categorizando-os segundo o tipo de informação que o nó representa e mostrando de que forma se dá a incorporação dos aspectos de mobilidade nos modelos compartimentais. Neste sentido, há trabalhos que modelam a propagação de doenças na rede relacionando com a teoria de redes sociais. Nesses trabalhos a dinâmica de propagação da doença na rede apresenta uma estrutura de rede na qual os indivíduos (nós) são conectados por contato e comunicação, e cada nó está em um dos estados do modelo compartimental adotado na rede. Normalmente, essas redes fazem o uso da rede de contatos para modelar o espalhamento da doença, analisando a disseminação da doença localmente. Em outros estudos os nós representam regiões ou grupos de indivíduos (em vez de indivíduos) e as arestas capturam o movimento de um local para outro. Esses trabalhos, em sua maioria, estão interessados em analisar qual a influência de um surto de uma região em outras regiões devido à movimentação de indivíduos.

Destacamos que o estudo de modelos baseados em rede tornou-se de fundamental importância, principalmente no tocante à interferência da mobilidade no estudo da COVID-19, fornecendo ferramentas para compreensão e prevenção de doenças infecciosas, além de permitir analisar o resultado das medidas de restrição adotadas pelos governos para mitigar a doença. Logo, este estudo coletou informações relevantes relacionadas a modelos epidemiológicos baseados em rede combinados com modelos compartimentais e possibili-

tou contribuir fornecendo uma análise preliminar da literatura, identificando lacunas de estudo que ainda não foram abordadas.

Como trabalho futuro, pretendemos continuar na exploração da literatura selecionada, eventualmente reorganizando ou estendendo as categorizações propostas. Além disso, consideramos utilizar cienciométrica, a fim de quantificar, analisar e avaliar tal literatura. Ainda como trabalho futuro planejamos realizar uma análise mais crítica dos modelos compartimentais baseados em rede, considerando principalmente suas limitações, produzir um paralelo com os modelos compartimentais.

## 5 Agradecimentos

Agradecemos o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ na realização do presente trabalho.

## Referências

- Barabási, A.-L. e Albert, R. (1999). “Emergence of scaling in random networks”. *science*, 286(5439):509–512.
- Batlle, P., Bruna, J., Fernandez-Granda, C., e Preciado, V. M. (2022). “Adaptive test allocation for outbreak detection and tracking in social contact networks”. *SIAM Journal on Control and Optimization*, 60:S274–S293.
- Benita, F. (2021). “Human mobility behavior in covid-19: A systematic literature review and bibliometric analysis”. *Sustainable Cities and Society*, 70:102916.
- Bertaglia, G. e Pareschi, L. (2021). “Hyperbolic compartmental models for epidemic spread on networks with uncertain data: Application to the emergence of covid-19 in italy”. *Mathematical Models e Methods in Applied Sciences*, 31:2495–2531.
- Caetano, C., Morgado, M. L., Patrício, P., Pereira, J. F., e Nunes, B. (2021). “Mathematical modelling of the impact of non-pharmacological strategies to control the covid-19 epidemic in portugal”. *Mathematics*, 9(10):1084.
- Cao, L. e Liu, Q. (2022). “Covid-19 modeling: A review”. *medRxiv*.
- Castells, M. (2010). “Globalisation, networking, urbanisation: Reflections on the spatial dynamics of the information age”. *Urban studies*, 47(13):2737–2745.
- Chang, F., Wu, F., Chang, F., e Hou, H. (2021). “Research on adaptive transmission and controls of covid-19 on the basis of a complex network”. *Computers & Industrial Engineering*, 162:107749.
- Costa, G. S., Cota, W., e Ferreira, S. C. (2020). “Outbreak diversity in epidemic waves propagating through distinct geographical scales”. *Physical Review Research*, 2.
- Ding, X., Huang, S., Leung, A., e Rabbany, R. (2021). “Incorporating dynamic flight network in seir to model mobility between populations”. *Applied Network Science*, 6.



- Goel, R., Bonnetain, L., Sharma, R., e Furno, A. (2021). “Mobility-based sir model for complex networks: with case study of covid-19”. *Social Network Analysis and Mining*, 11(1):1–18.
- Gösgens, M., Hendriks, T., Boon, M., Steenbakkers, W., Heesterbeek, H., Van Der Hofstad, R., e Litvak, N. (2021). “Trade-offs between mobility restrictions and transmission of sars-cov-2”. *Journal of the Royal Society Interface*, 18(175):20200936.
- Gozzi, N., Bajardi, P., e Perra, N. (2021). “The importance of non-pharmaceutical interventions during the covid-19 vaccine rollout”. *PLoS computational biology*, 17(9):e1009346.
- Grave, M., Viguerie, A., Barros, G. F., Reali, A., Andrade, R. F., e Coutinho, A. L. (2022). “Modeling nonlocal behavior in epidemics via a reaction-diffusion system incorporating population movement along a network”. *arXiv preprint arXiv:2205.04868*.
- Hall, P. G. e Pain, K. (2006). *The polycentric metropolis: Learning from mega-city regions in Europe*. Routledge.
- Kuhl, E. (2021). *Computational Epidemiology: Data-Driven Modeling of COVID-19*. Springer International Publishing, Cham.
- Kuzdeuov, A., Baimukashev, D., Karabay, A., Ibragimov, B., Mirzakhmetov, A., Nurpeissov, M., Lewis, M., e Varol, H. (2020). “A network-based stochastic epidemic simulator: Controlling covid-19 with region-specific policies”. *IEEE Journal of Biomedical and Health Informatics*, 24:2743–2754.
- Liu, S. e Yamamoto, T. (2022). “Role of stay-at-home requests and travel restrictions in preventing the spread of covid-19 in japan”. *Transportation Research part a-policy and practice*, 159:1–16.
- Meijers, E. (2007). “From central place to network model: theory and evidence of a paradigm change”. *Tijdschrift voor economische en sociale geografie*, 98(2):245–259.
- Nonato, L. G., Peixoto, P., Pereira, T., Sagastizábal, C., e Silva, P. J. (2022). “Robot dance: A mathematical optimization platform for intervention against covid-19 in a complex network”. *EURO Journal on Computational Optimization*, 10:100025.
- Scala, A., Flori, A., Spelta, A., Brugnoli, E., Cinelli, M., Quattrocioni, W., e Pammolli, F. (2020). “Time, space and social interactions: exit mechanisms for the covid-19 epidemics”. *Scientific Reports*, 10.
- Shi, S., Pain, K., e Chen, X. (2022). “Looking into mobility in the covid-19 ‘eye of the storm’: Simulating virus spread and urban resilience in the wuhan city region travel flow network”. *Cities*, 126:103675.
- Singh, A. e Arquam, M. (2022). “Epidemiological modeling for covid-19 spread in india with the effect of testing”. *Physica A: Statistical Mechanics and its Applications*, 592:126774.

- Tagliazucchi, E., Balenzuela, P., Travizano, M., Mindlin, G., e Mininni, P. D. (2020). “Lessons from being challenged by covid-19”. *Chaos, Solitons & Fractals*, 137:109923.
- Turner, S., Klimek, P., e Hanel, R. (2020). “A network-based explanation of why most covid-19 infection curves are linear”. *Proceedings of the National Academy of Sciences of the United States of America*, 117:22684–22689.
- Viguerie, A., Veneziani, A., Lorenzo, G., Baroli, D., Aretz-Nellesen, N., Patton, A., Yankeelov, T. E., Reali, A., Hughes, T. J., e Auricchio, F. (2020). “Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to covid-19, mathematical analysis, and numerical study”. *Computational Mechanics*, 66(5):1131–1152.
- Wang, X., Yang, L., Zhang, H., Yang, Z., e Liu, C. (2021). “Forecasting confirmed cases of the covid-19 pandemic with a migration-based epidemiological model”. *Statistics and its Interface*, 14:59–71.
- Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., Halloran, B., Li, Z., Cui, Y., Liu, H., et al. (2022). “Human mobility and covid-19 transmission: a systematic review and future directions”. *Annals of GIS*, 28(4):501–514.
- Zhang, T. e Li, J. (2021). “Understanding and predicting the spatio-temporal spread of covid-19 via integrating diffusive graph embedding and compartmental models”. *Transactions in GIS*, 25:3025–3047.



# Multiscale Hybrid-Hybrid-Mixed Method

Franklin da Conceição de Barros<sup>1</sup>, Alexandre Loreiro Madureira<sup>2</sup> e Frédéric Gerard  
Christian Valentin<sup>1,2</sup>

<sup>1</sup> *Laboratório Nacional de Computação Científica, Petrópolis/RJ, Brazil*

---

## Abstract

We propose a numerical method, the Multiscale Hybrid-Hybrid-Mixed method (MH<sup>2</sup>M), to solve elliptic partial differential equations with multiscale coefficients. It is derived from the Three-Field Domain Decomposition method and searches a non-conforming solution and two conform Lagrange multipliers: the trace and the flux. A series of static condensations transforms a saddle-point problem into an elliptic one, posed on the interfaces. At the discrete level, this drastically reduces the size of the global system. The matrix of the associated linear system is symmetric and positive definite, and can be solved by classical iterative schemes such as the Conjugated Gradient method. We present the well-posedness of the method and establish error estimates. We also perform numerical tests to confirm the theoretical predictions and compare the method with the FEM, MsFEM and MHM schemes.

**Keywords:** Numerical methods, Finite element methods, Multiscale Hybrid-Hybrid-Mixed method, Multiscale Hybrid-Mixed method.

---

## 1 SETTING AND PRELIMINARY RESULTS

The MH<sup>2</sup>M method [C.Barros et al., 2022] was introduced as a master's thesis work at the Laboratório Nacional de Computação Científica - LNCC, in Petrópolis,RJ. We present its development and its main properties. Our problem is posed on a open bounded domain  $\Omega \subset \mathbb{R}^d$ ,  $d = 2, 3$ , with a polygonal boundary  $\partial\Omega$ . Here we consider  $d = 2$  for simplicity. Given  $f \in L^2(\Omega)$ , we need to find  $u : \Omega \rightarrow \mathbb{R}$  such that

$$\begin{aligned} -\nabla \cdot (\mathcal{A}\nabla u) &= f, & \text{in } \Omega \\ u &= 0, & \text{on } \partial\Omega \end{aligned} \quad (1)$$

where  $\mathcal{A} \in [L^\infty(\Omega)]^{d \times d}$  is a symmetric tensor and there exists positive constants  $a_{\min}$  and  $a_{\max}$  such that

$$a_{\min}|\mathbf{v}|^2 \leq \mathcal{A}(\mathbf{x})\mathbf{v} \cdot \mathbf{v} \leq a_{\max}|\mathbf{v}|^2, \quad \mathbf{v} \in \mathbb{R}^d, \text{ a. e. } \mathbf{x} \in \Omega \quad (2)$$

We take a partition of domain  $\Omega$  as follows: for  $\mathcal{H} \in ]0, 1[$ , let  $\mathcal{T}_{\mathcal{H}}$  be a regular mesh of  $\Omega$  composed of elements  $K \in \mathcal{T}_{\mathcal{H}}$  and let  $\mathcal{E}_{\mathcal{H}}$  be the mesh skeleton composed of edges. For  $K \in \mathcal{T}_{\mathcal{H}}$ ,  $\mathbf{n}^K$  is the outward normal vector on  $\partial K$  and  $\mathbf{n}$  is the outward normal vector on  $\partial\Omega$ . Consider the broken spaces:

$$H^1(\mathcal{T}_{\mathcal{H}}) := \{v \in L^2(\Omega); v|_K \in H^1(K), K \in \mathcal{T}_{\mathcal{H}}\}, \quad \Lambda := \prod_{K \in \mathcal{T}_{\mathcal{H}}} H^{-\frac{1}{2}}(\partial K), \quad (3)$$

$$H_0^{1/2}(\mathcal{E}_{\mathcal{H}}) := \{v|_{\mathcal{E}_{\mathcal{H}}}; v \in H_0^1(\Omega)\}. \quad (4)$$

For  $w, v \in L^2(\Omega)$ ,  $\rho \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}})$  and  $\mu \in \Lambda$ , let:

$$(v, w)_{\mathcal{T}_{\mathcal{H}}} := \sum_{K \in \mathcal{T}_{\mathcal{H}}} \int_K wv \, dx, \quad \langle \mu, \rho \rangle_{\mathcal{E}_{\mathcal{H}}} = \sum_{K \in \mathcal{T}_{\mathcal{H}}} \langle \mu, \rho \rangle_{\partial K}, \quad (5)$$

where  $\langle \cdot, \cdot \rangle_{\partial K}$  is the dual product involving  $H^{-1/2}(\partial K)$  and  $H^{1/2}(\partial K)$ . Then, the equivalent mixed-hybrid formulation of (1) consists of finding  $u \in H^1(\mathcal{T}_{\mathcal{H}})$ ,  $\rho \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}})$  and  $\lambda \in \Lambda$  such that

$$\begin{aligned} (\mathcal{A}\nabla u, \nabla v)_{\mathcal{T}_{\mathcal{H}}} - \langle \lambda, v \rangle_{\mathcal{E}_{\mathcal{H}}} &= (f, v)_{\mathcal{T}_{\mathcal{H}}}, & \forall v \in H^1(\mathcal{T}_{\mathcal{H}}); \\ -\langle \mu, u \rangle_{\mathcal{E}_{\mathcal{H}}} + \langle \mu, \rho \rangle_{\mathcal{E}_{\mathcal{H}}} &= 0, & \forall \mu \in \Lambda; \\ \langle \lambda, \xi \rangle_{\mathcal{E}_{\mathcal{H}}} &= 0, & \forall \xi \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}}). \end{aligned} \quad (6)$$

This is the Three-field formulation Brezzi and Marini [1992] introduced to solve second order linear elliptic problems and allows to apply domain decomposition methods. The system (24) consists of a saddle-point problem posed on the interfaces, which is hard to solve numerically. Moreover, the system becomes large for problems with multiscale coefficients. To circumvent these difficulties, we decompose the function spaces to apply static condensations.

First we show that (1) and (24) are equivalent problems. In fact, if  $u$  is the weak solution of (1), then  $\lambda := \mathcal{A}\nabla u \cdot \mathbf{n}^K$  and  $\gamma := u|_{\mathcal{E}_{\mathcal{H}}}$  solve (24). Conversely, let  $u \in H^1(\mathcal{T}_{\mathcal{H}})$ ,  $\rho \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}})$  and  $\lambda \in \Lambda$  solve (24). It follows from the third equation and [Gatica, 2014, Lemma 3.4] that  $\lambda|_{\partial K} = \sigma \cdot \mathbf{n}^K$ , for  $\sigma \in H(\text{div}; \Omega)$ . The second equation together with [Gatica, 2014, Theorem 3.1] imply  $u \in H_0^1(\Omega)$ . Finally, the first equation of (24) and again [Gatica, 2014, Lemma 3.4] give us that  $u \in H_0^1(\Omega)$  is the weak solution of (1).

We decompose elements of  $H^1(\mathcal{T}_{\mathcal{H}})$  and  $\Lambda$  as "constant" plus "zero average", i.e.

$$H^1(\mathcal{T}_{\mathcal{H}}) = \mathbb{P}_0(\mathcal{T}_{\mathcal{H}}) \oplus \tilde{H}^1(\mathcal{T}_{\mathcal{H}}), \quad \text{and} \quad \Lambda = \Lambda^0 \oplus \tilde{\Lambda}, \quad (7)$$

where  $\mathbb{P}_0(\mathcal{T}_{\mathcal{H}})$  is the space of piecewise constants in each element and

$$\tilde{H}^1(\mathcal{T}_{\mathcal{H}}) := \left\{ v \in H^1(\mathcal{T}_{\mathcal{H}}); \int_{\partial K} v \, ds = 0, \forall K \in \mathcal{T}_{\mathcal{H}} \right\}; \quad (8)$$

$$\Lambda^0 := \text{span} \left\{ \mu^0 \in \Lambda; \langle \mu^0, v \rangle_{\partial K} := \int_{\partial K} v \, ds, v \in H^1(\mathcal{T}_{\mathcal{H}}), K \in \mathcal{T}_{\mathcal{H}} \right\}; \quad (9)$$

$$\tilde{\Lambda} := \Lambda \cap \mathbb{P}_0(\mathcal{T}_{\mathcal{H}})^\perp = \{ \lambda \in \Lambda; \langle \lambda, v \rangle_{\mathcal{E}_{\mathcal{H}}} = 0, v \in \mathbb{P}_0(\mathcal{T}_{\mathcal{H}}) \}. \quad (10)$$



We also define the zero average functional space on the element boundary

$$\tilde{H}^{1/2}(\partial K) := \left\{ \xi \in H^{1/2}(\partial K); \int_{\partial K} \xi \, ds = 0 \right\}, \quad (11)$$

for  $K \in \mathcal{T}_{\mathcal{H}}$ . Let the seminorm

$$|\lambda|_{-\frac{1}{2}, \partial K} := \sup_{\phi \in \tilde{H}^{\frac{1}{2}}(\partial K)} \frac{\langle \lambda, \phi \rangle_{\partial K}}{|\phi|_{\frac{1}{2}, \partial K}}, \quad |\lambda|_{\tilde{\Lambda}}^2 := \sum_{K \in \mathcal{T}_{\mathcal{H}}} |\lambda|_{-\frac{1}{2}, \partial K}^2. \quad (12)$$

Then we can write  $u = u^0 + \tilde{u}$ , where  $u^0 \in \mathbb{P}_0(\mathcal{T}_{\mathcal{H}})$  and  $\tilde{u} \in \tilde{H}^1(\mathcal{T}_{\mathcal{H}})$ , and also  $\lambda = \lambda^0 + \tilde{\lambda}$ , for  $\lambda^0 \in \Lambda^0$  and  $\tilde{\lambda} \in \tilde{\Lambda}$ . It follows from (24) that  $\lambda^0 \in \Lambda^0$  and  $u^0 \in \mathbb{P}_0(\mathcal{T}_{\mathcal{H}})$  solve

$$\begin{aligned} \langle \lambda^0, v^0 \rangle_{\mathcal{E}_{\mathcal{H}}} &= -(f, v^0)_{\mathcal{T}_{\mathcal{H}}}, \quad \forall v^0 \in \mathbb{P}_0(\mathcal{T}_{\mathcal{H}}); \\ \langle \mu^0, u^0 \rangle_{\mathcal{E}_{\mathcal{H}}} &= \langle \mu^0, \rho \rangle_{\mathcal{E}_{\mathcal{H}}}, \quad \forall \mu^0 \in \Lambda^0. \end{aligned} \quad (13)$$

Note that  $\lambda^0$  is defined by the first equation of (13). We obtain the piecewise constant  $u^0$  after the computation of  $\rho$ . Then  $(\tilde{u}, \tilde{\lambda}, \rho) \in \tilde{H}^1(\mathcal{T}_{\mathcal{H}}) \times \tilde{\Lambda} \times H_0^{1/2}(\mathcal{E}_{\mathcal{H}})$  solves

$$\begin{aligned} (\mathcal{A}\nabla\tilde{u}, \nabla\tilde{v})_{\mathcal{T}_{\mathcal{H}}} - \langle \tilde{\lambda}, \tilde{v} \rangle_{\mathcal{E}_{\mathcal{H}}} &= (f, \tilde{v})_{\mathcal{T}_{\mathcal{H}}}, & \forall \tilde{v} \in \tilde{H}^1(\mathcal{T}_{\mathcal{H}}), \\ -\langle \tilde{\mu}, \tilde{u} \rangle_{\mathcal{E}_{\mathcal{H}}} + \langle \tilde{\mu}, \rho \rangle_{\mathcal{E}_{\mathcal{H}}} &= 0, & \forall \tilde{\mu} \in \tilde{\Lambda}, \\ \langle \tilde{\lambda}, \xi \rangle_{\mathcal{E}_{\mathcal{H}}} &= -\langle \lambda^0, \xi \rangle_{\mathcal{E}_{\mathcal{H}}}, & \forall \xi \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}}). \end{aligned} \quad (14)$$

The first equation of (14) allows the introduction of local solvers. Let  $K \in \mathcal{T}_{\mathcal{H}}$ . For  $\tilde{\mu} \in \tilde{\Lambda}$  and  $f \in L^2(\Omega)$ , let  $T : \tilde{\Lambda} \rightarrow \tilde{H}^1(\mathcal{T}_{\mathcal{H}})$  and  $\tilde{T} : L^2(\Omega) \rightarrow \tilde{H}^1(\mathcal{T}_{\mathcal{H}})$  be such that

$$\int_K \mathcal{A}\nabla(T\tilde{\mu}) \cdot \nabla\tilde{v} \, dx = \langle \tilde{\mu}, \tilde{v} \rangle_{\partial K} \quad \text{and} \quad \int_K \mathcal{A}\nabla(\tilde{T}f) \cdot \nabla\tilde{v} \, dx = \int_K f\tilde{v} \, dx, \quad (15)$$

for all  $\tilde{v} \in \tilde{H}^1(K)$ . Then

$$\tilde{u} = T\tilde{\lambda} + \tilde{T}f. \quad (16)$$

Replacing in (14), we obtain the saddle-point problem of finding  $(\tilde{\lambda}, \rho) \in \tilde{\Lambda} \times H_0^{1/2}(\mathcal{E}_{\mathcal{H}})$  such that

$$\begin{aligned} -\langle \tilde{\mu}, T\tilde{\lambda} \rangle_{\mathcal{E}_{\mathcal{H}}} + \langle \tilde{\mu}, \rho \rangle_{\mathcal{E}_{\mathcal{H}}} &= \langle \tilde{\mu}, \tilde{T}f \rangle_{\mathcal{E}_{\mathcal{H}}}, & \forall \tilde{\mu} \in \tilde{\Lambda}; \\ \langle \tilde{\lambda}, \xi \rangle_{\mathcal{E}_{\mathcal{H}}} &= -\langle \lambda^0, \xi \rangle_{\mathcal{E}_{\mathcal{H}}}, & \forall \xi \in H_0^{1/2}(\mathcal{E}_{\mathcal{H}}). \end{aligned} \quad (17)$$

Next we show that the bilinear form  $\langle \cdot, T \cdot \rangle_{\mathcal{E}_{\mathcal{H}}}$  on  $\tilde{\Lambda}$  is coercive. For  $\tilde{\mu} \in \tilde{\Lambda}$ , we get:

$$\begin{aligned} \langle \tilde{\mu}, T\tilde{\mu} \rangle_{\mathcal{E}_{\mathcal{H}}} &= \sum_{K \in \mathcal{T}_{\mathcal{H}}} \langle \tilde{\mu}, T\tilde{\mu} \rangle_{\partial K} = \sum_{K \in \mathcal{T}_{\mathcal{H}}} \int_K \mathcal{A}\nabla(T\tilde{\mu}) \cdot \nabla(T\tilde{\mu}) \, dx \geq \sum_{K \in \mathcal{T}_{\mathcal{H}}} a_{\min} \|\nabla T\tilde{\mu}\|_{0,K}^2 \\ &\geq \sum_{K \in \mathcal{T}_{\mathcal{H}}} C \|T\tilde{\mu}\|_{1,K}^2 \geq \sum_{K \in \mathcal{T}_{\mathcal{H}}} C |\tilde{\mu}|_{-\frac{1}{2}, \partial K}^2 = C |\tilde{\mu}|_{\tilde{\Lambda}}^2, \end{aligned} \quad (18)$$



where we have used in the last two inequalities the Generalized Poincaré inequality and the injectivity of the operator  $T$ . It allows the introduction of the so called Dirichlet-to-Neumann operator Belishev and Sharafutdinov [2008]. We define:

$$G : \prod_{K \in \mathcal{T}_\mathcal{H}} H^{1/2}(\partial K) \longrightarrow \tilde{\Lambda}, \quad (19)$$

such that, for  $K \in \mathcal{T}_\mathcal{H}$  and given  $\phi \in H^{1/2}(\partial K)$ , if  $\tilde{\lambda}_\phi|_{\partial K} = G\phi|_{\partial K} \in \tilde{H}^{-1/2}(\partial K)$ , then

$$\int_K \mathcal{A} \nabla(T\tilde{\lambda}_\phi) \cdot \nabla T\tilde{\mu} \, dx = \langle \tilde{\mu}, T\tilde{\lambda}_\phi \rangle_{\partial K} = \langle \tilde{\mu}, \phi \rangle_{\partial K}, \quad \forall \tilde{\mu} \in \tilde{\Lambda}. \quad (20)$$

Then, it follows from (17) that

$$\tilde{\lambda} = G(\rho - \tilde{T}f). \quad (21)$$

A second static condensation in (17) implies the global problem that characterizes the MH<sup>2</sup>M method. It consists of finding  $\rho \in H_0^{1/2}(\mathcal{E}_\mathcal{H})$  such that

$$\langle G\rho, \xi \rangle_{\mathcal{E}_\mathcal{H}} = -\langle \lambda^0, \xi \rangle_{\mathcal{E}_\mathcal{H}} + \langle G\tilde{T}f, \xi \rangle_{\mathcal{E}_\mathcal{H}}, \quad \forall \xi \in H_0^{1/2}(\mathcal{E}_\mathcal{H}). \quad (22)$$

The well-posedness of (22) follows from the Theorem 1.1 below.

