

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

**Predição de um Modelo SIR com Taxa de
Infecção Variante no Tempo para Diferentes
Períodos**

por

Walter Mendes Haselein

Tese submetida como requisito parcial
para a obtenção do título de
Doutor em Matemática Aplicada

Prof. Dr. Diego Eckhard
Orientador

Porto Alegre, 30 de janeiro de 2025

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Mendes Haselein, Walter

Predição de um Modelo *SIR* com Taxa de Infecção Variante no Tempo para Diferentes Períodos / Walter Mendes Haselein.—Porto Alegre: PPGMAp da UFRGS, 2025.

112 p.: il.

Tese (doutorado)— Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada, Porto Alegre, 2025.

Orientador: Eckhard, Diego

Tese: Matemática Aplicada: Sistemas Não-Lineares,
Dissertação, Tese, Mestrado, Doutorado

Predição de um Modelo *SIR* com Taxa de Infecção Variante no Tempo para Diferentes Períodos

por

Walter Mendes Haselein
Tese submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de

Doutor em Matemática Aplicada

Linha de Pesquisa: Sistemas Não-Lineares

Orientador: Prof. Dr. Diego Eckhard

Banca examinadora:

Prof. Dr. João Frederico da Costa Azevedo Meyer
IMECC - UNICAMP

Prof. Dr. Alexandre Molter
UFPel

Profa. Dra. Cláudia Pio Ferreira
IBB - UNESP

Tese apresentada e aprovada em
30 de outubro de 2024.

Prof. Dr. Lucas da Silva Oliveira
Coordenador

Eu não quero acreditar. Eu quero saber.

Carl Sagan

AGRADECIMENTOS

Agradeço à minha família, aos meus professores e colegas, aos servidores do Instituto de Matemática e Estatística.

SUMÁRIO

LISTA DE FIGURAS	xiii
LISTA DE TABELAS	xvii
LISTA DE SIGLAS E SÍMBOLOS	xix
RESUMO	xxiii
ABSTRACT	xxv
1 INTRODUÇÃO	1
2 MODELOS MATEMÁTICOS PARA EPIDEMIOLOGIA	13
2.1 Modelo <i>SIS</i>	14
2.1.1 Solução Analítica	16
2.2 Modelo <i>SIR</i>	17
2.2.1 Solução Analítica	18
2.3 Modelo <i>SEIR</i>	19
2.3.1 Solução Analítica	20
3 MÉTODOS NUMÉRICOS	23
3.1 Métodos Numéricos para Equações Diferenciais Ordinárias	24
3.1.1 Método de Euler	26
3.1.2 Método de Runge-Kutta de Quarta Ordem	27
3.2 Métodos Numéricos para Otimização	29

3.2.1	Método de Newton	30
3.2.2	Método Simplex de Nelder-Mead	32
4	MODELO SIR COM β VARIANTE NO TEMPO	37
4.1	Simulação do Modelo	38
4.2	Identificação do Modelo	40
4.2.1	Dados	42
4.2.2	Otimização	45
4.2.3	Algoritmo de Otimização	46
4.3	Resultados da Identificação para as Zonas de Agrupamento	54
4.3.1	Identificação para R22 - Bagé	55
4.3.2	Identificação para R27 - Cachoeira do Sul	56
4.3.3	Identificação para R08 - Canoas	57
4.3.4	Identificação para R04,05 - Capão da Canoa	59
4.3.5	Identificação para R23, 24, 25, 26 - Caxias do Sul	60
4.3.6	Identificação para R12 - Cruz Alta	62
4.3.7	Identificação para R16 - Erechim	63
4.3.8	Identificação para R09 - Guaíba	65
4.3.9	Identificação para R13 - Ijuí	66
4.3.10	Identificação para R29,30 - Lajeado	67
4.3.11	Identificação para R07 - Novo Hamburgo	69

4.3.12	Identificação para R15, 20 - Palmeira das Missões	70
4.3.13	Identificação para R17, 18, 19 - Passo Fundo	71
4.3.14	Identificação para R21 - Pelotas	73
4.3.15	Identificação para R10 - Porto Alegre	74
4.3.16	Identificação para R28 - Santa Cruz	76
4.3.17	Identificação para R01,02 - Santa Maria	77
4.3.18	Identificação para R14 - Santa Rosa	79
4.3.19	Identificação para R11 - Santo Ângelo	81
4.3.20	Identificação para R06 - Taquara	82
4.3.21	Identificação para R03 - Uruguaiana	84
4.4	Resultado da Identificação para o Estado	86
5	PREDIÇÃO DO MODELO	91
5.1	Resultados	92
5.1.1	Predição para 7 dias ($T = 7$)	93
5.1.2	Predição para 14 dias ($T = 14$)	94
5.1.3	Predição para 30 dias ($T = 30$)	96
5.1.4	Predição para 60 dias ($T = 60$)	97
5.1.5	Predição para 180 dias ($T = 180$)	98
6	CONCLUSÃO	101
6.1	Trabalhos Futuros	103

6.1.1	Identificação de uma Taxa de Remoção Variante no Tempo . .	103
6.1.2	Modelagem com Vacinação	104
6.1.3	Técnicas de Controle	104
REFERÊNCIAS BIBLIOGRÁFICAS		112

LISTA DE FIGURAS

2.1	Modelo <i>SIS</i> com $\beta = 0.5$ e $\gamma = 0.1$	16
2.2	Modelo <i>SIR</i> com $\beta = 0.5$ e $\gamma = 0.1$	18
2.3	Modelo <i>SEIR</i> com $\beta = 0.5$, $\gamma = 0.1$ e $\mu = 0.7$	20
4.1	Simulação do modelo <i>SIR</i>	40
4.2	Número de casos diários no RS	44
4.3	Casos Acumulados no RS	44
4.4	Modelo <i>SIR</i> para R22 - Bagé	55
4.5	Taxas de Infecção para R22	56
4.6	Modelo <i>SIR</i> para R27 - Cachoeira do Sul	56
4.7	Taxas de infecção para R27	57
4.8	Modelo <i>SIR</i> para R08 - Canoas	58
4.9	Taxas de Infecção para R08	58
4.10	Modelo <i>SIR</i> para R04 e R05 - Capão da Canoa	59
4.11	Taxas de Infecção para R04 e R05	60
4.12	Modelo <i>SIR</i> para R23, R24, R25 e R26 - Caxias do Sul	61
4.13	Taxas de Infecção para R23, R24, R25 e R26	62
4.14	Modelo <i>SIR</i> para R12 - Cruz Alta	62
4.15	Taxas de Infecção para R12	63

4.16	Modelo SIR para R16 - Erechim	64
4.17	Taxas de Infecção para R16	64
4.18	Modelo SIR para R09 - Guaíba	65
4.19	Taxas de Infecção para R09	66
4.20	Modelo SIR para R13 - Ijuí	66
4.21	Taxas de Infecção para R13	67
4.22	Modelo SIR para R29 e R30 - Lajeado	68
4.23	Taxas de Infecção para R29 e R30	68
4.24	Modelo SIR para R07 - Novo Hamburgo	69
4.25	Taxas de Infecção para R07	70
4.26	Modelo SIR para R15 e R20 - Palmeira das Missões	71
4.27	Taxas de Infecção para R15 e R20	71
4.28	Modelo SIR para R17, R18 e R19 - Passo Fundo	72
4.29	Taxas de Infecção para R17, R18 e R19	73
4.30	Modelo SIR para R21 - Pelotas	73
4.31	Taxas de Infecção para R21	74
4.32	Modelo SIR para R10 - Porto Alegre	75
4.33	Taxas de Infecção para R10	76
4.34	Modelo SIR para R28 - Santa Cruz	77
4.35	Taxas de Infecção para R28	77

4.36	Modelo SIR para R01 e R02 - Santa Maria	78
4.37	Taxas de Infecção para R01 e R02 - Santa Maria	79
4.38	Modelo SIR para R14 - Santa Rosa	80
4.39	Taxas de Infecção para R14	80
4.40	Modelo SIR para R11 - Santo Ângelo	81
4.41	Taxas de Infecção para R11	82
4.42	Modelo SIR para R06 - Taquara	83
4.43	Taxas de Infecção para R06	83
4.44	Modelo SIR para R03 - Uruguaiana	84
4.45	Taxas de Infecção para R03	85
4.46	Comportamento Real x Simulado	87
4.47	Taxas de Infecção diárias para o RS	87
4.48	Simulação do Modelo com β variante no tempo	88
5.1	Erros percentuais para 7 dias no futuro	94
5.2	Erros percentuais para 14 dias no futuro	95
5.3	Erros percentuais para 30 dias no futuro	96
5.4	Erros percentuais para 60 dias no futuro	98
5.5	Erros percentuais para 180 dias no futuro	99

LISTA DE TABELAS

4.1	Divisão das Regiões do RS	43
4.2	Iterações do Algoritmo de Otimização	51
4.3	Pseudocódigo do Algoritmo de Otimização	52
4.4	População Total das Regiões e do RS	53

LISTA DE SIGLAS E SÍMBOLOS

Lista de Siglas

ETL	Erro de Truncamento Local
OMS	Organização Mundial da Saúde
PVI	Problema de Valor Inicial

Lista de Símbolos

\bar{x}	Centroide do Simplex
β	Taxa de Infecção
$\beta(t)$	Taxa de Infecção variante no tempo t
β_i	Taxa de Infecção no Período $t_{i-1} < t \leq t_i$
β_i^j	Parâmetro identificado na iteração j , alocado na posição i
δ	Taxa de Vacinação
γ	Taxa de Remoção
$\hat{Y}(t)$	Saída Simulada pelo Modelo
\mathcal{O}	Ordem do Erro de Aproximação
μ	Parâmetro de fase latente
τ	Tamanho do Período de Simulação
a	Coefficiente de Reflexão
b	Coefficiente de Expansão

c	Coeficiente de Contração
d	Coeficiente de Encolhimento
D_h^+	Aproximação da derivada por diferenças progressivas
D_h^-	Aproximação da derivada por diferenças regressivas
E	Expostos
e_T	Erro Percentual entre as previsões do modelo e os valores reais observados
h	Tamanho do Passo de Simulação
$H_f(x_k)$	Matriz Hessiana de f no ponto x_k
I	Infectados
$J(P)$	Matriz Jacobiana no ponto P
M	Dimensão do vetor de parâmetros a ser identificado
m	Critério de parada das iterações
N	População Total
P	População do local registrada nos dados
R	Removidos
R_0	Taxa de Reprodução da Infecção
S	Suscetíveis
T	Intervalo de tempo futuro na predição
$v(t)$	Função que representa atividade de vacinação
x_e	Ponto de Expansão do Simplex

x_r	Ponto de Reflexão do Simplex
x_{ce}	Ponto de Contração Externa do Simplex
x_{ci}	Ponto de Contração Interna do Simplex
$Y(t)$	Saída do Modelo obtida pelos dados

RESUMO

Epidemias afetam a humanidade com frequência. Quando sua propagação atinge todo o planeta, configura-se uma pandemia. No ano de 2020, o mundo enfrentou surtos de contaminação de uma doença respiratória causada por um novo coronavírus, o que exigiu das autoridades competentes a implementação de medidas de combate e prevenção. A compreensão da dinâmica de epidemias é essencial para prever cenários e fundamentar decisões, motivo pelo qual o uso de modelos matemáticos constitui uma ferramenta indispensável para descrever o comportamento dessas doenças. Entre os modelos mais utilizados, destacam-se os compartimentais, que dividem a população afetada em classes com características e inter-relações bem definidas. Esses modelos são estruturados com base em conjuntos de equações diferenciais que descrevem as taxas de variação das categorias populacionais que os compõem. Um exemplo notável é o modelo *SIR* (Suscetível-Infetado-Removido), que inclui parâmetros constantes representando as taxas de transmissão da doença e de remoção do grupo infectado. No entanto, o uso de valores constantes para descrever a taxa de transmissão ao longo de períodos prolongados pode não ser eficiente, dado que alterações no comportamento das pessoas, medidas governamentais e a evolução da própria doença podem modificar tal parâmetro nesse intervalo de tempo. Com base nessa premissa, este trabalho propôs uma abordagem para o modelo compartimental *SIR* que modifica o parâmetro de infecção, permitindo que, em vez de assumir um valor constante, seja variante no tempo. Essa modificação possibilitou a representação da atividade da COVID-19 em períodos mais extensos. Para validar o novo modelo, foram utilizados dados oficiais fornecidos pelo governo do Estado do Rio Grande do Sul. Por meio desses dados, aplicou-se um algoritmo de otimização destinado a minimizar o somatório dos quadrados dos erros entre o total de infectados registrado nos dados e o valor simulado pelo modelo. Posteriormente, os resultados obtidos foram comparados ao comportamento real da doença no estado, atestando a precisão do modelo. Além disso, com base nas taxas de infecção

estimadas pelo algoritmo, investigou-se a capacidade do modelo de prever a atividade da doença em cenários futuros. O cálculo do erro percentual entre o número total de infectados observado no período e a previsão do modelo revelou que, para intervalos de 7 ou 14 dias, o modelo apresentou resultados satisfatórios de predição, com erros percentuais inferiores a 0,13% e 0,6%, respectivamente.

ABSTRACT

Epidemics frequently affect humanity. When their spread reaches the entire planet, a pandemic is established. In 2020, the world faced outbreaks of a respiratory disease caused by a new coronavirus, which required competent authorities to implement measures for combating and preventing it. Understanding the dynamics of epidemics is essential for predicting scenarios and informing decisions, which is why the use of mathematical models is an indispensable tool for describing the behavior of these diseases. Among the most commonly used models, compartmental models stand out, which divide the affected population into classes with well-defined characteristics and interrelations. These models are structured based on sets of differential equations that describe the rates of change of the population categories they comprise. A notable example is the *SIR* (Susceptible-Infected-Removed) model, which includes constant parameters representing the disease transmission rates and the removal rate of the infected group. However, using constant values to describe the transmission rate over extended periods may not be efficient, as changes in people's behavior, government measures, and the evolution of the disease itself can alter this parameter over time. Based on this premise, this study proposed an approach to the *SIR* compartmental model that modifies the infection parameter, allowing it to vary over time instead of assuming a constant value. This modification enabled the representation of COVID-19 activity over extended periods. To validate the new model, official data provided by the government of the State of Rio Grande do Sul were used. Using this data, an optimization algorithm was applied to minimize the sum of squared errors between the total number of infected individuals recorded in the data and the value simulated by the model. Subsequently, the results obtained were compared to the actual behavior of the disease in the state, confirming the accuracy of the model. Additionally, based on the infection rates estimated by the algorithm, the model's ability to predict the disease's activity in future scenarios was investigated. The calculation of the percentage error between the total number

of observed infections during the period and the model's forecast revealed that, for 7-day or 14-day intervals, the model provided satisfactory prediction results, with percentage errors lower than 0.13% and 0.6%, respectively.

