

Universidade Federal do Rio de Janeiro  
Instituto de Matemática  
Departamento de Matemática Aplicada

*Relatório Final PIBIC -*  
**Modelagem de Dados Epidemiológicos por  
Equações Diferenciais Ordinárias Universais**

**Luan Lima Freitas**

Aluno do Bacharelado em Matemática Aplicada - UFRJ  
Bolsista PIBIC (04/2022 - 08/2022)

*Orientador:*

**Prof. Dr. Ricardo Rosa**

Departamento de Matemática Aplicada - IM/UFRJ

# **Conteúdo**

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Equações Diferenciais Ordinárias Universais</b>	<b>2</b>
<b>3</b>	<b>Dados e Modelos</b>	<b>3</b>
<b>4</b>	<b>Ajuste</b>	<b>4</b>
<b>5</b>	<b>Referências</b>	<b>6</b>

# 1 Introdução

Na vigência da bolsa de iniciação científica PIBIC (04/2022 - 08/2022), o aluno bolsista desenvolveu modelos e previsões para os dados da pandemia de COVID-19 na cidade do Rio de Janeiro, dando continuidade ao trabalho realizado durante a vigência da bolsa de iniciação científica INCTMat (10/2021 - 02/2022). Os modelos trabalhados consistiram de equações diferenciais ordinárias universais (UODEs) [1]. As UODEs se encontram no escopo do aprendizado científico de máquina (SciML), realizando a proposta de mesclar técnicas clássicas de modelagem matemática com recursos de aprendizado de máquina. Toda a parte computacional foi implementada no ecossistema de SciML da linguagem Julia. Este trabalho dá sequência às produções dos estudantes Gil Miranda [2] e Beatriz Farah [3] sob a orientação do professor Ricardo Rosa.

## 2 Equações Diferenciais Ordinárias Universais

Avanços recentes na área de aprendizado de máquina viabilizaram a utilização de técnicas de deep learning para a modelagem de fenômenos dos quais dispõe-se de grandes quantidades de dados. Uma vantagem desta abordagem é o aprendizado automático do conjunto de interações não-lineares, as quais podem ser tão complexas quanto numerosas. Por outro lado, esta estratégia é impraticável para a solução de problemas nos quais a disponibilidade de dados é exígua.

No extremo oposto do espectro encontram-se os modelos de equações diferenciais conhecidos como mecanicistas, na medida em que correspondem a uma tradução de fenômenos naturais (ou sociais) em um conjunto de leis explícitas, as quais representam a ação de mecanismos e exprimem um conhecimento consolidado pela literatura científica.

No intuito de conjugar as virtudes e mitigar as limitações de ambos os extremos surgem diversos modelos e métodos híbridos, dentre os quais estão as equações diferenciais ordinárias universais (UODEs), objeto do presente trabalho. Uma UODE é uma equação diferencial ordinária definida por

$$\mathbf{u}'(t) = f(\mathbf{u}, t, U_{\theta}(\mathbf{u}, t)),$$

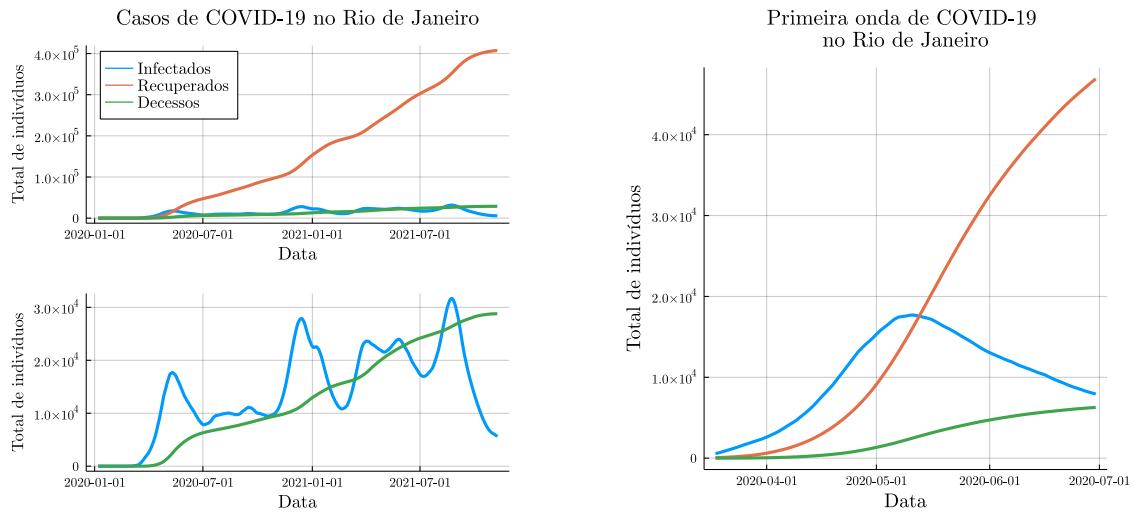
onde  $U_{\theta}$  é um aproximador universal, i.e., uma função capaz de aproximar qualquer função suficientemente regular. No nosso caso, o aproximador universal será dado por uma rede neural de vetor de pesos  $\theta$  que denotaremos por  $NN_{\theta}$ . A ideia por trás de uma UODE consiste em inscrever em uma estrutura a priori legada pela experiência científica um ou mais termos capazes de adquirir dos dados relações não-lineares potencialmente intrincadas e obscuras. Uma classe particular de UODE é formada pelas equações diferenciais ordinárias