**Theorem 1.1.** *The bilinear form  $\langle G\cdot, \cdot \rangle_{\mathcal{E}_\mathcal{H}} : H_0^{1/2}(\mathcal{E}_\mathcal{H}) \times H_0^{1/2}(\mathcal{E}_\mathcal{H}) \rightarrow \mathbb{R}$  is symmetric, bounded and coercive.*

It follows from (7),(16) and (21) that the exact solution is given by:

$$u = u^0 + TG\rho + (I - TG)\tilde{T}f. \quad (23)$$

**Remark 1.1.** *Note that the present method generalizes the MHM method Paredes and Valentin [2013b]. Of course, we get from the third equation of (24) that  $\lambda = \sigma \cdot \mathbf{n}$ , for  $\sigma \in H(\operatorname{div}; \Omega)$ . It means that the flow  $\lambda$  has null jump. Then, this equation is satisfied and, by taking the null jump functionals is  $\Lambda$  as the test function, we reduce the second equation and the sistem becomes: find  $(u, \lambda) \in H^1(\mathcal{T}_\mathcal{H}) \times \Lambda$  such that*

$$\begin{aligned} (\mathcal{A} \nabla u, \nabla v)_{\mathcal{T}_\mathcal{H}} - \langle \lambda, v \rangle_{\mathcal{E}_\mathcal{H}} &= (f, v)_{\mathcal{T}_\mathcal{H}}, & \forall v \in H^1(\mathcal{T}_\mathcal{H}); \\ -\langle \mu, u \rangle_{\mathcal{E}_\mathcal{H}} &= 0, & \forall \mu \in \Lambda; \end{aligned} \quad (24)$$

## 2 GALERKIN SCHEME

We approximate the global problem (22) by considering finite dimensional subspaces  $\Gamma_{\mathcal{H}_\Gamma} \subset H_0^{1/2}(\mathcal{E}_\mathcal{H})$  and  $\tilde{\Lambda}_{\mathcal{H}_\Lambda} \subset \tilde{\Lambda}$ , where  $\mathcal{H}_\Gamma, \mathcal{H}_\Lambda$  and also  $h$  are the refinement level of the finite dimensional spaces. Two compatibility conditions must be satisfied: define  $\tilde{V}_h \subset \tilde{H}^1(\mathcal{T}_\mathcal{H})$  such that

$$\tilde{\mu}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_\Lambda} \quad \text{and} \quad \langle \tilde{\mu}_{\mathcal{H}_\Lambda}, \tilde{v}_h \rangle_{\partial K} = 0, \quad \forall \tilde{v}_h \in \tilde{V}_h, \quad \forall K \in \mathcal{T}_\mathcal{H} \implies \tilde{\mu}_{\mathcal{H}_\Lambda} = 0. \quad (25)$$



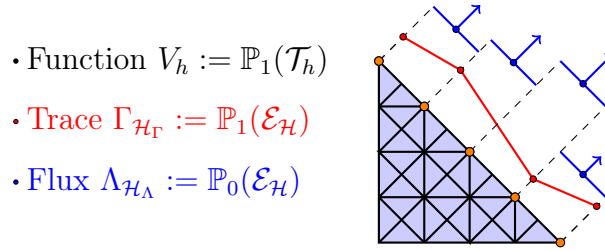
Let the discrete operator  $T_h : \tilde{\Lambda} \rightarrow \tilde{V}_h$  be such that, for a given  $\tilde{\mu} \in \tilde{\Lambda}$ ,

$$\int_K \mathcal{A} \nabla(T_h \tilde{\mu}) \cdot \nabla \tilde{v}_h \, dx = \langle \tilde{\mu}, \tilde{v}_h \rangle_{\partial K}, \quad \forall \tilde{v}_h \in \tilde{V}_h. \quad (26)$$

Consider  $\tilde{\Gamma}_{\mathcal{H}_\Gamma} := \Gamma_{\mathcal{H}_\Gamma} \cap \tilde{H}^{1/2}(\mathcal{E}_\mathcal{H})$ , where  $\tilde{H}_0^{1/2}(\mathcal{E}_\mathcal{H})$  stands for space of "zero mean" functionals on the elements border, and define  $\tilde{\Lambda}_{\mathcal{H}_0} \subset \tilde{\Lambda}$  such that

$$\tilde{\xi}_{\mathcal{H}_\Gamma} \in \tilde{\Gamma}_{\mathcal{H}_\Gamma} \quad \text{and} \quad \langle \tilde{\mu}_{\mathcal{H}_\Lambda}, \tilde{\xi}_{\mathcal{H}_\Gamma} \rangle_{\partial K} = 0, \quad \forall \tilde{\mu}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_0}, \quad \forall K \in \mathcal{T}_\mathcal{H} \implies \tilde{\xi}_{\mathcal{H}_\Gamma} = 0. \quad (27)$$

Figure 1 shows an example of compatible finite dimensional subspaces. We assume hence-



**Fig. 1:** Compatible finite dimensional subspaces.

forward that  $\tilde{\Lambda}_{\mathcal{H}_\Lambda}$  is a finite dimensional space such that  $\tilde{\Lambda}_{\mathcal{H}_0} \subset \tilde{\Lambda}_{\mathcal{H}_\Lambda} \subset \tilde{\Lambda}$ . Finally, we define the operator  $G_h : H_0^{1/2}(\mathcal{E}_\mathcal{H}) \rightarrow \tilde{\Lambda}_{\mathcal{H}_\Lambda}$ , the discrete counterpart to  $G$  defined in (19) as follows: for  $\phi \in H_0^{1/2}(\mathcal{E}_\mathcal{H})$ , let  $\tilde{\lambda}_\phi = G_h \phi$  such that, for  $K \in \mathcal{T}_\mathcal{H}$ ,

$$\int_K \mathcal{A} \nabla(T_h \tilde{\lambda}_\phi) \cdot \nabla(T_h \tilde{\mu}_{\mathcal{H}_\Lambda}) \, dx = \langle \tilde{\mu}_{\mathcal{H}_\Lambda}, T_h \tilde{\lambda}_\phi \rangle_{\partial K} = \langle \tilde{\mu}_{\mathcal{H}_\Lambda}, \phi \rangle_{\partial K}, \quad \forall \tilde{\mu}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_\Lambda}. \quad (28)$$

. The Galerkin scheme of (22) consists to finding  $\rho_{\mathcal{H}_\Gamma} \in \Gamma_{\mathcal{H}_\Gamma}$  such that

$$\langle G_h \rho_{\mathcal{H}_\Gamma}, \xi_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}} = -\langle \lambda^0, \xi_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}} + \langle G_h \tilde{T}_h f, \xi_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}}, \quad \forall \xi_{\mathcal{H}_\Gamma} \in \Gamma_{\mathcal{H}_\Gamma}. \quad (29)$$

The compatibility of the meshes of the flows and traces on the interfaces follows from the stability result below.

**Proposition 2.1.** *For  $K \in \mathcal{T}_\mathcal{H}$ , let  $\tilde{\Lambda}_{\mathcal{H}_0} \subset \tilde{\Lambda}$  introduced in (27). Then, there exists  $\gamma_K > 0$  independent of  $\mathcal{H}$  such that*

$$\sup_{\tilde{\mu} \in \tilde{\Lambda}} \frac{\langle \tilde{\mu}, \xi_{\mathcal{H}_\Gamma} \rangle_{\partial K}}{|\tilde{\mu}|_{-\frac{1}{2}, \partial K}} \leq \gamma_K \sup_{\tilde{\mu}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_0}} \frac{\langle \tilde{\mu}_{\mathcal{H}_\Lambda}, \xi_{\mathcal{H}_\Gamma} \rangle_{\partial K}}{|\tilde{\mu}_{\mathcal{H}_\Lambda}|_{-\frac{1}{2}, \partial K}}, \quad (30)$$

for all  $\xi_{\mathcal{H}_\Gamma} \in \Gamma_{\mathcal{H}_\Gamma}$ .

The following theorem ensures the well-posedness to (29).

**Theorem 2.1.** *Assume that (27) holds. Then,*

$$\langle G_h \tilde{\xi}_{\mathcal{H}_\Gamma}, \tilde{\xi}_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}} \geq \gamma^{-2} |\xi_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}}^2 \quad \text{and} \quad \langle G_h \xi_{\mathcal{H}_\Gamma}, \rho_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}} \leq C |\xi_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} |\rho_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} \quad (31)$$

for all  $\xi_{\mathcal{H}_\Gamma}, \rho_{\mathcal{H}_\Gamma} \in \Gamma_{\mathcal{H}_\Gamma}$ , where  $\gamma = \max\{\gamma_K; K \in \mathcal{T}_\mathcal{H}\}$ , and  $\gamma_K$  is the same as in (2.1).

Then approximated solutions are

$$\begin{aligned} u_h &= u^0 + T_h G_h \rho_{\mathcal{H}_\Gamma} + (I - T_h G_h) \tilde{T}_h f; \\ \lambda_{\mathcal{H}_\Lambda} &= \lambda^0 + G_h (\rho_{\mathcal{H}_\Gamma} - \tilde{T}_h f). \end{aligned} \quad (32)$$

**Remark 2.1.** *The exact solution (23) and its numerical approximation (32) are related. Of course, we define the discrete operator  $T_h$  (26) by taking as the test function in (7) a subspace  $V_h \subset H^1(\mathcal{T}_\mathcal{H})$ . The definition of the discrete operator  $G_h$  (28) follows from (19) replacing the space  $\Lambda$  by a subspace  $\tilde{\Lambda}_{\mathcal{H}_\Lambda} \subset \Lambda$ . In particular the solutions (23) and (32) match if the ranges  $T(\Lambda)$  and  $T_h(\Lambda)$  are the same, as well as the ranges  $G(H_0^{1/2}(\mathcal{E}_\mathcal{H}))$  and  $G_h(H_0^{1/2}(\mathcal{E}_\mathcal{H}))$ .*

### 3 MAIN RESULT

We derive estimates for the approximations errors  $\rho - \rho_{\mathcal{H}_\Gamma}$ ,  $\lambda - \lambda_{\mathcal{H}_\Lambda}$  and  $u - u_h$ . The next result is strongly based on the First Strang Lemma [Ern and Guermond, 1997].

**Theorem 3.1.** *Let  $(u, \lambda, \rho) \in H^1(\mathcal{T}_\mathcal{H}) \times \Lambda \times H_0^{1/2}(\mathcal{E}_\mathcal{H})$  be solution of the infinite dimensional variational formulation (24) and  $(u_h, \lambda_{\mathcal{H}_\Lambda}, \rho_{\mathcal{H}_\Gamma}) \in V_h \times \Lambda_{\mathcal{H}_\Lambda} \times \Gamma_{\mathcal{H}_\Gamma}$  its approximated solution from Galerkin scheme (32). Then, problem (29) is well-posed and*

$$|\rho - \rho_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} \leq \inf_{\phi_{\mathcal{H}_\Gamma} \in \Gamma_{\mathcal{H}_\Gamma}} \left\{ 2|\rho - \phi_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} + E(\phi_{\mathcal{H}_\Gamma}) \right\} + |(\tilde{T} - \tilde{T}_h)f|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} + E(\tilde{T}_h f); \quad (33)$$

$$|\lambda - \lambda_{\mathcal{H}_\Lambda}|_\Lambda \leq |\rho - \rho_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} + E(\rho_{\mathcal{H}_\Gamma}) + |(\tilde{T} - \tilde{T}_h)f|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} + E(\tilde{T}_h f); \quad (34)$$

$$|u - u_h|_{1, \mathcal{A}, \mathcal{T}_\mathcal{H}} \leq |\lambda - \lambda_{\mathcal{H}_\Lambda}|_\Lambda + \|T - T_h\| |\tilde{\lambda}_{\mathcal{H}_\Lambda}|_\Lambda + |(\tilde{T} - \tilde{T}_h)f|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}}. \quad (35)$$

where, for  $\phi \in H_0^{1/2}(\mathcal{E}_\mathcal{H})$ ,

$$E(\phi) := \inf_{\tilde{\mu}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_\Lambda}} \left\{ (C_T^{-1} + 1) |G\phi - \tilde{\mu}_{\mathcal{H}_\Lambda}|_\Lambda + C_T^{-1} S(\tilde{\mu}_{\mathcal{H}_\Lambda}) \right\}; \quad (36)$$

$$S(\tilde{\mu}_{\mathcal{H}_\Lambda}) := \sup_{\tilde{\eta}_{\mathcal{H}_\Lambda} \in \tilde{\Lambda}_{\mathcal{H}_\Lambda}} \frac{\langle \tilde{\mu}_{\mathcal{H}_\Lambda}, T\tilde{\eta}_{\mathcal{H}_\Lambda} \rangle_{\mathcal{E}_\mathcal{H}} - \langle \tilde{\mu}_{\mathcal{H}_\Lambda}, T_h\tilde{\eta}_{\mathcal{H}_\Lambda} \rangle_{\mathcal{E}_\mathcal{H}}}{|\tilde{\eta}_{\mathcal{H}_\Lambda}|_\Lambda}. \quad (37)$$

Finally, the following weak continuity

$$\langle \mu_{\mathcal{H}_\Lambda}, u_h - \rho_{\mathcal{H}_\Gamma} \rangle_{\mathcal{E}_\mathcal{H}} = 0, \quad \forall \mu_{\mathcal{H}_\Lambda} \in \Lambda_{\mathcal{H}_\Lambda}, \quad (38)$$

holds.

These estimates can be simplified by assuming that local problems have exact solution, by taking  $V_h = H^1(\mathcal{T}_\mathcal{H})$ , which implies  $T_h = T$  and  $\tilde{T}_h = T$ .

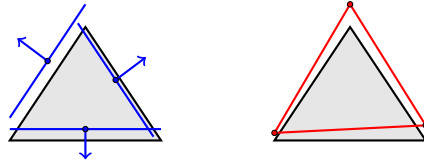


#### 4 THE MH<sup>2</sup>M METHOD WITH THE SIMPLEST ELEMENT

The Multiscale Hybrid-Hybrid-Mixed method is a technique to numerically solve the finite dimensional global problem (29) by using polynomial finite dimensional subspaces satisfying the compatibility conditions (25) and (27). Assuming  $V_h = H^1(\mathcal{T}_\mathcal{H})$ , we introduce below a pair of compatible finite element spaces. They are illustrated in Figure 2.

$$\begin{aligned} \Gamma_{\mathcal{H}_\Gamma}^1 &:= \left\{ \xi_{\mathcal{H}_\Gamma} \in H_0^{1/2}(\mathcal{E}_\mathcal{H}); \xi_{\mathcal{H}_\Gamma}|_F \in \mathbb{P}_1(F), \forall F \in \mathcal{E}_\mathcal{H} \right\}; \\ \Lambda_{\mathcal{H}_\Lambda}^0 &:= \prod_{K \in \mathcal{T}_\mathcal{H}} \left\{ \mu_{\mathcal{H}_\Lambda} \in L^2(\partial K); \mu_{\mathcal{H}_\Lambda}|_F \in \mathbb{P}_0(F), \forall F \in \partial K \right\}. \end{aligned} \quad (39)$$

We obtain in the Theorem below the optimal convergence rates.



**Fig. 2:** Representative functions of  $\Lambda_{\mathcal{H}_\Lambda}^0$  and  $\Gamma_{\mathcal{H}_\Gamma}^1$ .

**Theorem 4.1.** *Under assumptions of Theorem 3.1, considering  $u \in H^2(\mathcal{T}_\mathcal{H})$ , let  $V_h = H^1(\mathcal{T}_\mathcal{H})$ ,  $T_h = T$  and the finite element spaces  $\Gamma_{\mathcal{H}_\Gamma}^1$  and  $\Lambda_{\mathcal{H}_\Lambda}^0$  introduced in (39). Then, there exists constants  $C > 0$  independent of  $\mathcal{H}$  such that*

$$\begin{aligned} |\rho - \rho_{\mathcal{H}_\Gamma}|_{\frac{1}{2}, \mathcal{E}_\mathcal{H}} &\leq C\mathcal{H}\|f\|_{0, \Omega}, & |\lambda - \lambda_{\mathcal{H}_\Lambda}|_\Lambda &\leq C\mathcal{H}\|f\|_{0, \Omega}; \\ |u - u_h|_{1, \mathcal{A}, \mathcal{T}_\mathcal{H}} &\leq C\mathcal{H}\|f\|_{0, \Omega}. \end{aligned} \quad (40)$$

#### 5 ON THE RELATION BETWEEN MH<sup>2</sup>M AND MSFEM METHODS

We show that, under some conditions, the MH<sup>2</sup>M method and Multiscale Finite Element method (MsFEM) of Babuska and Osborn [1983] have the same nodal-value solutions. Let  $\mathcal{E} : H_0^{1/2}(\mathcal{E}_\mathcal{H}) \rightarrow H_0^1(\Omega)$  be a piecewise  $\mathcal{A}$ -harmonic function. Consider the infinite dimensional version of MsFEM of finding  $\rho \in H_0^{1/2}(\mathcal{E}_\mathcal{H})$  such that

$$\int_{\Omega} \mathcal{A}\nabla\mathcal{E}(\rho) \cdot \nabla\mathcal{E}(\xi) dx = \int_{\Omega} f\mathcal{E}(\xi) dx, \quad \xi \in H_0^{1/2}(\mathcal{E}_\mathcal{H}). \quad (41)$$

We gather from a discretization of (41):

$$(\mathcal{E}(\rho), \mathcal{E}(\xi))_{\mathcal{T}_\mathcal{H}} = \sum_{K \in \mathcal{T}_\mathcal{H}} \int_K \mathcal{A}\nabla TG\rho \cdot \nabla TG\xi dx = \langle G\rho, \xi \rangle_{\mathcal{E}_\mathcal{H}}, \quad (42)$$

that is, the left-hand side of the global problem (22). Then, we get by joining the right-hand side of (41) and (22) that

$$-\langle \lambda^0, \xi \rangle_{\mathcal{E}_\mathcal{H}} + \langle G\tilde{T}f, \xi \rangle_{\mathcal{E}_\mathcal{H}} = \int_{\Omega} f\mathcal{E}(\xi) dx. \quad (43)$$

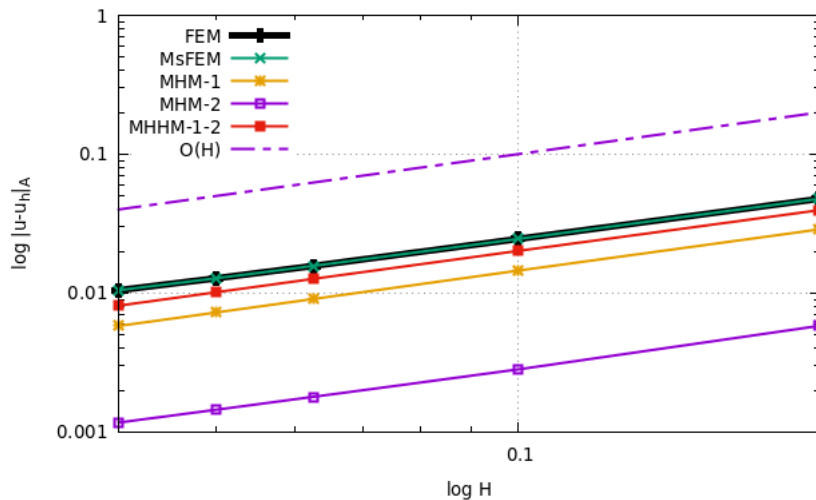
Assuming that there are exact solutions for second level problems, the solutions  $\rho$  from (22) and (41) coincide pointwise if  $\Lambda$  is an infinite dimensional space. For  $G = G_h$ , solutions match if range of  $\mathcal{E}$  is contained in range of  $TG_h$ . Moreover, for  $\mathcal{A} = \mathcal{I}$ , the solutions match, for  $V_h = \mathbb{P}_1(\mathcal{T}_h)$ .

## 6 NUMERICAL RESULTS

The following numerical experiments aim to access the analytical predictions and investigate the performance of the MHHM method compared to MHM, MsFEM and FEM, see Larson and Bengzon [2013], Paredes and Valentin [2013b], Paredes et al. [2016], Paredes and Valentin [2013a]. Let  $\Omega := ]0, 1[ \times ]0, 1[$  and the boundary value problem (1) with  $\mathcal{A} = I$  and  $f(x, y) := -2x(x - 1) - 2y(y - 1)$ , for  $(x, y) \in \Omega$ . This problem admits an analytical solution given by

$$u(x, y) = x(x - 1)y(y - 1), \quad (x, y) \in \Omega. \quad (44)$$

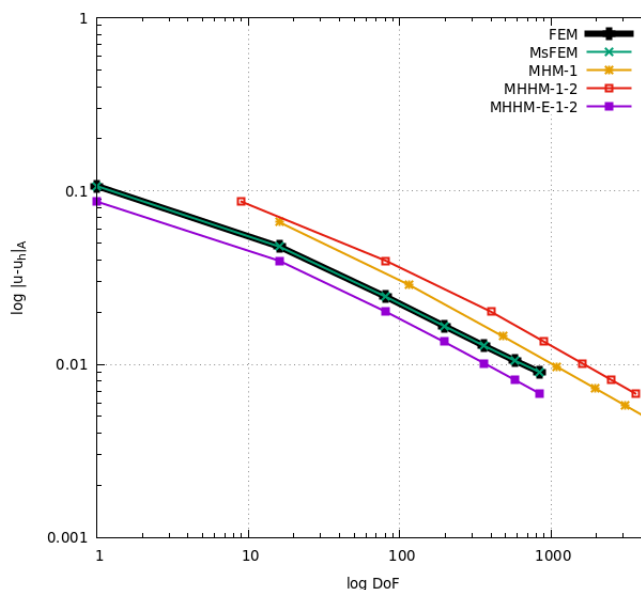
We denote MHHM-1-2 when the flux has one degree of freedom on each interface and trace has two degrees of freedom. We plot in Figure 3 the convergence curves in the energy norm  $|u - u_h|_{\mathcal{A},1,\mathcal{T}_h}$ . A computational cost graph is depicted in Figure 4. Here we denote MHHM-1-2 if the method solves a saddle-point problem (17), and we use MHHM-E-1-2 when method solves an elliptical problem (22). We observe that the accuracy of the MHHM method matches that of the MHM method with fewer global degrees of freedom.