1 INTRODUÇÃO

A humanidade convive com doenças infecciosas desde o início de sua história. Algumas dessas enfermidades podem acometer simultaneamente um grande número de indivíduos em uma determinada localidade. Quando ocorrem surtos de contágio de uma mesma doença em uma região específica, considera-se que tal região está enfrentando uma *epidemia*. Caso a disseminação dessa doença ultrapasse fronteiras e atinja toda a população mundial, configura-se um caso de *pandemia*.

A prevalência e os efeitos de uma epidemia podem variar amplamente. Em países com baixo nível de desenvolvimento, mesmo doenças consideradas de baixo risco podem resultar em consequências graves. O nível de desenvolvimento econômico de uma região determina, em grande medida, a qualidade dos serviços de saúde e a eficácia na vigilância e contenção de doenças infecciosas. Por esse motivo, ainda são registradas taxas alarmantes de mortalidade em epidemias de malária, tifo e cólera, entre outras enfermidades que seriam facilmente tratáveis [9]. Além disso, uma epidemia pode gerar impactos econômicos significativos em uma ou mais regiões. Em muitos casos, as medidas de controle adotadas afetam a circulação de pessoas e o funcionamento de estabelecimentos, ocasionando prejuízos econômicos durante o período de vigência dessas ações. Dessa forma, compreender os mecanismos associados a essas crises sanitárias torna-se essencial para definir estratégias eficazes de enfrentamento.

Em 2020, o mundo foi acometido pela pandemia do novo coronavírus. O vírus causador da COVID-19 teve sua origem na cidade de Wuhan, na China, onde os primeiros casos foram registrados ainda em 2019 [6]. No ano seguinte, a disseminação alcançou países de todos os continentes, obrigando as autoridades a implementar medidas de controle e prevenção. No Brasil, o primeiro caso da doença confirmado pelo Ministério da Saúde ocorreu em 26 de fevereiro de 2020, poucas semanas antes de a Organização Mundial da Saúde (OMS) anunciar que a Europa havia se tornado

o novo epicentro da pandemia. O primeiro óbito no país foi registrado em março de 2020. No Estado do Rio Grande do Sul, o primeiro caso de infecção foi registrado em 10 de março de 2020 [3], e o primeiro óbito no estado ocorreu em 25 de março de 2020, na cidade de Porto Alegre.

Os coronavírus constituem um grupo de vírus de RNA que causam doenças em mamíferos e aves. Em seres humanos, podem provocar infecções respiratórias leves, como uma espécie de gripe comum, ou graves, como a Síndrome Respiratória Aguda Grave (SRAG). Esta última é causada pelo SARS-CoV-2, o agente etiológico da COVID-19, responsável pela pandemia que assolou o mundo em 2020 e que, até o momento, já resultou na morte de milhões de pessoas em todo o planeta [4].

Os sintomas causados pelo novo coronavírus manifestam-se de forma variável entre os indivíduos, podendo oscilar entre quadros leves e severos [1]. Em média, o aparecimento dos sintomas ocorre três dias após o contágio, embora seja possível que um indivíduo apresente manifestações até sete dias após a exposição ao vírus [21]. O grupo sintomático constitui a maior parcela da população infectada. Entretanto, pelo menos um terço dos indivíduos contaminados permanece assintomático. Apesar disso, os indivíduos assintomáticos também são capazes de transmitir a doença [2].

A transmissão do novo coronavírus ocorre, predominantemente, pelo contato direto entre seres humanos, seja por vias respiratórias ou por superfícies contaminadas. Toda vez que uma pessoa infectada tosse, espirra, fala ou respira, partículas contendo o vírus são liberadas e podem infiltrar-se no organismo de indivíduos expostos. Boca, nariz e olhos constituem as principais portas de entrada do vírus [2]. Ademais, a interação direta não é a única via de transmissão da doença. Partículas virais podem permanecer depositadas em superfícies, como objetos, corrimãos e outros locais frequentemente tocados, possibilitando a infecção de pessoas saudáveis que entrem em contato com essas áreas.

Para controlar a disseminação da doença no Brasil, diferentes políticas de controle foram implementadas em cada estado. No Rio Grande do Sul, o sistema adotado foi denominado Programa de Distanciamento Controlado. Essa política baseava-se em bandeiras definidas semanalmente, com critérios e protocolos específicos a serem seguidos pelos diversos setores da sociedade. No modelo, o estado foi dividido em 21 regiões, considerando-se a velocidade de propagação do vírus e a capacidade de atendimento do sistema de saúde. Ao todo, foram levados em conta 11 indicadores para determinar a gravidade da situação em cada região. De acordo com o risco identificado pelos indicadores, atribuía-se uma cor (amarela, laranja, vermelha ou preta) para classificar a região. A cor amarela indicava uma situação mais amena, em que medidas flexíveis eram aplicadas ao setor econômico e à circulação de pessoas. À medida que o risco aumentava, a classificação progredia para as cores laranja, vermelha e, por fim, preta, que representava o perigo máximo, com altíssimo risco de contágio. Nesse último cenário, normas mais rigorosas eram adotadas para conter a disseminação do vírus [5].

Para a adoção de medidas de prevenção e combate, torna-se essencial compreender os fenômenos epidemiológicos. Nesse contexto, a modelagem matemática apresenta-se como uma ferramenta extremamente útil. Por meio de modelos matemáticos, é possível simular a dinâmica de sistemas que reproduzem, por exemplo, o comportamento de uma doença infecciosa. Dessa forma, torna-se viável analisar e interpretar os resultados, buscando identificar causas e delinear estratégias de controle de epidemias [37]. Quando devidamente calibrado, um modelo matemático oferece um recurso valioso para autoridades políticas e sanitárias, permitindo fundamentar a tomada de decisões e a implementação de medidas mais eficazes.

Na literatura, encontram-se diversos tipos de modelos epidemiológicos. Dentre os mais utilizados, destacam-se os modelos compartimentais, que segmentam a população em classes com características comuns e estabelecem relações entre elas [41]. Esses modelos podem ser ajustados conforme os fatores considerados, mas,

em geral, são constituídos por um conjunto de equações diferenciais que descrevem como as taxas de transição entre os compartimentos envolvidos evoluem ao longo do tempo. Cada equação do sistema apresenta parâmetros que desempenham um papel fundamental na dinâmica das doenças, ao indicar de que maneira os compartimentos se inter-relacionam. Um exemplo clássico é o modelo *SIR*, que classifica a população em três classes: suscetíveis, infectados e removidos. O modelo utiliza essas categorias e parâmetros constantes para descrever as variações no número de indivíduos em cada compartimento ao longo do tempo.

No artigo [17], o modelo *SIR* foi utilizado para descrever o comportamento da COVID-19 em diferentes comunidades. No estudo, a população total não foi considerada constante, o que resultou em uma dinâmica diferenciada na classe de indivíduos suscetíveis, que não diminuiu de forma monótona, como ocorre quando a população é definida como fixa. Assim, demonstrou-se que a classe de suscetíveis pode aumentar em períodos de surto, e investigou-se a eficácia de medidas governamentais em cada comunidade analisada. Os autores concluíram que, com o suporte de dados confiáveis, o modelo *SIR* é adequado para prever tendências epidêmicas relacionadas à disseminação da doença, uma vez que é capaz de acomodar surtos e ser ajustado aos dados registrados. Além disso, evidenciou-se que, ao comparar os dados publicados com as simulações, o modelo pode ser utilizado para avaliar a eficácia das intervenções governamentais.

Em [44], discute-se a dificuldade em obter predições eficazes da COVID-19 utilizando modelos epidemiológicos. Os autores avaliaram as razões pelas quais os resultados variavam tanto, concluindo que isso se deve à não identificação das calibrações necessárias nos parâmetros do sistema. Em seguida, avaliaram o uso de dois modelos compartimentais, *SIR* e *SEIR*, para representar as informações contidas nos dados de casos confirmados. Demonstraram que o modelo *SIR* apresenta um desempenho superior e que a utilização de modelos mais simples pode ser mais confiável do que o uso de modelos complexos. Para isso, modelaram os impactos

da quarentena na cidade de Wuhan, na China, e um possível segundo surto após o término da medida.

No trabalho [40], um modelo *SIR* estendido, com dois compartimentos adicionais, foi adotado para analisar a dinâmica da disseminação do novo coronavírus, levando em consideração o isolamento social. Nesse modelo, foram inseridas as classes de indivíduos isolados e de óbitos. Com base nessas equações, simulou-se o comportamento da doença a fim de investigar a melhor estratégia para a adoção dessa medida, avaliando o momento mais adequado para seu início, a duração do isolamento e o momento apropriado para sua interrupção. Os cenários considerados envolveram 50%, 70% e 80% da população isolada, com início da restrição de contato nos 5^o, 10^o, 15^o e 20^o dias após a confirmação do 50^o caso. Com base nos dados oficiais, os autores concluíram que o melhor período para o início da medida ocorre entre o 10^o e 15^o dias, uma vez que a atividade da doença permanece estável durante a fase de isolamento. Além disso, destacaram que a saída dessa condição deve ser gradual, após 30 dias, para evitar um novo pico de infecção.

No artigo [29], os autores desenvolveram um modelo compartimental com classes adicionais, considerando mortes e hospitalizações, a fim de descrever a disseminação do novo coronavírus na China. Com esse modelo, foi estudado o impacto das subnotificações e estimada a necessidade de leitos hospitalares. Utilizando dados governamentais, foram identificados os parâmetros do modelo e, com eles, investigados cenários nos quais diferentes percentuais de casos não detectados alterariam a magnitude da doença no país.

Em [16], os autores propõem um modelo *SIR* discretizado, com as taxas de infecção e remoção dependentes do tempo. Utilizando dados fornecidos pelas autoridades chinesas, eles demonstram que o modelo apresenta um erro inferior a 3% em cada dia da pandemia de COVID-19 simulada. Além disso, utilizam o modelo para analisar o impacto das infecções não detectadas na disseminação da doença. Adicionalmente, mostram que, por meio do raio espectral de uma matriz 2×2 ,

composta pelos parâmetros do modelo, é possível prever a ocorrência de surtos. Dessa forma, investigam países que poderiam estar à beira de um surto de COVID-19 em março de 2020. O artigo também demonstra a eficácia do distanciamento social no controle da propagação da doença.

No artigo [39], foi realizada uma pesquisa sobre diversas publicações envolvendo modelagem matemática e uso da inteligência artificial para entender a dinâmica da COVID-19. Ao todo, foram estudados 61 artigos de periódicos, nos quais se constatou o amplo uso de modelos compartimentais, especialmente os modelos *SIR* e *SEIR*. Quanto às implementações de IA, a utilização de redes neurais foi a mais destacada. Os autores constataram que tanto a modelagem matemática quanto a inteligência artificial têm se mostrado ferramentas confiáveis para o entendimento da pandemia.

No trabalho [26], de 2022, foram investigadas as razões pelas quais algumas previsões da COVID-19 falharam. Os autores concluíram que, entre outros motivos, a modelagem incorreta, a falta de transparência, dados com erros e a ausência de evidências sobre o efeito das intervenções estão entre as principais causas. Eles afirmam que a modelagem cuidadosa, respeitando os aspectos característicos da doença, e a validação do modelo são fundamentais para evitar erros. Além disso, defendem que decisões importantes sejam baseadas em evidências globais, não apenas em casos específicos.

O artigo [47] propôs simular um algoritmo de Monte Carlo de um modelo *SIR*, com ciclos de 1 dia, a fim de estimar os picos de demanda clínica e os cenários em que a capacidade hospitalar apresentava sobrecarga na região metropolitana da Filadélfia. Os parâmetros utilizados no modelo foram extraídos de outros artigos e, portanto, não foram identificados no trabalho.

No artigo [49], é investigada a mudança nas taxas de reprodução da COVID-19 em diferentes regiões do território chinês. Constatou-se que a taxa de re-

produção controlada, que expressa a disseminação da doença quando há intervenção das autoridades, foi muito inferior à taxa de reprodução básica, que mede o número médio de novas infecções quando não há interferências, o que indicou a eficácia das medidas adotadas pelo governo local. Para determinar esses valores, foram implementadas três abordagens: os métodos baseados na verossimilhança de Poisson, na taxa de crescimento exponencial e um estocástico com base no modelo *SIR*.

Em [18], é utilizado o modelo *SIR* convencional para determinar o valor da taxa de reprodução da infecção básica na Itália, durante o surto inicial da COVID-19, ao longo de duas semanas. Utilizando dados do Ministério da Saúde do país, os autores calcularam valores semelhantes aos observados no surto inicial ocorrido na China, na cidade de Wuhan. As taxas calculadas confirmam também evidências anteriores, encontradas na literatura, que apontam valores semelhantes a essa taxa quando se aplica o modelo *SIR* ao novo coronavírus.

Outra aplicação do modelo *SIR* na descrição da COVID-19 foi testada em [43], no qual se analisou a eficácia desse sistema de equações na reprodução da doença, qualitativa e quantitativamente, em diversos países. Mostrou-se que, ao reduzir o modelo sequencialmente à equação logística, a dinâmica foi retratada adequadamente com alto grau de correlação, permitindo sua utilização para estimativas preditivas. Além disso, avaliou-se o impacto das medidas de prevenção nos parâmetros envolvidos. Para isso, foram utilizados dados do Centro Europeu de Controle de Doenças. Verificou-se que a redução da taxa de infecção durante o surto pode originar-se não apenas da diminuição do contato, mas principalmente do efeito normal de desaceleração característico do intervalo completo do surto. Isso proporciona uma oportunidade para estimar com mais garantia a qualidade da previsão baseada na curva logística.

No artigo [45], adotou-se um modelo *SIR*, com taxa de infecção variante no tempo, para investigar a tendência epidêmica da COVID-19 na Itália. O objetivo era considerar diferentes medidas de intervenção na disseminação da doença. Para

isso, utilizaram-se dados de séries temporais da COVID-19, de 22 de janeiro a 02 de abril de 2020, e estimou-se a taxa de reprodução básica média no país, comparando o resultado obtido com a região de Hunan, na China, que possui população total semelhante. Métodos de Cadeia de Markov foram aplicados para a obtenção do número de reprodução básico da doença. Os autores concluíram que as políticas de controle adotadas na província chinesa também tiveram efeitos semelhantes na Itália. Em ambos os casos, notou-se uma queda considerável no número de infecções no período em que as medidas foram executadas.

Outro artigo que utilizou um modelo *SIR* com taxa variante no tempo foi [8], no qual os autores desenvolveram um método para estimar o número de reprodução básica da doença através do Filtro de Kalman. Utilizando uma versão discretizada do sistema, a proposta foi determinar esse valor em tempo real. Para isso, utilizaram dados sobre novos casos para construir uma série temporal e estimar quantos infectados havia em certo tempo t , para então aplicar o Filtro de Kalman e estimar a taxa de crescimento dessa série. Por fim, utilizando a referência teórica, foi possível obter a taxa de reprodução básica. As estimativas foram feitas para 124 países e permitiram avaliar as medidas que reduziram o valor desse número, tais como distanciamento social, *lockdowns*, etc.

