

# Compartmental Epidemiological Models for Covid-19: Sources of Uncertainty, Goodness-of-Fit and Goodness-of-Projections

Rafael da Silva Fernandes

**Abstract**—Compartmental Epidemiological Models have been widely used to describe the dynamics of infectious transmission diseases. These models describe the temporal progression of case counts related to an epidemic and evaluate hypotheses about the possible underlying mechanisms that explain patterns in the observed data, with distinct spatial and temporal scales. Furthermore, the models can vary in complexity, whether in terms of the number of parameters or the number of variables that characterize epidemiological states. Therefore, this approach is essential to select adequate models that characterize better transmission dynamics. This generates reliable estimates of the main parameters, which is a critical factor in minimizing uncertainty. This study aims to apply the main epidemiological compartmental models to the initial data of Covid-19 in the city of Parauapebas / PA - Brazil. A model selection criterion is proposed that, in addition to taking into account the quality of fit, mechanisms are added that capture information about past forecasts. Such a proposal contributes by providing subsidies for an adequate selection of the best model in order to enable more precise characterization of the transmission dynamics and with less loss of information.

**Index Terms**—Covid-19, Compartmental Epidemiological Models, Forecasting Epidemics, Goodness-of-fit, Goodness-of-projections, Parameter Estimation, Sources of Uncertainty.

## I. INTRODUÇÃO

A sociedade tem sido marcada por acontecimentos históricos que desafiam a sobrevivência humana. Atualmente, enfrenta-se uma nova doença infecciosa viral chamada COVID-19 (*Coronavirus Disease 2019*), descoberta em dezembro de 2019 na China [1] causada por um vírus chamado Coronavirus SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirys 2 of the Genus Betacoronavirus*). Este vírus se disseminou para mais de 160 países em pouco menos de 3 meses, sendo que no Brasil o primeiro caso foi confirmado em 26 de fevereiro de 2020, e em menos de um mês foram registrados 2,4 mil infecções. 2 meses após, o Brasil já apresentava quase 62 mil casos confirmados de infecção, 4.205 mortes e 30.152 mil recuperados [2].

Neste contexto, diversos trabalhos científicos têm sido realizados para compreensão desta pandemia. Dentre eles, boa parte se trata de modelagem matemática, principalmente, com o intuito de realizar previsões futuras quanto ao número de infectados e de óbitos. Tal preocupação, remete à premissa de

que o sistema de saúde poderia colapsar nos municípios e, com isso, acarretar o aumento do número óbitos, devido à falta de assistência hospitalar.

Os modelos compartimentais epidemiológicos analisam a disseminação de uma doença infecciosa por meio de uma população. Tais modelos variam em sua complexidade quanto ao número de variáveis e de parâmetros que caracterizam os estados dinâmicos do sistema. Essa estruturação é crítica, pois além de definir os compartimentos, implica também em quantificar os fatores que contribuem para o processo de transmissão. Logo, surge a importância em selecionar uma estrutura que melhor represente a realidade.

Usualmente, são usadas métricas de desempenho, por exemplo, o *Root Mean Square Error (RMSE)*, como critério para seleção de modelos [3]. No entanto, pode haver uma incerteza substancial sobre essa seleção e, para tanto, é apresentada em [4] uma abordagem baseada na informação perdida de um modelo, quando este é usado para descrever a realidade. Trata-se de uma abordagem que se baseia na qualidade de ajuste e tem como vantagem quantificar evidências sobre a perda de informação e probabilidade do modelo ser suportado pelos dados.

Frente ao exposto, este trabalho propõe um critério de seleção de modelos que, além da qualidade de ajuste, adiciona-se uma avaliação da qualidade das projeções passadas. Projeções passadas remetem às previsões realizadas num período anterior e que, posteriormente, são avaliadas. Especificamente, incorpora-se ao critério de seleção apresentado por [4], mecanismos capazes de fornecer informações dessas projeções passadas. Além disso, como forma de melhorar as condições de previsão, mecanismos subjacentes, que, identificam características de assinatura de padrões de crescimento epidêmico, são adicionados ao modelo compartimental geral.

Este trabalho contribui para o entendimento amplo sobre a dinâmica da transmissão da Covid-19, avaliando a progressão ou o reajuste ao longo do tempo, bem como quantifica a informação e seleciona o modelo com menor perda de informação dentre os modelos analisados.

Este trabalho apresenta ainda as fontes de incertezas associadas aos modelos compartimentais e discute, dentro da proposta apresentada, como modelos com complexidades diferentes são suportados pelos dados.

A principal motivação deste trabalho é dada pela dificuldade em determinar o modelo compartimental mais adequado para descrever a pandemia do Covid-19. Tal adequação remete a um *trade off* entre o nível de complexidade do modelo e a capacidade de parametrizar de forma confiável o modelo com os dados disponíveis, isto é, se ao incrementar parâmetros e

variáveis se obtêm melhores *insights* sobre a dinâmica de transmissão de doenças infecciosas, em contrapartida, isto pode acarretar a falta de identificabilidade dos parâmetros, resultante da falta de informações do fenômeno para estruturação do modelo.

O restante deste documento está organizado da seguinte forma. Na Seção II, sintetiza-se os principais modelos compartimentais; na Seção III as fontes de incertezas associadas aos dados e parâmetros são apresentadas. Na Seção IV, introduz a proposta do critério de seleção de modelos baseada em previsões passadas. Na Seção V é apresentado os resultados, análises pertinentes e demais discussões. Por fim, na Seção VI, as considerações finais.

II. MODELOS COMPARTIMENTAIS EPIDEMIOLÓGICOS

A modelagem matemática epidemiológica é bastante utilizada para auxiliar os responsáveis por um sistema de saúde em tomar decisões mais acertadas a respeito de uma pandemia. A aplicação destes modelos em diferentes etapas fornece características sobre a evolução da dinâmica de transmissão. Uma vez compreendido a dinâmica de transmissão, pode-se então realizar uma projeção de curto ou longo prazo.

Neste trabalho, apresenta-se os modelos compartimentais baseados em sistemas de equações diferenciais ordinárias. Tais modelos foram sintetizados em uma classe de modelos, podendo assim, variar sua complexidade com a presença ou não dos compartimentos e parâmetros. Tratam-se de modelos clássicos encontrados em diversos trabalhos na literatura, como [5], [6], [7].

A. O Modelo SEAIR

O modelo SEAIR representa o modelo compartimental epidemiológico que abrange um conjunto amplo de parâmetros e compartimentos e foi baseado no modelo SIAS, apresentado por [8], [9], [10] e [11].