**Fig. 3:** Mesh based energy error graph for the analytical solution problem.

## 7 CONCLUSION

The Multiscale Hybrid-Hybrid-Mixed method is derived from a hybrid-mixed three-field formulation and characterized by a symmetric elliptical problem. We relax the flux defining it on each element boundary, so that different flux meshes can be taken for each element. Although there is no conformity to the numerical solution, the trace and flux are conform, that is, our method preserves mass conservation for the flux. Continuous and



**Fig. 4:** Degrees of freedom by error graph for the analytical solution problem.

discrete inf-sup conditions hold. We introduce a pair of compatible finite element spaces and error estimates, as well as, numerical experiments to access analytical predictions.

## REFERENCES

- I. Babuska and J. E. Osborn. Generalized finite element methods: Their performance and their relation to mixed methods. *SIAM Journal on Numerical Analysis*, 20(3):510–536, 1983.
- Mikhail Belishev and Vladimir Sharafutdinov. Dirichlet to Neumann operator on differential forms. *Bulletin des Sciences Mathématiques*, 86(132):128–145, 2008.
- F. Brezzi and D. Marini. A three-field domain decomposition method. *American Mathematical Society*, (157):27–34, 1992.
- Franklin C.Barros, Alexandre L. Madureira, and Frédéric G.C. Valentin. The Multiscale Hybrid-Hybrid-Mixed method. Mestrado em modelagem computacional, Laboratório Nacional de Computação Científica, Petrópolis, RJ, 2022.
- Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*. Scuola Norm. Sup. Pisa, 1997.
- Gabriel N. Gatica. *A Simple Introduction to the Mixed Finite Element Method - Theory and Applications*. SpringerBriefs in Mathematics, Concepción, Chile, 2014. ISBN 9783319036946.
- M. G. Larson and F. Bengzon. *The Finite Element Method: Theory, Implementation, and Applications*. Springer-Verlag, Heidelberg New Yourk Dordrecht London, 2013.

- C. Harder; D. Paredes and F. Valentin. A family of multiscale hybrid-hybrid-mixed finite element methods for darcy equation with rough coefficients. *Journal of Computational Physics*, (245):107–130, 2013a.
- D. Paredes, F. Valentin, and H. Versieux. On the robustness of multiscale hybrid-mixed methods. *Mathematics of Computation, American Mathematical Society*, 86(304):525–548, 2016.
- R. Araya; C. Harder; D. Paredes and F. Valentin. Multiscale hybrid-mixed methods. *SIAM Journal on Numerical Analysis*, 51(6):3505–3531, 2013b.



# Uso do Critério de Informação de Akaike para Estimativa das Ordens de Trajetórias de Ratos Reais e Virtuais no Labirinto em Cruz Elevado

Leonardo Rocha de Freitas Dias<sup>1</sup>, Renato Tinós<sup>2</sup> e Ariadne de Andrade Costa<sup>1</sup>

<sup>1</sup> Grupo de Redes Complexas Aplicadas de Jataí (GRAJ), Universidade Federal de Jataí (UFJ), Jataí/GO, Brasil

<sup>2</sup> Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP), Universidade de São Paulo (USP), Ribeirão Preto/SP, Brasil

---

## Abstract

Processos estocásticos são dados por uma sequência de estados que representa a evolução temporal de um sistema. A ordem do processo mostra quantos estados – atual ( $k = 1$ ) e anteriores ( $k = 2, 3, \dots$ ) – são usados para tomada de decisão do próximo estado. Nesse artigo, propõe-se o uso do Critério de Informação de Akaike para estimativa das ordens das trajetórias de ratos reais e virtuais (obtidos por inteligência artificial) no labirinto em cruz elevado. Para isso, os ratos foram divididos em três categorias: controle e com efeito de drogas ansiolíticas e ansiogênicas. Foram comparadas as ordens dos processos estocásticos compreendidos pelas trajetórias de ratos reais e virtuais. Curiosamente, embora os resultados comportamentais gerais para ambos os tipos de ratos sejam semelhantes, as ordens dos processos diferem. Esse resultado mostra a complexidade da simulação do comportamento exploratório de ratos, trazendo uma perspectiva diferente da existente na literatura.

**Keywords:** Labirinto em cruz elevado, Rato, Redes neurais artificiais, Processo estocástico, Critério de informação de Akaike

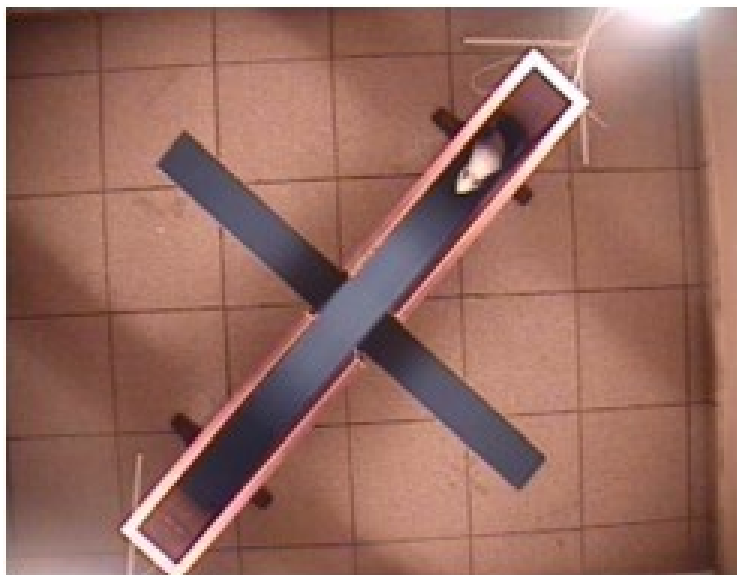
---

## 1 INTRODUÇÃO

Em meados dos anos 50, Montgomery criou um labirinto elevado em Y que visava explorar o comportamento de ratos quando expostos a áreas abertas e fechadas no labirinto [Montgomery, 1955]. Então Pellow [Pellow et al., 1985] criou o labirinto em cruz elevado (LCE),



que é usado nos dias de hoje principalmente nas áreas de Psicobiologia e Farmacologia. O aparato recebe esse nome porque contém dois braços abertos e dois braços fechados, em formato de cruz, e é mantido elevado do solo (Fig. 1).



**Fig. 1:** Labirinto em cruz elevado com um rato Wistar durante um experimento. Fonte: Acervo próprio.

O experimento geralmente é realizado com ratos ou camundongos, colocando-os no aparato, por exemplo, sob efeito de drogas para testar se há aumento/diminuição de ansiedade. Drogas que causam esses efeitos são chamadas respectivamente de ansiogênicas e ansiolíticas. Os resultados são comparados com os de roedores controle, que não estão sob o efeito de drogas durante o experimento – nesses casos, o animal recebe a aplicação de uma solução salina para que sofra também a injeção de uma solução com uma seringa antes do teste.

Com o avanço da tecnologia, testes no LCE em ambiente virtual vêm sendo utilizados, nos quais ratos virtuais são usados para reprodução e investigação do comportamento de ratos reais [Salum et al., 2000, Miranda et al., 2009, Shimo et al., 2010, de Andrade Costa and Tinos, 2016]. O uso desses modelos vem sendo proposto principalmente porque os ratos reais utilizados em testes são usualmente eutanasiados após os experimentos, o que acontece porque o comportamento do animal é alterado quando ele é reapresentado ao aparato experimental.

Ratos controle tendem a passar mais tempo nos braços fechados do que abertos, se deslocar menos e chegar menos vezes às extremidades. Isso ocorre porque os riscos associados aos braços abertos são maiores do que aos braços fechados. Com isso, quando o animal está sob efeito de drogas ansiogênicas (que aumentam a ansiedade) esse comportamento tende a ser intensificado. Porém o oposto acontece quando há inserção de alguma droga ansiolítica (que diminui a ansiedade) no animal [Montgomery, 1955].

O LCE é dividido em posições para se analisar as trajetórias dos animais. Tais trajetórias são então descritas como sequências discretas de posições, que podem ser analisadas como processos estocásticos [Haccou et al., 1983, Garcia-Perez et al., 2005, Yang



and Chao, 2005, Tejada et al., 2010]. No caso do LCE especificamente, alguns trabalhos partiram da suposição de que a ordem do processo estocástico das trajetórias de ratos é 1, isto é, que são processos Markovianos, e foram calculados os vetores estacionários que dão a probabilidade de se encontrar o animal em cada posição do labirinto [Tejada et al., 2010, de Andrade Costa and Tinos, 2016]. Contudo, nunca foram de fato estimadas as ordens dos processos estocásticos decorrentes das trajetórias de ratos no LCE e nos diferentes aparatos usualmente utilizados em Farmacologia e Psicobiologia.

Neste trabalho faz-se a análise de trajetórias de ratos reais sem efeitos de drogas e sob efeitos de drogas ansiogênicas e ansiolíticas. Também foram analisadas trajetórias de ratos virtuais, obtidos por simulações computacionais baseadas em redes neurais artificiais (RNAs) otimizadas por algoritmo genético (AG) [Costa et al., 2012, 2013, 2014]. Esses ratos virtuais simulam o comportamento de ratos reais nessas três condições. A utilidade dessas análises pode ser a proposição de um método quantitativo simples para comparar o efeito de drogas e possivelmente também de doenças neurológicas no uso de memória para tomada de decisão do deslocamento de roedores no LCE.

## 2 MÉTODOS

### 2.1 Experimentos com ratos reais

O testes com ratos reais consistem em posicionar o rato no centro e deixá-lo percorrer o LCE por um período de 5 min (300 s). Esse intervalo é definido, pois após esse período o animal perde o interesse em explorar o ambiente e seu comportamento muda significativamente. A trajetória do animal é filmada para posterior análise.

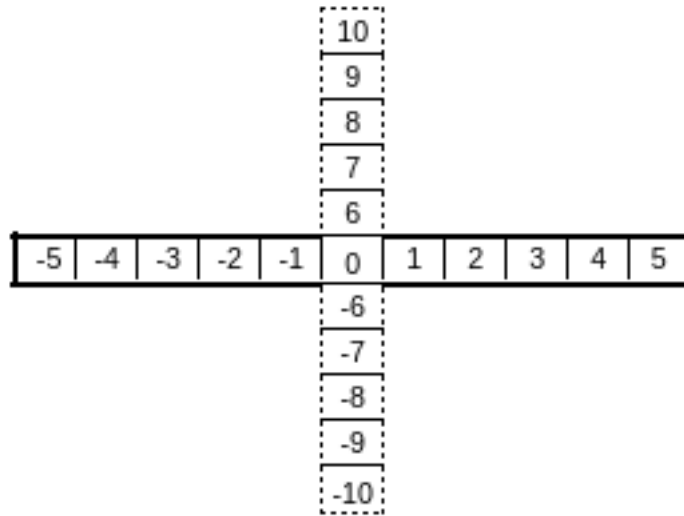
Os dados de experimentos com ratos reais utilizados neste trabalho foram cedidos pelo Prof. Silvio Morato de Carvalho, da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP) da Universidade de São Paulo (USP). Os dados são provenientes de testes com ratos controle (sem efeitos de drogas), ratos sob a influência de drogas ansiogênicas (semicarbazida e pentilenotetrazol) em diferentes dosagens e ratos sob influência de drogas ansiolíticas (clordiazepóxido e midazolam), conforme mostra a Tabela 1. O número de ratos por subcategoria varia, pois esses dados são provenientes de experimentos realizados como parte de pesquisas com finalidades diferentes.

**Tabela 1:** QUANTIDADE DE RATOS REAIS PARA CADA TIPO DE DROGA ESTUDADO.

	Ansiogênico							Controle	Ansiolítico		
	PTZ (mg/kg)			SMZ (mg/kg)					-	CDZ (mg/kg)	MDZ (mg/kg)
	10	20	30	10	20	40	80			5	1
Ratos	12	12	9	11	11	11	11	30	12	13	

### 2.2 Labirinto em cruz elevado virtual

Nos testes com ratos virtuais, um LCE virtual como o representado pela Fig. 2 é simulado. Os ratos virtuais percorrem o LCE virtual por 300 passos de tempo, assumindo a equivalência de 1 passo de tempo de simulação como 1 s nos experimentos com ratos reais.



**Fig. 2:** Representação do labirinto em cruz elevado virtual. As linhas tracejadas representam o braço aberto e as linhas em negrito, os braços fechados.

### 2.3 Modelo computacional – ratos virtuais

O modelo computacional usado para as simulações foi escrito em linguagem computacional C/C++ e se baseia em RNAs otimizadas por AG. O AG é composto por 1000 indivíduos otimizados por 15.000 gerações. Cada indivíduo é um cromossomo que contém os pesos (números reais) associados a uma rede de Elman contendo 6 sensores na camada de entrada, 4 neurônios recorrentes na camada intermediária e 4 neurônios na camada de saída. Todos os sensores assumem valores binários, sendo 3 deles para detecção de parede nos lados direito, esquerdo e à frente do rato virtual a uma distância próxima (adjacente à posição atual). Os outros 3 sensores são longa distância, detectando a presença de paredes a três posições da atual, nas mesmas três direções – direita, esquerda e frente. Os 4 neurônios da camada intermediária voltam como recorrências para a camada de entrada. Por fim, os neurônios de saída representam as quatro ações possíveis para cada rato virtual a cada passo de tempo, quais sejam: ficar parado, seguir em frente, rotacionar 90° para a direita ou para a esquerda sem sair da posição atual.

Em cada geração, os ratos virtuais percorrem o aparato virtual por 300 passos de tempo. Uma função de *fitness* é calculada para cada rato virtual ao percorrer o labirinto. Esse *fitness* que cada rato recebe é inspirado em características evolutivas associadas ao comportamento de ratos no LCE. A função contém um termo positivo (recompensa) e outro negativo (punição), conforme a Eq. 1:

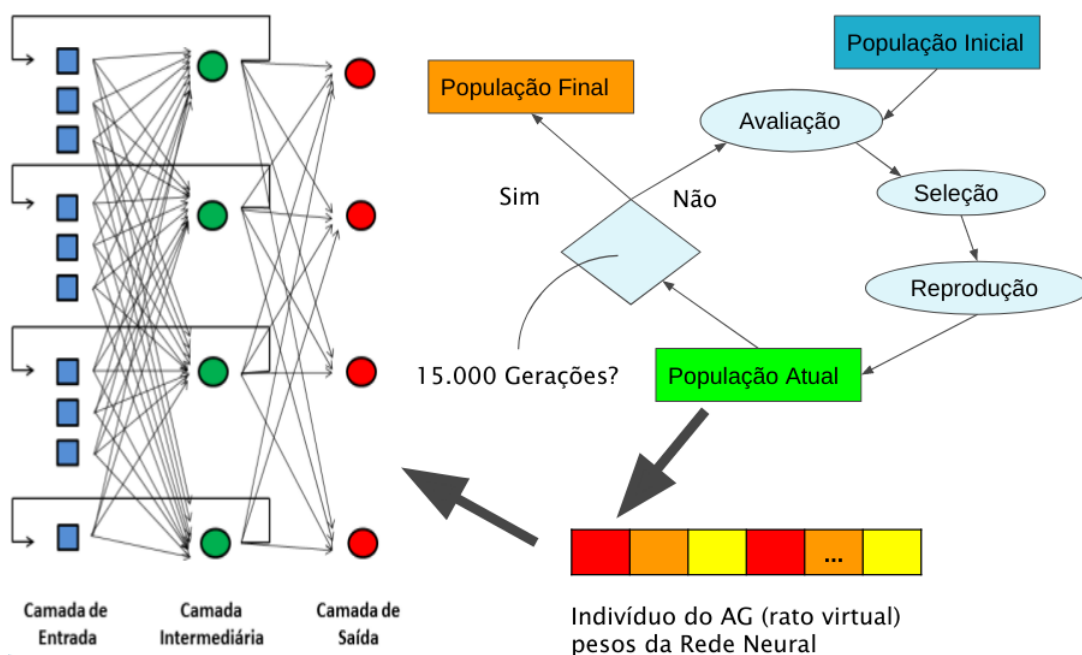
$$f(\mathbf{x}) = \sum_{t=1}^{300} (r(\mathbf{x}, t) - \beta \cdot s(\mathbf{x}, t)). \quad (1)$$

Em  $r(\mathbf{x}, t)$ , o rato virtual recebe um ponto por passo de tempo se estiver em uma posição não visitada nos últimos 3 passos anteriores. Por  $s(\mathbf{x}, t)$ , o agente é punido em um ponto por passo de tempo probabilisticamente, conforme sua posição atual. A chance de punição é 0,011 nos braços fechados, 0,012 no centro e 0,015 nos braços abertos. Isso reflete a preferência natural do rato real pelos braços fechados.



Os dois indivíduos com maior *fitness* são selecionados diretamente para a próxima geração (por elitismo). Como a quantidade de indivíduos em cada geração é mantida fixa, os demais indivíduos a comporem a próxima geração são selecionados por torneio. No torneio, pares aleatórios de indivíduos são sorteados e, com probabilidade de 75%, aqueles com maior *fitness* são selecionados; caso contrário, seguem aqueles com menor *fitness*. Os indivíduos selecionados passam então por recombinação de 1 ponto, com 60% de chance, e cada alelo pode passar por mutação com 5% de probabilidade. A mutação é gaussiana, com média 0 e desvio 0,05.

Esse processo completa os indivíduos da próxima geração e após as 15.000 gerações a execução é finalizada, conforme o mostra o esquema da Fig 3. Os dados do rato virtual com maior *fitness* são registrados. As trajetórias analisadas correspondem a 30 ratos virtuais, cada um com o melhor *fitness* de uma execução completa do AG.



**Fig. 3:** Modelo computacional. A população inicial do AG é gerada aleatoriamente. A sequência, vista na parte de cima à direita, mostra as etapas seguintes do AG: avaliação, seleção e reprodução, que originam a próxima geração. Esse processo é repetido até que o critério de 15.000 gerações seja atingido. A população do AG é composta por cromossomos de números reais, que representam os pesos das conexões entre os elementos da rede neural (representada à esquerda).

#### 2.4 Processos estocásticos

Um processo estocástico pode ser compreendido como uma sequência cronológica de observações de um determinado fenômeno. As cadeias de Markov são processos estocásticos com tempo discreto, para os quais a probabilidade da próxima ação do agente depende somente de seu estado atual. Esses processos são considerados de ordem um, por a próxima ação depender de um único estado (o atual) [Markov, 1971]. Contudo, processos estocásticos podem ser de ordem zero, isto é, não possuir nenhuma dependência entre o estado atual ou anteriores e o estado seguinte (processo completamente aleatório). Mas

os processos estocásticos podem ainda ser de ordens maiores que um, quando os estados seguintes dependem de um ou mais estados passados além do estado atual.

No presente trabalho, os processos estocásticos estudados são as sequências de posições discretas (ou seja, as trajetórias) de ratos virtuais em um LCE virtual e de ratos reais em um LCE. É importante destacar que cada posição é incluída uma única vez na sequência mesmo que o tempo de permanência do animal ou do agente virtual seja longa naquele local; assim, uma nova posição é inserida na sequência sempre que houver a transição entre posições distintas.

### 2.5 Análises

Para realizar as estimativas das ordens dos processos estocásticos foi utilizado um pacote implementado em Python [Singer et al., 2013, 2014, 2015]. O pacote é válido para processos dados por sequências de números inteiros, como é o caso das sequências obtidas pelos deslocamentos dos ratos virtuais no LCE virtual esquematizado na Fig. 2. O LCE real foi segmentado de maneira semelhante ao LCE virtual durante as análises.

Para as análises estatísticas, foi utilizado o Critério de Informação de Akaike (AIC) [Akaike, 1974]. Este método é baseado em teoria da informação e relaciona verossimilhança e a quantidade de informação que se perde quando aproximamos os dados com um modelo candidato. Aqui testou-se as ordens de 0 a 3 para os processos estocásticos das trajetórias dos ratos, tanto reais como virtuais, e a ordem que obteve menor perda foi a estimada para cada trajetória testada.

## 3 RESULTADOS E DISCUSSÕES

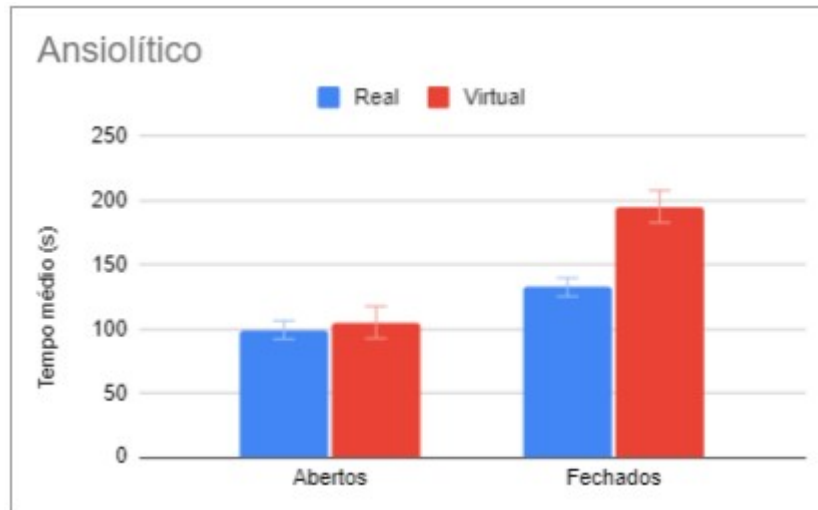
Inicia-se os resultados mostrando-se comparações do tempo médio que os ratos reais e virtuais passam nos braços abertos e fechados do LCE. Este é o ponto de partida, pois o tempo médio é uma das principais medidas usadas para análises do comportamento exploratório de ratos. O que distingue as simulações dos ratos virtuais é apenas o parâmetro  $\beta$ , que balanceia a punição e a recompensa no cálculo da função de *fitness* do agente virtual. Para  $\beta = 2,2$ , tem-se a simulação de um rato sob efeito de droga ansiolítica Clordiazepóxido na concentração de 1 mg/kg (Figura 4). Já a Figura 5, compara ratos virtuais, com  $\beta = 3,2$ , a ratos reais controle. Por fim, na Figura 6 exibem-se os tempos médios de permanência de ratos reais sob efeito de droga ansiogênica Semicarbazida na concentração de 20 mg/kg e ratos virtuais ( $\beta = 6,3$ ). Em todos os casos têm-se poucas amostras e as distribuições são assimétricas. Assim, o teste estatístico usado para validar os resultados obtidos para os ratos virtuais, em relação aos ratos reais, foi o Mann-Whitney, com significância  $p = 0,05$ . Em todos os casos, as distribuições dos dois grupos (ratos reais e virtuais) são as mesmas.

Como esperado, ratos controle passam mais tempo nos braços fechados do que nos braços abertos e no centro do LCE. O mesmo ocorre e é acentuado para ratos mais ansiosos (sob drogas ansiogênicas). Este é o oposto do observado para ratos sob efeito de drogas ansiolíticas.

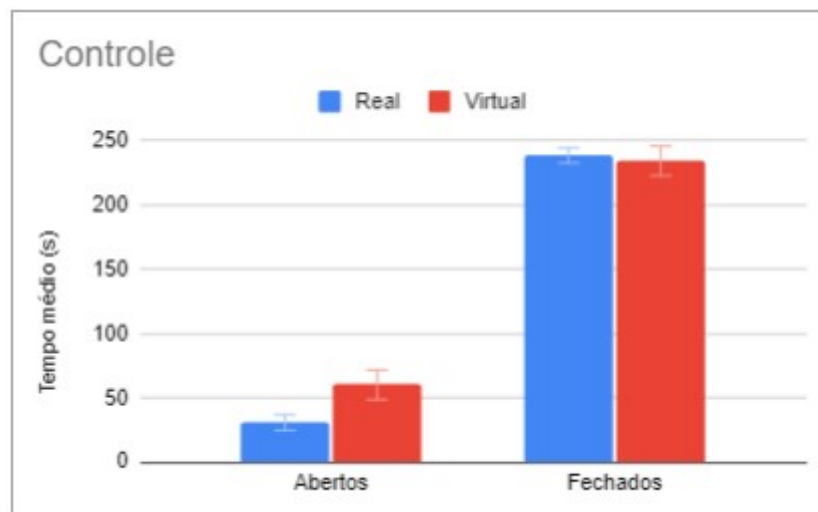
Passando da etapa inicial de validação dos dados de ratos virtuais, segue-se com a apresentação das ordens estimadas para os ratos reais controle e com diferentes dosagens das drogas ansiolíticas e ansiogênicas estudadas (Tabela 2). Pela distinção entre as médias



e considerando os erros padrões, sugere-se que haja uma diferença estatística entre as categorias. Uma questão importante é que se pode estabelecer uma sequência crescente entre as médias das ordens estimadas para as categorias estudadas: ansiogênica, controle e ansiogênica. Tal resultado indica que os ratos menos ansiosos estariam usando mais informações do ambiente e/ou memória como critério para a tomada de decisão em seu percurso no LCE. Quanto maior a ansiedade, menos informação é usada. Uma justificativa preliminar para este fato é que animais com mais medo/ansiedade agem de forma mais impulsiva [Jakuszkowiak-Wojten et al., 2015].