Em [35], é feita uma modificação no modelo convencional *SIR* para considerar os casos de infecção não relatados. Assim, adicionou-se um novo parâmetro ao sistema e, utilizando dados oficiais, identificaram-se seus valores para casos de gripe e COVID-19 no condado de Missoula, em Montana, EUA. O intervalo estipulado para o novo parâmetro foi $[0, 1]$, estabelecendo-se que, quando a maioria dos casos é relatada, o valor da constante adicional é próximo ou igual a 1. Por outro lado, quando há muita subnotificação, o parâmetro de relato aproxima-se de 0. Para a identificação com os dados da COVID-19, foi constatado que o valor ficou próximo de 1, e assim concluiu-se que o modelo *SIR* convencional seria apropriado para descrever o comportamento da doença na região. Já para os dados de gripe, os

valores encontrados se aproximavam de 0, indicando alta subnotificação, o que levou os autores à conclusão de que o modelo *SIR* tradicional, sem o novo parâmetro, é incapaz de retratar o comportamento real dessa enfermidade.

No trabalho de [14], desenvolveu-se um modelo compartimental derivado do *SIR*, incluindo a classe de óbitos, denotada por D , e considerando a população em R como sendo composta apenas por recuperados. Com o novo modelo *SIRD*, adotou-se também sua forma discretizada, a fim de considerar os parâmetros envolvidos no novo sistema como variáveis em função do tempo. O modelo foi aplicado na descrição da pandemia de COVID-19 na Itália, com o intuito de verificar a evolução da doença no país, levando em conta as diferentes medidas de prevenção praticadas pelo governo. Além de observar a eficácia de algumas medidas, o modelo permitiu também notar o intervalo de tempo no qual as políticas começaram a surtir efeito. Outra constatação do trabalho é que, com o passar do tempo, a taxa de recuperação começou a aumentar e a de mortalidade a diminuir. Com essa informação, os autores sugeriram como possível causa a melhor resposta do sistema de saúde, além de mudanças na letalidade do vírus.

Em [27], é utilizado um modelo *SIR* discretizado para estimar a prevalência do vírus da COVID-19, combinando diferentes fontes de dados disponíveis, como óbitos, casos confirmados, testes positivos, etc. Por meio de uma estrutura bayesiana, calibrou-se o modelo para projetar o número de novas infecções nos estados de Ohio e Indiana, nos Estados Unidos. Como resultado, os autores puderam constatar que o número de infectados foi subestimado, sobretudo no início da pandemia.

Com base nesses estudos, é possível afirmar que, embora viáveis para retratar certos aspectos da pandemia, os modelos compartimentais apresentaram limitações quanto à sua implementação. Os sistemas tradicionais, que assumem valores constantes para os parâmetros, esbarram na configuração da doença a longo prazo, sendo mais indicados para períodos curtos, nos quais a COVID-19 alterou

pouco seu comportamento. Já os modelos que utilizaram parâmetros variantes no tempo focaram sua análise nas medidas de contenção dos governos. Entretanto, seria possível aplicar um modelo compartimental para descrever o comportamento do novo coronavírus em períodos ainda mais prolongados? E o que esses modelos podem nos dizer sobre a tendência futura da pandemia? Seria possível que um modelo simples e determinístico nos fornecesse um prognóstico do que ocorrerá nos próximos dias, semanas ou meses?

A partir dessas questões, propôs-se neste trabalho uma adaptação para o modelo *SIR*. A proposta consiste em alterar o parâmetro que representa a taxa de infecção, de modo que ele seja variante no tempo, expresso por uma função constante por partes. Dessa forma, pode-se denotar essa taxa por meio de um vetor. O objetivo é que cada entrada desse vetor denote uma taxa de infecção específica para o período considerado. Assim, se for analisado um intervalo longo de tempo, não é necessário que um único valor represente todo o período. Por exemplo, se o tempo analisado for de um mês, pode-se obter um vetor com 30 valores, em que cada um indica a taxa de infecção diária. Desse modo, é possível que o modelo consiga expressar o comportamento da doença em intervalos mais prolongados, para os quais se dispõe de dados.

Para validar o modelo desenvolvido, foram utilizados dados oficiais fornecidos pela Secretaria de Saúde do Rio Grande do Sul, os quais registravam o número de infectados pelo novo coronavírus no estado. Com base nesses dados, identificou-se o vetor das taxas de infecção em dois cenários distintos.

No primeiro cenário, calcularam-se taxas semanais, ou seja, uma taxa de infecção para cada semana analisada, abrangendo as 21 regiões de agrupamento em que o território estadual foi subdividido. O objetivo dessa análise foi comparar os resultados do modelo com o comportamento real da pandemia durante a vigência do Programa de Distanciamento Controlado. No segundo cenário, identificaram-se

taxas diárias para o estado como um todo, com o propósito de examinar a atividade da pandemia em uma base diária.

Em ambas as situações, foi aplicado um algoritmo de otimização, o qual minimizou a soma dos quadrados das diferenças entre o número total de infectados registrado nos dados oficiais e aquele estimado pelo modelo. Essa abordagem permitiu ajustar os parâmetros de forma a representar com precisão a evolução da pandemia.

A identificação de um vetor com um grande número de entradas pode apresentar elevada complexidade. Isso ocorre porque o algoritmo de otimização requer valores iniciais para ajustar a solução até alcançar a ótima. No entanto, como o problema é não convexo, há a possibilidade de que os valores inicialmente atribuídos dificultem a convergência para a solução desejada, além de acarretarem um alto custo computacional.

Para contornar essas dificuldades, torna-se essencial inicializar o algoritmo com valores relativamente próximos à solução pretendida. Considerando-se o elevado número de componentes no vetor, foi desenvolvida uma estratégia iterativa para determinar os valores iniciais. Essa estratégia consiste em iniciar o algoritmo com um vetor contendo apenas um elemento e, a cada iteração, duplicar o número de entradas do vetor. Os valores iniciais de cada iteração são definidos com base nos resultados obtidos na etapa anterior. Esse processo é repetido até que o número total de entradas duplicadas exceda o número de entradas desejado.

Ao final do processo, as entradas excedentes são descartadas, e o algoritmo é executado até atingir a solução ótima. A validação do modelo foi realizada por meio da comparação de gráficos que representam o comportamento real da pandemia e o simulado pelo modelo identificado.

Após o desenvolvimento e a validação do modelo por meio da identificação de parâmetros, investigou-se sua capacidade de prever cenários futuros. Além

de descrever a evolução da doença em um período conhecido, avaliou-se a acurácia do modelo na previsão do comportamento do coronavírus no estado em tempos posteriores.

Para essa análise, considerou-se que o comportamento da doença seria desconhecido a partir de um determinado ponto, simulando-se a evolução do número total de casos a partir desse momento. Na simulação, foi utilizada uma taxa de infecção constante, previamente identificada pelo modelo com taxa variável no tempo. Em seguida, calculou-se o erro percentual entre os resultados previstos e os dados observacionais reais, analisando-se as condições nas quais o modelo apresentou eficácia.

Este trabalho está disposto da seguinte forma: no capítulo 2, são apresentados modelos epidemiológicos compartimentais comumente utilizados na literatura, destacando a configuração dos sistemas de equações diferenciais de cada modelo. A seguir, no capítulo 3, encontram-se os métodos numéricos estudados para a obtenção dos resultados, abordando técnicas de simulação numérica para equações diferenciais e também de otimização, necessárias para a identificação dos parâmetros. Logo após, no capítulo 4, propõe-se o uso de um novo modelo *SIR*, com a taxa de infecção variante no tempo, para simular a atividade da pandemia em um intervalo prolongado. A identificação dos parâmetros para validação do modelo, as simulações e os resultados são apresentados. No capítulo 5, é feita a investigação sobre a capacidade do modelo em prever o comportamento da pandemia em períodos futuros. Por fim, no capítulo 6, são feitas considerações sobre os resultados obtidos neste trabalho e sobre possíveis trabalhos futuros.

Parte dos resultados dessa tese compôs um artigo que foi aceito e apresentado em [24]. Outra parte dos resultados produziu artigo submetido à revista *Trends in Computational and Applied Mathematics (TCAM)*, que também está publicado em [23]

2 MODELOS MATEMÁTICOS PARA EPIDEMIOLOGIA

Modelos epidemiológicos são ferramentas essenciais para a compreensão e prevenção de padrões de doenças em populações humanas ou animais. Além de serem cruciais para entender a propagação de doenças infecciosas, eles também são aplicados em uma variedade de outros contextos de saúde pública, podendo ajudar a prever tendências futuras, identificar grupos de alto risco, avaliar o impacto de intervenções e políticas de saúde e guiar a alocação eficiente de recursos [9].

Existem diferentes tipos de modelos epidemiológicos, cada um com suas próprias vantagens e limitações. Modelos determinísticos, que utilizam equações diferenciais para descrever as mudanças nas populações ao longo do tempo, são úteis para compreender a dinâmica geral de uma doença. Por outro lado, modelos estocásticos incorporam aleatoriedade e incerteza, tornando-os mais adequados para simular eventos raros ou situações em que a variabilidade é importante [10].

Os modelos epidemiológicos compartimentais constituem uma classe importante dentro do arsenal de ferramentas utilizadas na epidemiologia. Esses modelos, amplamente utilizados na literatura, dividem a população em classes com características comuns [37]. Esses compartimentos representam os indivíduos envolvidos no processo e o estado no qual se encontram. As principais classes adotadas nesses modelos são [22]:

- Suscetíveis (S): formada pelas pessoas não infectadas que estão aptas a contrair a doença quando expostas ao agente responsável pela transmissão;

- Expostos (E): fração do grupo suscetível que já foi exposta ao contágio mas que ainda não apresentou sintomas ou a manifestação da doença. Nesse estágio, a doença pode ser transmitida ou não, dependendo do caso;
- Infectados (I): Indivíduos que estão doentes e são capazes de infectar algum suscetível;
- Removidos (R): Pessoas que foram infectadas mas já não são mais portadoras da doença, seja por motivo de cura ou morte. Também não podem mais transmiti-la e são imunes a uma nova infecção.

Com base nessa classificação, considera-se a população total como a soma dos indivíduos pertencentes a cada uma dessas classes. Assim, denotando essa quantidade total por N , tem-se que $N = S + E + I + R$. Contudo, nem todos esses grupos precisam ser contemplados na modelagem. Geralmente, a escolha das classes a serem utilizadas leva em consideração a doença em questão, de acordo com as formas de contágio e a maneira como ela se propaga na população [37].

2.1 Modelo SIS

O Modelo SIS é útil para compreender a dinâmica de doenças que não conferem imunidade duradoura após a recuperação, podendo auxiliar os pesquisadores no desenvolvimento de estratégias de controle e prevenção mais eficazes para essas doenças [9]. Nesse modelo, os indivíduos podem se infectar com a doença, recuperar-se e voltar ao estado suscetível, estando sujeitos a serem reinfectados, como ilustrado na representação em (2.1). Nesse cenário, não há pessoas removidas no sistema. Além disso, a classe de expostos é desconsiderada, pois assume-se que a manifestação dos sintomas ocorre em um curto intervalo de tempo após o contágio. Assim, o uso desse tipo de modelo é mais adequado para representar o

comportamento de enfermidades com baixas taxas de mortalidade e com períodos de imunidade inexistentes ou de curta duração, como as gripes comuns [25].



O modelo *SIS* é descrito através de equações diferenciais que levam em conta as taxas de transmissão da doença e de recuperação. Estas equações, que podem ser vistas em (2.2), descrevem como a proporção de indivíduos suscetíveis e infectados na população muda ao longo do tempo.

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t)I(t) + \gamma I(t) \\ \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t), \end{cases} \quad (2.2)$$

Na primeira equação, o primeiro termo, $-\beta S(t)I(t)$, representa a dinâmica de transmissão da doença, em que β é a taxa de transmissão e $S(t)I(t)$ corresponde ao número de interações entre indivíduos suscetíveis e infectados que podem resultar em transmissão. O segundo termo, $\gamma I(t)$, descreve o processo pelo qual os indivíduos infectados, $I(t)$, recuperam-se, sendo γ a taxa de recuperação.

Na segunda equação, o primeiro termo indica o número de novos infectados, que aumenta à mesma taxa em que a população deixa a condição de suscetível. O segundo termo representa os indivíduos recuperados, que deixam o estado de infectado na mesma proporção em que retornam à condição de suscetível.

A população total é expressa por $N = S(t) + I(t)$. A Figura 2.1 apresenta uma representação gráfica do modelo.

Uma das principais características do modelo *SIS* é a existência de um equilíbrio endêmico, no qual a doença permanece circulando indefinidamente em uma população, mantendo uma prevalência constante. Esse equilíbrio é determinado pela relação entre a taxa de transmissão da doença e a taxa de recuperação.

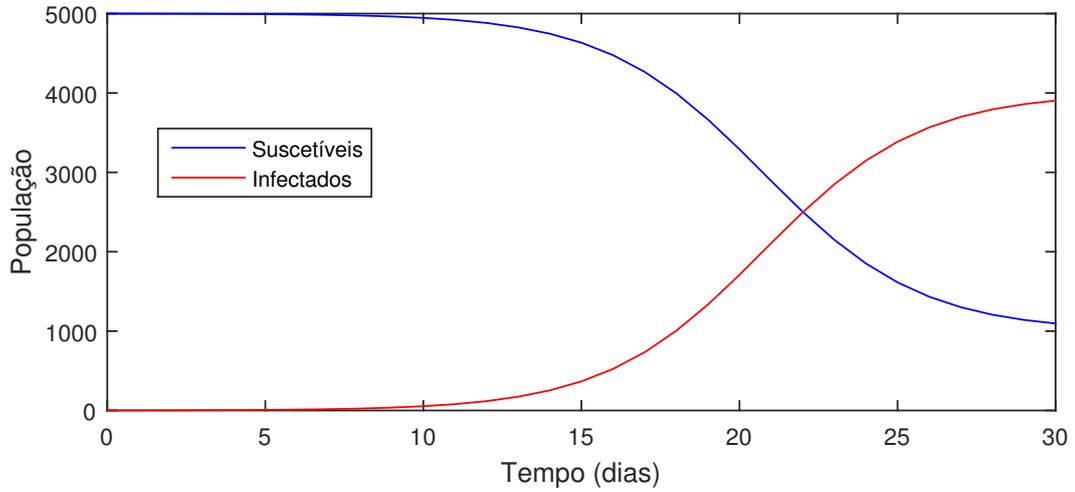


Figura 2.1: Modelo *SIS* com $\beta = 0.5$ e $\gamma = 0.1$

2.1.1 Solução Analítica

Outra maneira de compreender a evolução de uma pandemia em seu estágio inicial é por meio do cálculo de uma solução analítica aproximada da segunda equação diferencial do sistema. Para isso, admite-se $S = N - I$ e considerando que, no instante $t = 0$, $I(0) = I_0 > 0$, de (2.2), segue:

$$\frac{dI(t)}{dt} = \beta SI - \gamma I \quad (2.3)$$

$$\frac{dI(t)}{dt} = \beta(N - I)I - \gamma I \quad (2.4)$$

$$\frac{dI(t)}{dt} = \beta I \left[\left(N - \frac{\gamma}{\beta} \right) - I \right], \quad (2.5)$$

que caracteriza uma equação *logística* cuja solução pode ser obtida via separação de variáveis. Usando a condição inicial, obtém-se:

$$I(t) = \frac{\beta N - \gamma}{\beta + \left[(\beta N - \gamma) \frac{1}{I_0} - \beta \right] e^{-(\beta N - \gamma)t}} \quad (2.6)$$

Com isso, quando $t \rightarrow \infty$, $I \rightarrow N - \frac{\gamma}{\beta}$ e, conseqüentemente, o ponto de equilíbrio é assintoticamente estável. Em outras palavras, o modelo indica que, com o passar do tempo, o número de infectados tende a se aproximar cada vez mais de $N - \frac{\gamma}{\beta}$. Além

disso, pode-se observar pela solução analítica que, se a razão entre as taxas γ e β for maior que o tamanho da população N , a propagação da doença tende a cessar, já que $I(t) \rightarrow 0$ quando $N < \frac{\gamma}{\beta}$.