neurais (NODEs) [4]. Uma NODE é uma UODE dada por

$$\mathbf{u}'(t) = NN_{\theta}(\mathbf{u}, t).$$

### 3 Dados e Modelos

Os dados disponibilizados na internet pela Secretaria Municipal de Saúde da Prefeitura do Rio de Janeiro foram manipulados para a obtenção das curvas de casos ativos, recuperados e mortos durante a pandemia de COVID-19 na cidade. Em seguida, foi selecionado um período correspondente a uma “onda” da pandemia para ser modelado por meio das técnicas mencionadas acima, qual seja, o período de 18/03/2020 a 30/06/2020 (cf. **Figura 1**).



**Figura 1:** Média móvel dos casos de COVID-19 na cidade do Rio de Janeiro.

Foi tomado como base dos experimentos o modelo compartmental SIRD. O modelo SIRD apresenta quatro compartimentos ( $S$  = “suscetíveis”,  $I$  = “infectados”,  $R$  = “recuperados” e  $D$  = “decessos”) e a evolução do número de casos em cada compartimento é dada pelo sistema de equações

$$\begin{cases} \dot{S} = -\beta \frac{I}{N} S \\ \dot{I} = \beta \frac{I}{N} S - (\gamma_R + \gamma_D) I \\ \dot{R} = \gamma_R I \\ \dot{D} = \gamma_D I \end{cases}$$

onde  $N = S + I + R + D$  e  $\{\beta, \gamma_I, \gamma_R\} \subset [0, +\infty)$ .

Tipicamente, a dinâmica mais árdua de se modelar em uma pandemia é a conversão dos indivíduos suscetíveis em infectados mediante a interação entre os dois grupos, no

nosso caso representada pelo *termo de infecção*  $\beta \frac{I}{N} S$ . As demais dinâmicas podem muito satisfatoriamente ser modeladas por termos lineares, havendo inclusive em alguns casos a possibilidade de estimar seus respectivos parâmetros por meio de investigações empíricas supplementares. Desta feita, propomos substituir inteira ou parcialmente o termo de infecção do modelo por redes neurais, dando assim gênese a UODEs. No nosso estudo foram contempladas duas possibilidades:

- Substituir o termo de infecção por  $NN_{\theta}(S/N, I)$ . A UODE assim definida foi nomeada de SIRD UODE  $\beta$ SI, onde a terminação  $\beta$ SI denota o termo substituído por uma rede neural.
- Substituir o termo de infecção por  $NN_{\theta}(S, I, R, D) \frac{I}{N} S$ . A UODE assim definida foi nomeada de SIRD UODE  $\beta$ , por motivo análogo ao especificado acima.