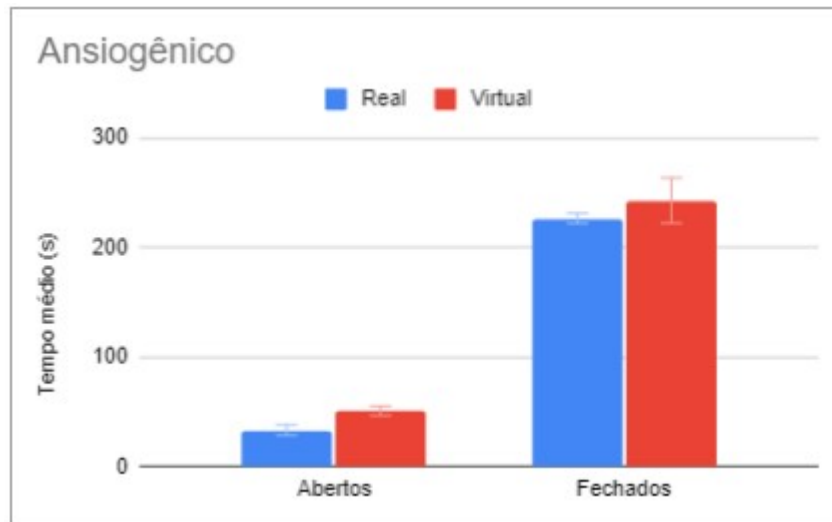
**Fig. 4:** Tempo médio em cada tipo de braço para os ratos reais e virtuais sob efeito de droga ansiolítica clordiazepóxido na concentração de 1 mg/kg. Significância estatística do Mann-Whitney teste  $p = 0,05$ .



**Fig. 5:** Tempo médio em cada tipo de braço para os ratos reais e virtuais controles. Mann-Whitney teste com  $p = 0,05$ .

A ordem do processo estocástico estimada para cada rato virtual é apresentada na Tabela 3. Os dados dos ratos virtuais utilizados nessa análise são os mesmos das Fig.

ras 4-6. Como o número de amostras é reduzido e os resultados obtidos foram binários, não encontramos um teste estatístico apropriado para comparação das categorias, mas considerando as médias e erros padrões sugere-se que as categorias sejam estatisticamente distintas umas das outras.



**Fig. 6:** Tempo médio em cada tipo de braço para os ratos reais e virtuais sob efeito de droga ansiogênica semicarbazida 20 mg/kg. Teste Mann-Whitney com  $p = 0,05$ .

Nota-se que as médias das ordens estimadas para os processos estocásticos dos ratos virtuais apresentaram uma sequência diferente da dos ratos reais. Neste caso, a média dos ratos virtuais simulando efeitos de drogas ansiogênicas é menor do que a de drogas ansiolíticas. Além disso, todos os ratos virtuais controle tiveram a ordem da trajetória estimada em um, resultando em uma média 1 (ainda maior que a da droga ansiogênica).

Essas análises sugerem que a média de tempo gasta em cada tipo de braço pode não ser suficiente para caracterizar o comportamento de ratos no LCE. As diferenças obtidas para as ordens das trajetórias de ratos reais e virtuais podem refletir o uso de memória operacional e informações ambientais durante o percurso. Assim, os resultados apontam para uma diferença na tomada de decisão das diferentes categorias de ratos (controle e sob efeito de drogas) tanto para o grupo de ratos reais quanto dos virtuais.

Ressalta-se que processos estocásticos são caracterizados por ordens representadas unicamente por números inteiros. Sendo assim, as médias e erros padrões (números reais) foram calculados apenas como uma forma de se mensurar a diferença entre as categorias estudadas.

O estudo da ordem dos processos estocásticos pode ajudar a verificar variações no uso de memória de trabalho de roedores no LCE. Os resultados sugerem que os ratos reais sob efeito de drogas ansiolíticas são menos impulsivos e, estando mais tranquilos, podem fazer maior uso da memória ambiental.

Uma grande dificuldade ao lidar com animais é a baixa amostragem e o fato que após os experimentos os animais são, em muitos casos, sacrificados. Cada animal pode participar de apenas um experimento no aparato, pois eles perdem o interesse exploratório caso haja uma reexposição. Desta forma, vale destacar que a falta de amostras grandes



para a análise não é exclusiva de nossas análises, mas do uso de animais em experimentos de modo geral.

**Tabela 2:** ORDENS DOS PROCESSOS ESTOCÁSTICOS ESTIMADAS PARA OS RATOS SOB EFEITO DE DIFERENTES DOSAGENS DE ANSIOGÊNICOS (PENTILENOTETRAZOL-PTZ E SEMICARBAZIDA-SCZ), ANSIOLÍTICOS (CLORDIAZEPÓXIDO-CDZ E MIDAZOLAM-MDZ) E RATOS CONTROLE (SEM USO DE DROGAS), BEM COMO AS RESPECTIVAS MÉDIAS E ERROS PADRÕES.

Rato	Ansiogenico							Controle	Ansiolitico	
	PTZ (mg/kg)			SMZ (mg/kg)				-	CDZ (mg/kg)	MDZ (mg/kg)
	10	20	30	10	20	40	80	-	5	1
1	1	0	1	1	0	0	1	1	1	1
2	1	0	1	0	0	1	0	1	1	1
3	0	1	0	0	1	1	1	1	1	1
4	0	1	0	1	1	0	0	1	1	1
5	0	0	1	0	0	0	0	1	1	1
6	1	0	0	1	0	0	0	1	1	1
7	0	0	0	0	0	1	0	1	1	1
8	0	0	0	0	0	0	1	0	1	0
9	0	0	1	1	1	0	0	1	1	1
10	0	1		1	0	0	1	1	1	0
11	0	1		1	1	1	1	0	1	1
12	1	0						1	1	1
13								1		1
14								1		
15								0		
16								0		
17								1		
18								0		
19								0		
20								0		
21								0		
22								1		
23								1		
24								0		
25								0		
26								0		
27								0		
28								1		
29								0		
30								0		
Média	0,33	0,33	0,44	0,55	0,36	0,36	0,45	0,53	1	0,85
E. Padrão	0,14	0,14	0,18	0,16	0,15	0,15	0,16	0,09	0	0,11



**Tabela 3:** ORDENS DOS PROCESSOS ESTOCÁSTICOS ESTIMADAS PARA OS 30 RATOS VIRTUAIS ANSIOGÊNICOS, CONTROLE E ANSIOLÍTICOS, COM AS RESPECTIVAS MÉDIAS E ERROS PADRÕES.

Rato	Ansiogênico ( $\beta = 2, 2$ )	Controle ( $\beta = 3, 2$ )	Ansiolítico ( $\beta = 6, 3$ )
1	1	1	0
2	1	1	1
3	1	1	0
4	1	1	0
5	1	1	1
6	1	1	0
7	0	1	0
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	0
12	1	1	1
13	1	1	1
14	1	1	0
15	1	1	1
16	1	1	0
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0
21	1	1	0
22	1	1	0
23	0	1	1
24	1	1	1
25	1	1	0
26	1	1	1
27	1	1	1
28	1	1	1
29	1	1	1
30	1	1	1
Média	0,93	1	0,53
EP	0,05	0	0,09

#### 4 CONCLUSÃO

Nesta pesquisa buscou-se obter a ordem do processo estocástico da trajetória de animais controle e sob efeito de drogas ansiolíticas e ansiogênicas reais e simulados. Os resultados mostram que é possível distinguir o nível de ansiedade do animal utilizando o Critério de Informação de Akaike.



A comparação dos dados de ratos reais com ratos virtuais é feita no intuito de compreender o funcionamento dos ratos virtuais e como aprimorar não apenas esse modelo, mas outros sistemas artificiais que atuem em situações de tomada de decisão. Assim, houve diferenças nas ordens dos processos estocásticos estimados para os ratos virtuais em relação aos ratos reais, embora os resultados comumente observados (tempo de permanência em cada tipo de braço) sejam estatisticamente equivalentes. Desta forma, mostra-se a importância de futuros estudos para a compreensão do funcionamento interno das RNAs dos ratos virtuais e da sua tomada de decisão. Tais análises poderão trazer contribuições no entendimento do papel dos neurônios e da tomada de decisão de ratos virtuais, bem como inferências sobre relações entre o funcionamento do sistema nervoso e tomadas de decisão de ratos reais.

## 5 Agradecimentos

LRF Dias agradece a bolsa de pesquisa concedida por Rafael Grisotto. R Tinós agradece o CNPq (#306689/2021-9), a FAPESP (#2021/09720-2), o Center for Artificial Intelligence (C4AI-USP) com suporte da FAPESP (#2019/07665-4) e a IBM Corporation. AA Costa agradece o suporte do CNPq (Universal #405508/2021-2). Todos os autores são gratos ao Prof. Silvio Morato pela concessão dos dados de ratos reais.

## Referências

- Hirotsugu Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotsugu Akaike*, pages 215–222. Springer, 1974.
- Ariadne Costa, Patrícia Vargas, and Renato Tinós. Using explicit averaging fitness for studying the behaviour of rats in a maze. In *ECAL 2013: The Twelfth European Conference on Artificial Life*, pages 940–946. MIT Press, 2013.
- Ariadne A Costa, Antonio C Roque, Silvio Morato, and Renato Tinós. A model based on genetic algorithm for investigation of the behavior of rats in the elevated plus-maze. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 151–158. Springer, 2012.
- Ariadne A Costa, Silvio Morato, Antonio C Roque, and Renato Tinós. A computational model for exploratory activity of rats with different anxiety levels in elevated plus-maze. *Journal of neuroscience methods*, 236:44–50, 2014.
- Ariadne de Andrade Costa and Renato Tinos. Investigation of rat exploratory behavior via evolving artificial neural networks. *Journal of neuroscience methods*, 270:102–110, 2016.
- Elizabeth Garcia-Perez, Alberto Mazzoni, Davide Zoccolan, Hugh PC Robinson, and Vincent Torre. Statistics of decision making in the leech. *Journal of Neuroscience*, 25(10):2597–2608, 2005.
- Patsy Haccou, Herman Dieneske, and Evert Meelis. Analysis of time-inhomogeneity in markov chains applied to mother-infant interactions of rhesus monkeys. *Animal Behaviour*, 31(3):927–945, 1983.

- Katarzyna Jakuszkowiak-Wojten, Jerzy Landowski, Mariusz S Wiglusz, and Wiesław Jerzy Cubała. Impulsivity in anxiety disorders. a critical review. *Psychiatria Danubina*, 27(suppl 1):452–455, 2015.
- Andrey Andreyevich Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic probabilistic systems*, 1:552–577, 1971.
- DA Miranda, CA Conde, C Celis, and SP Corzo. Modelado del comportamiento de ratas en laberinto en cruz elevado basado en redes neuronales artificiales. *Revista Colombiana de Física*, 41(2):406, 2009.
- KC Montgomery. The relation between fear induced by novel stimulation and exploratory drive. *Journal of comparative and physiological psychology*, 48(4):254, 1955.
- Sharon Pellow, Philippe Chopin, Sandra E File, and Mike Briley. Validation of open: closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *Journal of neuroscience methods*, 14(3):149–167, 1985.
- Cristiane Salum, Silvio Morato, and Antônio Carlos Roque-da Silva. Anxiety-like behavior in rats: a computational model. *Neural Networks*, 13(1):21–29, 2000.
- Helder K Shimo, Antônio C Roque, Renato Tinós, Julián Tejada, and Sílvio Morato. Use of evolutionary robots as an auxiliary tool for developing behavioral models of rats in an elevated plus-maze. In *2010 Eleventh Brazilian Symposium on Neural Networks*, pages 217–222. IEEE, 2010.
- Philipp Singer, Thomas Niebler, Markus Strohmaier, and Andreas Hotho. Computing semantic relatedness from human navigational paths: A case study on wikipedia. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 9(4):41–70, 2013.
- Philipp Singer, Denis Helic, Behnam Taraghi, and Markus Strohmaier. Detecting memory and structure in human navigation patterns using markov chain models of varying order. *PloS one*, 9(7):e102070, 2014.
- Philipp Singer, Denis Helic, Andreas Hotho, and Markus Strohmaier. Hyptrails: A bayesian approach for comparing hypotheses about human trails on the web. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1003–1013, 2015.
- Julián Tejada, Geraldine G Bosco, Silvio Morato, and Antonio C Roque. Characterization of the rat exploratory behavior in the elevated plus-maze with markov chains. *Journal of neuroscience methods*, 193(2):288–295, 2010.
- Hsin-Chou Yang and Anne Chao. Modeling animals’ behavioral response by markov chain models for capture–recapture experiments. *Biometrics*, 61(4):1010–1017, 2005.



# Modelagem da Dinâmica de uma Válvula Solenoide para Composição de Máquina Virtual (MV) em Funções de Blowout Preventer (BOP)

Lucas Constantino Mendonça<sup>1</sup>, Yasmin Nunes Alves<sup>2</sup>, Luiz Antônio de Oliveira Chaves<sup>3</sup>, Iara Tammela<sup>4</sup>, Rodolfo Cardoso<sup>5</sup> e Danilo Colombo<sup>6</sup>

<sup>1</sup> *Universidade Federal Fluminense, EEIMVR, Volta Redonda/RJ, Brasil*

<sup>2</sup> *Universidade Federal Fluminense, ICT, Rio das Ostras/RJ, Brasil*

<sup>3</sup> *Universidade Federal Fluminense, ICT, Rio das Ostras/RJ, Brasil*

<sup>4</sup> *Universidade Federal Fluminense, ICT, Rio das Ostras/RJ, Brasil*

<sup>5</sup> *Universidade Federal Fluminense, ICT, Rio das Ostras/RJ, Brasil*

<sup>6</sup> *CENPES - Petrobras, Av. Horácio Macedo 950, Rio de Janeiro/RJ, Brasil*

---

## Abstract

Due to the occurrence of accidents owing systemic failures during maritime drilling operations of oil wells, the study of well safety equipment Blowout Preventer (BOP) becomes crucial for accident prevention. In many cases, computational modeling is used as an instrument for investigating dynamic behavior and proposition of standardization profiles to support decision-takers in the diagnosis of the “health” of a given equipment. The present work depicts a Virtual Machine (VM) model developed with the purpose of simulating the dynamics of solenoid valve functions in the electrical-hydraulic circuit of BOP for activation of the Annular function and future application in the study of failure scenarios. Qualitative and quantitative analyses of the results provided by the simulator were performed in order to obtain a validation of the model. To this end, data from pressure and volume curves provided by drilling contractors and metrics such as MSE, MAE, MAPE and RMSE were used as validation instruments. The results obtained were satisfactory, showing that the physical model developed has a high rate of confidence when compared to actual data.

**Keywords:** Computational modeling, Blowout Preventer (BOP), Solenoid valves, Virtual Machine (MV), Model simulation validation.

---

## 1 INTRODUÇÃO

As operações de perfuração marítima de poços de petróleo envolvem riscos de acidentes com formação de *blowouts* que ocorrem por meio de falhas sistêmicas. Nesses eventos, a erupção de fluidos oriundos de reservatórios de alta pressão, seguido do desdobramento de total descontrole, em razão de falhas do sistema, provocam a perda de ativos, capital e danos ambientais de múltiplas formas [Drægebø, 2014].

O principal equipamento para evitar *blowouts* é o *Blowout Preventer* (BOP), que contempla diversos dispositivos formados por linhas hidráulicas, válvulas gavetas (*Blind Shear Ram*), elementos vedadores de espaço anular (*Annular*) e dispositivos elétricos que estão interconectados para enviar e receber sinais distribuídos em subsistemas. Possuem também, funções específicas de nível de importância dinâmica para abertura, fechamento, além de intertravamento e bloqueio [Tammela et al., 2020a] [Huse and Alme, 2013].

Assim, os relatórios de registros de confiabilidade demonstram que válvulas reguladoras de pressão, solenoides e válvulas *Sub-Plate Mounted* (SPM) são componentes críticos em razão dos mecanismos de falhas e dos impactos na condição do equipamento para executar funções de segurança que são vitais [Huse and Alme, 2013] [Drægebø, 2014] [BSEE, 2016].

Nesse contexto, estudos de análise de confiabilidade e modos de falhas, com a identificação de mecanismos de degradação, são cada vez mais realizados para buscar alternativas de tornar o BOP um sistema com maior operacionalidade, para reduzir a ocorrência de eventos acidentais [Martins et al., 2018] [Drægebø, 2014] [Huse and Alme, 2013].

A evolução dessa abordagem atual está na análise de estados da condição do componente com o emprego da metodologia *Condition-based Maintenance* (CBM). A metodologia representa a combinação de monitoramento de condição e/ou inspeção e/ou testes, além das análises e ações de manutenção que fornece um padrão da capacidade dinâmica de resposta para execução de funções [Tammela et al., 2020b] [Marinho et al., 2022] [Martins et al., 2018].

Uma das formas de testar os níveis de degradação de um componente físico, seja no domínio hidráulico, mecânico, térmico ou elétrico, são por meio da modelagem computacional, seguida da simulação. Os modelos de domínios multifísicos podem ser obtidos de forma virtual e buscam representar de forma fiel os subsistemas conforme os fenômenos envolvidos para simulação dinâmica e extrair informações na previsão, diagnóstico e associação de efeitos [Khaled et al., 2020].

Dessa forma, a pesquisa busca construir um modelo de Máquina Virtual (MV) com a finalidade de simular a dinâmica das funções da válvula solenoide no circuito elétrico-hidráulico do BOP para ativação da função do *Annular* e futura aplicação no estudo de cenários de falhas. A modelagem computacional é empregada como instrumento para investigação do comportamento dinâmico e proposição de perfis de padronização para subsidiar o tomador de decisão no diagnóstico da “saúde” do componente.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 *Compensated Chamber Solenoid Valves (CCSV)*

O mecanismo de acionamento do BOP é dinamicamente desenvolvido por meio de um sistema hidráulico com modos de ativação elétrico por válvulas solenoides.



Uma válvula solenoide pode ser definida em alguns casos como uma *Compensated Chamber Solenoid Valve* (CCSV) que consiste em uma câmara em aço inoxidável de compensação de pressão preenchida com fluido dielétrico com atuadores solenoides protegidos para executar funções específicas. O componente tem a função eletro-hidráulica e atua como interface entre o sistema de controle e o piloto hidráulico, sendo responsável pelo acionamento das SPMs [NOV, 2015].

A CCSV basicamente possui dois componentes principais; o solenoide e o *fluid end* para acionamento. Desta forma, a válvula tem a finalidade de atuar após a energização da haste móvel por atuação hidráulica do fluido na válvula piloto para permitir a passagem de um fluido piloto [NOV, 2015].

## 2.2 *Sub-Plate Mounted (SPM)*

A SPM atua como uma válvula direcional utilizada para fornecer a pressão hidráulica para várias funções do *Stack*, consistindo em duas ou mais subplacas conectadas para formar uma pilha de válvulas com entrada e descarga comum. Dependendo da função, a SPM pode ser do tipo normalmente fechada (NC), normalmente aberta (NO), normalmente fechada bidirecional (NCB), de dupla ação bidirecional (DAB) ou de dupla ação normalmente fechada bidirecional (DANCB). Quando a SPM é “piloted” hidráulicamente para abrir ou fechar no ponto de distribuição POD ativo, ela direciona o fluido de controle regulado para um BOP *stack* ou função do *Lower Marine Riser Package* (LMRP) específica ou ventila para o mar. Por outro lado, quando um comando de anulação é recebido pelos PODs *Multiplex BOP Control System* (MUX) submarinos, a CCSV é desenergizada. A CCSV então fecha e libera a pressão piloto da válvula SPM. Na sequência, a válvula SPM fecha ou abre, direcionando o fluido de controle regulado para uma BOP *stack* específico, função LMRP ou ventila para o mar [BSEE, 2016].

## 2.3 *Integração entre CCSV e SPM*

A CCSV pode ser configurada como uma função direta em conjunto com uma válvula SPM. Deste modo, a válvula é operada por uma solenoide cujo sinal de comando é controlado a partir dos painéis de controle auxiliares ou do *Driller* (sondador). As válvulas SPM montadas no POD e montadas na *stack* são então controladas pela CCSV por meio da pressão piloto para controlar o fluxo de pressão no circuito de forma a atuar no componente do equipamento BOP [NOV, 2012].

As CCSVs não configuradas em conjunto com uma válvula SPM operam outras válvulas SPM montadas na *stack* ou dispositivos finais. Essas funções são chamadas de funções “diretas” porque transmitem sinais do piloto diretamente através do POD. Nesses casos, a CCSV é conectada a uma válvula SPM configurada para permitir o fluxo contínuo o tempo todo. Com efeito, a CCSV está operando outra função SPM montada na *stack* da mesma forma que a CCSV independente [NOV, 2012].

## 2.4 *Execução de Funções com a CCSV/SPM*

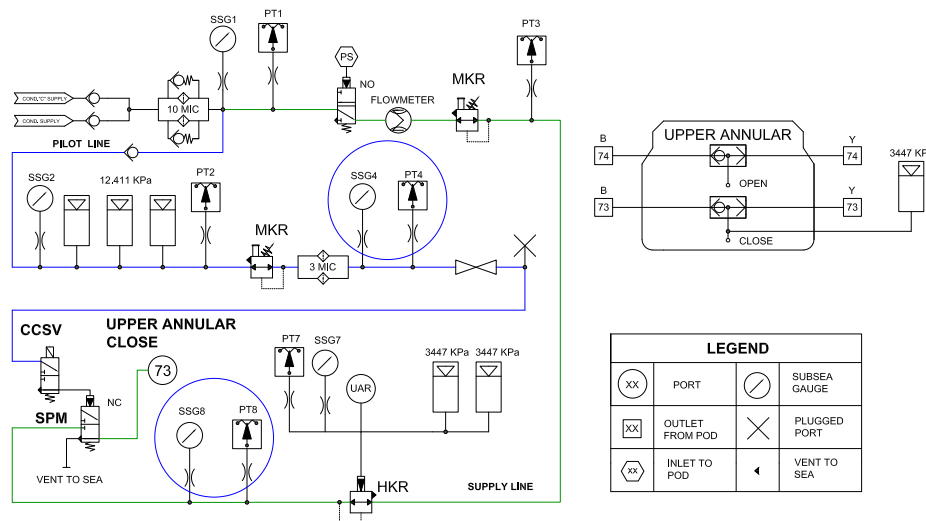
Conforme exposto por BSEE [2016], a SPM pode ser destinada para execução de uma determinada função do BOP. As válvulas SPM de controle direcionam o fluido de suprimento para a função desejada do componente do BOP e ventilam o fluido do lado oposto

ao daquele componente. As válvulas são operadas por piloto a partir da superfície com auxílio de mangueiras piloto no umbilical. Em geral, duas válvulas SPMs são requeridas para operar cada função. Por outro lado, os componentes NO e NC operados por piloto e retorno por mola podem requerer apenas uma SPM. Dentre as diversas funções do BOP, podem ser citadas as funções: *Upper Annular Open*, *Upper Annular Close*, *Lower Annular Open*, *Lower Annular Close*, *Upper Blind Shear Ram (UBSR) – HP Close*, *Close* e *Open*. Um diagrama hidráulico esquemático para execução da função *Upper Annular Close* pelas válvulas CCSV e SPM é apresentado na Fig. (1). De maneira habitual, essas funções são monitoradas através de curvas de pressão que são obtidas dos dispositivos medidores de pressão PT-4 e PT-8 localizados em pontos estratégicos do sistema hidráulico do BOP conforme apresentado na Fig. (1) que fornece destaque para os medidores de pressão delimitados por círculos azuis. Por vezes, esses dispositivos de medição podem realizar mais de uma função. Além de desempenharem a função de sensor de pressão, podem atuar também como transdutores de pressão, transformando o sinal de pressão obtido em um sinal elétrico [Seborg et al., 2016]. Esses instrumentos de medição possuem também uma determinada resolução “capacidade de leitura” que pode afetar a interpretação dos dados em alguns casos. Os sensores/transdutores PT-4 e PT-8 possuem uma resolução de 2 segundos e não são capazes de fornecer uma boa leitura para funções que são executadas em um curto intervalo de tempo ( $\leq 1$  segundo) como no caso da solenoide, em que sua energização/desenergização promove efeitos na leitura da pressão que podem não ser captados com a resolução adequada. O medidor de pressão PT-4 por estar localizado na linha piloto, se torna ideal para avaliação dos efeitos surgentes durante a ativação da CCSV/SPM. Já o medidor de pressão PT-8 por estar localizado na linha de suprimento se torna mais adequado para avaliação dos efeitos surgentes na execução da função *Upper Annular Open/Close*, função esta responsável pela abertura/fechamento do elemento de vedação (*Annular*) de poço do BOP. Sendo assim, em conjunto, os dados obtidos dos medidores de pressão fornecem informações significativas para o monitoramento do sistema hidráulico do BOP durante a execução de suas funções.