2.2 Modelo *SIR*

O modelo epidêmico *SIR*, proposto por Kermack e McKendrick (1927), é um dos mais populares na descrição de doenças com transmissão comunitária, aquelas cuja origem da infecção em uma pessoa é indeterminada, pois o agente transmissor circula entre toda a população [9]. Em sua dinâmica, admite-se que um indivíduo infectado só possa ser removido do sistema, desconsiderando-se a possibilidade de reinfecção. Sendo assim, o modelo pode ser representado esquematicamente da seguinte forma:



Nesse panorama, a população total N é dada pela soma $S(t) + I(t) + R(t) = N$. A configuração do modelo *SIR* é expressa pelo conjunto de equações diferenciais ordinárias:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N} \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (2.8)$$

Note que, na primeira equação do sistema, que denota a taxa de variação da população suscetível, a contaminação desses indivíduos ocorre pela interação entre as classes S e I . Assim, considera-se que há uma probabilidade de que o indivíduo suscetível contraia a doença quando interagir com algum infectado capaz de transmiti-la. Essa taxa de infecção da população suscetível por um infectado é

expressa pelo parâmetro β . Dessa forma, à mesma taxa com que uma parcela da população deixa o grupo de suscetíveis, ela ingressa no de infectados. Da mesma maneira, uma parte da população infectada deixa essa classe, após um determinado período, e passa a pertencer ao grupo dos removidos. Assim como no modelo *SIS*, os parâmetros referentes à taxa de infecção e de remoção são constantes positivas, representadas por β e γ , respectivamente. Na Figura 2.2, é apresentado um exemplo gráfico do modelo.

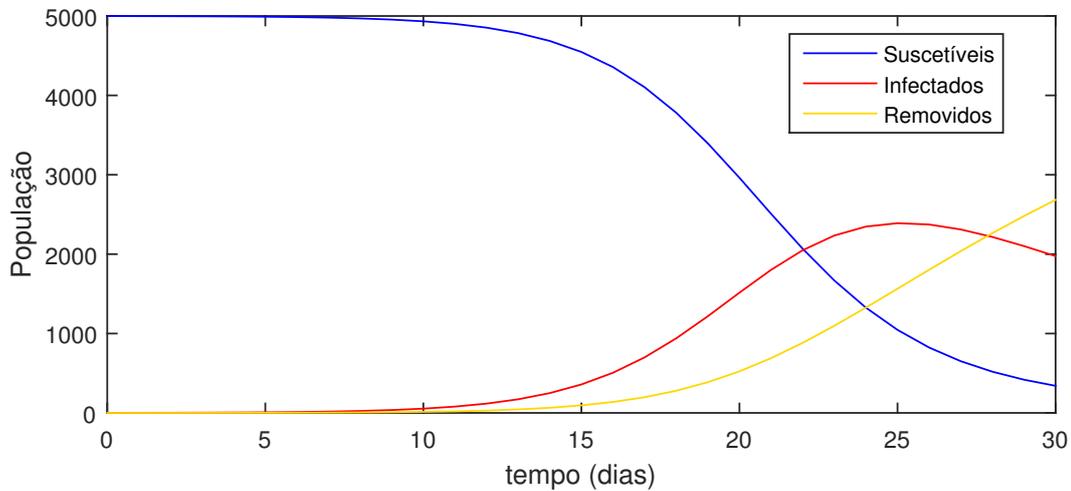


Figura 2.2: Modelo *SIR* com $\beta = 0.5$ e $\gamma = 0.1$

2.2.1 Solução Analítica

É possível também avaliar a evolução da epidemia em um modelo *SIR* quando, no início do processo ($t = 0$), o número de pessoas suscetíveis é praticamente toda a população ($S \approx N$). Sendo $I(0) = I_0 > 0$ tem-se que, na segunda equação:

$$\frac{dI}{dt} = \frac{\beta IN}{N} - \gamma I = I(\beta - \gamma), \quad (2.9)$$

cuja solução pode ser facilmente determinada pelo método da separação de variáveis:

$$I(t) = I_0 e^{(\beta - \gamma)t} = I_0 e^{\gamma t \left(\frac{\beta}{\gamma} - 1\right)} \quad (2.10)$$

Tomando $R_0 = \frac{\beta}{\gamma}$, segue que:

- Se $R_0 > 1$, $I(t) \rightarrow \infty$, ou seja, a epidemia permanece em crescimento;
- Se $R_0 < 1$, $I(t) \rightarrow 0$, isto é, a infecção tende a se encerrar com o tempo.

A constante R_0 é denominada taxa de reprodução da infecção e o período médio de infecção é dado por $1/\gamma$.

2.3 Modelo *SEIR*

Um dos mais populares para descrever fenômenos epidemiológicos, o modelo *SEIR* é comumente utilizado para analisar dados de infecção durante os diferentes estágios de sua propagação e avaliar a efetividade de medidas de contenção, como os *lockdowns* [22]. Nesse modelo, uma pessoa suscetível, quando exposta ao vírus, pode não apresentar sinais de infecção imediatamente, permanecendo assintomática por um período específico. Esse intervalo é denominado fase latente [31]. Após essa fase, cuja duração varia conforme a doença, o indivíduo permanece infectado até alcançar a recuperação ou evoluir para o óbito. O modelo também desconsidera a possibilidade de reinfeção, conforme retratado no esquema (2.11):



A versão clássica do modelo *SEIR* é composta por quatro equações diferenciais ordinárias que descrevem o comportamento de cada uma das classes admitidas. Além das constantes de infecção β e de remoção γ , um novo parâmetro, referente à fase latente, μ , é acrescentado ao sistema. Esse parâmetro dimensiona o atraso entre o contato infeccioso e o aparecimento dos sintomas que configuram a

infecção. Dessa forma, as equações são caracterizadas da seguinte forma:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta SI \\ \frac{dE(t)}{dt} = \beta SI - \mu E \\ \frac{dI(t)}{dt} = \mu E - \gamma I \\ \frac{dR(t)}{dt} = \gamma I, \end{cases} \quad (2.12)$$

com $N = S(t) + E(t) + I(t) + R(t)$. Para exemplificar graficamente o comportamento do modelo $SEIR$, observa-se um exemplo na Figura 2.3.

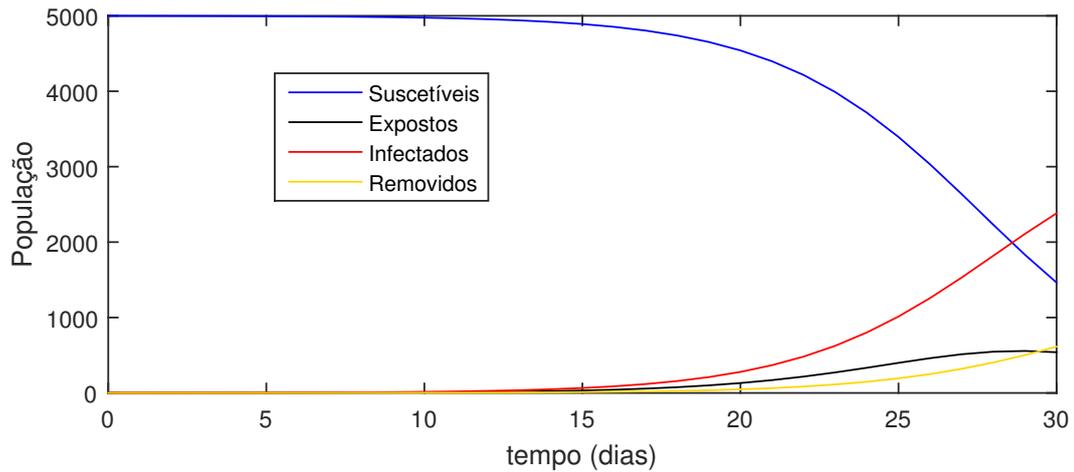


Figura 2.3: Modelo $SEIR$ com $\beta = 0.5$, $\gamma = 0.1$ e $\mu = 0.7$

2.3.1 Solução Analítica

Para observar a evolução da pandemia na perspectiva do modelo $SEIR$, é necessário a análise de dois grupos que afetam a dinâmica de infecção quando $S \approx N$. Desse modo, pode-se investigar a solução analítica das classes de Expostos e Infectados para verificar a tendência de progressão da epidemia ao longo do tempo.

Assim, se no instante $t = 0$ tem-se $I(0) = I_0$ e $E(0) = E_0$, considera-se o sistema:

$$\begin{cases} \frac{dE(t)}{dt} = \beta NI - \mu E \\ \frac{dI(t)}{dt} = \mu E - \gamma I \end{cases} \quad (2.13)$$

A partir daí, toma-se:

$$y'(t) = \begin{bmatrix} \frac{dE(t)}{dt} \\ \frac{dI(t)}{dt} \end{bmatrix}, \quad A = \begin{bmatrix} -\mu & \beta N \\ \mu & -\gamma \end{bmatrix}, \quad y(t) = \begin{bmatrix} E(t) \\ I(t) \end{bmatrix}$$

e resolve-se o problema de valor inicial (PVI):

$$\begin{cases} y(t) = Ay(t) \\ y(0) = y_0, \end{cases} \quad (2.14)$$

onde

$$y_0 = \begin{bmatrix} \frac{dE(0)}{dt} \\ \frac{dI(0)}{dt} \end{bmatrix} = \begin{bmatrix} E_0 \\ I_0 \end{bmatrix}$$

Calculando os autovalores $\lambda = \{\lambda_1, \lambda_2\}$:

$$\begin{aligned} \begin{vmatrix} -\mu - \lambda & \beta N \\ \mu & -\gamma - \lambda \end{vmatrix} &= \mu\gamma + \mu\lambda + \gamma\lambda + \lambda^2 - \mu\beta N \\ &= \lambda^2 + \lambda(\mu + \gamma) + \mu(\gamma - \beta N) \end{aligned}$$

Então,

$$\lambda = \frac{-(\mu + \gamma) \pm \sqrt{(\mu + \gamma)^2 - 4\mu(\gamma - \beta N)}}{2}$$

Sabe-se que a solução do sistema de EDO é do tipo $y(t) = e^{\lambda t}v$, onde v é um autovetor associado ao autovalor λ . Sendo assim, observa-se as seguintes possibilidades:

- se $N < \frac{\gamma}{\beta}$, segue que $4\mu(\gamma - \beta N) > 0$. Assim, se $(\mu + \gamma)^2 > 4\mu(\gamma - \beta N)$, $\sqrt{(\mu + \gamma)^2 - 4\mu(\gamma - \beta N)} < \sqrt{(\mu + \gamma)^2} = (\mu + \gamma)$. Isso implica que $\lambda_1, \lambda_2 < 0$, o que faz $y(t) \rightarrow 0$ quando $t \rightarrow \infty$. No caso em que $(\mu + \gamma)^2 < 4\mu(\gamma - \beta N)$, os autovalores serão complexos de parte real negativa, o que indica o mesmo comportamento com o passar do tempo. Consequentemente, projeta-se o fim da epidemia;

- Se $N > \frac{\gamma}{\beta}$, segue que $4\mu(\gamma - \beta N) < 0$. Desse modo, $\sqrt{(\mu + \gamma)^2 - 4\mu(\gamma - \beta N)} > \sqrt{(\mu + \gamma)^2} = (\mu + \gamma)$. Portanto, ao menos um dos autovalores λ será positivo. Logo, $y(t)$ tem crescimento exponencial e a epidemia permanece em progressão.

Como visto, os modelos epidemiológicos compartimentais baseiam-se em sistemas de equações diferenciais não lineares. Sendo assim, é conveniente que, para simular a dinâmica de uma doença por meio desses modelos, sejam utilizados métodos computacionais. No próximo capítulo, serão apresentadas algumas técnicas que podem ser empregadas nas simulações.

3 MÉTODOS NUMÉRICOS

O aprimoramento dos métodos numéricos tem desempenhado um papel essencial na integração entre a matemática e suas aplicações nas ciências e tecnologias. Com o aumento da capacidade computacional, a utilização de simulações computacionais de modelos matemáticos tem se expandido, consolidando-se como uma prática amplamente adotada em diversas áreas do conhecimento humano [34]. Essas técnicas computacionais visam obter aproximações numéricas para problemas que não podem ser resolvidos de forma analítica, uma situação frequente em modelagens matemáticas de fenômenos reais. Nesse contexto, a escolha do método a ser utilizado exige a consideração de aspectos como precisão, eficiência computacional e o impacto do erro de simulação nos resultados [42].

De forma geral, os algoritmos de métodos numéricos podem ser classificados em três categorias: diretos, iterativos e recursivos [28]. Os métodos diretos proporcionam uma solução aproximada em um número finito de passos. Por sua vez, os métodos iterativos realizam melhorias sucessivas na precisão, aproximando-se progressivamente da solução exata à medida que as iterações avançam. Para determinar o momento de encerrar as iterações, é usual a definição de um critério de parada ou a aplicação de um teste de convergência que estabeleça o número adequado de passos. Já os métodos recursivos são caracterizados por sub-rotinas que chamam a si mesmas, dividindo o problema em subproblemas semelhantes, cuja solução é interrompida com base em um critério de término predefinido.

As técnicas de aproximação numérica têm sido adaptadas para resolver uma ampla gama de problemas matemáticos, tais como a resolução de equações algébricas, a obtenção de soluções de sistemas lineares e não lineares, a interpolação de pontos, o cálculo de derivadas e integrais, e a determinação de soluções para equações diferenciais ordinárias e parciais [34]. No caso específico dos métodos de otimização, o objetivo é identificar a melhor solução para um determinado

problema dentro de um conjunto de soluções viáveis, denominado conjunto factível. Para alcançar os resultados apresentados neste trabalho, foram implementados métodos para a obtenção da solução numérica de um sistema de equações diferenciais e algoritmos de otimização.