O modelo geral aqui apresentado é subdividido nos seguintes compartimentos:

- (**S**) – **Suscetíveis**: que representam os indivíduos que são suscetíveis de serem infectados;
- (**E**) – **Expostos**: são indivíduos infectados, mas não infectantes ou não ativos e remete ao período de incubação do vírus no indivíduo;
- (**A**) – **Assintomáticos**: são indivíduos infectados, mas que não apresentam sintomas. Porém, ainda podem ser infectantes;
- (**I**) – **Infectados**: indivíduos que estão infectados com a doença, apresentam os sintomas e são capazes de transmitir a indivíduos suscetíveis;
- (**D**) – **Óbitos**: Indivíduos infectados e que morreram em decorrência da doença;
- (**R**) – **Recuperados**: Indivíduos que se recuperaram da doença e se tornam imunes.

Tem-se que a população  $N = S + E + A + I + D + R$  e o número de indivíduos em cada compartimento no tempo  $t$  deve ser inteiro, de modo que o modelo possa ser descrito pelo seguinte sistema de equações diferenciais ordinárias:

$$\begin{cases} \frac{dN}{dt} = N(t) - \frac{\delta}{1-\delta} \gamma I(t) \\ \frac{dS}{dt} = -(\beta I^\alpha(t) + \mu A^\eta(t)) \frac{S(t)}{N(t)} \\ \frac{dE}{dt} = (\beta I^\alpha(t) + \mu A^\eta(t)) \frac{S(t)}{N(t)} - \kappa E(t) \\ \frac{dA}{dt} = (1-\rho)\kappa E(t) - \sigma A(t) \\ \frac{dI}{dt} = \rho \kappa E(t) - \frac{\gamma}{(1-\delta)} I(t) \\ \frac{dD}{dt} = \frac{\delta}{1-\delta} \gamma I(t) \\ \frac{dR}{dt} = \gamma I(t) + \sigma A(t) \end{cases} \quad (1)$$

Os parâmetros  $\beta$  e  $\mu$  representam as **taxas de transmissão da doença**, que são as taxas médias em que ocorrem contatos entre dois indivíduos quaisquer. Sendo que  $\beta$  é a taxa para infectados e  $\mu$  a taxa para assintomático.

Os parâmetros  $\gamma$  e  $\sigma$  são chamados de **taxa de recuperação**, e representa a taxa com que um indivíduo infectado ou assintomático, respectivamente, é recuperado.

O parâmetro  $\delta$  é a **probabilidade de mortalidade**, isto é, a probabilidade de um indivíduo infectado ir a óbito devido à doença, antes mesmo de se recuperar.

O parâmetro  $\kappa$  representa a taxa de mudança de exposto para infeccioso, determinado pelo **período de incubação do vírus** no indivíduo.

Frequentemente, quando se aplicam modelos matemáticos com o intuito de realizar previsões de epidemias, pressupõe-se que, a fase inicial de crescimento, segue uma dinâmica de crescimento exponencial. Essa suposição pode ser conveniente, pois ajuda a orientação sobre o tipo e a intensidade das intervenções, incluindo necessidades de infraestrutura de saúde para diagnóstico, isolamento de indivíduos infecciosos e atividades de rastreamento de contatos [12].

Assim, como forma de refinar o modelo e possibilitar a caracterização da dinâmica de crescimento subexponencial, durante a fase inicial de uma de doença infecciosa, dois mecanismos podem ser incorporados ao sistema de equações. O primeiro se refere à premissa de mistura homogênea da população, logo, é incorporado um **parâmetro de lei de potência  $\alpha$  e  $\eta$**  como expoente, fazendo  $I(t) \rightarrow I^\alpha(t)$  e  $A(t) \rightarrow A^\eta(t)$ . Para  $\alpha = \eta = 1$ , trata-se de um caso específico, onde configura uma mistura homogênea e,

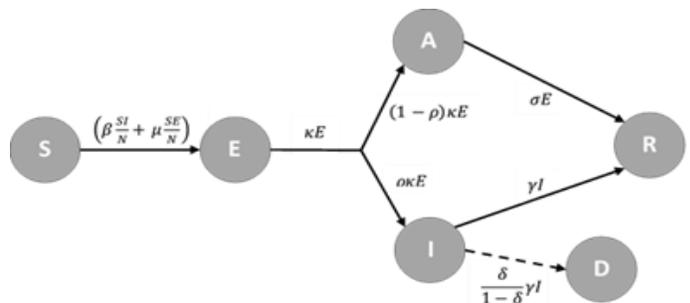


Fig. 1. Diagrama Compartimental do Modelo SEAIR.

para  $\alpha = \eta < 1$ , indica uma mistura não homogênea da população e que induz a um crescimento mais lento (subexponencial) [12]. O segundo mecanismo, baseado na premissa do efeito crescente das mudanças de comportamento da população e das intervenções de controle que ocorrem na população, modela o comportamento dinâmico por meio de uma taxa de transmissão  $\beta$  dependente do tempo, fazendo  $\beta \rightarrow \beta(t)$ .

O diagrama da representa a dinâmica do modelo *SEAIDR*. Neste aspecto, uma proporção  $\rho$  dos indivíduos Expostos migram para o compartimento dos Infectados, de modo que uma proporção de  $(1 - \rho)$  para o compartimento dos Assintomáticos. Posteriormente, indivíduos migram dos compartimentos dos Infectados e Assintomáticos a uma taxa de recuperação  $\gamma$  e  $\sigma$ , respectivamente, de maneira que parte dos Infectados pode ir a óbito a uma proporção de  $\gamma\delta/(1 - \delta)$ , migrando para o compartimento dos Óbitos.

Por fim, cabe mencionar que o modelo *SEAIDR* pode abarcar outros parâmetros, tornando-o um modelo ainda mais amplo. Tais parâmetros poderiam representar a dinâmica de vitalidade, por exemplo, conforme mostrado em [11].

### B. Modelos Restritos

Os modelos restritos representam casos específicos do modelo *SEAIDR* podendo ser tratado como uma classe ou família de modelos. Desta forma, sua restrição ou complexidade se faz pela quantidade de parâmetros e compartimentos presentes, sendo o modelo *SEAIDR* o mais amplo, contendo todos os parâmetros e compartimentos, e o modelo *SIR* o mais restrito, contendo menos parâmetros e menos compartimentos.

O **modelo *SIR*** é o modelo compartimental mais simples apresentado na literatura considerando uma população constante e que, após a infecção, os indivíduos recuperados ficam imunes à novas infecções. Inicialmente publicado por [7], o modelo *SIR* foi o ponto de partida para estudos das propriedades das epidemias. Observa-se que o modelo *SIR* é um caso específico do modelo *SEAIDR* composto pelos parâmetros  $\beta, \alpha$  e  $\gamma$  podendo variar e os demais como constantes.