### 2.5 *Máquina Virtual (MV) e Simulação*

No ramo da engenharia e afins, indústrias dependem do desenvolvimento de modelos matemáticos que sejam capazes de representar um determinado fenômeno físico ou químico por meio de equações suficientemente precisas para permitir a simulação de processos em mecanismos específicos, especialmente em componentes que envolvem riscos de falhas em unidades químicas, marítimas, automotivos e aeroespacial.

Nesse contexto, as Máquinas Virtuais (MVs); *softwares* computacionais capazes de simular o comportamento de um determinado sistema, são cada vez mais desenvolvidas e empregadas para construção de cenários de falhas e investigação dos padrões de falhas através da simulação. Um conjunto de *softwares* comerciais e livres podem ser utilizados para prototipagem virtual como o Amesim, Matlab®/Simulink, SimulationX, Scilab/XCOS, OpenModelica, entre outros. A principal vantagem desses *softwares* é a existência de uma ampla biblioteca de modelos de componentes e funções para construção de sistemas e subsistemas, seja elétrico, hidráulico, mecânicos, físicos ou químicos [Chaves et al., 2021] [Smith and Nair, 2005].



**Fig. 1:** Diagrama Hidráulico Simplificado do Modelo.

### 3 METODOLOGIA

A metodologia aplicada para a pesquisa tem caráter exploratório de abordagem mista, em que através da análise teórica do sistema foi possível realizar a modelagem computacional, assim como a validação do simulador elaborado. O desenvolvimento executado no presente trabalho consistiu em cinco etapas.

Na primeira etapa foi realizada uma análise teórica objetivando o entendimento do funcionamento dos componentes CCSV e SPM, assim como dos fenômenos físicos procedentes do funcionamento dos mesmos. Para isso, foram realizados estudos em manuais técnicos e reuniões com especialistas.

Em uma segunda etapa, foram obtidos e processados os dados de monitoramento e execução das funções analisadas, sendo esses fornecidos pelas *drilling contractors*, o que possibilitou o início da etapa 3 que consistiu na análise dos dados coletados e agrupamento de curvas de pressão e volume ao decorrer do tempo com comportamento similar utilizando a técnica de clusterização, tornando possível a determinação de limites superiores, limites inferiores, médias, desvios padrão e a identificação de um padrão de curva de pressão e volume ao longo do tempo para a execução das funções CCSV/SPM.

Na etapa 4 ocorreu o desenvolvimento do modelo no ambiente Simulink/Simscape do *software* Matlab®, baseado na pesquisa teórica e análise de dados executadas nas etapas anteriores, gerando o modelo simulado, que teve sua validação realizada na etapa seguinte através da comparação qualitativa dos perfis das curvas simuladas e reais, assim como uma análise quantitativa executada por meio de métricas estatísticas para validação do modelo.

### 4 DADOS EXPERIMENTAIS

Os dados para validação foram coletados do sistema real e disponibilizados pelas *drilling contractors*, tratados, compilados e compactados numa base de dados desenvolvida na ferramenta *Power BI*.

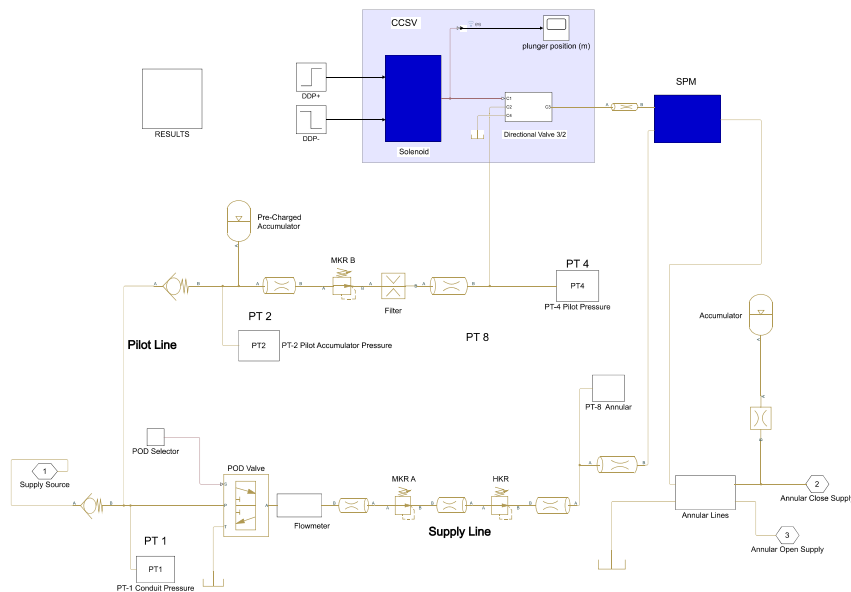


As informações de dados experimentais do modelo da CCSV/SPM, foram obtidos para comparação, assim como a percepção dos perfis de curvas de pressão obtidas pelos sensores/transdutores de pressão PT-4 e PT-8 e volume de fluido ao longo do tempo. Dessa forma, os dados específicos serviram de *input* para a análise estatística executada no Matlab®, o que possibilitou a clusterização das curvas de acionamento das funções e a definição do perfil de curva padrão.

O procedimento empregado utilizou as variáveis de pressão e volume para delimitação de limites inferior, superior e a média, dados utilizados como parâmetros de validação para máquina virtual, tanto qualitativo quanto de forma quantitativa, quando comparados com a curva gerada como resposta da máquina virtual.

## 5 DESENVOLVIMENTO DA MÁQUINA VIRTUAL

O modelo desenvolvido dos componentes CCSV e SPM para execução das funções do BOP visa contemplar os efeitos físicos: acionamento eletromagnético da CCSV, acionamento hidráulico da SPM, deslocamento do *plunger* da CCSV, deslocamento do precursor da SPM e forças restauradoras. Para tal, o pacote Simulink/Simscape da Matlab® foi utilizado. O pacote consiste em um ambiente que dispõe de blocos pré-construídos que representam componentes físicos e lógicos, o que permite uma construção de sistemas dinâmicos de forma modular. O conjunto de blocos construídos no ambiente do Simulink que representa o modelo da CCSV e SPM pode ser observado na Fig. (2).



**Fig. 2:** Modelo da CCSV/SPM em Conjunto com *Annular* no Simulink.

### 5.1 Equações do bloco Solenoide

As equações do bloco do Simulink/Simscape/Matlab® que representam o comportamento dinâmico da válvula solenoide com retorno por mola são apresentadas conforme as Eq.



(1), Eq. (2) e Eq. (3) [Mathworks, 2022]:

$$V = L \frac{di}{dt} + i \frac{dL}{dt} \quad (1)$$

Como  $L$  depende unicamente de  $x$  a Eq. (1) pode ser reescrita como:

$$V = L \frac{di}{dt} + i \frac{dL}{dx} \frac{dx}{dt} \quad (2)$$

A derivada de  $L$  com relação a  $x$  pode ser obtida dos dados de *force-stroke* do fabricante através da relação expressa pela Eq. (3).

$$F = 0.5i^2 \frac{dL}{dx} \quad (3)$$

Onde  $L$  é a indutância,  $x$  a posição do *plunger*,  $i$  a corrente,  $V$  a tensão e  $F$  é a força magnética da solenoide.

## 6 VALIDAÇÃO

A etapa de validação do modelo CCSV/SPM na execução da função *Upper Annular Close* foi composta por uma análise mista em duas etapas.

Para isso, foi desenvolvida em uma primeira etapa uma avaliação qualitativa baseada na comparação dos perfis reais e simulados das curvas obtidas pelos medidores de pressão PT-4 e PT8, identificando que a curva simulada para ambos apresentam perfis similares às curvas reais, respeitando os limites de máximo e mínimo para cada intervalo de tempo da execução da função dos componentes.

Em uma segunda etapa, as análises quantitativas objetivaram avaliar a variabilidade da curva simulada em comparação à curva real. Dentro desse contexto, foram utilizadas as métricas Erro Absoluto Médio (MAE); Erro Quadrático Médio (MSE); Raiz do Erro Quadrático Médio (RMSE) e Erro Percentual Médio Absoluto (MAPE) para verificar a acurácia das curvas. Onde a primeira métrica representa a média dos erros em valores absolutos, a segunda calcula o quadrado da média de diferença entre o valor simulado com o real, a terceira é obtida a raiz quadrada da média dos erros quadráticos, e a última dada pela média do erro percentual em valor absoluto [Hair et al., 2009] [dos Santos et al., 2011] [Moreno et al., 2013] [Hora and Campos, 2015].

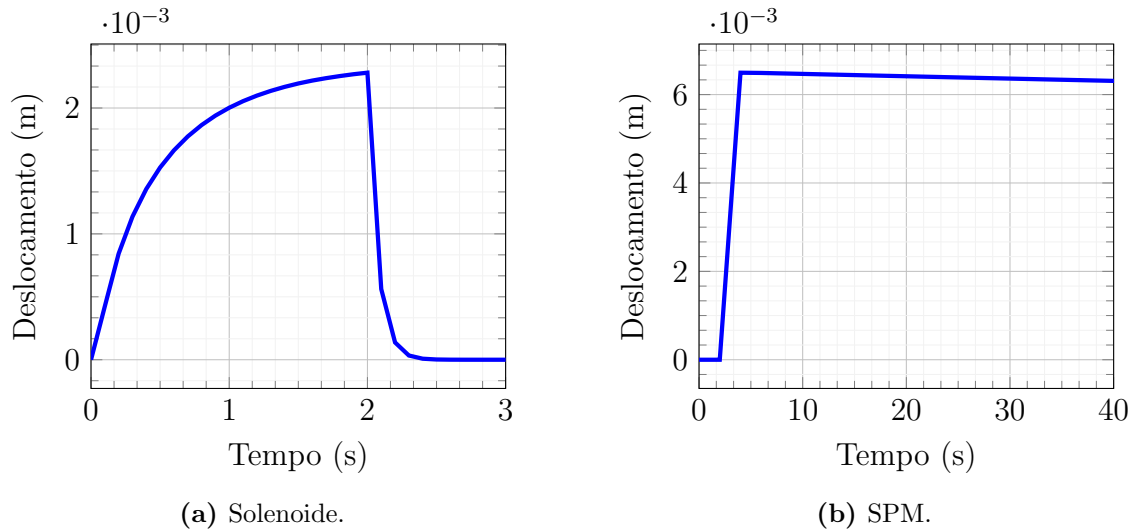
Além das métricas descritas, foi executada a verificação da correlação entre a curva simulada e os agrupamentos clusterizados previamente obtidos para cada uma das funções analisadas. Em adição, o modelo foi validado em conjunto com a função *close* do *Annular Upper* que é um dos diversos componentes críticos presentes no sistema hidráulico do BOP que foi modelado e desenvolvido em paralelo.

## 7 RESULTADOS E DISCUSSÃO

### 7.1 Resultados Qualitativos

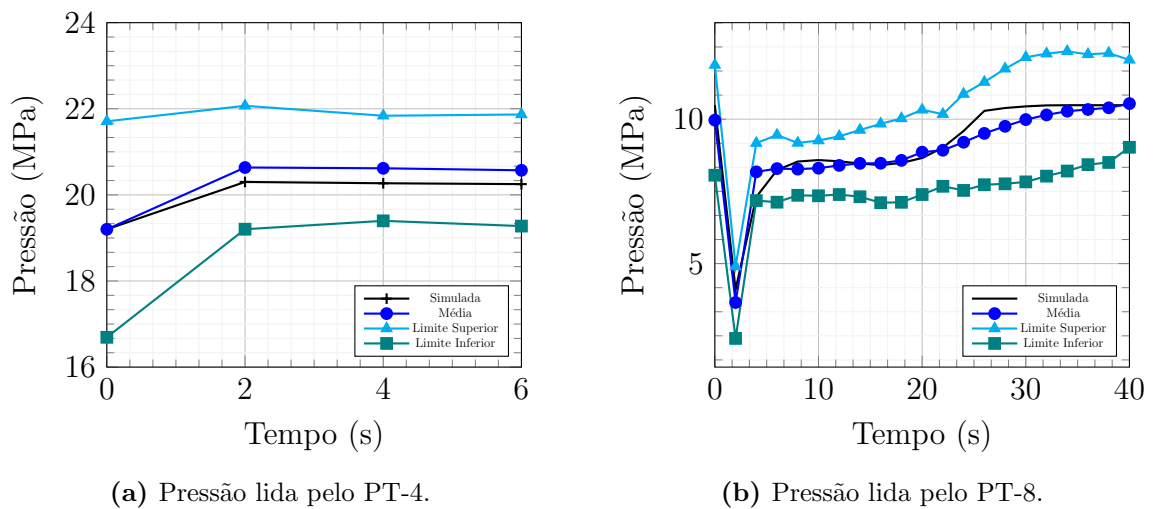
Na validação qualitativa, foi considerado um tempo de execução das funções para análise gráfica que fosse capaz de caracterizar o perfil das curvas em análise. Desse modo, foram analisados os 6 (seis) primeiros segundos para a curva de pressão fornecida pelo sensor

PT-4 e 40 (quarenta) segundos para curva de pressão fornecida pelo sensor PT-8, representando o intervalo de interesse. A partir da simulação, além das curvas de pressão foram obtidos os gráficos referentes a execução da solenoide e deslocamento do percursor da SPM, como apresentado na Fig. (3).



**Fig. 3:** Deslocamento do Percursor da Solenoide e da SPM.

Dando continuidade a análise qualitativa, com enfoque para estudo de avaliação do real perfil, a curva simulada PT-4 obtida foi comparada de forma clusterizada com as curvas: real média, real superior e real inferior conforme registrado na Fig. (4.a). Os resultados obtidos são específicos no sentido de demonstrar nessa fase de testes experimentais o desempenho do simulador, a existência de alta concordância entre as curvas que mostram o perfil de variação de pressão no intervalo de 0 a 6 segundos após o acionamento.

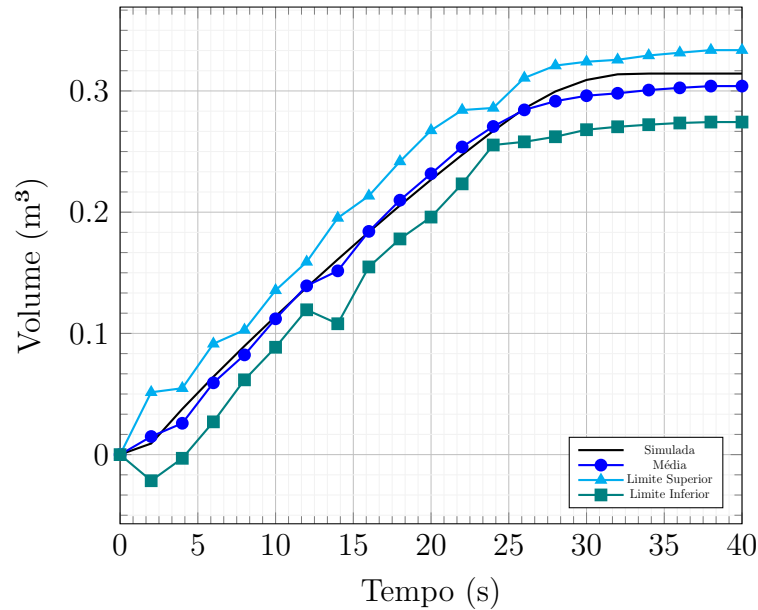


**Fig. 4:** Curvas de Pressão Reais e Simulada.



Em uma segunda etapa da análise qualitativa, foram verificados os impactos do modelo na execução da função *Close Upper* do *Annular*. Na Figura (4.b) são exibidas as curvas de pressão simulada e do sensor PT-8 com o perfil da curva simulada dentro da faixa de limite inferior e superior e sobreposta à curva do modelo físico real.

Por fim, a Figura (5) mostra a semelhança entre o perfil de curva simulada e o perfil real do volume de fluido em relação ao tempo de execução das funções e representa a alta concordância entre os resultados obtidos pelo simulador e os dados reais.



**Fig. 5:** Curvas de Volume de Fluido Reais e Simulada.

Uma vez que os resultados gráficos constataram que o simulador apresentou os resultados esperados, se prosseguiu para execução da etapa de validação quantitativa.

## 7.2 Resultados Quantitativos

O diagnóstico quantitativo de validação estatística foi desenvolvido para os dados relacionados ao sensor PT-4 e PT-8 através da análise de correlação e das métricas MAE, MSE, RMSE e MAPE. As equações das métricas descritas são apresentadas, respectivamente nas Eq. (4), Eq. (5), Eq. (6), Eq. (7).

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| \quad (4)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad (5)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N |PE_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (7)$$

Onde  $N$  é o número de variáveis e  $PE_i = e_i/x_i$ .

No que diz respeito à correlação, as curvas simuladas foram comparadas com as médias reais obtidas. Para a curva de pressão do sensor PT-4 obteve-se correlação total (100%) da curva simulada, tanto com relação à curva real do grupo padrão quanto à curva média. Já para a curva de pressão do sensor PT-8 o resultado da correlação em comparação com a curva do grupo padrão foi de 0,983 e de 0,752 com a média.

De modo subsequente foram analisadas as métricas MAE, MSE, RMSE e MAPE, para as mesmas funções. Nessa etapa foram comparadas as curvas de forma individual com a curva simulada. Tal análise consistiu em verificar individualmente a adequação do perfil dinâmico de simulação para cada uma das funções. Como critério de validação, uma curva pode ser considerada significativa quando além do MAPE estiver abaixo de 15%, houver pelo menos outras duas métricas dentro de um *range* (intervalo) que representa “alta acurácia”. As métricas com um *range* moderado são avaliadas dentro do contexto de cada função para verificar a significância da curva simulada. A Tabela (1) apresenta o intervalo de aceitação das métricas analisadas em 3 classes de acurácia de validação do modelo:

**Tabela 1:** INTERVALO DE CRITÉRIO DE ACURÁCIA PARA VALIDAÇÃO DO MODELO POR MÉTRICAS ESTATÍSTICAS PARA MEDIDAS DOS SENSORES PT-4 E PT-8

Métrica	Alta acurácia	Média acurácia	Baixa acurácia
MAE	< 100	100 < $x$ < 150	> 150
MSE	< 10.000	10.000 < $x$ < 22.500	> 22.500
RMSE	< 100	100 < $x$ < 150	> 150
MAPE	< 15%	-	-

O parecer relacionando o sensor/transdutor PT-4 à função *Upper Annular Close* observou 42 curvas, obtendo como não significativos 12 itens e 71,42% das curvas com alta acurácia para as métricas. Para as curvas obtidas pelo sensor PT-8, foram analisadas 49 curvas, com apenas 3 curvas de baixa representatividade. Os resultados indicaram alta acurácia para comparação das curvas reais com a curva simulada com adequação de 93,87% das curvas.

## 8 CONCLUSÃO

Uma vez realizada as análises de validação do modelo, foi possível verificar que os resultados qualitativos obtidos mostraram a existência de alta concordância entre as curvas de respostas de simulação de ativação e desenergização da CCSV/SPM do BOP e os dados reais dos perfis das curvas de pressão e volume no tempo fornecidas pelas *drilling contractors*.



No que tange a validação quantitativa, as curvas lidas pelo sensor PT-4 relacionadas a execução da função *Upper Annular Close* mostraram que de acordo com as métricas 71,42% das curvas apresentaram alta acurácia. Contudo, uma vez que o sensor PT-4 (localizado na linha de “pilotagem”) lê os dados durante a execução da CCSV, que por sua vez possui duração de execução muito curta (instantânea), a resolução do sensor torna a leitura real falha, impactando a acurácia dos resultados. Já para as curvas lidas pelo sensor PT-8 para a execução da mesma função, os resultados indicaram uma adequação de 93,87% das curvas reais com a curva simulada e de 100% do grupo padrão, uma vez que as curvas registradas pelo sensor PT-8 são mais longas, possuindo um função com maior tempo de execução. Dessa forma, o sensor PT-8 captura perfis de curvas que são medidas de forma mais assertiva fornecendo uma melhor acurácia dos resultados.

Sendo assim, validações qualitativas e quantitativas realizadas do comportamento dinâmico do sistema demonstram que o modelo desenvolvido atende aos requisitos necessários para representação do modelo físico real. Para trabalhos futuros e em andamento se busca determinar e compreender os modos de falhas do equipamento através da simulação de falhas na Máquina Virtual desenvolvida, o que irá permitir um diagnóstico da “saúde” do equipamento em questão.

## 9 Agradecimentos

Agradecemos à Petrobras por todo apoio fornecido ao projeto “Desenvolvimento de Metodologia para Manutenção Baseada em Condição (CBM) em Blowout Preventers (BOPs)”, tornando capaz a construção do presente trabalho.

## Referências

- BSEE. Report blowout preventer control system reliability. Technical report, Bureau of Safety and Environmental Enforcement, 2016.
- L. A. O. Chaves, L. N. de Almeida, M. C. Amaral, L. C. Mendonça, and L. A. S. B. Leite. Modelagem e simulação de sistemas dinâmicos com máquinas virtuais para prototipagem de sistemas de segurança em processos de engenharia. *IV Simpósio de Engenharia, Gestão e Inovação*, 4, 2021.
- R. S. dos Santos, L. C. Costa, G. C. Sediyaama, B. G. Leal, R. A. de Oliveira, and F. B. Justino. Avaliação da relação seca/produzividade agrícola em cenário de mudanças climáticas. *Revista Brasileira de Meteorologia*, 26(2):313–321, 2011.
- E. Drægebø. *Reliability Analysis of Blowout Preventer Systems: A comparative study of electrohydraulic vs. all-electric BOP technology*. Marine Technology, Trondheim, Norway, 2014.
- J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham. *Análise multivariada de dados*. Bookman, Porto Alegre, 2009.
- J. Hora and P. Campos. A review of performance criteria to validate simulation models. *Expert Systems*, 32(5):578–595, 2015.