3.1 Métodos Numéricos para Equações Diferenciais Ordinárias

As equações diferenciais desempenham um papel crucial na modelagem matemática de fenômenos naturais abordados pela ciência. A formulação dessas equações decorre da relação entre variáveis e suas taxas de variação instantâneas, permitindo a representação matemática do problema em análise. Contudo, em muitas situações, os fenômenos reais são descritos por expressões cujas soluções analíticas não podem ser obtidas. Nesse caso, a equação diferencial que descreve o evento pode ser simplificada, viabilizando sua resolução analítica. Alternativamente, métodos numéricos podem ser empregados para aproximar a solução do problema original, gerando resultados mais próximos dos dados observados no fenômeno estudado [13]. Na maioria dos casos, essas abordagens buscam determinar a solução de um problema de valor inicial, ou seja, uma solução aproximada da equação diferencial que satisfaça uma condição inicial predefinida.

A estimação da solução numérica de um problema de valor inicial para uma equação diferencial requer expressões que aproximem as derivadas das funções. Dada uma função f , a sua derivada no ponto t_0 pode ser definida como [34]:

$$f'(t_0) = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h} \quad (3.1)$$

Portanto, pode-se obter uma aproximação para $f'(t_0)$ simplesmente computando o valor do quociente para um h pequeno. Dessa forma, para $h > 0$,

tem-se:

$$D_h^+ f(t_0) = \frac{f(t_0 + h) - f(t_0)}{h} \approx f'(t_0) \quad (3.2)$$

Se $h < 0$:

$$D_h^- f(t_0) = \frac{f(t_0) - f(t_0 - h)}{h} \approx f'(t_0) \quad (3.3)$$

Essas aproximações para a derivada são denominadas diferenças progressivas, se $h > 0$, e diferenças regressivas, quando $h < 0$ [13]. Como cada um desses métodos fornece uma estimativa aproximada da derivada, é esperado que esses resultados contenham erros. Portanto, para avaliar a qualidade da estimativa, é importante compreender os erros associados calculados em cada caso.

O erro que cada aproximação produz no cômputo da derivada pode ser observado por meio da expansão das funções em série de Taylor. Dentre várias representações, pode-se definir essa expansão por [7]:

$$f(t_0 + h) = \sum_{k=0}^{\infty} \frac{f^{(k)}(t_0)}{k!} h^k = f(t_0) + hf'(t_0) + h^2 \frac{f''(t_0)}{2} + \dots + h^k \frac{f^{(k)}(t_0)}{k!} + \dots (3.4)$$

Para truncar a série no termo de ordem 2, considera-se o Teorema de Taylor, o qual mostra a existência de $\xi \in (t_0, t_0 + h)$ tal que [34]:

$$f(t_0 + h) = f(t_0) + hf'(t_0) + h^2 \frac{f''(\xi)}{2} \quad (3.5)$$

Então, isolando o termo $f'(t_0)$, obtém-se:

$$f'(t_0) = \frac{f(t_0 + h) - f(t_0)}{h} - h \frac{f''(\xi)}{2} \quad (3.6)$$

A estimativa de erro é dada a partir da diferença entre a solução analítica exata e a solução simulada. Portanto, o erro de aproximação para diferenças progressivas é determinado da seguinte forma:

$$f'(t_0) - D_h^+ f(t_0) = \frac{f(t_0 + h) - f(t_0)}{h} - h \frac{f''(\xi)}{2} - \left(\frac{f(t_0 + h) - f(t_0)}{h} \right)$$

$$= h \frac{f''(\xi)}{2} = \mathcal{O}(h),$$

Denota-se por \mathcal{O} a ordem do erro de aproximação. Nesse contexto, observa-se que o erro é de ordem h , ou seja, proporcional ao valor de h . Em outras palavras, ao reduzir h em 10 vezes, espera-se uma diminuição correspondente do erro, também em 10 vezes. O mesmo raciocínio aplica-se às diferenças regressivas na determinação da ordem do erro.

É importante destacar que ambos os métodos mencionados são sensíveis ao tamanho do intervalo de diferenciação escolhido. Adicionalmente, deve-se considerar que os cálculos são realizados em máquinas com capacidade de armazenamento finito, o que pode ocasionar erros numéricos caso os valores utilizados ultrapassem os limites suportados pelo sistema computacional. Assim, embora seja esperado que o erro diminua com a redução de h , é prudente adotar cautela na escolha desse parâmetro.

3.1.1 Método de Euler

O método de Euler é uma das técnicas de aproximação mais utilizadas para se determinar a solução de um problema de valor inicial de uma equação diferencial de primeira ordem [13]. Dado um PVI definido por:

$$\begin{cases} \frac{dy}{dt} = f(t, y(t)) \\ y(t_0) = y_0 \end{cases} \quad (3.7)$$

o procedimento consiste em aproximar a derivada presente na equação através de uma das técnicas de diferenciação apresentadas anteriormente. Considerando diferenças progressivas, tem-se:

$$\frac{dy}{dt} \approx \frac{y(t+h) - y(t)}{h} = f(t, y(t))$$

Assim, basta isolar o termo $y(t+h)$ e obtém-se:

$$y(t+h) = y(t) + h \cdot f(t, y(t)) \quad (3.8)$$

Esse método é conhecido como Método de Euler explícito, que estima o valor da função no ponto $t+h$ a partir de seu valor no ponto t . Como a condição inicial é conhecida, os pontos subsequentes podem ser determinados, com as distâncias entre eles definidas pelo valor atribuído a h . Para justificar a aplicação do método, é necessário analisar sua consistência e o erro gerado em sua utilização.

O erro de truncamento local (ETL) é definido como a diferença entre a solução analítica e a solução numérica, considerando apenas uma iteração e supondo que o passo anterior forneça a solução exata. Quando $ETL = \mathcal{O}(h^{p+1})$, o método é classificado como consistente de ordem p [34]. Assim, é possível determinar tanto a consistência quanto a ordem do erro do Método de Euler. A solução analítica no ponto t_0+h pode ser obtida por meio de uma expansão em série de Taylor, conforme apresentado em (3.5), enquanto a solução numérica é estimada de acordo com (3.8). Portanto,

$$ETL = \left(y(t_0) + hy'(t_0) + h^2 \frac{y''(\xi)}{2} \right) - (y(t_0) + hf(t_0, y(t_0))) \quad (3.9)$$

como $y'(t_0) = f(t_0, y(t_0))$, segue que:

$$ETL = h^2 \left(\frac{y''(\xi)}{2} \right) = \mathcal{O}(h^2) \quad (3.10)$$

Logo, o método de Euler é consistente de Ordem 1.

3.1.2 Método de Runge-Kutta de Quarta Ordem

Dada a equação diferencial $y'(\tau) = f(\tau, y(\tau))$, aplica-se a integração em ambos os lados da igualdade, no intervalo $t \leq \tau \leq t+h$, tal que:

$$y(t+h) = y(t) + \int_t^{t+h} f(\tau, y(\tau)) d\tau = y(t) + h \int_0^1 f(t+h\tau, y(t+h\tau)) d\tau \quad (3.11)$$

a partir disso, pode-se aproximar a segunda integral através de um somatório [42], como segue:

$$y(t+h) = y(t) + h \int_0^1 f(t+h\tau, y(t+h\tau))d\tau = y(t) + h \sum_{i=1}^m b_i k_i \quad (3.12)$$

onde

$$k_1 = f(t, y(t)) \quad (3.13)$$

$$k_i = f\left(y(t) + h \sum_{j=1}^{i-1} a_{ij} k_j, t + c_i h\right), \text{ com } i = 2, \dots, m \quad (3.14)$$

Dessa forma, basta calcular os valores de b_i , c_i e a_{ij} que melhor aproximem a solução numérica ao resultado analítico. Esse esquema de aproximação caracteriza os métodos de Runge-Kutta. Na literatura, são apresentadas as condições para determinar os valores dessas constantes [12, 13, 34, 42]. Nota-se que, ao considerar apenas um ponto e atribuir $b_1 = 1$, obtém-se uma expressão equivalente ao Método de Euler. Para o caso em que se utilizam 4 pontos, a expressão obtida é apresentada a seguir [34]:

$$y(t+h) = y(t) + \frac{k_1 + 2k_2 + 2k_3 + k_4}{6} \quad (3.15)$$

onde

$$\begin{aligned} k_1 &= hf(t, y(t)) \\ k_2 &= hf\left(t + \frac{h}{2}, y(t) + \frac{k_1}{2}\right) \\ k_3 &= hf\left(t + \frac{h}{2}, y(t) + \frac{k_2}{2}\right) \\ k_4 &= hf(t+h, y(t) + k_3) \end{aligned}$$

Esse método é denominado Runge-Kutta de quarta ordem, pois possui $ETL = \mathcal{O}(h^5)$ e, portanto, tem ordem de consistência igual a 4, o que indica uma precisão superior em relação ao Método de Euler previamente descrito.

Ambos os métodos apresentados também podem ser aplicados para determinar as soluções de um sistema de equações diferenciais ordinárias. Dado um sistema do tipo:

$$\begin{cases} y_1'(t) = f_1(t, y_1, y_2, \dots, y_n) \\ y_2'(t) = f_2(t, y_1, y_2, \dots, y_n) \\ \vdots \\ y_n'(t) = f_n(t, y_1, y_2, \dots, y_n) \end{cases} \quad (3.16)$$

com as condições iniciais $y_1(t_0) = y_{1,0}$, $y_2(t_0) = y_{2,0}$, \dots , $y_n(t_0) = y_{n,0}$, pode-se reescrever o sistema mediante uma notação vetorial, sendo:

$$y'(t) = \begin{bmatrix} y_1'(t) \\ y_2'(t) \\ \vdots \\ y_n'(t) \end{bmatrix}, \quad f(t, y(t)) = \begin{bmatrix} f_1(t, y(t)) \\ f_2(t, y(t)) \\ \vdots \\ f_n(t, y(t)) \end{bmatrix}, \quad y_0 = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ \vdots \\ y_{n,0} \end{bmatrix}.$$

Dessa forma, a resolução do sistema (3.16) apresenta-se como uma solução para o PVI:

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases} \quad (3.17)$$

que pode ser estimada a partir dos métodos apresentados.

3.2 Métodos Numéricos para Otimização

Os problemas de otimização têm como objetivo encontrar a melhor solução possível para um conjunto de parâmetros que atenda às condições do fenômeno investigado, identificando os valores máximos e mínimos de uma função objetivo [33]. As técnicas de otimização são geralmente empregadas quando não é possível calcular analiticamente os valores dessa função devido à complexidade do problema ou ao

elevado número de soluções possíveis. De forma geral, um problema de otimização pode ser definido como segue:

$$\text{Minimizar } f(x) \text{ sujeita a } x \in D. \quad (3.18)$$

A função f é denominada função objetivo, e (3.18) é considerado factível se o conjunto D for não vazio [30], ou seja, se houver uma solução na região considerada. Neste contexto, admite-se apenas o caso de minimização, pois maximizar $f(x)$ em uma região é equivalente a minimizar $-f(x)$ na mesma região.

Uma solução de mínimo global para o problema (3.18) é qualquer ponto $\hat{x} \in D$ que satisfaça:

$$f(\hat{x}) \leq f(x), \forall x \in D \quad (3.19)$$

Soluções de mínimo local configuram-se na existência de uma vizinhança U tal que dado um \hat{x} em U ,

$$f(\hat{x}) \leq f(x), \forall x \in D \cap U \quad (3.20)$$

Se $D = \mathbb{R}^n$, o problema é classificado como de otimização irrestrita [30]. Quando f é diferenciável, uma condição necessária para que um ponto \hat{x} seja um mínimo local de f [33, 38] é que $\nabla f(\hat{x}) = 0$. A partir da função objetivo, diferentes métodos de otimização podem ser aplicados, os quais se diferenciam pela utilização ou não de derivadas.

3.2.1 Método de Newton

O método de Newton geralmente é utilizado para resolver uma equação do tipo:

$$F(x) = 0, \quad (3.21)$$

onde $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Sendo x^* uma solução aproximada da equação (3.21), pode-se obter uma expansão de primeira ordem em série de Taylor para $F(x)$ em torno de x^* . Dessa forma,

$$F(x) \approx F(x^*) + (x - x^*)F'(x^*) \quad (3.22)$$

Assim, substituindo a expansão de (3.22) em (3.21):

$$F(x^*) + (x - x^*)F'(x^*) = 0 \implies x = x^* - (F'(x^*))^{-1}F(x^*) \quad (3.23)$$

A equação (3.23) provê uma iteração para o método de Newton, que pode ser expressa como:

$$x_{k+1} = x_k - (F'(x_k))^{-1}F(x_k), \text{ onde } k \geq 0 \quad (3.24)$$

sendo x_0 um valor inicial conhecido. Se $F : \mathbb{R} \rightarrow \mathbb{R}^n$, redefine-se a equação (3.24) na forma:

$$x_{k+1} = x_k - (J_F(x_k))^{-1}F(x_k) \quad (3.25)$$

na qual $J_F(x_k)$ representa a matriz Jacobiana de F calculada no ponto x_k .

Muitas situações em diferentes áreas do conhecimento são modeladas por sistemas de equações não lineares. Nesse contexto, o método de Newton tem sido amplamente utilizado no meio científico. No entanto, sua eficácia depende da escolha de um intervalo que contenha a solução e no qual a função e sua derivada sejam contínuas [38]. Outra aplicação desse método ocorre na resolução de problemas de otimização [38], como o descrito em (3.18). Por ser uma técnica de rápida convergência para sistemas não lineares, sua implementação pode produzir bons resultados na resolução de $\nabla f(x) = 0$.

Conforme apresentado anteriormente, uma condição necessária para que \hat{x} seja um mínimo local de uma função objetivo f é que o valor da derivada nesse

ponto seja nulo. Em outras palavras, para determinar o valor de mínimo, é necessário resolver a equação $\nabla f(\hat{x}) = 0$. Assim, definindo $F = \nabla f(\hat{x})$, o método de Newton (3.25) pode ser aplicado para encontrar o valor de \hat{x} que minimize a função objetivo. Nesse caso, como F corresponde ao gradiente da função objetivo f , a Matriz Jacobiana de F é equivalente à Matriz Hessiana de f no processo iterativo:

$$x_{k+1} = x_k - (H_f(x_k))^{-1} \nabla f(x_k) \quad (3.26)$$

onde $H_f(x_k)$ representa a matriz Hessiana de f no ponto x_k . Vale ressaltar que o método não favorece pontos de mínimo em relação aos de máximo, sendo, portanto, necessário atentar para o intervalo e o ponto inicial escolhidos [30].

3.2.2 Método Simplex de Nelder-Mead

Em contraste com métodos que utilizam derivadas para otimização, como o método de Newton, foram desenvolvidas técnicas que dispensam o cálculo de derivadas para resolver o problema. Entre essas técnicas, destacam-se os algoritmos de busca direta, processos iterativos que testam um conjunto de pontos na função objetivo para determinar o próximo passo da iteração e aproximar o valor de mínimo [15]. Nesse contexto, um dos métodos mais utilizados para solucionar problemas de otimização irrestrita é o método simplex de Nelder-Mead, reconhecido por sua simplicidade e facilidade de implementação [36].