Aspecto importante a observar, é que o parâmetro  $\kappa$  é comumente dado por  $1/T_i$  e  $T_i$  é o tempo médio de incubação, logo, não haveria tempo de incubação para o modelo *SIR*, fazendo com a dinâmica de transmissão seja pela migração do indivíduo do compartimento de Suscetíveis para o de Infectados diretamente. Na ausência do compartimento de expostos, como o modelo *SAIR* e *SAIDR*, o parâmetro  $\rho$  não é incorporado,

TABELA 1  
MODELOS RESTRITOS E SUAS COMPLEXIDADES

Modelos	$N_c$	$N_p$	$\beta$	$\alpha$	$\mu$	$\eta$	$\kappa$	$\rho$	$\sigma$	$\gamma$	$\delta$
<i>SIR</i>	3	3	1	1	0	0	0	0	0	1	0
<i>SIDR</i>	4	4	1	1	0	0	0	0	0	1	1
<i>SEIR</i>	4	4	1	1	0	0	1	0	0	1	0
<i>SEIDR</i>	5	5	1	1	0	0	1	0	0	1	1
<i>SAIR</i>	4	6	1	1	1	1	0	0	1	1	0
<i>SAIDR</i>	5	7	1	1	1	1	0	0	1	1	1
<i>SEAIDR</i>	6	9	1	1	1	1	1	1	1	1	1

uma vez que, a incidência dos casos ocorre diretamente nos

compartimentos (*A*) e (*I*) respectivamente.

Desta forma, ao variar a complexidade do modelo *SEAIDR* pelo número de parâmetros ( $N_p$ ) e compartimentos ( $N_c$ ) envolvidos, pode-se apresentar o conjunto de modelos pela

Tabela 1. Neste caso, o número 1 (um) representa a presença do parâmetro no modelo e 0 (zero) sua ausência.

### III. FONTES DE INCERTEZAS

Um dos maiores desafios da modelagem matemática é a estimação dos parâmetros. Uma primeira desvantagem, é que, se um determinado conjunto de dados possui apenas uma quantidade finita de informações, logo, cada vez que uma estimativa do parâmetro é feita, a informação restante é perdida. Outra desvantagem se trata do número de parâmetros envolvidos, acarretando numa estimação conjunta dos parâmetros. Assim, um método de estimação que ajusta os parâmetros aos dados amostrais com uma melhor qualidade é crucial para uma caracterização mais acertada. Por fim, uma terceira desvantagem, trata-se do sistema de equações que descrevem os modelos compartimentais, como é o caso dos modelos supracitados, que são resolvidos numericamente por não possuírem solução analítica.

Além disso, se os dados não representarem o fenômeno em estudo de forma confiável, qualquer estimação dos parâmetros bem feita será inútil.

#### A. Dados Utilizados para Estimação

O conjunto dos dados que consiste este trabalho é referente ao número de notificados infectados, óbitos e recuperados apresentados no boletins informativos oficiais pelo município de Parauapebas/PA – Brasil, por meio do site [13], cuja publicação ocorre diariamente. Os dados podem ser visualizados na Fig. 2 e compreendem o período entre 28/03/2020 a 31/05/2020, totalizando 65 dados notificados. Estes dados representam o primeiro estágio de uma curva epidêmica, caracterizado pelo rápido crescimento inicial do número de infectados.

Devido à preocupação por parte dos sistemas de saúde aos números de infectados e óbitos, sabe-se que a informação do número de recuperados geralmente não é levada em consideração. No Brasil, por exemplo, os dados informados por

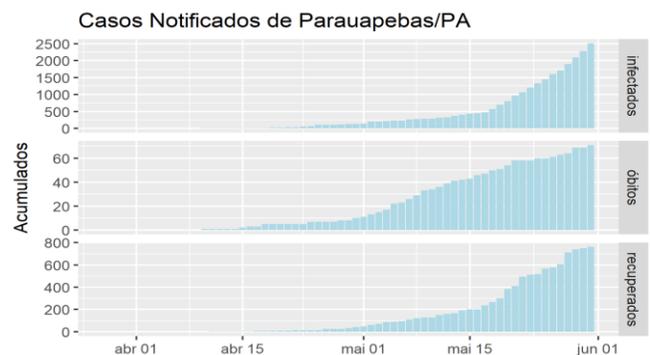


Fig. 2. Gráficos dos números acumulados de casos notificados de infectados, óbitos e recuperados no município de Parauapebas – Brasil.

[14], do número de indivíduos recuperados, foram inicialmente lançados em 19/04/2020, 52 dias após o primeiro caso confirmado.

Para o modelo *SEAIR* ainda é um desafio conseguir os dados referentes aos estados epidemiológicos:

- **dos suscetíveis**, devido ao desconhecimento quanto à auto imunidade de uma parcela da população e à premissa de uma população totalmente suscetível;
- **dos expostos**, que devido à incubação do vírus no indivíduo, este, por sua vez, não procura o sistema de saúde para realização dos testes;
- **dos assintomáticos**, que, devido ao baixo número de testes realizados no Brasil, gera uma subnotificação dos dados. Esta situação é bastante aceita no cenário atual, sendo que, sua quantificação é difícil e agravada pelos indivíduos assintomáticos, estes que, não procuram o sistema de saúde para realização dos testes.

#### B. Identificabilidade dos Parâmetros

Uma questão importante, ao se analisar a aplicação de diferentes modelos matemáticos a um mesmo conjunto de dados, é saber se os parâmetros são identificáveis a partir dos dados disponíveis. Assim, um parâmetro é identificável quando seu intervalo de confiança está em uma faixa finita de valores [3] e seu comportamento é estável entre os modelos. A estabilidade é avaliada quando os parâmetros não se alteram significativamente com a variação da complexidade do modelo.

Na prática, um parâmetro pode não ser identificável devido a dois fatores principais: a quantidade e qualidade dos dados disponíveis ou o número de parâmetros estimados em conjunto a partir dos dados disponíveis. Conforme apresentado em [3], este problema pode ser corrigido:

- a. Pelo emprego de um modelo alternativo de menor complexidade;
- b. Coletando mais dados sobre outros estados do sistema para caracterizar melhor os recursos dinâmicos do sistema;
- c. Acrescentando mais observações aos dados para restringir melhor os parâmetros do modelo e/ou;
- d. Reduzir o número de parâmetros estimados em conjunto.

Desta forma, para o item (a) acima, este trabalho analisa o impacto de diferentes modelos, conforme apresentado na Seção II e

Tabela 1. Para o item (b), é acrescentado os dados do número de recuperados. O item (c), corresponde a proposição deste trabalho a ser apresentado na Seção IV, neste ponto, interessado em avaliar a inserção de novos dados. E para o item (d) é apresentado, na subseção a seguir, uma estratégia para a estimação dos parâmetros.

#### C. Método de Estimação dos Parâmetros

Uma maneira adequada de reduzir a estimação conjunta dos parâmetros, conforme tratada em [16], é ajustar os modelos de menor complexidade aos dados originais e fazer inferência multimodal adequada. Esta estratégia é útil, pois permite compreender a perda de indenticabilidade do parâmetro com

o aumento da complexidade e facilita a compreensão das informações nos dados.

Neste trabalho, os parâmetros são estimados por meio da curva não linear do quadrado mínimo, ajustada aos dados observados. Para esse fim, é utilizado o algoritmo de Levenberg-Marquardt, publicado originalmente por [16] e aperfeiçoado por [17]. As estimações são realizadas usando o pacote *minpack.lm* do *Software R*.