- J. R. Huse and I. A. Alme. Bop reliability monitored real time. Technical report, SPE (Society Petroleum Engineers) International, 2013.
- N. Khaled, B. Pattel, and A. Siddiqui. *Digital twin development and deployment on the cloud: developing cloud-friendly dynamic models using Simulink®/Simscape™ and Amazon AWS*. Academic Press, an imprint of Elsevier, London, United Kingdom, 2020. ISBN 9780128216460.
- R. M. F. Marinho, R. Cardoso, I. Tammela, D. Colombo, and F. E. Souza. Integrando conceitos de mantabilidade e regras de interpretação na cbm: revisão sistemática da literatura integrating maintainability concepts and interpretation rules into cbm: a systematic literature review. *Brazilian Journal of Development*, 8(4):22805–22819, 2022.
- F. B. Martins, R. Cardoso, I. Tammela, D. Colombo, and B. A. de Matos. Applying CBM and PHM concepts with reliability approach for Blowout Preventer (BOP): a literature review. *Brazilian Journal of Operations and; Production Management*, 15(1):78–95, 2018.
- Mathworks. Documentation solenoid. <https://www.mathworks.com/help//simscape/ug/solenoid.html>, 2022. Accessed: 2022-11-14.
- J. J. Moreno, A. P. Pol, and A. S. Abad. Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, pages 500–506, 2013.
- NOV. Multiplex control pod and bop stack control system 1. Technical report, National Oiwell Varco, Houston, Texas, 2012.
- NOV. Acoustic control pod. Technical report, National Oiwell Varco, Houston, Texas, 2015.
- D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle III. *Process dynamics and control*. John Wiley & Sons, 2016.
- J. Smith and R. Nair. *Virtual Machines: Versatile Platforms for Systems and Processes*. The Morgan Kaufmann Series in Computer Architecture and Design. Morgan Kaufmann, 2005. ISBN 1558609105; 9781558609105.
- I. Tammela, R. Cardoso, M. C. Amaral, F. S. Machado, L. A. O. Chaves, T. A. Rodrigues, R. M. F. Marinho, I. S. Pinto, F. B. Martins, and D. Colombo. Análise dos modos, mecanismos e causas de falha para componentes e partes do blowout preventer. *Proceedings of the Rio Oil & Gas Expo and Conference, Rio de Janeiro, RJ, Brazil*, pages 1–13, 2020a.
- I. Tammela, R. Cardoso, G. V. Silva, L. A. S. Silva, E. B. M. Meza, R. B. Narcizo, G. J. M. Aguiar, and D. Colombo. Construção de dicionário de variáveis de monitoramento do sistema bop para auxiliar implantação de cbm. *Proceedings of the Rio Oil & Gas Expo and Conference, Rio de Janeiro, RJ, Brazil*, 04 2020b.



# Discrepância de Modelo: Estudo Inicial com Regressão por Processo Gaussiano

Lucas de Oliveira Sá<sup>1</sup>, Leonardo Tavares Stutz<sup>1</sup> e Daniel Alves Castello<sup>2</sup>

<sup>1</sup> *Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo/RJ, Brasil*

<sup>2</sup> *Universidade Federal do Rio de Janeiro, Instituto Alberto Luiz Coimbra, Rio de Janeiro/RJ, Brasil*

---

## Resumo

Este trabalho apresenta um estudo inicial sobre a discrepância de modelo na avaliação da resposta de um sistema dinâmico com amortecimento viscoelástico por meio de um modelo de amortecimento viscoso. É empregada uma abordagem que visa o desacoplamento dos problemas de inferência sobre o parâmetro desconhecido do modelo e sobre a discrepância em duas etapas. Na primeira etapa, é realizada a estimação do parâmetro por meio de uma abordagem bayesiana *likelihood-free*. Na segunda etapa, a amostra da distribuição a *posteriori* do parâmetro do modelo é utilizada para inferir sobre a discrepância de modelo e a resposta preditiva, empregando regressão por processo gaussiano em conjunto com o Método de Monte Carlo via amostragem de importância. O sistema estudado apresenta um grau de liberdade e seu amortecimento viscoelástico é dado pelo modelo de variáveis internas *Anelastic Displacement Fields*. O problema inverso de estimação de parâmetros é resolvido por meio do método *Sequential Monte Carlo*. A resposta preditiva a *posteriori* estimada é avaliada com base na sua capacidade em ajustar os dados de teste. A discrepância obtida via regressão por processo gaussiano é comparada com a discrepância exata que, no caso estudado, é conhecida.

**Palavras-Chave:** Discrepância de modelo, Amortecimento viscoelástico, Abordagem bayesiana *likelihood-free*, Regressão por processo gaussiano, Método de Monte Carlo via amostragem de importância

---

## 1 INTRODUÇÃO

Os materiais viscoelásticos recebem essa denominação por apresentarem o comportamento combinado de um sólido elástico ideal e de um fluido viscoso ideal [Trindade and Almeida, 2006]. Existem diversas metodologias para representar o comportamento desses materiais. O modelo de variáveis internas *Anelastic Displacement Fields* (ADF) é um exemplo dessas metodologias e contabiliza o deslocamento total como a soma de uma parcela elástica com



uma parcela anelástica (não-elástica), associada a fenômenos de relaxação [Lesieutre and Bianchini, 1995]. O deslocamento anelástico, por sua vez, pode ser entendido como uma soma, em que cada parcela se refere a um campo anelástico específico. Quanto mais fraca for a dependência do comportamento viscoelástico em relação à frequência, maior será o número de campos anelásticos, e vice-versa [Lesieutre, 1992].

É importante destacar que existem diversas fontes de incerteza na modelagem e que estas podem influenciar os resultados. Caso os parâmetros do modelo sejam desconhecidos, é necessário empregar alguma técnica para estimá-los. Por outro lado, nenhum modelo é perfeito. Mesmo em um cenário hipotético em que não haja incertezas paramétrica e experimental, a resposta simulada não é a resposta verdadeira do sistema estudado. Isto ocorre devido à existência da discrepância de modelo [Kennedy and O'Hagan, 2001]. A discrepância de modelo sempre está presente e isso pode ocorrer com diferentes graus e por diversas razões. Seja devido à ausência de determinados aspectos físicos na modelagem com o emprego de hipóteses simplificadoras, seja pelo uso de aproximações numéricas para que o modelo possa ser implementado computacionalmente. O fato é que a discrepância não deve ser ignorada, pois a ausência de um tratamento adequado para esta fonte de incertezas pode conduzir a resultados incorretos [Gardner et al., 2021].

Embora não seja uma tarefa simples caracterizá-la de maneira precisa, uma vez que é necessário um conhecimento profundo sobre o modelo empregado e o fenômeno físico real para poder conceber sua forma funcional, existem diferentes estratégias para lidar com a modelagem da discrepância de modelo [Ling et al., 2014]. Como a discrepância de modelo é representada na formulação matemática da resposta do sistema como um termo aditivo, corretivo e não-intrusivo, o desafio de sua modelagem não está apenas no desconhecimento de sua forma funcional, mas também na necessidade de realizar a estimação simultânea dos valores dos parâmetros de interesse do modelo e de hiperparâmetros associados à discrepância. O excesso de parâmetros desconhecidos, aliado à quantidade limitada de informação, pode ocasionar problemas de não-identificabilidade. Em outras palavras, a resposta experimental pode ser ajustada por diferentes combinações de parâmetros e hiperparâmetros [Arendt et al., 2012]. Isso ocorre, por exemplo, com abordagens bayesianas hierárquicas que empregam uma distribuição a *priori* não-informativa para a discrepância, lidando com funções de verossimilhança insensíveis à variação dos parâmetros do modelo. Esse problema se torna mais grave em casos de distribuições a *priori* não-informativas para os parâmetros do modelo [Brynjarsdóttir and OHagan, 2014].

Existem algumas possíveis soluções para o problema de não-identificabilidade, como o emprego de distribuições a *priori* informativas. Também pode-se citar a abordagem apresentada por Gardner et al. [2021], na qual é empregada uma estratégia de desacoplamento dos problemas de inferência em duas etapas. A primeira etapa consiste na estimação dos parâmetros do modelo e é realizada com uma abordagem bayesiana *likelihood-free*, *Bayesian History Matching* neste caso. Na segunda etapa, é realizada a inferência da discrepância e da resposta preditiva por meio de regressão por processo gaussiano, combinada com o Método de Monte Carlo via amostragem de importância, empregando a amostra da distribuição a *posteriori* dos parâmetros do modelo obtida na etapa anterior.

As abordagens bayesianas ditas *likelihood-free* buscam a amostragem da distribuição a *posteriori* dos parâmetros do modelo sem a necessidade de sucessivas avaliações da função de verossimilhança que, quando desconhecida ou custosa em termos computacionais, invi-



abiliza a aplicação de abordagens bayesianas convencionais [Sisson et al., 2018]. Existe um grupo particular dentre estas abordagens, denominado *Approximate Bayesian Computation* (ABC). A abordagem ABC padrão realiza um mapeamento para uma estatística de resumo mais simples e de dimensão inferior à amostra dos dados observados [Beaumont, 2019]. A ideia básica é empregar essa estatística em uma métrica que substitui a função de verossimilhança, em conjunto com uma tolerância pequena o suficiente para produzir uma boa aproximação para a distribuição a *posteriori* dos parâmetros. Entretanto, para muitos problemas práticos, é difícil obter estatísticas suficientes ou mesmo altamente informativas. Em alguns campos de aplicação, existe um histórico de desenvolvimento de estatísticas de resumo. No entanto, também é possível que estatísticas de resumo definidas empiricamente possam ser utilizadas [Ke and Tian, 2019].

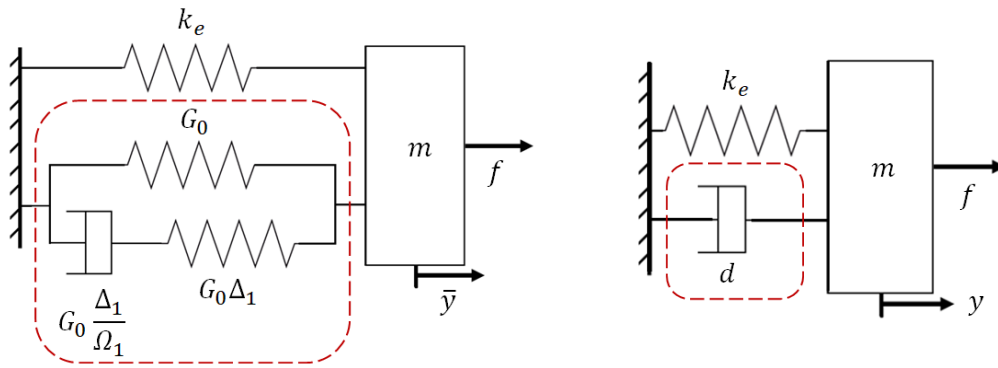
Entre as abordagens ABC, pode-se citar o método *Sequential Monte Carlo* (SMC-ABC). Introduzido por Sisson et al. [2007], é baseado na amostragem de importância sequencial. Realiza ajustes sucessivos, a cada iteração, de uma densidade de núcleo para amostrar a distribuição a *posteriori* dos parâmetros do modelo. Essa densidade de núcleo é empregada como distribuição proposta para o passo seguinte.

O trabalho está dividido em quatro capítulos. O Capítulo 2 apresenta, brevemente, a formulação matemática empregada. No Capítulo 3, são discutidos os resultados obtidos. Por fim, as conclusões são apresentadas no Capítulo 4.

## 2 FORMULAÇÃO MATEMÁTICA

### 2.1 Modelos físico-matemáticos

Neste trabalho, é considerado um sistema mecânico de um grau de liberdade e amortecimento viscoelástico dado pelo modelo ADF com um único campo anelástico, conforme apresentado na Fig. 1a. Nestas condições, o ADF é análogo à forma de Maxwell do modelo Sólido Linear Padrão [Lin, 2020].



(a) Sistema com amortecimento viscoelástico. (b) Sistema com amortecimento viscoso.

**Fig. 1:** Representações esquemáticas dos modelos empregados.

A dissipação de energia é representada pelo amortecedor viscoelástico, delimitado pela linha tracejada, no qual  $G_0$  é o módulo de relaxação estático,  $\Delta_1$  é a resistência à relaxação e  $\Omega_1$  é o inverso do tempo de relaxação. É importante notar que o amortecedor viscoelástico também contribui com rigidez ao sistema, com suas duas molas lineares. A inércia do sistema é representada pela massa concentrada  $m$ . Uma mola linear de cons-

tante elástica  $k_e$  também contribui para a rigidez do sistema. O deslocamento horizontal é dado por  $\bar{y}$  e a excitação externa consiste na carga pontual  $f$ .

A Figura 1b apresenta o modelo do sistema dinâmico de um grau de liberdade com amortecimento viscoso. Este modelo consiste em uma simplificação na qual o amortecedor viscoelástico (Fig. 1a) é substituído por um amortecedor viscoso, cuja constante de amortecimento  $d$  é desconhecida. Dessa forma, ela é um parâmetro a ser estimado por meio da solução de um problema inverso.

A equação de movimento do sistema apresentado na Fig. 1a é dada por

$$m\ddot{\bar{y}}(t) + k_e\bar{y}(t) + G_0(1 + \Delta_1)\bar{y}(t) - G_0\Delta_1\xi_1(t) = f(t) \quad (1)$$

$$\frac{\Delta_1}{\Omega_1}\dot{\xi}_1(t) + \Delta_1\xi_1(t) = \Delta_1\bar{y}(t) \quad (2)$$

$$\bar{y}(0) = \dot{\bar{y}}(0) = 0 \quad (3)$$

$$\xi_1(0) = 0 \quad (4)$$

O termo  $\xi_1$  é uma variável interna para a representação da evolução temporal do campo anelástico [Vasques et al., 2010].

A equação de movimento do sistema da Fig. 1b é dada por

$$m\ddot{y}(t) + d\dot{y}(t) + k_e y(t) = f(t) \quad (5)$$

$$y(0) = \dot{y}(0) = 0 \quad (6)$$

Onde  $y$  é o deslocamento horizontal calculado a partir do modelo de amortecimento viscoso.

Tomando-se o sistema descrito pelo modelo de amortecimento viscoelástico, de resposta exata  $\bar{y}$ , tem-se que, ao considerar um ruído experimental aditivo  $\varepsilon$ , a resposta experimental  $y_{exp}$  pode ser expressa por

$$y_{exp}(t) = \bar{y}(t) + \varepsilon(t) \quad (7)$$

Ao tomar-se o modelo de amortecimento viscoso, a Eq. (7) pode ser escrita na forma

$$y_{exp}(t) = y(t) + \delta(t) + \varepsilon(t) \quad (8)$$

Onde  $\delta$  é a discrepância do modelo de amortecimento viscoso. Considera-se que a discrepância e o erro experimental são independentes [Kennedy and O'Hagan, 2001]. É importante destacar que, por ser modelada como um termo corretivo e não-intrusivo, a discrepância de modelo atua como uma compensação sobre a resposta do modelo. Isto significa que ela não realiza nenhuma mudança interna na formulação físico-matemática do mesmo.



## 2.2 Regressão por processo gaussiano

Com a regressão por processo gaussiano, pode-se expressar a distribuição a *posteriori* da resposta preditiva  $f(y, t_*) = y(t_*) + \delta(t_*)$  [Gardner et al., 2021] em instantes  $t_*$ , empregando um processo gaussiano de média nula e função de covariância arbitrária como uma distribuição a *priori* de  $f(y, t)$ , em instantes  $t$ , nos quais a resposta experimental é considerada [Rasmussen and Williams, 2006]. Ou seja,

$$f(y, t) \sim PG(0, v((y, t), (y', t'))) \quad (9)$$

Neste trabalho, a função de covariância é exponencial quadrática [Gardner et al., 2020]

$$v((y, t), (y, t')) = \phi_1^2 \exp\left(-\frac{(y(t) - y(t'))^2}{2\phi_2} - \frac{(t - t')^2}{2\phi_3}\right) \quad (10)$$

Onde  $\Phi = [\phi_1, \phi_2, \phi_3]^T$  é o vetor de hiperparâmetros,  $\phi_1$  é o desvio-padrão do processo gaussiano,  $\phi_2$  e  $\phi_3$  são comprimentos de correlação ao quadrado.

Ao considerar que o erro experimental é de distribuição normal de média nula e desvio-padrão constante  $\sigma$ , pode-se tomar dados de treinamento (avaliados em  $t$ ) e dados de teste (avaliados em  $t_*$  e sobre os quais se deseja fazer predições) e expressá-los como [Rasmussen and Williams, 2006]

$$\begin{bmatrix} \mathbf{y}_{exp}(\mathbf{t}) \\ \mathbf{f}(\mathbf{y}, \mathbf{t}_*) \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma^2 \mathbf{I}_{\mathbf{t},\mathbf{t}} & \mathbf{K}_{\mathbf{t},\mathbf{t}_*} \\ \mathbf{K}_{\mathbf{t}_*,\mathbf{t}} & \mathbf{K}_{\mathbf{t}_*,\mathbf{t}_*} \end{bmatrix}\right) \quad (11)$$

Onde  $\mathbf{y}_{exp}(\mathbf{t})$  corresponde à resposta experimental nos instantes de treinamento, armazenados no vetor  $\mathbf{t}$ , e  $\mathbf{f}(\mathbf{y}, \mathbf{t}_*)$  corresponde à resposta corrigida  $f$  avaliada nos instantes de teste, armazenados no vetor  $\mathbf{t}_*$ . As matrizes  $\mathbf{K}_{\mathbf{t},\mathbf{t}}$ ,  $\mathbf{K}_{\mathbf{t},\mathbf{t}_*}$ ,  $\mathbf{K}_{\mathbf{t}_*,\mathbf{t}}$  e  $\mathbf{K}_{\mathbf{t}_*,\mathbf{t}_*}$  são calculadas a partir de  $v((\mathbf{y}, \mathbf{t}), (\mathbf{y}, \mathbf{t}))$ ,  $v((\mathbf{y}, \mathbf{t}), (\mathbf{y}, \mathbf{t}_*))$ ,  $v((\mathbf{y}, \mathbf{t}_*), (\mathbf{y}, \mathbf{t}))$  e  $v((\mathbf{y}, \mathbf{t}_*), (\mathbf{y}, \mathbf{t}_*))$  respectivamente. O termo  $\mathbf{I}_{\mathbf{t},\mathbf{t}}$  é uma matriz identidade cuja ordem é a dimensão do vetor  $\mathbf{t}$ .

Assim, a distribuição preditiva a *posteriori* condicional de  $f$ , obtida por meio de regressão por processo gaussiano, é expressa como

$$p(\mathbf{f}(\mathbf{y}, \mathbf{t}_*) | \mathbf{y}_{exp}(\mathbf{t}), d, \Phi) = N(\mathbf{E}[\mathbf{f}], \mathbf{V}[\mathbf{f}]) \quad (12)$$

$$\mathbf{E}[\mathbf{f}] = \mathbf{K}_{\mathbf{t}_*,\mathbf{t}} (\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma^2 \mathbf{I}_{\mathbf{t},\mathbf{t}})^{-1} \mathbf{y}_{exp}(\mathbf{t}) \quad (13)$$

$$\mathbf{V}[\mathbf{f}] = \mathbf{K}_{\mathbf{t}_*,\mathbf{t}_*} - \mathbf{K}_{\mathbf{t}_*,\mathbf{t}} (\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma^2 \mathbf{I}_{\mathbf{t},\mathbf{t}})^{-1} \mathbf{K}_{\mathbf{t},\mathbf{t}_*} \quad (14)$$

## 2.3 Predição da resposta corrigida

De posse das Equações (12 - 14), é realizado um processo de marginalização, seguido pelo emprego do Método de Monte Carlo via amostragem de importância, em que são empregadas a amostra a *posteriori* de  $d$  e uma amostra da distribuição a *priori* de  $\Phi$ . Assim, obtém-se uma aproximação da distribuição preditiva a *posteriori* de  $\mathbf{f}$  como um processo gaussiano cuja esperança e variância podem ser estimadas por meio de médias ponderadas por pesos de importância [Gardner et al., 2021].

Assim, o vetor esperança  $\mathbf{E}[\mathbf{f}]$  e a matriz de covariância  $\mathbf{V}[\mathbf{f}]$  podem ser expressos como

$$\mathbf{E}[\mathbf{f}] \cong \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} w^{(j,k)} \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) \quad (15)$$

$$\begin{aligned} \mathbf{V}[\mathbf{f}] \cong & \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} w^{(j,k)} \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}_*}^{(j,k)} - \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{K}_{\mathbf{t}, \mathbf{t}_*}^{(j,k)} \right. \\ & \left. + \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) \right)^T \right) - \mathbf{E}[\mathbf{f}] \mathbf{E}[\mathbf{f}]^T \end{aligned} \quad (16)$$

Onde o peso de importância  $w^{(j,k)}$  é dado por

$$w^{(j,k)} = \frac{p(\mathbf{y}_{exp}(\mathbf{t}) \mid d^{(j)}, \Phi^{(j,k)})}{\sum_{j=1}^{N_j} \sum_{k=1}^{N_k} p(\mathbf{y}_{exp}(\mathbf{t}) \mid d^{(j)}, \Phi^{(j,k)})} \quad (17)$$

O numerador apresentado na Eq. (17) é a distribuição marginal do vetor  $\mathbf{y}_{exp}$  mostrado na Eq. (11), que consiste em uma distribuição normal multivariada de média nula e matriz de covariância dada por  $\mathbf{K}_{\mathbf{t}, \mathbf{t}} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}}$ . Como pode ser notado, os pesos de importância consistem em valores normalizados desta distribuição, considerando cada uma das  $N_j$  partículas  $d^{(j)}$  da amostra da distribuição a *posteriori* aproximada do parâmetro  $d$  e os  $N_k$  vetores  $\Phi^{(j,k)}$  da amostra da distribuição a *priori* dos hiperparâmetros.

#### 2.4 Predição da discrepância de modelo

A distribuição preditiva a *posteriori* da discrepância do modelo de amortecimento viscoso pode, por sua vez, ser expressa como

$$p(\delta(\mathbf{t}_*) \mid \mathbf{y}_{exp}(\mathbf{t}), d, \Phi) = N(\mathbf{E}[\delta], \mathbf{V}[\delta]) \quad (18)$$

$$\mathbf{E}[\delta] \cong \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} w^{(j,k)} \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) - \mathbf{y}^{(j)}(\mathbf{t}_*) \right) \quad (19)$$

$$\begin{aligned} \mathbf{V}[\delta] \cong & \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} w^{(j,k)} \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}_*}^{(j,k)} - \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{K}_{\mathbf{t}, \mathbf{t}_*}^{(j,k)} \right. \\ & \left. + \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) - \mathbf{y}^{(j)}(\mathbf{t}_*) \right) \right. \\ & \left. \left( \mathbf{K}_{\mathbf{t}_*, \mathbf{t}}^{(j,k)} \left( \mathbf{K}_{\mathbf{t}, \mathbf{t}}^{(j,k)} + \sigma^2 \mathbf{I}_{\mathbf{t}, \mathbf{t}} \right)^{-1} \mathbf{y}_{exp}(\mathbf{t}) - \mathbf{y}^{(j)}(\mathbf{t}_*) \right)^T \right) - \mathbf{E}[\delta] \mathbf{E}[\delta]^T \end{aligned} \quad (20)$$

Onde  $\mathbf{y}^j(\mathbf{t}_*)$  é vetor de resposta do modelo de amortecimento viscoso avaliado em  $\mathbf{t}_*$  e considerando  $d = d^{(j)}$ .