O método de Nelder-Mead busca minimizar o valor de uma função de n variáveis sem a necessidade de informações sobre suas derivadas. A partir de um conjunto de $n + 1$ pontos no espaço \mathbb{R}^n , forma-se uma figura geométrica chamada simplex, cujos vértices correspondem a esses $n + 1$ pontos. A estratégia do método consiste em avaliar, a cada iteração, o valor da função objetivo em cada vértice, descartando o maior valor e atualizando o simplex por meio de processos de reflexão, expansão, contração ou encolhimento [15].

Seja $f(x)$ uma função real com $x \in \mathbb{R}^n$. Considera-se as seguintes constantes:

- a - coeficiente de reflexão;
- b - coeficiente de expansão;
- c - coeficiente de contração;
- d - coeficiente de encolhimento;

Tais constantes devem satisfazer as condições [36]:

$$a > 0, b > 1, 0 < c < 1, 0 < d < 1 \quad (3.27)$$

O algoritmo de Nelder-Mead padrão costuma adotar os seguintes valores para os coeficientes[32, 36]:

$$a = 1, b = 2 \text{ e } c = \frac{1}{2} = d. \quad (3.28)$$

Define-se também o centroide \bar{x} dos n melhores pontos, ou seja, dos pontos que não foram descartados:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (3.29)$$

O centroide é a base para realizar as operações do método. Cada uma delas visa gerar um novo ponto, que pode substituir o descartado na próxima iteração. As operações são descritas como segue:

1. Ordenação: em cada etapa ordena-se os vértices de modo que satisfaçam $f(x_1) \leq f(x_2) \leq f(x_3) \leq \dots \leq f(x_{n+1})$.
2. Reflexão: reflete o vértice descartado sobre a face oposta do simplex, computando o ponto de reflexão x_r por:

$$x_r = \bar{x} + a(\bar{x} - x_{n+1}); \quad (3.30)$$

Se $f(x_1) < f(x_r) < f(x_n)$, substitui-se x_{n+1} pelo ponto x_r na próxima iteração.

3. Expansão: Se $f(x_r) < f(x_1)$, calcula-se o ponto de expansão x_e , dado por:

$$x_e = \bar{x} + b(x_r - \bar{x}) = \bar{x} + ab(\bar{x} - x_{n+1}) \quad (3.31)$$

se $f(x_e) < f(x_r)$, toma-se x_e em substituição a x_{n+1} na próxima iteração. Caso contrário, o ponto tomado é x_r .

4. Contração: Se $f(x_r) > f(x_n)$, realiza uma contração entre \bar{x} e o melhor ponto entre x_r e x_{n+1} .

a) contração externa: Se $f(x_n) \leq f(x_r) < f(x_{n+1})$, isto é, x_r é estritamente melhor que x_{n+1} , calcula-se:

$$x_{ce} = \bar{x} + c(x_r - \bar{x}) = \bar{x} + ac(\bar{x} - x_{n+1}) \quad (3.32)$$

Se $f(x_{ce}) < f(x_r)$, aceita-se x_{ce} para a próxima iteração. Senão, efetua-se a operação de encolhimento, descrita mais adiante.

b) contração interna: Se $f(x_r) \geq f(x_{n+1})$, computa-se:

$$x_{ci} = \bar{x} - c(\bar{x} - x_{n+1}) \quad (3.33)$$

Se $f(x_{ci}) < f(x_{n+1})$, admite-se o ponto x_{ci} na próxima iteração. Do contrário, efetua-se a operação de encolhimento.

5. Encolhimento: Calcula-se o valor de f nos n pontos dados por $m_i = x_1 + d(x_i - x_1)$, para $i = 2, \dots, n + 1$. Então, cria-se um novo simplex, cujos vértices serão dados por x_1, m_2, \dots, m_{n+1} . Cabe salientar que os novos vértices não estarão ordenados de acordo com seu valor na função objetivo.

No procedimento descrito anteriormente, observa-se que, ao determinar x_r e comparar seu valor na função objetivo com o valor dos pontos do simplex na mesma função, obtém-se uma boa minimização caso o ponto refletido resulte no menor valor entre os calculados. Assim, é conveniente expandir o simplex nessa direção, através do ponto x_e . Se isso não ocorrer, ou seja, se x_r não fornecer um bom valor para a função objetivo, é possível que o simplex já esteja próximo do ponto de mínimo procurado. Nesse caso, processam-se as operações de contração. Se, ainda assim, não se obtiver um novo valor para um vértice que seja melhor que o do ponto a ser descartado, é provável que o ponto de mínimo esteja no interior do simplex, e, portanto, realiza-se a operação de encolhimento. Naturalmente, o algoritmo irá convergir para a solução desejada, e pode-se definir um critério de parada monitorando, por exemplo, o tamanho do simplex [32].

4 MODELO SIR COM β VARIANTE NO TEMPO

A taxa de infecção no modelo *SIR* convencional representa como um indivíduo infectado transmite a doença para a população suscetível. Esse parâmetro é essencial para compreender a disseminação de uma enfermidade ao longo do tempo. Entretanto, a atribuição de um valor fixo a essa taxa por um período prolongado pode gerar imprecisões na simulação do comportamento da doença. Tal limitação decorre da dependência do modelo em relação à interação entre as classes suscetíveis e infectadas, interação que pode sofrer alterações ao longo do tempo.

Mudanças comportamentais nos indivíduos, motivadas pela propagação da doença, podem alterar a probabilidade de infecção. Além disso, medidas de contenção adotadas pelas autoridades podem reduzir o contato entre as classes, enquanto mutações no agente causador da doença podem aumentar ou diminuir sua contagiosidade. Esses fatores contribuem para a variabilidade da taxa de infecção.

Com base nisso, propõe-se uma modificação no modelo *SIR*, permitindo que a taxa de infecção β varie ao longo do tempo. Essa abordagem possibilita uma análise mais precisa da propagação da doença em intervalos diários, semanais ou mensais, permitindo prever mudanças em seu comportamento. O modelo revisado é configurado da seguinte maneira:

$$\begin{cases} \frac{dS(t)}{dt} = -\frac{\beta(t)I(t)S(t)}{N} \\ \frac{dI(t)}{dt} = \frac{\beta(t)I(t)S(t)}{N} - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t), \end{cases} \quad (4.1)$$

onde $\beta(t)$ é a taxa de infecção variante no tempo t .

A proposta deste trabalho é definir a função $\beta(t)$ como uma função constante por partes, permitindo sua parametrização, definida por:

$$\beta(t) = [\beta_1 \ \beta_2 \ \cdots \ \beta_{k-1} \ \beta_k] , \quad (4.2)$$

onde

$$\beta(t) = \begin{cases} \beta_1, & 0 < t \leq t_1 \\ \beta_2, & t_1 < t \leq t_2 \\ \beta_3, & t_2 < t \leq t_3 \\ \vdots & \\ \beta_k, & t_{k-1} < t \leq t_k \end{cases} , \quad (4.3)$$

em que k é o número de dias, semanas, meses, etc.

4.1 Simulação do Modelo

A simulação do modelo requer a aplicação de um método computacional para obter a solução numérica do sistema de equações diferenciais que caracteriza o modelo *SIR*. Neste estudo, foi utilizado o método de Runge-Kutta de quarta ordem, adequado para a resolução de problemas do tipo $y'(t) = f(t, x)$. O período de amostragem adotado foi fixo, simulando o comportamento do sistema em intervalos diários, de forma compatível com a estrutura dos dados utilizados.

Além disso, considerando que o período médio de infecção é dado por $1/\gamma$, fixou-se o valor dessa constante com base na estimativa média de 10 dias [6]. Assim, assumiu-se $\gamma = 0,1$, o que configura o modelo *SIR* simulado da seguinte

maneira:

$$\begin{cases} \frac{dS(t)}{dt} = -\frac{\beta(t)I(t)S(t)}{N} \\ \frac{dI(t)}{dt} = \frac{\beta(t)I(t)S(t)}{N} - 0.1I(t) \\ \frac{dR(t)}{dt} = 0.1I(t) \end{cases} \quad (4.4)$$

Define-se como saída o acumulado de pessoas infectadas, que pode ser determinada por:

$$\hat{Y}(t) = \int_0^\tau \frac{\beta(t)I(t)S(t)}{N} dt = N - S(t), \quad (4.5)$$

onde τ é o tamanho do período considerado. Com isso, para a simulação de (4.4) pelo método Runge-Kutta, considerou-se:

$$\frac{dy}{dt} = \begin{bmatrix} \frac{dS(t)}{dt} \\ \frac{dI(t)}{dt} \\ \frac{dR(t)}{dt} \end{bmatrix} \quad (4.6)$$

e também:

$$f(t, y(t)) = \begin{bmatrix} -\frac{\beta(t)I(t)S(t)}{N} \\ \frac{\beta(t)I(t)S(t)}{N} - 0.1I(t) \\ 0.1I(t) \end{bmatrix} \quad (4.7)$$

Nos resultados apresentados neste trabalho, o tamanho do passo adotado no algoritmo foi $h = 1$, em função da natureza dos dados disponíveis. Esses dados foram amostrados em uma base diária, justificando a escolha do passo h para preservar o intervalo temporal utilizado na agregação das informações. Além disso, como a finalidade é representar o comportamento de uma epidemia em escala diária,

essa decisão assegura maior aderência do modelo ao fenômeno estudado, permitindo uma descrição mais compatível com a frequência de coleta dos dados.

O período de iteração n foi definido conforme a granularidade desejada. Para os resultados que analisaram a taxa de infecção semanal, utilizou-se $n = 7$. Já para a identificação diária, adotou-se $n = 1$.

Para exemplificar a simulação do modelo, o algoritmo foi executado em um caso hipotético. Considerou-se uma população de 500 habitantes e analisou-se a disseminação de uma doença diariamente nesse cenário. O período simulado foi de 30 dias, com o vetor β composto por 30 elementos. Essas entradas, que representam as taxas de infecção diárias, foram definidas aleatoriamente, limitando-se a valores positivos. Dessa forma, a atividade da doença simulada nesse cenário hipotético pode ser observada na figura 4.1.

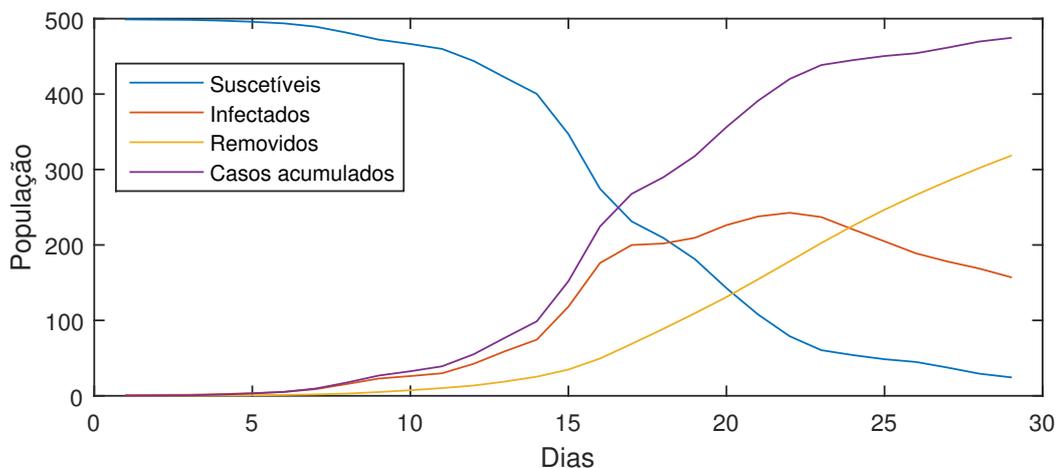


Figura 4.1: Simulação do modelo *SIR*

4.2 Identificação do Modelo

A utilização de um modelo matemático para descrever um fenômeno epidemiológico requer um conhecimento prévio adequado sobre sua dinâmica ou a

utilização de dados reais para determinar os valores de seus parâmetros. Contudo, o número de parâmetros e a qualidade dos dados empregados na sua identificação influenciam diretamente na precisão do modelo [19]. Assim, a escolha do sistema deve considerar as informações disponíveis sobre a doença estudada.

A pandemia do novo coronavírus apresentou um cenário de constantes mudanças. Novas informações sobre seu comportamento foram obtidas ao longo do tempo, e muitas particularidades da doença ainda podem ser desconhecidas. Nesse contexto, para empregar dados dessa doença na construção de um modelo que represente sua dinâmica, é fundamental dispor de informações consistentes e confiáveis. Entre as diversas informações coletadas, o número de casos registrados destaca-se como uma alternativa relevante.

A identificação de parâmetros é uma técnica utilizada para validar modelos a partir de dados obtidos em experimentos práticos [11]. Primeiramente, define-se um vetor contendo os parâmetros que se deseja identificar. Para determinar seus valores, utiliza-se um algoritmo de otimização que compara o comportamento do fenômeno estudado com o comportamento simulado pelo modelo. O algoritmo ajusta iterativamente os valores desses parâmetros para que o modelo se aproxime o máximo possível da realidade. Um dos procedimentos amplamente utilizados em problemas de otimização é o método dos mínimos quadrados.

O método dos mínimos quadrados tem como objetivo calcular um vetor θ que minimize o somatório de uma função $F(\theta)$ elevada ao quadrado [48]. Para ajustar um modelo matemático que descreva um fenômeno cujas informações de saída são conhecidas, a função F é definida como a diferença entre a saída estimada pelo modelo e a saída observada nos dados. Considerando $Y(t)$ a saída obtida pelos dados e $\hat{Y}(t)$ a saída simulada pelo modelo, a função a ser otimizada é expressa por:

$$J(\beta(t)) = \sum_{t=1}^n (Y(t) - \hat{Y}(t, \beta(t)))^2 = (Y - \hat{Y}(\beta(t)))^T (Y - \hat{Y}(\beta(t))), \quad (4.8)$$

onde n é o número de amostras do experimento, J é dita função objetivo, a que se quer minimizar, e:

$$\beta(t) = \arg \min J(\beta) \quad (4.9)$$

No Matlab, o método de otimização dos mínimos quadrados pode ser implementado utilizando a função *fminsearch*, baseada no método Simplex de Nelder-Mead, ou a função *lsqnonlin*, fundamentada no método de Newton. Neste trabalho, ambos os métodos foram aplicados para avaliar qual deles apresentava melhor desempenho. Os resultados obtidos foram semelhantes, razão pela qual optou-se por apresentar os correspondentes à solução obtida com a função *fminsearch*.

No caso deste trabalho, a saída do modelo corresponde ao número acumulado de infectados. O vetor a ser otimizado, de modo que o modelo represente com maior precisão o comportamento real desse total de casos, é o das taxas de infecção $\beta(t)$. O número de amostras, n , refere-se ao número de dias considerados, visto que os dados foram registrados diariamente.