#### IV. CRITÉRIOS PARA SELEÇÃO DE MODELOS

Sob o pressuposto de um fenômeno epidêmico, diferentes modelos compartimentais podem assumir diferentes formas. Deve-se utilizar algum critério para seleção do modelo mais adequado ao fenômeno em estudo, principalmente quanto aos problemas advindos da não identificabilidade prática dos parâmetros envolvidos.

##### A. Critério de Seleção de Modelos Baseado em Qualidade de Ajuste.

Baseando-se na qualidade o ajuste, os autores em [4], descrevem uma abordagem formal, fundamentalmente sólida, de desenvolvimento de um critério de seleção de modelos que quantifica a evidência de cada modelo aplicado em um único conjunto de dados. O procedimento consiste no escalonamento da *Root Mean Square Error (RMSE)* e é formulado da seguinte maneira:

$$AIC_r = 2 \times \log(RMSE) + 2N_p + \frac{2N_p(N_p + 1)}{(n - N_p - 1)} \quad (2)$$

$AIC_r$  representa o Critério de Informação de Akaike com correção de viés em estudos comportamentais e quantifica a perda de informação do  $r$ -ésimo modelo em relação aos dados observados,  $n$  representa o tamanho do conjunto dos dados e  $N_p$  o número de parâmetros do modelo. Operacionalmente, ao calcular os  $AIC_r$  para cada um dos  $r$  modelos, seleciona-se o modelo com menor valor de  $AIC_r$ . Conforme [4], como forma de obter um critério em escala contínua de informações e que, podem ser interpretados independentes da escala de medição, é introduzido o critério de seleção que classifica o modelo de acordo com a perda de informação  $AIC_r$  em relação ao melhor modelo, assim temos:

$$\Delta_r = AIC_r - AIC_{\text{mínimo}} \quad (3)$$

$\Delta_r$  representa o índice da perda de informação do modelo  $r$  em relação ao melhor modelo  $AIC_{\text{mínimo}}$ , ou seja, um valor muito alto representa um modelo que perde muita informação sobre a realidade em relação aos demais modelos.

A probabilidade relativa do modelo  $r$ , dado o conjunto de dados, fornece uma força formal de evidência dada pela quantidade  $l_r = \exp(-\Delta_r/2)$  e permite declarações de evidência, como, por exemplo, um modelo é tantas vezes melhor suportado pelos dados que outro modelo.

Ainda pode-se determinar a probabilidade de cada modelo  $r$ , dado os dados e os modelos, e é formulada como:

$$W_r = \frac{\exp\left(-\frac{\Delta_r}{2}\right)}{\sum_{r=1}^R \exp\left(-\frac{\Delta_r}{2}\right)} \quad (4)$$

Assim interpretamos  $W_r$  como a probabilidade relativa do modelo, o que permite declarações de evidência em que um modelo é mais provável que outro, em relação aos dados. Essa equação fornece um valor de  $W_r$  dentro do intervalo  $[0,1]$ , podendo o  $r$ -ésimo modelo assumir qualquer valor no intervalo. Pelo fato de que cada  $W_r$  ser uma partição do espaço que contém todos os possíveis valores da qualidade do ajuste,  $\sum_{r=1}^R W_r = 1$ .

Finalmente, pode-se ainda tomar as razões das probabilidades do modelo para quaisquer pares de modelos, assim, a evidência quantificada é representada tanto pelas probabilidades do modelo quanto pela razão de evidência.

#### B. Critério de Seleção de Modelos Baseado em Previsões Passadas.

Na incerteza quanto às previsões epidêmicas, é comum selecionar modelos que melhor se ajustaram aos dados conforme apresentação anterior. No entanto, durante a ocorrência de uma epidemia, os modelos são sempre reajustados a cada nova inserção de dados. Pressupõe-se, então, que, num determinado tempo passado  $t^-$ , seja possível avaliar como o modelo se comportou aos dados posteriores.

À medida que mais dados da fase inicial de crescimento são empregados para estimação dos parâmetros, conforme [3], a incerteza nas estimativas dos parâmetros não é apenas reduzida, mas as estimativas de parâmetros são mais restritas em torno de seus valores reais. No entanto, com a inserção de novos compartimentos, pode ocorrer o efeito contrário pois, devido à falta de dados dos estados epidemiológicos, os parâmetros podem ser calibrados gerando uma variabilidade alta.

Desta forma, analisando-se todas as previsões realizadas no passado e a qualidade de suas projeções, será então possível gerar um vetor de erros ao longo do tempo de reajuste. Pretende-se com este mecanismo, diminuir a incerteza dos parâmetros para um dado modelo, conforme já mencionado na Seção II. Reaplicando este procedimento a todos os modelos apresentados na

Tabela 1, pretende-se aqui gerar um mecanismo de comparação dos modelos, isto é, aplica-se o procedimento de comparação adotado na seção anterior; através das equações (2), (3) e (4), podendo assim avaliar o modelo que melhor se ajustou ao fenômeno observado ao longo do tempo. Logo, teremos um critério para determinar o melhor modelo.

Podemos descrever, então, uma estrutura capaz fornecer informações das projeções passadas, conforme descritas nos passos a seguir:

**Passo 1:** Determina-se uma data passada  $t^- > t_0$  em que se queira avaliar a evolução do modelo ao longo do tempo. Seja  $t = t_1, \dots, t_n$  os tempos diários da notificação dos dados e  $t_n$  o valor do último dado notificado, a diferença  $t_n - t^- = k$  representa a quantidade de avaliações das previsões passadas.

Um valor  $t^-$  para os primeiros casos de infecção é ainda prematura para conferir o status de epidemia. Um valor de  $t^-$

igual a  $t_n$  será incapaz de avaliar as projeções passadas do modelo. Portanto, recomenda-se um valor de  $1 < t^- < t_n$ .

**Passo 2:** Seja  $R$  o conjunto da classe de modelos apresentada na

Tabela 1, calibra-se os respectivos parâmetros de cada modelo  $r \in R$  em cada unidade tempo  $k \in [t^-, t_n)$ . A informação gerada será denominada **RMSE<sub>fit</sub>** ou **RMSE** ajustado e dada pelo vetor coluna  $e_f^r$ .

**Passo 3:** De forma análoga ao Passo 2, calcula-se o quanto as projeções passadas, em cada período  $k$ , foi capaz de prever os dados observados posteriores até  $t_n$ . Essa informação será denominada **RMSE<sub>proj</sub>** ou **RMSE** projetado e dada pelo vetor coluna  $e_p^r$ .

**Passo 4:** A Matriz **RMSE** ( $RMSE_{k \times r}$ ) será formada pela junção dos vetores coluna  $e^r = e_f^r + e_p^r$  para cada modelo  $r$  e seus elementos serão formados por  $e_k^r$ , que representa a soma do **RMSE<sub>fit</sub>** e **RMSE<sub>proj</sub>** para o  $r$ -ésimo modelo no  $k$ -ésimo período.