## 4 Ajuste

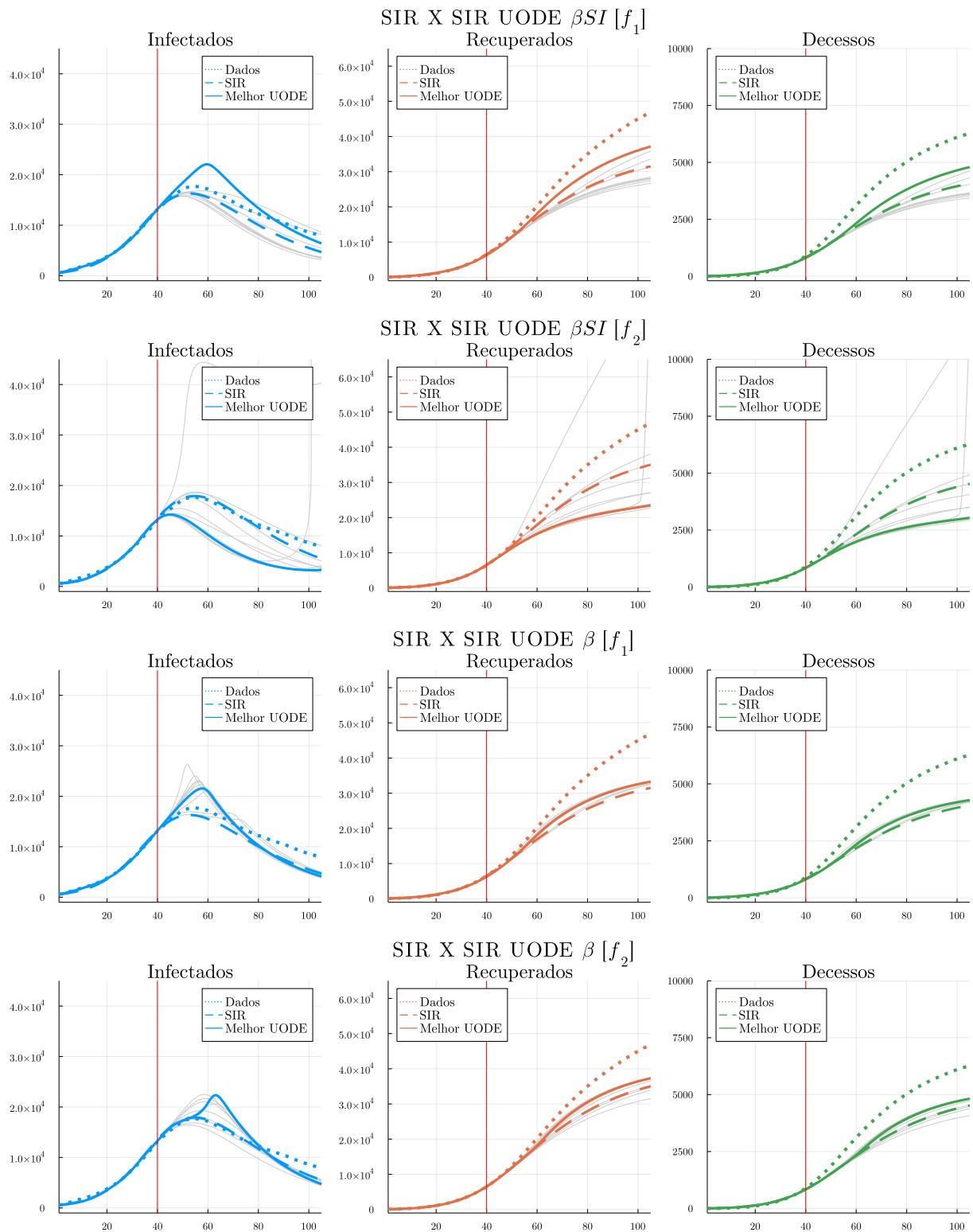
Os parâmetros dos modelos definidos acima foram ajustados aos dados coletados. Sejam  $I_i$ ,  $R_i$  e  $D_i$  os números de infectados, recuperados e decessos, respectivamente, no dia  $i$ . Além disso, fixada uma UODE  $U$  e dado um vetor de parâmetros para  $U$   $\mathbf{p}$ , sejam  $I_i(\mathbf{p})$ ,  $R_i(\mathbf{p})$  e  $D_i(\mathbf{p})$  os números de infectados, recuperados e decessos, respectivamente, determinados por  $U$  e  $\mathbf{p}$  no dia  $i$ . Os ajustes foram realizados minimizando-se a função objetivo

$$f_1(\mathbf{p}) = \sum_{i=1}^{40} [(I_i(\mathbf{p}) - I_i)^2 + (R_i(\mathbf{p}) - R_i)^2 + (D_i(\mathbf{p}) - D_i)^2]$$

através de um método de busca por mínimos locais para diferentes valores iniciais de  $\mathbf{p}$ . Observou-se que o modelos ajustados diferiram significativamente para valorores iniciais de  $\mathbf{p}$  diferentes. Os resultados do processo de otimização são reproduzidos abaixo, destacando-se o modelo ajustado para o qual a função objetivo atingiu seu menor valor (cf. **Figura 2**). Após o intervalo inicial de dias dedicado ao ajuste do modelo, é possível observar como cada modelo ajustado prevê a evolução subsequente da pandemia. Para efeito de comparação, também foi reproduzido o modelo SIRD ajustado.

Levando em consideração que as figuras numéricas de cada compartimento diferem substancialmente em escala, foi realizado um novo ajuste, minimizando-se uma função objetivo com pesos para cada compartimento dada por

$$f_2(\mathbf{p}) = \frac{\sum_{i=1}^{40} (I_i(\mathbf{p}) - I_i)^2}{\max_{i=1,\dots,40} I_i} + \frac{\sum_{i=1}^{40} (R_i(\mathbf{p}) - R_i)^2}{\max_{i=1,\dots,40} R_i} + \frac{\sum_{i=1}^{40} (D_i(\mathbf{p}) - D_i)^2}{\max_{i=1,\dots,40} D_i}$$



**Figura 1:** Média móvel dos casos de COVID-19 na cidade do Rio de Janeiro. Média móvel dos casos de COVID-19 na cidade do Rio de Janeiro. Média móvel dos casos de COVID-19 na cidade do Rio de Janeiro. Média móvel dos casos de COVID-19 na cidade do Rio de Janeiro.

## 5 Referências

- [1] <https://landscan.ornl.gov>
- [2] <https://github.com/ModSiming/EpiSiming>