4.2.1 Dados

Para realizar a identificação dos parâmetros do modelo, foram utilizados dados fornecidos pela Secretaria de Saúde do Estado do Rio Grande do Sul, disponíveis no site <https://ti.saude.rs.gov.br/covid19/>. Esses dados estão organizados por pessoa contaminada, apresentando a data de registro do caso e informações adicionais, como sexo, idade, entre outras. O mapeamento foi realizado por meio da divisão do Estado em 21 regiões, cada uma com um centro de saúde base localizado em uma cidade do respectivo território, conforme apresentado na Tabela 4.1. Nesses centros, são registradas as informações e implementadas medidas de combate à pandemia de acordo com a situação epidemiológica de cada região.

Região	Código
Bagé	R22
Cachoeira do Sul	R27
Canoas	R08
Capão da Canoa	R04,05
Caxias	R23-26
Cruz Alta	R12
Erechim	R16
Guaíba	R09
Ijuí	R13
Lajeado	R29,30
Novo Hamburgo	R07
Palmeira das Missões	R15,20
Passo Fundo	R17-19
Pelotas	R21
Porto Alegre	R10
Santa Cruz	R28
Santa Maria	R01,02
Santa Rosa	R14
Santo Ângelo	R11
Taquara	R06
Uruguaiana	R03

Tabela 4.1: Divisão das Regiões do RS

Com base nesses registros, é possível construir gráficos que mostram o número de casos detectados diariamente, bem como o total acumulado de infecções. Para exemplificar, utilizou-se o *software* Matlab para analisar a configuração diária da pandemia no Estado, identificando períodos mais críticos, de crescimento e de decréscimo no número de casos, entre outras observações. Esses comportamentos estão ilustrados nas Figuras 4.2 e 4.3.

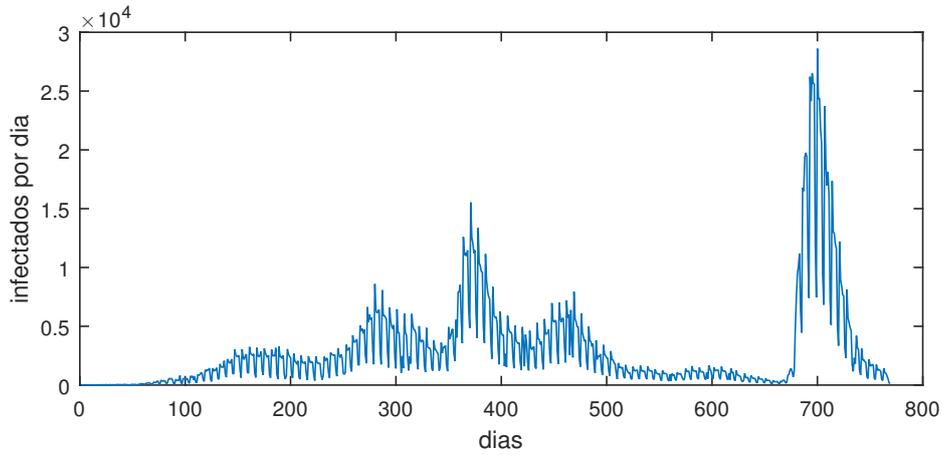


Figura 4.2: Número de casos diários no RS

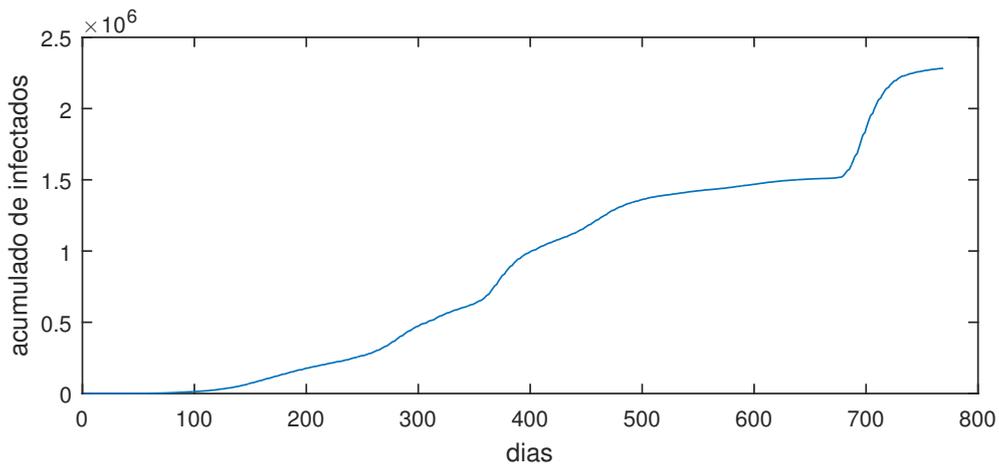


Figura 4.3: Casos Acumulados no RS

O gráfico da Figura 4.2 evidencia aumentos expressivos no número de casos de COVID-19 em determinados períodos, os quais podem ser associados à disseminação de diferentes variantes do vírus. Entre essas, destaca-se a variante Ômicron, cuja introdução resultou em um crescimento acentuado na parte final do período analisado. Esse comportamento está em conformidade com características específicas da Ômicron, como sua maior transmissibilidade, que contribuiu significativamente para o aumento no número de infecções, especialmente em populações amplamente expostas ao vírus e com alta circulação de pessoas.

Com esses dados, torna-se possível realizar a identificação do modelo por meio do algoritmo de otimização, cujos resultados serão apresentados nas seções subsequentes.

4.2.2 Otimização

A otimização dos valores do vetor de parâmetros, referentes às taxas de infecção do coronavírus, foi realizada de duas maneiras. Primeiramente, aplicou-se o algoritmo para identificar taxas semanais nas regiões de agrupamento em que o Rio Grande do Sul foi dividido, considerando um período de 50 semanas. Posteriormente, com base nos resultados obtidos, executou-se o mesmo algoritmo para modelar o comportamento diário da COVID-19 no Estado como um todo, ao longo de 770 dias.

Para a aplicação do algoritmo, é necessário estabelecer condições iniciais para as variáveis de estado que compõem o modelo. Assim, os valores definidos foram:

$$\begin{bmatrix} S(0) \\ I(0) \\ R(0) \end{bmatrix} = \begin{bmatrix} N - 1 \\ 1 \\ 0 \end{bmatrix}, \quad (4.10)$$

onde N é a população total da região estudada.

O problema de otimização abordado apresenta elevada complexidade devido à sua natureza não convexa e à dimensão infinita do espaço de solução. A ausência de convexidade implica na possibilidade de múltiplos mínimos locais, dificultando a identificação de um ponto de mínimo global. Ademais, a dimensão infinita do espaço de solução amplia a dificuldade de análise e resolução, exigindo métodos mais sofisticados para lidar com as variáveis envolvidas.

Para resolver o problema de dimensão infinita, adotou-se uma parametrização para a função $\beta(t)$. Essa parametrização consistiu em representar $\beta(t)$ como

uma função constante por partes, dividindo o intervalo de definição em subintervalos e atribuindo a cada um deles um valor constante, conforme descrito em (4.3). Essa abordagem transformou o problema original em um de dimensão finita, restringindo o número de parâmetros necessários para descrever $\beta(t)$. Como resultado, tornou-se viável aplicar métodos de otimização em um espaço finito, simplificando a análise e a busca por soluções, sem comprometer significativamente a qualidade da aproximação do problema original.

Para tratar o problema de não convexidade, buscou-se inicializar o vetor de parâmetros com valores ajustados previamente para estarem relativamente próximos da solução. Com esse propósito, foi desenvolvido um algoritmo iterativo que inicia com um vetor contendo apenas um parâmetro. A cada iteração, o número de entradas no vetor era duplicado, permitindo a identificação gradual de mais parâmetros. Esse processo iterativo foi mantido até que o vetor alcançasse o número desejado de parâmetros. Essa estratégia assegurou uma aproximação progressiva e controlada da solução, minimizando os riscos de convergência para ótimos locais e aumentando a eficiência da busca em um problema não convexo.

4.2.3 Algoritmo de Otimização

O algoritmo proposto é constituído por múltiplas etapas e tem como objetivo identificar um vetor de parâmetros de dimensão M . Inicialmente, define-se o número de parâmetros M , correspondente ao tamanho do vetor a ser identificado. Por exemplo, ao determinar taxas de infecção diárias para um período de 770 dias, o vetor conterá 770 entradas, ou seja, $M = 770$. Alternativamente, para taxas semanais ao longo de 350 dias, o vetor terá 50 elementos, cada um representando uma semana, resultando em $M = 50$.

Na sequência, estabelece-se um critério de parada m , definido como:

$$m = \min\{p \in \mathbb{N} \mid 2^p > M\},$$

ou seja, o menor valor inteiro positivo de m tal que $2^m > M$. Para o caso em que $M = 770$, obtém-se $m = 10$, enquanto que para $M = 50$, tem-se $m = 6$.

O processo iterativo do algoritmo inicia-se com a identificação de um vetor contendo apenas um elemento. Em cada iteração subsequente, o vetor é duplicado, e o resultado da iteração anterior é utilizado como condição inicial para a próxima. O algoritmo está projetado para realizar $m + 1$ iterações. Na última iteração, descarta-se qualquer elemento excedente, ajustando o vetor à dimensão M , de acordo com a solução final pretendida.

Para ilustrar o funcionamento do método descrito, apresenta-se o passo a passo do algoritmo aplicado ao caso em que se deseja identificar um vetor contendo 770 entradas ($M = 770$). Esse exemplo demonstra de forma prática cada etapa do processo, incluindo a determinação do critério de parada m e as iterações subsequentes até a obtenção do vetor final.

1. Definição da dimensão M do vetor:

O objetivo é identificar um vetor com 770 entradas, correspondente às taxas de infecção diárias para um período de 770 dias. Assim, $M = 770$.

2. Determinação do critério de parada m :

Define-se m como o menor inteiro que satisfaça $2^m > M$, ou seja:

$$m = \min\{p \in \mathbb{N} \mid 2^p > M\}$$

Para $M = 770$ tem-se:

$$2^9 = 512 < 770 \text{ e } 2^{10} = 1024 > 770$$

Logo, $m = 10$, e o algoritmo executará $m + 1 = 11$ iterações.

3. Iterações do Algoritmo:

A cada iteração j , o vetor identificado é representado por β_i^j , onde i indica a posição do elemento no vetor na iteração j . Cada elemento β_i^j representa a taxa de infecção para um intervalo específico de dias no período total de $2^m = 1024$ dias.

O vetor inicial utilizado na otimização de cada iteração j é formado a partir do vetor identificado na etapa anterior ($j - 1$), duplicando os valores obtidos. Na primeira iteração ($j = 1$), o vetor inicial foi definido como $\beta(t) = [1]$. A seguir, descrevem-se os detalhes de cada iteração:

- **Iteração 1** ($M = 1$): Definido o valor inicial nesta iteração como sendo $[1]$, o vetor resultante contém apenas o elemento β_1^1 , que representa a taxa de infecção para todo o período do dia 1 ao 1024. Como o período efetivo é de 770 dias, o excedente de 254 dias é descartado. Assim, β_1^1 corresponde ao intervalo do dia 1 ao 770, isto é,

$$\beta(t) = [\beta_1^1], \quad 0 < t \leq 770$$

- **Iteração 2** ($M = 2$): O vetor inicial nesta iteração é $[\beta_1^1 \ \beta_1^1]$, ou seja, o vetor identificado no passo anterior, duplicado. Então, o vetor resultante será dado por $[\beta_1^2 \ \beta_2^2]$. O período total de 1024 dias é dividido em dois intervalos de 512 dias: β_1^2 representa a taxa de infecção do dia 1 ao 512; β_2^2 cobre o intervalo do dia 513 ao 1024, mas admite-se apenas até o dia 770. Assim, β_2^2 corresponde efetivamente ao intervalo do dia 513 ao 770, como segue:

$$\beta(t) = \begin{cases} \beta_1^2, & 0 < t \leq 512 \\ \beta_2^2, & 512 < t \leq 770 \end{cases}$$

- **Iteração 3** ($M = 4$): Nesta iteração, o vetor inicial é dado por $[\beta_1^2 \ \beta_1^2 \ \beta_2^2 \ \beta_2^2]$, definido ao duplicar-se o resultado obtido na etapa antecedente. Após executar o algoritmo de otimização, o vetor resultante possui quatro elementos, dado por $[\beta_1^3 \ \beta_2^3 \ \beta_3^3 \ \beta_3^4]$. O período de 1024 dias é dividido em quatro intervalos de 256 cada, onde: β_1^3 equivale à taxa de infecção do dia 1 ao 256; β_2^3 do dia 257 ao 512; β_3^3 do dia 513 ao 768; β_4^3 do dia 769 ao 1024, porém como descarta-se o excedente, esse valor corresponde aos dias 769 e 770. Desse modo, tem-se:

$$\beta(t) = \begin{cases} \beta_1^3, & 0 < t \leq 256 \\ \beta_2^3, & 256 < t \leq 512 \\ \beta_3^3, & 512 < t \leq 768 \\ \beta_4^3, & 768 < t \leq 770 \end{cases}$$

- **Iteração 4** ($M = 8$): Duplicando os valores obtidos na iteração anterior, define-se $[\beta_1^3 \ \beta_1^3 \ \beta_2^3 \ \beta_2^3 \ \beta_3^3 \ \beta_3^3 \ \beta_4^3 \ \beta_4^3]$ como sendo o vetor inicial para essa etapa do algoritmo. Agora, o período dos 1024 dias é dividido em 8 intervalos de 128 dias cada, e as taxas identificadas ficam distribuídas como:

- β_1^4 : dia 1 ao 128;
- β_2^4 : dia 129 ao 256;
- β_3^4 : dia 257 ao 384;
- β_4^4 : dia 385 ao 512;
- β_5^4 : dia 513 ao 640;
- β_6^4 : dia 641 ao 768;
- β_7^4 : dia 769 ao 896;
- β_8^4 : dia 897 ao 1024;

Contudo, como o período desejado é 770 dias, o valor de β_8^4 não é utilizado nessa representação, e o β_7^4 efetivamente corresponde apenas aos dias 769 e 770. Com isso,

$$\beta(t) = \begin{cases} \beta_1^4, & 0 < t \leq 128 \\ \beta_2^4, & 128 < t \leq 256 \\ \beta_3^4, & 256 < t \leq 384 \\ \beta_4^4, & 384 < t \leq 512 \\ \beta_5^4, & 512 < t \leq 640 \\ \beta_6^4, & 640 < t \leq 768 \\ \beta_7^4, & 768 < t \leq 770 \end{cases}$$

- **Iterações subsequentes** ($M = 16, 32, \dots, 512$): Em cada iteração j , o vetor contém 2^{j-1} elementos, definidos por $\beta_1^j, \beta_2^j, \dots, \beta_{2^{j-1}}^j$, representando taxas de infecção para os intervalos de tamanho $1024/2^{j-1}$. O vetor inicial utilizado na otimização de cada iteração é o vetor identificado na iteração anterior, duplicado. Apenas os intervalos dentro do período efetivo dos 770 dias são considerados, descartando os excedentes.