**Passo 5:** Utilizando as equações (2), (3) e (4) geram-se a Matriz de Informações  $AIC_{k \times r}$ , Matriz de Índices  $\Delta_{k \times r}$  e Matriz de Evidências  $W_{k \times r}$ .

**Passo 6:** Por fim, como forma de avaliar o comportamento dos  $N_p$  parâmetros, ao longo dos  $k$  períodos e para os  $r$  modelos, podemos assim gerar  $N_p$  matrizes de parâmetros, denominada Matriz de Parâmetros  $P_{r \times k} \forall N_p$ . Em particular, estamos interessados nos parâmetros que podem ser estimados diretamente dos dados disponíveis.

Assim, o modelo que apresentar evidências mais fortes ao longo dos períodos avaliados, será considerado o melhor dentre os modelos analisados, dado o conjunto de dados.

Particularmente, a proposta apresentada inclui dois mecanismos para captar a informação das projeções, sendo: a inclusão da medida de desempenho das projeções  $e_p^r$  e que visa quantificar a perda de informação após o ajuste e, a inclusão da reaplicação do procedimento nos  $k$  períodos envolvidos, visando, assim, restringir melhor os parâmetros do modelo e avaliar as premissas subjacentes à inserção de novos dados.

Observe que as premissas subjacentes à fase de crescimento inicial de uma epidemia, diz respeito ao crescimento do **RMSE<sub>fit</sub>** e o decrescimento do **RMSE<sub>proj</sub>** ao longo do tempo.

## V. RESULTADOS E DISCUSSÕES

Nesta etapa da pesquisa já sabemos da busca dos modelos compartimentais epidemiológicos em caracterizar a dinâmica de transmissão de doenças infecciosas. Para tanto, é realizado um estudo de caso aplicando diferentes modelos a um mesmo conjunto de dados observados.

TABELA 2  
NÚMERO DE PARÂMETROS,  $RMSE$  E OS CRITÉRIOS DE SELEÇÃO  $AIC_r$ ,  $\Delta_r$  E  $W_r$   
PARA CADA MODELO APLICADO

Modelos	$N_p$	$RMSE$	$AIC_r$	$\Delta_r$	$W_r$
<i>SIR</i>	3	42.67109	494.3513	31.44193	0.000000
<i>SIDR</i>	4	41.78921	493.9096	31.00029	0.000000
<i>SEIR</i>	4	41.17510	491.9850	29.07571	0.000000
<i>SEIDR</i>	5	32.33322	462.9093	0.000000	<b>0.993478</b>
<i>SAIR</i>	6	52.11917	527.4075	64.49820	0.000000
<i>SAIDR</i>	7	33.62803	472.9617	10.05237	0.006521
<i>SEAIDR</i>	9	38.98866	497.4979	34.58861	0.000000

O cenário do estudo de caso foi escolhido diante da proximidade ao sistema de saúde e do auxílio dado às implementações e medidas de controle adotadas pelo município.

#### A. Análises Preliminares

Inicialmente faz-se necessário aplicar os modelos (Tabela 1) ao conjunto total dos dados coletados (ver Fig. 2) para ter um entendimento preliminar e compreender as evidências quantificadas para os diferentes modelos quanto ao ajuste. A análise preliminar consiste em avaliar, a partir dos critérios de seleção discutidas na subseção IV.A, como as estratégias para minimizar as fontes de incertezas se relacionam com as complexidades dos modelos, número de parâmetros e número de compartimentos.

Para tanto, calibra-se os parâmetros dos  $r$  modelos aos dados notificados. O número de infectados inicial é  $I_0 = 1$  e adota-se a premissa de uma população totalmente suscetível com  $S_0 = N - I_0$ . Outros valores utilizados e predeterminados foram, os parâmetros  $\kappa = 1/T_i$  e  $T_i = 7$  dias,  $\rho = 1/6$  e  $6$  é a estimativa da proporção de subnotificados e  $\sigma = 1/T_t$  e  $T_t = 14$  dias. Observa-se na Tabela 2 que o valor de  $AIC_r$  indica que o modelo *SEIDR* é o melhor no sentido de minimizar a perda de informações em relação aos dados observados.

Os modelos com  $\Delta_r > 14$  possuem evidências fracas e não possuem essencialmente nenhum suporte empírico, conforme discutido em [4]. No entanto, é necessário utilizar valores de acordo com o fenômeno em estudo e, assim, é possível estabelecer três subgrupos de modelos, sendo aqueles com:

- $\Delta_r > 20$ : São os modelos *SIR*, *SIDR*, *SEIR*, *SAIR*, *SEAIDR* e não têm essencialmente nenhum suporte empírico. Os modelos *SIR*, *SEIR* e *SAIR* perdem informações com a ausência do compartimento ( $D$ ) que é contrária a premissa de utilização de mais dados sobre os estados epidêmicos. Estes modelos também são de menor complexidade de parâmetros e, apesar de corroborar para minimização das incertezas, ainda sim perderam informações.
- $10 < \Delta_r < 20$ : Trata-se do modelo *SAIDR*. Nota-se imediatamente a presença do compartimento ( $D$ ), no entanto, a presença do compartimento ( $A$ ) faz aumentar a perda de informações. Neste caso, tais modelos tem relativamente pouco suporte empírico e as análises tornam-se inconclusivas. Pode-se observar a mesma relação do compartimento ( $A$ ) nos modelos *SAIR* – *SEAIDR*.
- $\Delta_r = 0$ : Representa o melhor modelo de acordo com o critério  $AIC_r$ . Neste caso, torna-se o modelo base para as análises anteriores, ou seja, a perda de informações dos demais modelos, ocorre em relação ao melhor modelo. Logo,

é possível inferir que o modelo *SEIDR* perde menos informações, o que reforça com a presença do compartimento ( $D$ ) e a ausência do compartimento ( $A$ ). Além do mais, a presença do compartimento ( $E$ ), ao comparar *SIR* – *SEIR* e *SIDR* – *SEIDR*, resulta na melhora do valor de  $RMSE$ .

Outra quantidade que permite declarações de evidências, para um conjunto de dados, é a força de evidência. Podemos, por exemplo, querer comparar o melhor modelo *SEIDR* com o modelo *SAIDR*, fazendo o seguinte cálculo:

$$l_{SEIDR} = \exp\left(-\frac{1}{2} \times (AIC_{SEIDR} - AIC_{SAIDR})\right)$$

Assim é possível declarar que o modelo *SEIDR* é 152.35 vezes melhor suportado pelos dados do que o segundo melhor modelo *SAIDR*. Além do mais, o cálculo de  $W_r$  permite afirmar que a probabilidade de *SEIDR* ser suportado pelos dados é  $P(SEIDR|dados) = 0.9935$ , enquanto que  $P(SAIDR|dados) = 0.0065$ .