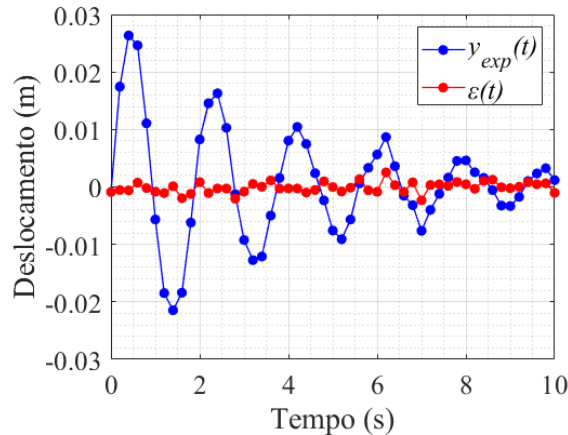


Para a amostragem da distribuição a *posteriori* do parâmetro  $d$ , o algoritmo SMC-ABC é empregado, baseado nas publicações de Toni et al. [2009] e Lintusaari et al. [2016]. Recomendamos a leitura destes artigos para mais detalhes. Neste trabalho, é considerada uma métrica estatística baseada na soma do erro quadrático entre a resposta experimental e a resposta do modelo [Abdessalem et al., 2018].

### 3 RESULTADOS E DISCUSSÕES

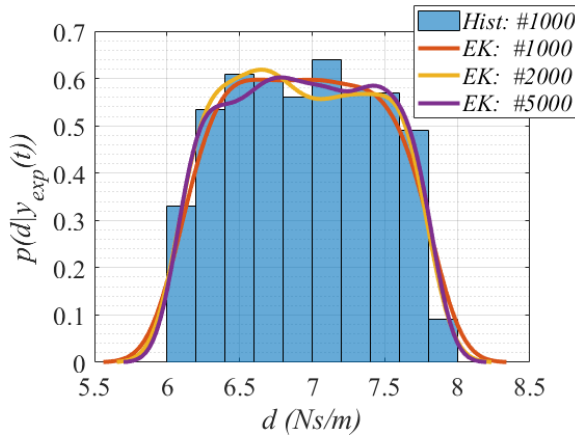
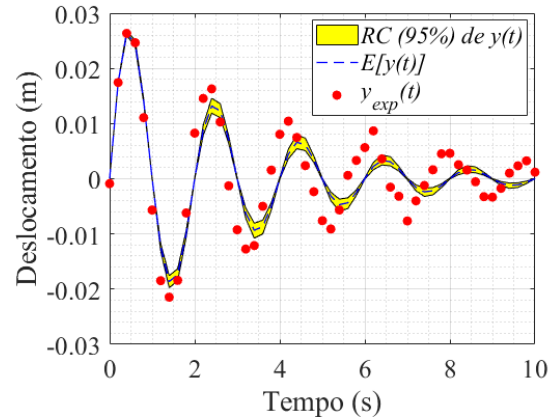
Neste trabalho, a resposta experimental é a resposta impulsiva obtida a partir das Eqs.(1,2), considerando  $G_0 = 10$  N/m,  $\Delta_1 = 10$ ,  $\Omega_1 = 20$  rad/s,  $k_e = 100$  N/m e  $m = 10$  kg, com ruído  $\varepsilon$  aditivo e não-correlacionado de distribuição normal de média nula e desvio-padrão  $\sigma = 10^{-3}$  m. As Equações (1 - 4, 7) são empregadas para construir a resposta experimental sintética  $y_{exp}(t_*)$ , para  $t_* \in [0, 10]$  s em intervalos de 0,1 s (10 amostras por segundo). Esta resposta corresponde aos dados de teste em  $\mathbf{y}_{exp}(\mathbf{t}_*)$  e é utilizada como meio de validação da predição de  $f(y, t_*)$ . Os dados de treinamento são respostas associadas a  $t \in [0, 10]$  s em intervalos de 0,2 s (cinco amostras por segundo), denotados por  $\mathbf{y}_{exp}(\mathbf{t})$ . Os dados de treinamento são empregados na estimação do parâmetro  $d$  e na predição da resposta  $\mathbf{f}(\mathbf{y}, \mathbf{t}_*)$ .

A resposta sintética empregada na etapa de treinamento, bem como o ruído associado, são apresentados na Fig. 2.



**Fig. 2:** Resposta experimental sintética de treinamento e ruído.

Com base na Figura 2, é possível notar que o ruído pode ser considerado de baixa amplitude em comparação com a resposta experimental de treinamento. Isso representa o contexto em que o experimento é bem controlado. Desse modo, espera-se que o ruído não tenha impacto relevante sobre os resultados e que a discrepância de modelo seja a principal fonte de incertezas. Além disso, é importante destacar que o comportamento oscilante com amplitude decrescente da resposta poderia ser reproduzido pelo modelo do sistema com amortecimento viscoso apenas se  $0 < d < 2\xi\sqrt{k_e m}$ , caso no qual este modelo representa um sistema sub-amortecido. Dito isto, para o caso estudado,  $d \in (0; 63, 24)$  Ns/m. Esta informação é utilizada na escolha do suporte da distribuição a *priori* do parâmetro  $d$ . Então, a densidade de probabilidade a *posteriori* aproximada de  $d$  é obtida considerando a distribuição a *priori*  $d \sim U[0, 63]$  Ns/m. Isto pode ser visto na Fig. 3.

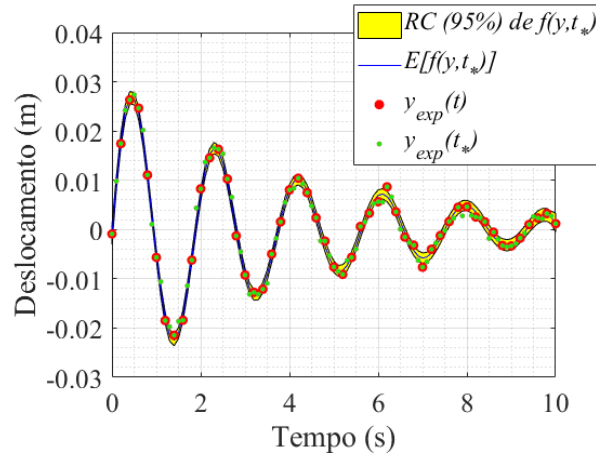
(a) Densidade *a posteriori* aproximada do parâmetro  $d$ .(b) Região de 95% de credibilidade da resposta *a posteriori* do modelo.**Fig. 3:** Solução do problema inverso com SMC-ABC.

Na Figura 3a, a densidade de probabilidade *a posteriori* aproximada para o parâmetro  $d$  é apresentada por meio de suas amostras nas formas de um histograma (*Hist.*) e três curvas de estimativas de densidade por *kernel* (*EK*), considerando diferentes quantidades de partículas em cada amostra: 1000, 2000 e 5000. É possível notar que o aumento da quantidade de partículas não traz alterações significativas na curva aproximada da distribuição *a posteriori*. Então, considerando a amostra de 1000 partículas neste trabalho, a densidade *a posteriori* aproximada de  $d$  apresenta o suporte  $[6, 8]$  Ns/m e a esperança *a posteriori*  $E[d|y_{exp}(\mathbf{t})]$ , corresponde a, aproximadamente, 6,95 Ns/m.

Tomando a amostra da densidade *a posteriori* aproximada do parâmetro  $d$ , pode-se avaliar a propagação de incerteza do modelo empregado por meio da região de 95% de credibilidade (RC) da resposta do modelo de amortecimento viscoso, conforme pode ser visto na Fig. 3b. Em todo este trabalho, a região de credibilidade é composta por 95% da amostra e suas fronteiras são definidas a partir dos percentis de 2,5% e 97,5%. É possível perceber que este modelo não ajusta a resposta experimental de treinamento de maneira satisfatória. Isso ocorre devido à discrepância de modelo que, neste caso, apresenta um grau de severidade considerável. Assim, ela não pode ser ignorada.

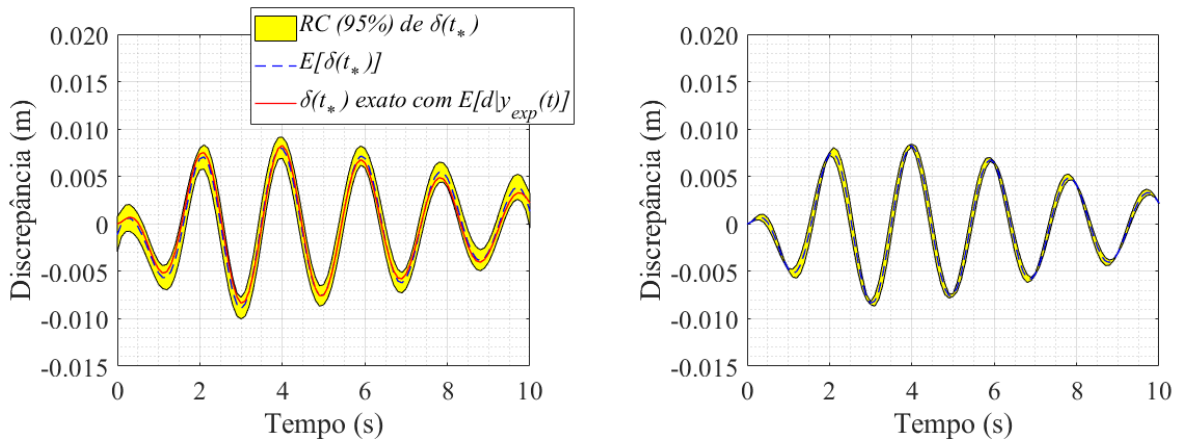
Dito isto, considerando a resposta experimental de treinamento e a amostra da densidade *a posteriori* aproximada e aplicando regressão por processo gaussiano para a predição de  $\mathbf{f}(\mathbf{y}, \mathbf{t}_*)$ , obtém-se a RC (95%) da resposta preditiva *a posteriori*, apresentada na Fig. 4. Para os hiperparâmetros, são empregadas densidades *a priori* não-informativas, dadas por  $\phi_i \sim U[0, 5], i = 1, 2, 3$ , formando uma amostra de 1000 vetores  $\Phi$ .

Na Figura 4, nota-se que a RC (95%) apresenta curvas que proporcionam ajustes satisfatórios para os dados de teste. Além disso, nos primeiros cinco segundos, a área da RC é maior nas proximidades dos mínimos e máximos locais, indicando um maior nível de incerteza. Para instantes posteriores a cinco segundos, as incertezas se tornam mais significativas com o decorrer do tempo, ampliando a área da RC. Isto ocorre devido ao fato de o ruído experimental apresentar influência crescente com o tempo, dado o comportamento oscilatório em torno de zero com amplitude decrescente na resposta.



**Fig. 4:** Região de 95% de credibilidade da resposta preditiva a *posteriori* via regressão por processo gaussiano.

A discrepância inferida via regressão por processo gaussiano é apresentada na Fig. 5a. A curva exata da discrepância é obtida por meio das Eqs. (1-8), considerando  $E[d|\mathbf{y}_{exp}(\mathbf{t})]$  como estimativa pontual para  $d$ . É importante destacar que a discrepância de modelo via regressão por processo gaussiano apresenta incerteza considerável (área da RC), decorrente das influências das incertezas do parâmetro do modelo e dos hiperparâmetros do processo gaussiano, além do próprio ruído experimental.



**(a)** Região de 95% de credibilidade da discrepância de modelo via regressão por processo gaussiano.

**(b)** Região de 95% de credibilidade da discrepância exata via amostra a *posteriori* do parâmetro  $d$ .

**Fig. 5:** Discrepância do modelo de amortecimento viscoso.

Além disso, percebe-se que a RC da discrepância acomoda a curva da resposta exata, acompanhando suas variações suaves, máximos e mínimos locais. A curva média se aproxima da curva exata. Isso mostra que o procedimento empregado é capaz de fornecer uma representação satisfatória da discrepância de modelo. Este fato também pode ser constatado ao comparar a Fig. 5a com a Fig. 5b, na qual a RC (95%) da discrepância exata é apresentada, tendo em vista as Eqs. (1-8) e a amostra da densidade a *posteriori* aproximada do parâmetro  $d$ .



## 4 CONCLUSÕES

Neste trabalho, foi realizado um estudo inicial e simplificado sobre a discrepância de modelo, empregando regressão por processo gaussiano para inferir sobre o comportamento de um sistema dinâmico com amortecimento viscoelástico por meio de um modelo de amortecimento viscoso. Alguns aspectos da modelagem foram apresentados de forma resumida. O modelo escolhido é simples, conhecido e de fácil compreensão. Os autores empregaram uma abordagem que visa desacoplar os problemas de inferência conforme mencionado neste artigo. Foram realizadas previsões sobre a resposta corrigida do modelo e sobre a sua discrepância de modelo. Todos os resultados foram satisfatórios.

Com esses resultados, os autores puderam obter uma compreensão inicial sobre o emprego da técnica de regressão por processo gaussiano na inferência da discrepância de modelo, permitindo a possibilidade de seu emprego posterior na realização de estudos da discrepância em contextos de maior complexidade.

## 5 Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores também agradecem pelo apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

## Referências

- A. Abdessalem, N. Dervilis, D. Wagg, and K. Worden. Model selection and parameter estimation in structural dynamics using approximate bayesian computation. *Mechanical Systems and Signal Processing*, 99:306–325, 2018. © 2017 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).
- P. D. Arendt, D. W. Apley, and W. Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design - Transactions of the ASME*, 134(10), 2012.
- M. A. Beaumont. Approximate bayesian computation. *Annual Review of Statistics and Its Application*, 6(1):379–403, 2019.
- J. Brynjarsdóttir and A. OHagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.
- P. Gardner, C. Lord, and R.J. Barthorpe. Bayesian history matching for structural dynamics applications. *Mechanical Systems and Signal Processing*, 143:106828, 2020. © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).
- P. Gardner, T.J. Rogers, C. Lord, and R.J. Barthorpe. Learning model discrepancy: A gaussian process and sampling-based approach. *Mechanical Systems and Signal Processing*, 152:107381, 2021. © 2021 The Author(s). Pu-



- blished by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).
- Y. Ke and T. Tian. *Approximate Bayesian Computational Methods for the Inference of Unknown Parameters*, pages 515–529. Springer International Publishing, 2019.
- M. C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- G. A. Lesieutre. Finite elements for dynamic modeling of uniaxial rods with frequency-dependent material properties. *International Journal of Solids and Structures*, 29(12):1567–1579, 1992.
- G. A. Lesieutre and E. Bianchini. Time Domain Modeling of Linear Viscoelasticity Using Anelastic Displacement Fields. *Journal of Vibration and Acoustics*, 117(4):424–430, 1995.
- C. Y. Lin. Alternative form of standard linear solid model for characterizing stress relaxation and creep: Including a novel parameter for quantifying the ratio of fluids to solids of a viscoelastic solid. *Frontiers in Materials*, 7, 2020.
- Y. Ling, J. Mullins, and S. Mahadevan. Selection of model discrepancy priors in bayesian calibration. *Journal of Computational Physics*, 276(C):665–680, 2014.
- J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander. Fundamentals and Recent Developments in Approximate Bayesian Computation. *Systematic Biology*, 66(1):e66–e82, 2016.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The M.I.T Press, 2006.
- S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.
- S. A. Sisson, Y. Fan, and M. A. Beaumont. *Overview of ABC from: Handbook of Approximate Bayesian Computation*. CRC Press, 2018.
- T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6(31):187–202, 2009.
- M. Trindade and S. Almeida. Análise custo-benefício de modelos de amortecimento para estruturas com elementos viscoelásticos. 06 2006.
- C. M. A. Vasques, R.A.S. Moreira, and J. Rodrigues. Viscoelastic damping technologies—part i: Modeling and finite element implementation. *Journal of Advanced Research in Mechanical Engineering*, 1, 2010.

---

## Resumos



# Técnicas de Processamento de Linguagem Natural Para Detecção de Discurso de Ódio

Cássia Claudiane Silva da Rosa<sup>1</sup> and Renato Porfirio Ishii<sup>2</sup>

<sup>1</sup> *Faculdade de Computação/UFMS, Campo Grande/MS, Brasil*

---

## Abstract

As inúmeras publicações nocivas geradas entre grandes quantidades de dados, que são expelidos diariamente nas mídias sociais, fazem com que seja necessário adotar tecnologias automatizadas de moderação de conteúdo online. A classificação de sentenças e a análise de sentimentos são técnicas de Processamento de Linguagem Natural (PLN) utilizadas para detecção de discurso de ódio em plataformas de mídias sociais como o Facebook, Twitter e Instagram. Entretanto, ainda existem dificuldades que diminuem a eficácia dessas ferramentas na língua portuguesa. Pesquisas anteriores como a de Leite et al. [2020] evidenciam como modelos de PLN têm sua acurácia elevada quando são treinados com conjuntos de dados focadas no domínio do idioma português. Neste trabalho, é proposta a criação de um corpus linguístico de larga escala para o português do Brasil composto por tweets coletados após as eleições presidenciais de 2022. Experimentos utilizando o modelo pré-treinado Bertimbau foram feitos a partir do fine-tuning com a base de dados pré-existente “hatecheck-portuguese” de Röttger et al. [2022].

**Keywords:** Processamento de linguagem natural, Discurso de ódio, Mineração de dados

---

## Referências

- Joao A Leite, Diego F Silva, Kalina Bontcheva, and Carolina Scarton. Toxic language detection in social media for brazilian portuguese: New dataset and multilingual analysis. *arXiv preprint arXiv:2010.04543*, 2020.
- Paul Röttger, Haitham Seelawi, Debora Nozza, Zeerak Talat, and Bertie Vidgen. Multilingual hatecheck: Functional tests for multilingual hate speech detection models. *arXiv preprint arXiv:2206.09917*, 2022.



## Consensus docking to ensemble-docking using multiple programs: exemple with *Drosophila melanogaster* dopamine transporter.

Fabiani Fernanda Triches<sup>1</sup>, Francieli Triches<sup>2</sup>, Cilene Lino de Oliveira<sup>1</sup>.

<sup>1</sup>Laboratory of Behavioral Neurobiology, CCB, Federal University of Santa Catarina, Brazil.

<sup>2</sup>Department of Mathematics, CFM, Federal University of Santa Catarina, Brazil.

---

### Abstract

The main goal of this study is to validate a consensus docking calculation to ensemble-docking using multiple programs. The ensemble-docking will be done with 11 *Drosophila melanogaster* dopamine transporter (dDAT), 10 downloaded from RCSB-PDB, and 1 theoretical dDAT was downloaded from AlphaFold and 26 antidepressants downloaded from ZINC and 1550 decoys downloaded from DUD-E. We are using three different programs, Autodock Vina 1.5.7, DockThor e Gold 2021.3.0. A consensus docking technique combined the score functions of different programs into one. The consensus will be divided into three steps. The first step of consensus is to transform the  $\Delta G$  (to Autodock Vina and DockThor) and fitness (to Gold) for each ligand-dDAT pair in a standard metric. For this, we will test two different forms of ordination. The first form, the ligand with the best interaction, receives the 1<sup>o</sup> position, and the last interaction receives the 1576<sup>o</sup> position. For that, we will use the following equation:  $\text{SUMPRODUCT}((x \leq 1^{\circ} \text{ column element} : \text{last column element}) / \text{COUNTIF}(1^{\circ} \text{ column element} : \text{last column element}; 1^{\circ} \text{ column element} : \text{last column element}))$ . In this way, we lost the scale between the elements, it is understood here that the difference between elements 1 and 2 is the same difference between elements 2 and 3, and so on. In the other form, in order to estimate the magnitude of the difference between the ordered elements, we use the formula of the linear function:  $f(x) = (x - \text{min}) / (\text{min} - \text{max})$ . Where the “x” is the  $\Delta G$  or fitness, the “min” is the result of the ligand with the minimum value (worse affinity), and the “max” is the result of the ligand with the maximum value (best affinity). In the linear function, the resulting values range from 0 to 1, with 0 being the ligand with the best interaction and 1 being the ligand with the worst interaction. Regardless of the ordering of the ligands, the subsequent steps are the same. In the second step, a consensus will be made between the programs of each ligand-dDAT pair. For this, the exponential average calculation between the ranks of each ligand-dDAT pair in each program is applied. The last step is a simple average between the posts of each ligand docked with all dDATs (TRICHES et al., 2022). In this way, we hope to find antidepressants in the top positions of the ranking, followed by decoys. We hope that by using the linear function in the first step, it will be possible to capture the difference between the ranks of the ligands. Finally, we hope that the consensus will give a more reliable result when compared to the data on the inhibition constant ( $K_i$ ) in the literature obtained from *in vitro* studies.

Keywords: affinity, antidepressants, *Drosophila melanogaster*, monoamines transporters.

---

Referências: TRICHES, F. F.; TRICHES, F.; DE OLIVEIRA, C. L. "Consensus combining outcomes of multiple ensemble dockings: examples using dDAT crystalized complexes." *MethodsX* 9 (2022).

Contato: Fabiani Triches, [fabianitriches@gmail.com](mailto:fabianitriches@gmail.com).



# A molecular dynamics approach to spermine and spermidine SI-PPCs interactions with heparin and DNA

Frederico Henrique do C. Ferreira<sup>1</sup> and Luiz Antônio S. Costa<sup>1\*</sup>

<sup>1</sup> NEQC/Universidade Federal de Juiz de Fora, Juiz de Fora, MG-Brasil

---

## Abstract

Since the discovery of anticancer activity of cisplatin many direct analogues have been studied and developed. Those analogues share the same mechanism of action and aim a higher effectiveness with lower side effects. BBR3464 was the first non-analogue to be tested in humans and it started a new way of interaction called clamps and forks: a noncovalent cyclic hydrogen interaction stabilized by 6 and 8 atoms, respectively, among phosphates/sulphates groups in biomolecules and the PtN<sub>4</sub> centres. After that, many compounds that follow the same mechanism of action have been designed, the so called substitution inert polynuclear platinum complexes [Farrell, 2015, N. M. P. Rosa, 2022]. The main idea of this project is to substitute the simple diamine linking chains in those complexes for both spermine and spermidine, which have more amines in the structure, and, thus, may increase the interaction of the metallic complexes with biological macromolecules. It was performed molecular dynamics simulations on systems containing both DNA/HS and a SI-PPC and the results were analysed in terms of atomic fluctuations, interaction energies and solvent accessible surface area in order to evaluate the atomic-scale variables in the studied arrangements. Those results may corroborate to those available in the experimental literature and might help explain these results.

**Keywords:**SI-PPCs, Heparin, Molecular Dynamics, DNA, Polinuclear Platinum Complexes

---

## References

- N. P. Farrell. Multi-platinum anti-cancer agents. substitution-inert compounds for tumor selectivity and new targets. *Chem. Soc. Rev.*, 44:8773–8785, 2015.
- et. al. N. M. P. Rosa. Substitution-inert polynuclear platinum complexes and glycosaminoglycans: A molecular dynamics study of its non-covalent interactions. *Journal of Inorganic Biochemistry*, 232:111811, 2022.

Contato: Frederico Henrique do Carmo Ferreira, frederico.h@outlook.com



# Segmentação de pragas e doenças em folhas de café utilizando redes convolucionais

Humberto da Silva Neto<sup>1</sup>, Shirley Peroni Neves Cani<sup>1</sup> e Mariana Rampinelli Fernandes<sup>1</sup>

<sup>1</sup> Instituto Federal do Espírito Santo/ES, Brasil

---

## Abstract

Este trabalho tem como objetivo desenvolver uma ferramenta baseada em visão computacional e aprendizado de máquina capaz de identificar e estimar a severidade de pragas e doenças em folhas de café. As doenças abordadas neste trabalho são bicho-mineiro, ferrugem, mancha de olho pardo e phoma, cujos sintomas manifestam-se principalmente nas folhas. A base utilizada para teste é a BRACOL [Krohling, 2019], composta por 100 imagens de folhas saudáveis e 400 com sintomas, incluindo exemplos de folhas com múltiplas doenças. Para tanto, propõe-se explorar e comparar modelos de segmentação semântica como U-Net, PSPNet e DeepLab, treinados em outras bases de imagens, em conjunto com técnicas de *transfer learning*, que possibilita a rede aprender melhores representações ainda que a base de dados seja pequena, e *data augmentation*, que auxilia a reduzir o erro de generalização da maioria dos modelos de visão computacional. Como resultados, espera-se aperfeiçoar o trabalho de Esgario et al. [2022] ao desenvolver um único modelo de rede neural convolucional capaz de segmentar e classificar esses sintomas, exprimindo a magnitude do espalhamento da doença, almejando reduzir significativamente o custo computacional do procedimento, sem perdas significativas na precisão do classificador, permitindo assim embarcar o modelo em um aplicativo de *smartphone* que possa identificar e estimar a severidade de doenças em plantas de café sem a necessidade de conexão com internet.