Por fim, se a evidência para cada modelo, aplicado ao conjunto de dados analisados, pode ser quantificada usando  $AIC_r$ ,  $\Delta_r$  e  $W_r$ , então, pelo critério de seleção de modelos baseado no ajuste, temos evidências fortes que o modelo *SEIDR* é o mais adequado, dentre os modelos analisados.

É importante fazer uma ressalva pois, o critério de seleção discutido, pode sugerir que o modelo *SEIDR* se ajusta melhor aos dados, o que é incorreto conforme sinalizado em [4]. Modelos com mais parâmetros tendem a se ajustar melhor aos dados e o conceito de parcimônia entra aqui. Além do que, para o contexto de modelos compartimentais epidemiológicos, e com o intuito de evitar que esta falha ocorra, é necessário analisar as premissas subjacentes do modelo e as fontes de incertezas. Logo, dado que os modelos *SEIDR*, *SAIDR* e *SEAIDR* abarcam os 3 conjuntos de dados analisados, o conceito de parcimônia se fez presente, indicando como melhor, dentre estes, o modelo com menor complexidade.

#### B. Análise da Qualidade de Previsões Passadas

O procedimento utilizado para a análise preliminar possibilitou compreender a importância de ser criterioso ao selecionar o melhor modelo. Como a quantificação das evidências apresentadas fornece um arcabouço melhor para tais análises, podemos, então, reaplicar este procedimento em períodos passados, e assim, avaliar a progressão da dinâmica de transmissão ao longo do tempo.

Foi apresentado anteriormente que modelos aplicados numa quantidade maior de dados representativos dos estados epidemiológicos, geram evidências mais fortes, além do que, corrobora com as estratégias para diminuir o problema de não indentificabilidade prática dos parâmetros, ou seja, diminuir as fontes de incertezas dos parâmetros.

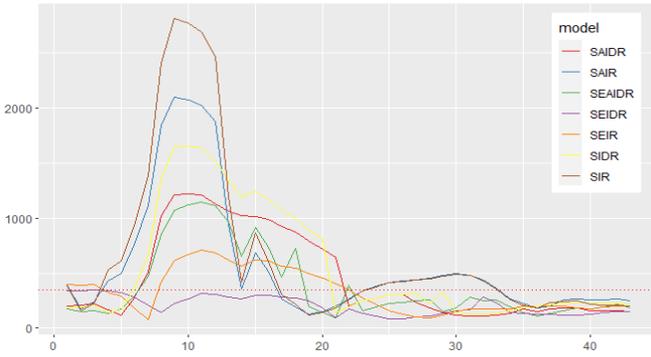


Fig. 3. Gráfico de linhas das matrizes  $RMSE_{k \times r}$ ,  $AIC_{k \times r}$ ,  $\Delta_{k \times r}$  e  $W_{k \times r}$ .

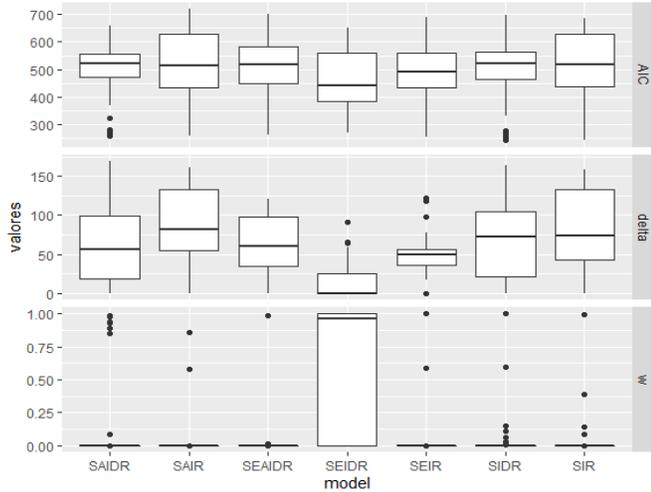


Fig. 4. Gráfico box-plot das matrizes  $AIC_{k \times r}$ ,  $\Delta_{k \times r}$  e  $W_{k \times r}$ .

A partir deste ponto, pode-se aplicar a proposta deste trabalho ao incluir à qualidade das projeções passadas executando os 6 passos apresentados na subseção IV.B, sendo o valor de  $t^- = 22$  e  $k = 43$ . No entanto, será omitido as matrizes  $RMSE_{k \times r}$ ,  $AIC_{k \times r}$ ,  $\Delta_{k \times r}$  e  $W_{k \times r}$ .

Para a matriz  $RMSE_{k \times r}$ , observa-se na Fig. 3, que o modelo *SEIDR* está predominantemente abaixo de um  $RMSE < 350$  (ver linha pontilhada vermelha), enquanto que para os demais modelos, curva tem seu valor  $RMSE > 500$  em algum momento. Estes valores de  $RMSE$  indicam um efeito de deterioração das projeções, ou seja, estes modelos realizaram projeções muito acima ou abaixo dos valores posteriormente observados, principalmente no intervalo entre  $5 \leq k < 15$ , em que ocorre um pico na maioria dos modelos.

Devido ao grande número de períodos, um gráfico box-plot é mais bem aproveitado para realizar uma análise destas matrizes, conforme podemos observar na Fig. 4. É notório a predominância do modelo *SEIDR* que se mantém mais estável ao longo dos  $k$  períodos analisados. O box-plot para a matriz  $AIC_{k \times r}$  indica que o modelo perde, levemente, menos informações em média que os demais modelos. O box-plot para a matriz  $\Delta_{k \times r}$ , indica que o modelo *SEIDR* foi classificado mais vezes como melhor opção, neste caso mais próximo do valor zero. A Matriz  $W_{k \times r}$ , que indica a probabilidade do modelo ser suportado, claramente indica o modelo *SEIDR* como o mais provável dentre os modelos.

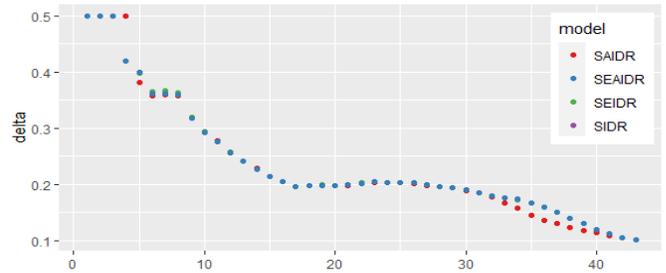


Fig. 5. Gráfico de linhas para o comportamento do parâmetro  $\delta$  ao longo dos  $k$  períodos analisados referente aos modelos com o compartimento ( $D$ ).

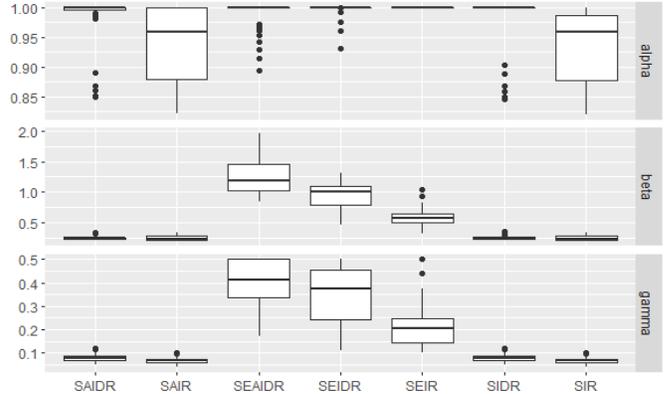


Fig. 6. Gráfico box-plot da matriz de parâmetros de cada modelo.

Por fim, ainda se faz necessário avaliar o comportamento dos parâmetros ao longo dos  $k$  períodos. Estamos interessados especificamente nos parâmetros  $\beta, \alpha, \gamma$  e  $\delta$ , pois estes estão relacionados diretamente aos dados analisados, sendo os demais parâmetros provenientes de valores predeterminados.

Para análise dos parâmetros  $\beta, \alpha, \gamma$ , mais uma vez recorre-se ao gráfico de box-plot, conforme Fig. 6, ao qual destaca-se inicialmente que os parâmetros  $\beta$  e  $\gamma$  tem suas estimativas alteradas nos modelos que possuem a presença do compartimento ( $E$ ) e, para o parâmetro  $\delta$ , utiliza-se o gráfico de pontos que, conforme a Fig. 5, indica uma estabilidade entre os modelos e seu valor decresce ao longo do período.

É razoável acreditar que o parâmetro  $\gamma$  esteja incorporando parte do parâmetro  $\kappa$ . Tal fato pode ser confirmado ao calcular o tempo médio de transmissão  $T_t = 1/\gamma$ , logo, para o modelo *SEIR* tem-se  $\bar{T}_t \cong 4.3$  e para *SEIDR* e *SEAIDR*  $\bar{T}_t \cong 2.65$ , enquanto que, para os modelos com a ausência do compartimento ( $E$ ), tem-se valores compatíveis com a literatura com um  $\bar{T}_t$  em torno de 12 a 16 dias.

Para o parâmetro  $\beta$ , este aumento de valor ao adicionar o compartimento ( $E$ ), representa uma dinâmica de crescimento inicial mais acelerado para estes modelos, o que, além de gerar perda de informação devido à falta de informação do estado epidemiológico, caracteriza falta de identificabilidade prática para este parâmetro.

Para o parâmetro  $\alpha$  interessa-se em capturar assinatura de mistura não homogênea da população, ao qual foi identificada apenas pelos modelos *SIR*, *SAIR*. Tais modelos possuem em comum a ausência dos compartimentos ( $E$ ) e ( $D$ ), que são compartimentos que melhoraram em geral as evidências

analisadas. Ainda assim, esta alteração dos parâmetros indica não identificabilidade prática.

## VI. CONSIDERAÇÕES FINAIS

Neste trabalho, foi proposto um critério de seleção de modelos que, além da qualidade de ajuste, adicionou-se mecanismos que capturam informações sobre as previsões passadas. Sob o pressuposto de que o aumento na complexidade de um modelo pode gerar melhor suporte dos dados analisados e, em contrapartida, pode incorrer na falta de identificabilidade prática dos parâmetros, resultante da falta de dados disponíveis para análise, o critério apresentado mostrou-se válido em avaliar conjuntamente este *trade off*. Desta forma, ao construir matrizes de evidências quantificadas, tornou-se capaz de fornecer subsídios para selecionar o melhor modelo. Tal seleção é crítica, pois o melhor modelo fornece informações melhores e, assim, caracterizam melhor a evolução da dinâmica de transmissão da epidemia ao longo do tempo.

A utilização de um modelo global, dado pelo modelo *SAEIDR*, a princípio pode parecer desvantajoso, principalmente devido à falta de precisão dos parâmetros estimados, uma vez que estes são estimados separadamente e a informação restante é reduzida. No entanto, ao avaliar modelos aninhados restritivamente por seus parâmetros, esta abordagem torna-se útil e facilita a compreensão das informações nos dados.

Além disso, os mecanismos adicionados, tais como o  $RMSE_{proj}$  e reaplicação aos  $k$  períodos, tratam-se de estratégias que visam a corrigir os problemas advindos da não identificabilidade prática dos parâmetros. No entanto, mesmo com este cuidado, os parâmetros  $\beta, \alpha, \gamma$  apresentaram este fenômeno, principalmente devido à presença do compartimento ( $E$ ) e da respectiva falta dos seus dados para validação. Essa análise é crucial, pois demonstra haver uma incerteza considerável na seleção do melhor modelo baseado nos dados, isto é, se este conjunto de dados fosse reamostrado por *bootstrap*, por exemplo, pode ocorrer de outro modelo ser estimado como o melhor. Apesar dos autores em [16] mencionarem que essa conclusão refere-se às probabilidades do modelo, dada pela matriz  $W_{k \times r}$ , que sejam menores, por exemplo, abaixo de 50%, no contexto deste trabalho essa análise torna-se relevante. Uma alternativa para contornar este problema pode ser visto em [3], pelo qual é descrito uma abordagem paramétrica de *bootstrap* para quantificar a incerteza dos parâmetros e construir intervalos de confiança. Essa análise encontra-se fora de escopo deste trabalho, ficando como proposta futura.

Por fim, esforços na melhoria da modelagem da transmissão de doenças são úteis para o entendimento da evolução da dinâmica de transmissão de uma epidemia, principalmente no atual contexto da Covid-19, preocupada em realizar previsões mais acertadas. Tal preocupação, remete à premissa de que o sistema de saúde poderia colapsar nos municípios e, com isso, acarretar no aumento do número óbitos devido à falta de assistência hospitalar.