**Keywords:** Visão computacional, Inteligência artificial, Folha de café, Pragas e doenças.

---

## Referências

- José G.M. Esgario, Pedro B.C. de Castro, Lucas M. Tassis, and Renato A. Krohling. An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. *Information Processing in Agriculture*, 9(1):38–47, 2022. ISSN 22143173. doi: 10.1016/j.inpa.2021.01.004.
- Renato A. Krohling. Bracol - a brazilian arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests, 2019. URL <https://data.mendeley.com/datasets/yy2k5y8mxg/1>.



# A container-based environment for the parametric simulation of job schedulers

João Pedro Macleure Nunes dos Santos<sup>1</sup> and Antônio Tadeu Azevedo Gomes<sup>1</sup>

<sup>1</sup> *Laboratório Nacional de Computação Científica, Petrópolis/RJ, Brasil*

---

## Abstract

High performance computing (HPC) enables intensive computations, which allows simulations and data analyses in different areas of knowledge. Maximizing the throughput of jobs in HPC clusters accelerates science discoveries and improves resource utilization. Managing resources and job scheduling policies is key to achieve better usage of computational pools [Prabhakaran, 2016]. Because of that, resource management systems (RMSs) are deployed in production HPC clusters. Crucially, these systems allow flexible associations of processing nodes to job queues, being highly configurable in this respect. Such flexibility comes at the cost of difficulties in determining how to best fit resource usage and slowdown job waiting time in queues [Simakov et al., 2017]. Simulating the behavior of RMSs allows administrators to change and test configurations without impacting the production HPC cluster. In this work, we focus on Slurm, an open source, fault-tolerant, and highly scalable RMS [Yoo et al., 2003]. Slurm is used by many HPC infrastructures on the Top500 List (<http://top500.org>). A common approach to simulating Slurm is to adapt its source code, maintaining its overall functionalities [Jokanovic et al., 2018] and making the realization process a stochastic problem [Agrawala et al., 1976]. However, we believe that this approach is limited on the given data based on resources and queue configurations of specific HPC clusters. The capability of generating various queue configurations is useful in generating data for any simulation scenario [Barford and Crovella, 1998]. Container technologies allow software stacks to be delivered independent from the host machine, having minimal or no performance impact compared to bare metal applications [Torrez et al., 2019]. This work aims at proposing a container-based environment for the parametric simulation of RMS deployments, allowing various target queue configurations to be simulated through the execution of standard benchmark jobs. We expect that this work will help HPC providers to manage their computational pools for optimal usage and to plan the capacity of future deployments.

**Keywords:**Simulation, Job Scheduling, HPC

---



## Referências

- Ashok K Agrawala, Jeffrey M Mohr, and Raymond M Bryant. An approach to the workload characterization problem. *Computer*, 9(6):18–32, 1976.
- Paul Barford and Mark Crovella. Generating representative web workloads for network and server performance evaluation. In *Proceedings of the 1998 ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*, pages 151–160, 1998.
- Ana Jokanovic, Marco D’Amico, and Julita Corbalan. Evaluating Slurm simulator with real-machine Slurm and vice versa. In *2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)*, pages 72–82. IEEE, 2018.
- Suraj Prabhakaran. Dynamic resource management and job scheduling for high performance computing. Master’s thesis, Technische Universität Darmstadt, 2016.
- Nikolay A Simakov, Martins D Innus, Matthew D Jones, Robert L DeLeon, Joseph P White, Steven M Gallo, Abani K Patra, and Thomas R Furlani. A Slurm simulator: Implementation and parametric analysis. In *International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems*, pages 197–217. Springer, 2017.
- Alfred Torrez, Timothy Randles, and Reid Priedhorsky. HPC container runtimes have minimal or no performance impact. In *2019 IEEE/ACM International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE-HPC)*, pages 37–42. IEEE, 2019.
- Andy B Yoo, Morris A Jette, and Mark Grondona. Slurm: Simple linux utility for resource management. In *Workshop on job scheduling strategies for parallel processing*, pages 44–60. Springer, 2003.



# Serviços de Atualização no Gateway BioinfoPortal: Suporte ao Bancos de dados de Proveniência

Marco Cabral<sup>1</sup>, Marcelo Galheigo<sup>1</sup>, Antônio Gomes<sup>1</sup> e Kary Ocaña<sup>1</sup>

<sup>1</sup> *National Laboratory of Scientific Computing, Petrópolis/RJ, Brazil*

---

## Abstract

Os *gateways* científicos são ambientes centrados no usuário que integram diversas tecnologias, reunindo recursos avançados de computação, armazenamento, bancos de dados, sistemas de gerenciamento, sistemas de envio de trabalhos e software científico em ambiente *Web* de trabalho unificado. Os *gateways* incluem interfaces *Web*, *APIs* e *middleware* que fornecem acesso a *software* e dados [2]. A implementação dessas tecnologias depende amplamente do domínio científico ao qual um *gateway* oferece suporte. Neste trabalho, apresentamos um portal científico de bioinformática chamado BioinfoPortal [1], que está acoplado ao ambiente do supercomputador Santos Dumont (SDumont), principal recurso computacional do Sistema Nacional de Computação de Alto Desempenho (SINAPAD). O BioinfoPortal utiliza, via RESTful *Web Services*, o *middleware* CSGrid para permitir a extração, gerenciamento e processamento de dados em cada envio de trabalho. Este trabalho propõe a integração dos dados dos usuários, das execuções das aplicações e dos recursos da computação de alto desempenho (CAD) em um banco de dados de proveniência centralizado. Esses dados podem ser usados, por exemplo, para controlar o *cache*, detectar erros de transmissão, configurar um trabalho usando um aplicativo específico (por exemplo, o *software* de bioinformática RAxML), executar o trabalho com eficiência nos recursos CAD e recuperar resultados. O banco de dados foi projetado para ter dados de proveniência adicionais e serviços de gerenciamento de metadados. A implementação de serviços e do banco de dados atualizado permite melhorias no desempenho e funcionalidade do BioinfoPortal em termos de armazenamento, velocidade e funcionalidades relacionadas às interfaces de gerenciamento de arquivos, envio de trabalhos e contabilidade. A integração dos sistemas ao banco de dados centralizado visa tornar mais eficiente a coleta e o gerenciamento dos dados do BioinfoPortal.

**Keywords:** Computação Web e Aplicações, Computação Científica em Larga Escala, Bioinformática

---

**References**

- [1] K. Ocaña, M. Galheigo, C. Osthoff, L. Gadelha, F. Porto, A. Gomes, D. Oliveira, and A. Vasconcelos. BioinfoPortal: A scientific gateway for integrating bioinformatics applications on the Brazilian national HPC network. *FGCS*, 107:192–214, 2020.
- [2] M. Pierce, M. Miller, E. Brookes, M. Wong, E. Afgan, Y. Liu, S. Gesing, M. Dahan, T. Walker, and S. Marru. Towards a science gateway reference architecture. *CEUR Workshop Proc.*, 2357, 2019. 10th Int. Workshop on Science Gateways, IWSG 2018.



## Modelagem computacional de drenagem urbana: Estudo para sub-bacia do Jandiá em Macapá-AP.

Mauricio Dias Da Conceição Neto<sup>1</sup>, Luana Oliveira de Sá<sup>2</sup>

<sup>1</sup> Universidade Federal do Amapá (UNIFAP), Universidade de Marília SP (UNIMAR)

<sup>2</sup> Universidade Federal do Amapá (UNIFAP)

**Introdução:** Cerca de 30% da área da bacia amazônica é classificada como áreas de inundação episódica, periódica ou permanente [Junk et al. 2011]. A proteção dessas áreas é relevante para o abastecimento e manutenção da qualidade da água dos canais, onde os compartilhamentos terrestres e aquáticos estão relacionados. Mas estes ambientes urbanos amazônicos têm sido severamente impactados, tornando-se susceptíveis a alagamentos, exigindo a gestão de risco hidrológico pois causam danos econômicos à sociedade e ao meio ambiente [Miguez et al., 2018]. No campo da engenharia, o risco urbano está associado a eventos de chuva/marés [Souza et al., 2021]. **Objetivo:** Assim, o objetivo da pesquisa foi simular eventos hidrológicos extremos através da modelagem computacional utilizando os softwares de Modelo de Gestão de Drenagem Urbana - SWMM concomitantemente com o auxílio da ferramenta de geoprocessamento Qgis e avaliar os impactos em zona urbana crítica de Macapá. Isto é, se o uso e ocupação da sub-bacia do Jandiá é fator determinante de geração de episódios de alagamentos, especialmente sobre a infraestrutura próxima do canal. **Descrição da experiência:** As seguintes etapas metodológicas foram utilizadas no presente estudo: a) definição de fluxos de escoamento. Nesta etapa é fundamental o recorte do Modelo Digital do Terreno (MDT) (Rodrigues, 2021; Cordeiro 2019). (Figura 1); b) identificar a sub-bacia hidrográfica inserida na área e classificar a direção do escoamento na drenagem (segmentos de fluxo); c) delimitar o trecho da sub-bacia e o domínio espacial das quadras nelas contidas. **Resultados:** As simulações mostraram-se úteis para traçar perfis longitudinais de alagamentos críticos (“nós” críticos de transbordamento), abrangendo tanto o período chuvoso quanto no período seco. As simulações com o SWMM sugeriram que o transbordamento do Canal do Jandiá ocorre com impacto significativo na circunvizinhança durante todo o período anual (chuva e maré). Assim, a diretriz de ocupação do solo em Macapá não é suficiente para prevenir alagamentos, visto que estes significam alertas frequentes de riscos hidrológicos extremos evidenciados na simulação com o SWMM.

---

**Keywords:** Drenagem urbana, simulação computacional, SWMM, hidráulica.

---

## Referências

- [1] JUNK, W.J., et al., A classification of major naturally occurring Amazonian lowland wetlands. *Wetlands*, 31: 623–640, 2011.
- [2] MIGUEZ, M. G., et al. 2018. *Gestão de Riscos e Desastres Hidrológicos*, RJ. Elsevier, 2018.
- [3] SOUSA, T. S. Et al. Risco de alagamento influenciado por fatores ambientais em zonas urbanas de Macapá e Santana - AP. *Revista Ibero-americana de Ciências Ambientais*, v. 12, p. 245-259, 2021 *Amazônica*, 35(2):197-205.
- [4] RODRIGUES, GISELE; SANTINI JUNIOR, Maurício. Avaliação de técnicas de emprego compensatórias nas sub-bacia urbana Ribeirão do Santa Rita. *MONOGRAFIA ACADÊMICA, FERNANDOPÓLIS, SÃO PAULO*, p. 1-75, 10 mar. 2021.
- [5] CORDEIRO, ANDREW. Utilização do SWMM - STORM WATER MANAGEMENT MODEL como ferramenta de planejamento de drenagem urbana. *MONOGRAFIA ACADÊMICA, Caruaru*, p.1-75, 14 jan. 2019.



# Indentificação de Transientes astrofísicos Utilizando Redes Neurais Convolucionais

Phelipe Darc<sup>1</sup> and Clécio R. Bom<sup>1</sup>

<sup>1</sup> Centro Brasileiro de Pesquisa físicas, CBPF, RJ, Brasil

<sup>2</sup> Laboratório Nacional de Ciência da Computação, Petrópolis/RJ, Brasil

---

## Abstract

Neste trabalho é apresentado o uso de Redes Neurais Convolucionais para separar transientes astrofísicos de artefatos. Transientes são objetos astronômicos com luminosidade variável, como supernovas, kilonovas, entre outros. Identificar esses transientes no céu é historicamente feito por astrônomos treinados utilizando inspeção visual, facilitada pelo uso de algoritmos de diferenciação de imagem. Esses algoritmos subtraem de uma imagem contendo um candidato a transiente (*search*) uma imagem capturada anteriormente (*template*), procurando por locais de concentração de fluxo que poderiam indicar a presença de um transiente. Um dos problemas mais comuns desse método é o chamado *Bad-Subtraction*, que ocorre quando o *search* e *template* estão desalinhados, gerando artefatos que podem ser confundidos com transientes (falsos positivos). Para corrigir esse problema, usamos duas redes neurais convolucionais no processo de diferenciação (Shandonay et al. [2022]): a primeira tem como objetivo identificar as imagens que sofreram *Bad-Subtraction*, enquanto a segunda identifica os transientes nas imagens que sobraram. Com isso, nosso objetivo é automatizar a identificação de transientes no pipeline para busca de contrapartida ótica de ondas gravitacionais do Dark Energy Survey (DES-GW) e acelerar a detecção fazendo com que o número de imagens que requerem inspeção visual diminua. A primeira rede performou diminuindo em 95% a quantidade de *Bad-Subtraction*, e mantendo uma taxa de verdadeiros positivos de 98.3% enquanto que a segunda rede atuando em seguida obteve 94.5% de taxa de verdadeiros positivos. O que mostra a capacidade da nossa rede de filtrar artefatos sem perder Transientes.

**Keywords:** Deep Learning, CNN, Astronomia, Transientes, Astrofísica

---

## Referências

Adam Shandonay, Robert Morgan, Keith Bechtol, Clécio R Bom, Brian Nord, Alyssa Garcia, Ben Henghes, Kenneth Herner, Megan Tabbutt, Antonella Palmese, et al. Expediting decam multimessenger counterpart searches with convolutional neural networks. *The Astrophysical Journal*, 925(1):44, 2022.

Contato: Phelipe Darc, phelipedarc@gmail.com



# Uso de Modelagem Hidrológica visando a Mitigação dos Efeitos de Desastres Naturais

Renata Nalim Basilio Tissi<sup>1</sup>, Wagner Rambaldi Telles<sup>2</sup> e Antônio Silva Neto<sup>3</sup>

<sup>1</sup>*Instituto Federal Fluminense, Campos dos Goytacazes, RJ, Brasil*

<sup>2</sup>*Universidade Federal Fluminense, Santo Antônio de Pádua, RJ, Brasil*

<sup>3</sup>*Universidade do Estado de Rio de Janeiro, Nova Friburgo, RJ, Brasil*

---

## Resumo

O Brasil é um dos países com mais desastres de classificação hidrológica e, a nível nacional, a região Sul e Sudeste são as mais afetadas [SCHOEN et al., 2015]. Os desastres naturais envolvendo os recursos hídricos podem ocorrer, dentre outras maneiras, de forma espontânea. Esses tipos de desastres infelizmente não têm como parar, pois, é algo natural e, muitas vezes, o máximo que se há de conseguir é estimar, prever, seguir protocolos para minimizar os impactos. No entanto, as ações antrópicas não devem ser desconsideradas. Indiretamente essas ações poderão acelerar tais eventos na natureza. Por sofrer ações antrópicas discriminatórias e irresponsáveis, a natureza talvez responda de forma violenta e, em muitos casos, os prejuízos poderão se perpetuar por anos. Diante de um cenário de desastres envolvendo recursos hídricos que vem a cada dia se tornando mais frequentes, surgem a modelagem matemática e a simulação hidrológica com o propósito de tentar minimizar os impactos causados por tais eventos. Neste contexto, dentro dos softwares de modelagem computacional disponíveis para execução da modelagem hidrológica, há a plataforma MOHID, sistema de modelagem numérica tridimensional. Assim, o objetivo deste trabalho é o estudo dos efeitos decorrentes de desastres naturais. Embora se sabe que mesmo sendo a modelagem e a simulação ferramentas importantes no processo da mitigação, ambas sozinhas não serão suficientes para conter os desastres naturais. Considerando que não é possível eliminar por completo o risco de inundações, medidas devem ser previstas em todo o ciclo da gestão de risco [MACHADO et al., 2022]. Um fator importante e que pode fazer toda a diferença nesse processo, seja talvez a sensibilização da humanidade em prol da vida e manutenção do meio ambiente.

**Palavras-chave:** Recursos hídricos, MOHID, Modelo matemático

---

## Referências

- R. K. Machado; A. K. B. Oliveira; M. G. Miguez. Avaliação Dinâmica do Risco a Inundações através da Modelagem Computacional. Disponível em: <[https://files.abrhidro.org.br/Eventos/Trabalhos/189/XIV-ENAU\\_IV-SRRU0040-1-0-20220726-153904.pdf](https://files.abrhidro.org.br/Eventos/Trabalhos/189/XIV-ENAU_IV-SRRU0040-1-0-20220726-153904.pdf)>. Acesso em: 16 out. 2022.
- C. Schoen; G. A. Piazza; J. G. C. Sperb; K. L. Hermann; M. P. Serbent; A. Pinheiro. A inserção da variável social na modelagem de desastres naturais. XXI Simpósio Brasileiro de Recursos Hídricos. Brasília, DF (2015). Disponível em: <<https://files.abrhidro.org.br/Eventos/Trabalhos/4/PAP019348.pdf>>. Acesso em: 16 out. 2022.



# FLASH Radiotherapy RNA-Seq Data Mining

Samella Salles<sup>1</sup> and Kary Ocaña<sup>2</sup>

<sup>1</sup> *Laboratório Nacional de Computação Científica, Petrópolis/RJ, Brazil*

---

## Abstract

Radiotherapy is one of the major cancer treatment strategies. However, its efficacy is limited by the development of radioresistance and damage to normal tissues. Many efforts have been made to improve radiotherapy efficacy [Kim et al., 2019]. Recently a novel modality of radiation delivery has emerged. FLASH radiotherapy (FLASH-RT) allows for radiation delivery at ultra-high dose rates, which are much higher than conventional radiotherapy (CONV-RT). Studies have demonstrated a potential protective role of FLASH-RT in normal tissues, despite maintaining levels of tumor control at least equivalent to those of CONV-RT [Friedl et al., 2022]. Still, the mechanisms of FLASH-RT are not fully elucidated. This work aims to explore FLASH-RT by examining gene expression patterns with RNA-Seq data through the use of data mining techniques and High-Performance Computing (HPC). This research could provide a better comprehension of FLASH-RT biological mechanisms by uncovering its differential gene expression and activated pathways. FLASH-RT may become one of the main radiotherapy technologies in clinical practice in the future, being a promising new area of study [Lin et al., 2021].

**Keywords:** FLASH Radiotherapy, RNA-Seq, High-Performance Computing, Data Mining

---

## Referências

- Anna A Friedl, Kevin M Prise, Karl T Butterworth, Pierre Montay-Gruel, and Vincent Favaudon. Radiobiology of the flash effect. *Medical Physics*, 49(3):1993–2013, 2022.
- Wanyeon Kim, Sungmin Lee, Danbi Seo, Dain Kim, Kyeongmin Kim, EunGi Kim, JiHoon Kang, Ki Moon Seong, HyeSook Youn, and BuHyun Youn. Cellular stress responses in radiotherapy. *Cells*, 8(9):1105, 2019.
- Binwei Lin, Feng Gao, Yiwei Yang, Dai Wu, Yu Zhang, Gang Feng, Tangzhi Dai, and Xiaobo Du. Flash radiotherapy: history and future. *Frontiers in Oncology*, page 1890, 2021.

Contato: Samella Salles, [ssalles@posgrad.lncc.br](mailto:ssalles@posgrad.lncc.br)





# Um dispositivo de Biometria para a segurança residencial

Vicente Soares<sup>1</sup> and Aline Lunkes<sup>1</sup>

<sup>1</sup> *Instituto Politécnico do Rio de Janeiro - UERJ, Nova Friburgo/RJ, Brasil*

---

## Abstract

Uma das maiores preocupações em uma residência é a invasão por intrusos, assim, devemos encontrar soluções eficientes e baratas para a segurança. Porém, os dispositivos de monitoramento atuais são caros ou possuem uma grande dificuldade de implementação, ver Liu and Silverman [2001] e Hemalatha [2020]. Pensando nisso, criamos uma alternativa com baixo custo, eficiente, autossuficiente e de fácil instalação para o usuário. O projeto consiste em fabricar um leitor de biometria com Arduino, podendo ser colocado em qualquer portão elétrico, sem a necessidade de alterar o ambiente no qual ele será instalado. Além disso, o próprio usuário conseguirá fazer a instalação e a configuração do leitor de biometria. Pode, também, ser instalado em uma empresa de pequeno porte. Neste projeto utilizamos um recurso que consiste no cadastro instantâneo de pessoas. No primeiro momento cadastramos o usuário principal, como primeiro cadastro, em seguida, os demais usuários. A tela contém todas as informações necessárias para orientar o usuário. As vantagens do nosso leitor é o aparelho não possui acesso à internet e os códigos ficam gravados no Arduino e não tem como outro usuário, além do usuário principal, ter acesso ou modificar o código e o baixo custo tanto financeiro como de implementação e manutenção. Esse projeto ainda se encontra em um formato de prototipagem, o objetivo é transformá-lo em um produto, e a criação de um aplicativo de celular, gerando um controle remoto dos usuários ao acesso do portão e para auxiliar no cadastro, provendo mais informações das pessoas que possuem a digital cadastrada.

**Keywords:** Leitor de biometria, Arduino, Segurança

---

## Referências

- S Hemalatha. A systematic review on fingerprint based biometric authentication system. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–4, 2020. doi: 10.1109/ic-ETITE47903.2020.342.
- S. Liu and M. Silverman. A practical guide to biometric security technology. *IT Professional*, 3(1):27–32, 2001. doi: 10.1109/6294.899930.



# Human Skin Color Diversity with CUDAMCML

Victor Porto Gontijo de Lima<sup>1</sup> and Lilian Tan Moriyama<sup>1</sup>

<sup>1</sup> Instituto de Física de São Carlos - Universidade de São Paulo, São Carlos/SP, Brasil

---

## Abstract

In Biomedical Optical techniques, light interaction with biological tissues is used both for therapeutic and diagnostics purposes. In order to achieve the best outcomes, it is essential to understand how light propagates through the target tissue. In the last years, GPU-accelerated software for Monte Carlo simulations of photon transportation like CUDAMCML [Alerstam, 2009] have been shown to be powerful tools to study light propagation through turbid media, such as biological tissues. To perform simulations with CUDAMCML, one need to assign the number of layers, the optical properties (absorption coefficient, scattering coefficient, refractive index and anisotropy factor) and the thickness of each layer. For the human skin, the wavelength dependent absorption coefficient can be modeled by the melanin volume fraction (in the epidermis), the blood volume fraction (in the dermis), the melanin blend ratio and the blood oxygenation ratio [Donner and Jensen, 2006]. This work, thus, aims to evaluate if this modelling consistently reproduces human skin color diversity with CUDAMCML. The surface diffuse reflectance spectra output of a four-layered skin model were transformed to sRGB and CIELAB coordinates, which were then compared to colorimetric experimental measures reported in the litterature [Alaluf et al., 2002].

**Keywords:** Monte Carlo simulations, Skin Color, Colorimetry.

---

## Referências

Simon Alaluf, Derek Atkins, Karen Barrett, Margaret Blount, Nik Carter, and Alan Heath. The impact of epidermal melanin on objective measurements of human skin colour. *Pigment Cell Research*, 15(2):119–126, 2002.

Erik Alerstam. CUDAMCML: User manual and implementation notes. 2009.

Craig Donner and Henrik Wann Jensen. A spectral BSSRDF for shading human skin. *Eurographics Symposium on Rendering*, 2006.

Contato: Victor Porto, victorporto@ifsc.usp.br