## REFERÊNCIAS

- [1] F. Wu, S. Ahao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes e Y.-Z. Zhang, "A New Coronavirus associated with human respiratory disease in China," *Nature*, pp. 265-269, 2020.
- [2] M. d. S. d. Brasil, "Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil pelo Ministério da Saúde," 2020. [Online]. Available: <https://covid.saude.gov.br/>.
- [3] G. Chowell, "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts," *Infectious Disease Modelling*, vol. 2, no 3, pp. 379-398, 2017.
- [4] K. P. Burnham, D. R. Anderson e K. P. Huyvaert, "AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons," *Behavioral Ecology and Sociobiology*, vol. 65, pp. 23-35, 2011.
- [5] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no 4, pp. 599-653, 2000.
- [6] "Mathematical models of infectious disease transmission," *Nature Reviews Microbiology*, vol. 6, pp. 477-487, 2008.
- [7] W. O. Kermack e A. G. McKendrick, "A Contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London Series A Containing Papers of a Mathematical and Physical Character*, vol. 115, no 772, pp. 700-721, 1987.
- [8] S. B. Bastos e D. O. Cajueiro, "Modeling and forecasting the Covid-19 pandemic in Brazil," Universidade de Brasília, Brasília, 2020.
- [9] J. Arino, P. V.-D. Driessche, J. Watmough e J. Wu, "How will country-based mitigation measures influence the course of the covid-19 epidemic?," *Journal of Theoretical Biology*, vol. 253, pp. 118-130, 2008.
- [10] M. Longini-Jr, M. E. Halloran, A. Nizam e Y. Yang, "Containing pandemic influenza with antiviral agents," *American Journal of Epidemiology*, vol. 159, pp. 623-633, 2004.
- [11] M. Robinson e N. I. Stilianakis, "A model for the emergence of drug resistance in the presence of asymptomatic infections," *Mathematical Biosciences*, vol. 243, pp. 163-177, 2013.
- [12] G. Chowell, L. Sattenspiel, S. Bansal e C. Viboud, "Mathematical models to characterize early epidemic growth: A review," *Physics of Life Reviews*, vol. 18, pp. 66-97, 2016.
- [13] P. d. Parauapebas, "Parauapebas contra o Coronavírus," Prefeitura Municipal de Parauapebas, [Online]. Available: <https://parauapebas.pa.gov.br/index.php/parauapebas-contra-o-coronavirus.html>.
- [14] M. d. S. d. Brasil, "Coronavírus Brasil," Ministério da Saúde do Brasil, [Online]. Available: <https://covid.saude.gov.br/>.
- [15] A. M. Silva, C. P. Zingano, J. P. Zingano e P. R. Zingano, "Observações sobre previsões da evolução da Covid-19 por modelos matemáticos, II," 01 06 2020. [Online]. Available: <https://www.ufrgs.br/ime/observacoes-sobre-previsoes-da-evolucao-da-covid-19-por-modelos-matematicos-ii/>.
- [16] K. Levenberg, "A Method fo the Solution of Certain Non-Linear Problems in Least Squares," *Quarterly of Applied Mathematics*, vol. 2, pp. 164-168, 1944.
- [17] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal on Applied Mathematics*, vol. 11, pp. 431-441, 1963.
- [18] K. P. Burnham e D. R. Anderson, "Multimodel Inference: Understanding AIC and BIC in Model Seleccion," *Sociological Methods & Research*, vol. 33, no 2, pp. 261-304, 2004.
- [19] G. Chowell e C. Viboud, "Is it growing exponentially fast? - Impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics," *Infectious Disease Modelling*, vol. 1, pp. 71-78, 2016.
- [20] M. J. Keeling e P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton: Princeton University Press, 2011.

- [21] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk e R. M. Eggo, "Early dynamics of transmission and control of COVID-19: a mathematical modelling study," *Lancet Infectious Diseases*, vol. 20, pp. 553-558, 2020.
- [22] T. A. Brubaker e K. R. O'Keefe, "Nonlinear Parameter Estimation," *Analytical Chemistry*, vol. 51, no 13, pp. 1385-1388, 1979.
- [23] D. J. Daley e J. Gani, *Epidemic Modelling: An Introduction*, Cambridge: Cambridge Studies in Mathematical Biology, 2001.
- [24] W.-H. LI e K. MAKOVA, "Least Square Analysis," em *Brenner's Online Encyclopedia of Genetics*, Academic Press, 2001, pp. 14106-14110.
- [25] A. Zellner e H. Theil, "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, vol. 1962, pp. 54-78, 1962.
- [26] L. J. S. Allen e A. M. Burgin, "Comparison of deterministic and stochastic SIS and SIR models in discrete time," *Mathematical Biosciences*, vol. 163, pp. 1-33, 2000.
- [27] M. H. Holmes, *Introduction to Numerical Methods in Differential Equations*, New York: Springer-Verlag, 2000.
- [28] S. A. Mulaik, L. R. James, J. V. Alstine, N. Bennett, S. Lind e D. C. Stilwell, "Evaluation of Goodness-of-Fit Indices for Structural Equation Models," *Psychological Bulletin*, vol. 3, no 105, pp. 430-445, 1989.
- [29] M. A. Capistrán, M. A. Moreles e B. Lara, "Parameter Estimation of Some Epidemic Models. The Case of Recurrent Epidemics Caused by Respiratory Sincytial Virus," *Bulletin of Mathematical Biology*, vol. 71, pp. 1890-1901, 2009.
- [30] P. R. Almeida, "Modelos Epidêmicos SIR, Contínuos e Discretos e Estratégias de Vacinação," Universidade Federal de Viçosa, Viçosa, 2014.
- [31] J. Nocedal e S. J. Wright, *Numerical Optimization*, New York: Springer, 2006.
- [32] M. Y. Li, J. R. Graef, L. Wang e J. Karsai, "Global dynamics of a SEIR model with varying total population size," *Mathematical Biosciences*, nº 160, pp. 191-213, 1999.
- [33] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang e J. Shaman, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-Cov2)," *Science*, vol. 368, pp. 489-493, 2020.
- [34] C. Viboud, L. Simonsen e G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks," *Epidemics*, vol. 15, pp. 27-37, 2016.
- [35] G. Chowell, D. H. Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen e C. Viboud, "Using Phenomenological Models to Characterize Transmissibility and Forecasting Patterns and Final Burden of Zika Epidemics," *PLOS Currents Outbreaks*, no 1, 2016.
- [36] S. Vaid, C. Cakan e M. Bhandari, "Using machine learning to estimate unobserved COVID-19 infections in North America," *The Journal of Bone & Joint Surgery*, no 13, 2020.
- [37] J. Hellewell, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russel, J. D. Munday e A. J. Kucharski, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *The Lancet Global Health*, vol. 8, nº 4, pp. e488-e496, 2020.
- [38] M. d. S. Brasil, "Concluída a maior pesquisa brasileira sobre a COVID-19," Ministério da Saúde, 02 07 2020. [Online]. Available: <https://www.saude.gov.br/noticias/agencia-saude/47155-concluida-a-maior-pesquisa-brasileira-sobre-a-covid-19>. [Acesso em 31 07 2020].
- [39] "Instituto Brasileiro de Geografia e Estatística," IBGE, 2019. [Online]. Available: <https://www.ibge.gov.br/cidades-estados/pa/parauapebas.html>. [Acesso em 31 07 2020].
- [40] B. Efron, *The jackknife, the bootstrap, and other resampling plans.*, Filadélfia: SIAM, 1982.
- [41] A. Mellan, H. H. Hoeltgebaum, S. Mishra e C. Whittaker, "Estimating COVID-19 cases and reproduction number in Brazil," *Imperial College London*, p. Report #21, 2020.



**Rafael S. Fernandes** Doutor em Engenharia de Produção pela Universidade Federal de Minas Gerais (UFMG), e atualmente professor adjunto em Engenharia de Produção na Universidade Federal Rural da Amazônia (UFRA) *campus* Parauapebas. Possui projetos de pesquisa na área de Confiabilidade, Gestão da Qualidade e Pesquisa Operacional.