4. Ajuste final:

Na penúltima iteração ($j = 10$), são identificados 512 elementos ($\beta_1^{10}, \beta_2^{10}, \dots, \beta_{512}^{10}$), correspondendo a 512 intervalos de 2 dias cada. Dessa forma:

- Para a última iteração ($j = 11$), o vetor inicial é formado duplicando os valores identificados na penúltima, gerando um vetor de 1024 elementos dado por:

$$\beta(t) = [\beta_1^{10} \ \beta_1^{10} \ \beta_2^{10} \ \beta_2^{10} \ \dots \ \beta_{512}^{10} \ \beta_{512}^{10}]$$

- Como o vetor final precisa conter exatamente 770 elementos, o excedente de 254 é descartado antes da otimização. Assim, o vetor inicial nesta última iteração contém os primeiros 770 elementos do vetor duplicado.
- Após a última execução do algoritmo, os 770 elementos finais β_1^{11} , $\beta_2^{11}, \dots, \beta_{770}^{11}$ são identificados, correspondendo a cada um dos dias do período de análise dos 770 dias.

O resumo do procedimento descrito pode ser visto na Tabela 4.2

Iteração	Número de Parâmetros	Tamanho do Período	Vetor Inicial	Vetor Final
1	1	1024 dias	[1]	$[\beta_1^1]$
2	2	512 dias	$[\beta_1^1 \beta_1^1]$	$[\beta_1^2 \beta_1^2]$
3	4	256 dias	$[\beta_1^2 \beta_1^2 \beta_2^2 \beta_2^2]$	$[\beta_1^3 \beta_2^3 \beta_3^3 \beta_4^3]$
4	8	128 dias	$[\beta_1^3 \beta_1^3 \beta_2^3 \beta_2^3 \beta_3^3 \beta_3^3 \beta_4^3 \beta_4^3]$	$[\beta_1^4 \beta_2^4 \dots \beta_7^4 \beta_8^4]$
\vdots	\vdots	\vdots	\vdots	\vdots
j	2^{j-1}	$1024/2^{j-1}$ dias	$[\beta_1^{j-1} \beta_1^{j-1} \dots \beta_{2^{j-1}}^{j-1} \beta_{2^{j-1}}^{j-1}]$	$[\beta_1^j \beta_2^j \dots \beta_{2^{j-1}}^{j-1}]$
\vdots	\vdots	\vdots	\vdots	\vdots
10	512	2 dias	$[\beta_1^9 \beta_1^9 \dots \beta_{256}^9 \beta_{256}^9]$	$[\beta_1^{10} \dots \beta_{512}^{10}]$
11	770 (1024 - excedente)	1 dia	$[\beta_1^{10} \beta_1^{10} \dots \beta_{385}^{10} \beta_{385}^{10}]$	$[\beta_1^{11} \dots \beta_{770}^{11}]$

Tabela 4.2: Iterações do Algoritmo de Otimização

É possível também definir um pseudocódigo do algoritmo de otimização descrito anteriormente, conforme 4.3 :

Nas próximas seções serão apresentados os resultados da identificação do modelo *SIR*, com taxa de infecção variante no tempo, utilizado para descrever o comportamento do novo coronavírus nas regiões de agrupamento em que o Rio Grande do Sul foi dividido, bem como sua dinâmica no Estado como um todo. Em cada caso, será realizado um comparativo entre os resultados obtidos pelo modelo e o desenvolvimento real da doença. Para tanto, foram analisados os gráficos gerados pela simulação e os obtidos a partir dos dados observados.

Pseudocódigo do Algoritmo de Otimização
<p>Entrada:</p> <ul style="list-style-type: none"> - M: número de parâmetros desejados - Função _ Otimização: função que realiza a otimização para identificar os parâmetros <p>Saída:</p> <ul style="list-style-type: none"> - Vf: vetor final com M elementos identificados <ol style="list-style-type: none"> 1. Inicializar o vetor inicial $V_i = [1]$ 2. Calcular m como menor inteiro tal que $2^m < M$ 3. Definir $\text{PeriodoTotal} = 2^m$ 4. Definir $\text{Nro_iterações} = m + 1$ 5. Para $j=1$ até Nro_iterações fazer: <ul style="list-style-type: none"> - $\text{Tamanho_vetor} = 2^{\{j-1\}}$ - Dividir o PeríodoTotal em Tamanho_vetor intervalos iguais - Se $j = \text{Nro_iterações}$: Remove o excedente fora do intervalo de M dias; $\text{Tamanho_vetor} = M$ - Otimizar valores do vetor: $\text{VetorIdentificado} = \text{Função_Otimização}(V_i, \text{intervalos})$ - Se $j < \text{Nro_iterações}$: Formar V_i duplicando os elementos do VetorIdentificado 6. Retornar $V_f = \text{VetorIdentificado}$

Tabela 4.3: Pseudocódigo do Algoritmo de Otimização

A população total de cada zona, bem como do Estado, contida nos dados utilizados para identificar os parâmetros, está apresentada na Tabela 4.2.3. Esses números foram empregados como condições iniciais para as variáveis de estado, conforme discutido na seção anterior. Nas simulações, foram desconsideradas as dinâmicas populacionais, como crescimento ou declínio demográfico, além de processos de migração e outras interações entre os indivíduos de cada região.

Código da Região	População
R22	188571
R27	203390
R08	778554
R04,05	397129
R23-26	1228814
R12	151898
R16	233010
R09	413416
R13	229282
R29,30	356164
R07	829897
R15,20	345622
R17-19	660667
R21	878788
R10	2369149
R28	350710
R01,02	559270
R14	223720
R11	279720
R06	234957
R03	466912
RS (total)	11379640

Tabela 4.4: População Total das Regiões e do RS

4.3 Resultados da Identificação para as Zonas de Agrupamento

A validação do modelo foi realizada considerando as 21 regiões de agrupamento em que o estado do Rio Grande do Sul foi subdividido. Essa abordagem possibilitou analisar a aplicabilidade do modelo em diferentes contextos regionais, levando em conta suas características individuais. A identificação realizada para todas as regiões permitiu verificar a capacidade do modelo de capturar as dinâmicas específicas de cada área, proporcionando uma análise mais robusta e representativa.

Além disso, o fato de poucas regiões apresentarem comportamentos similares reforça a importância de ampliar a divisão regional. Um maior número de regiões pode fornecer mais dados e variações contextuais, o que contribui para testar a eficácia do modelo em cenários distintos. Essa ampliação aumenta a confiabilidade dos resultados e demonstra que o modelo possui potencial para ser aplicado em uma diversidade de situações, validando sua versatilidade e precisão.

Com base nisso, foi realizada a identificação do modelo para todas as 21 zonas de agrupamento. O período analisado abrangeu 50 semanas, ou 350 dias. Como as medidas de combate à propagação da doença nessas zonas eram definidas semanalmente pelas autoridades, optou-se por identificar taxas semanais para esse caso, permitindo observar a eficiência dessas medidas. Assim, a otimização determinou um vetor com 50 elementos, em que cada entrada de $\beta(t)$ corresponde a uma semana. Esses valores foram representados em gráficos para viabilizar a análise pretendida.

4.3.1 Identificação para R22 - Bagé

A região R22 tem como cidade-base Bagé, mas inclui também outros 5 municípios. Com base nos dados dessa região, foi executado o algoritmo de otimização, permitindo a identificação do vetor β , que representa as taxas de infecção semanais. Na Figura 4.4, apresenta-se a comparação entre a saída do modelo, que calcula o número acumulado de infecções, e o total acumulado informado pelos dados. Observa-se que o comportamento do modelo é bastante semelhante ao dos dados reais, evidenciando sua eficácia em descrever a atividade da doença.

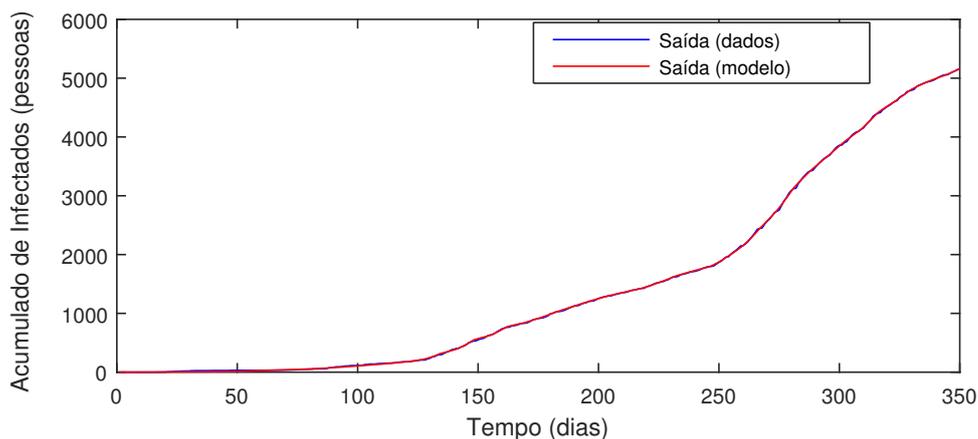


Figura 4.4: Modelo SIR para R22 - Bagé

Na Figura 4.5, ilustra-se a variação dessas taxas ao longo do período analisado, correspondente a 50 semanas.

Destaca-se um crescimento acentuado entre a 16^a e a 19^a semana, intervalo no qual a doença atingiu seu pico de infecção na região. Nas semanas subsequentes, observa-se uma redução nos valores das taxas, indicando, possivelmente, a implementação de medidas de contingência.

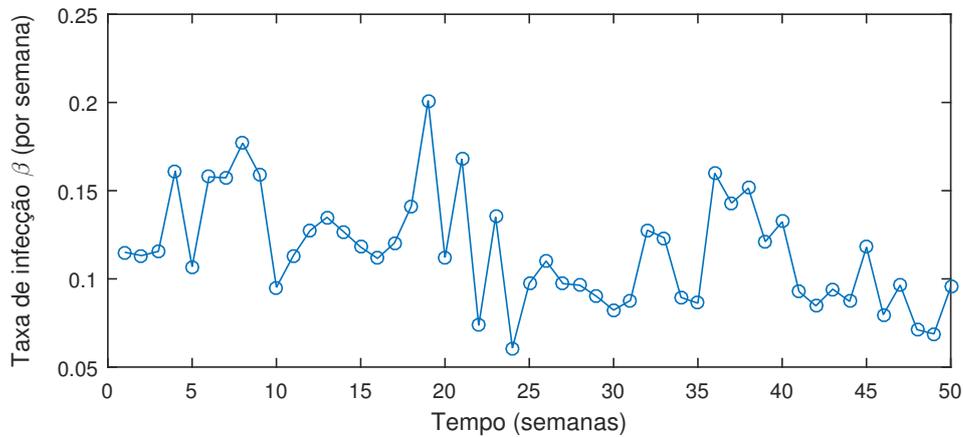


Figura 4.5: Taxas de Infecção para R22

4.3.2 Identificação para R27 - Cachoeira do Sul

A zona de agrupamento que tem como cidade-base Cachoeira do Sul abrange 12 municípios. Utilizando os dados dessas cidades, realizou-se a identificação do modelo para a região, conforme apresentado na comparação dos gráficos na Figura 4.6.

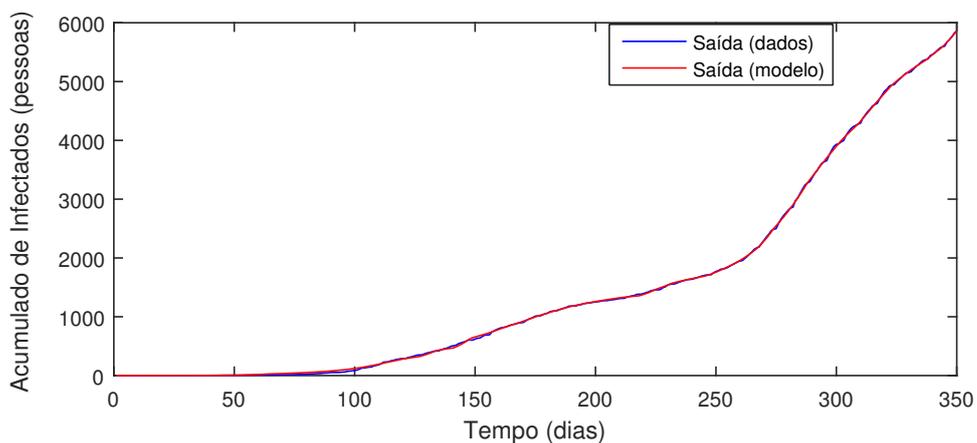


Figura 4.6: Modelo SIR para R27 - Cachoeira do Sul

Na figura, observa-se que a identificação foi novamente capaz de reproduzir com eficiência a evolução dos casos acumulados, evidenciada pela proximidade

entre as curvas. Os valores das taxas de infecção identificados estão representados na Figura 4.7.

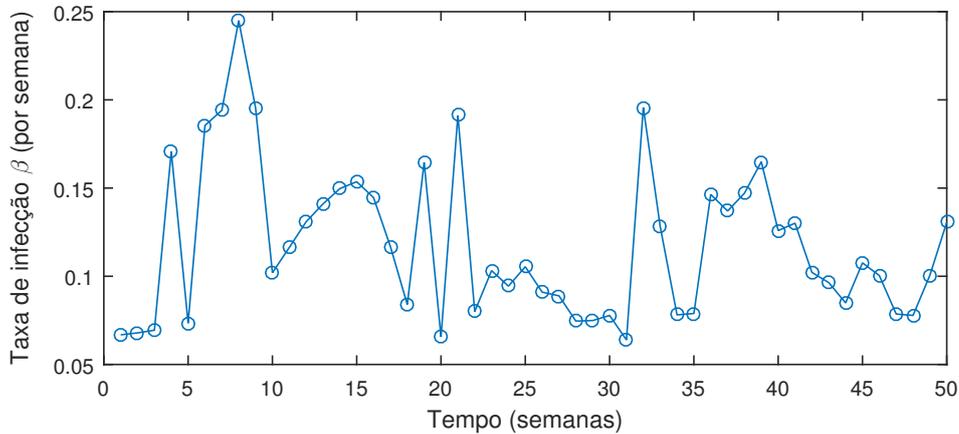


Figura 4.7: Taxas de infecção para R27

Nessa região, verifica-se que a maior probabilidade de infecção ocorreu na 8ª semana, embora outros períodos de maior risco tenham sido observados, como nas semanas 21 e 33. A redução nas semanas subsequentes a esses picos pode ser atribuída, possivelmente, à implementação de medidas de controle.

4.3.3 Identificação para R08 - Canoas

A região R08 é composta pela cidade de Canoas, onde se localiza o centro de saúde-base, e outros 17 municípios em suas proximidades. O resultado da identificação do modelo aplicado a essa região está ilustrado no gráfico da Figura 4.8.

Mais uma vez, percebe-se a semelhança entre as curvas simulada pelo modelo e aquelas geradas a partir dos dados, indicando uma boa precisão na identificação. Os valores das taxas de infecção identificadas estão representados na Figura 4.9.

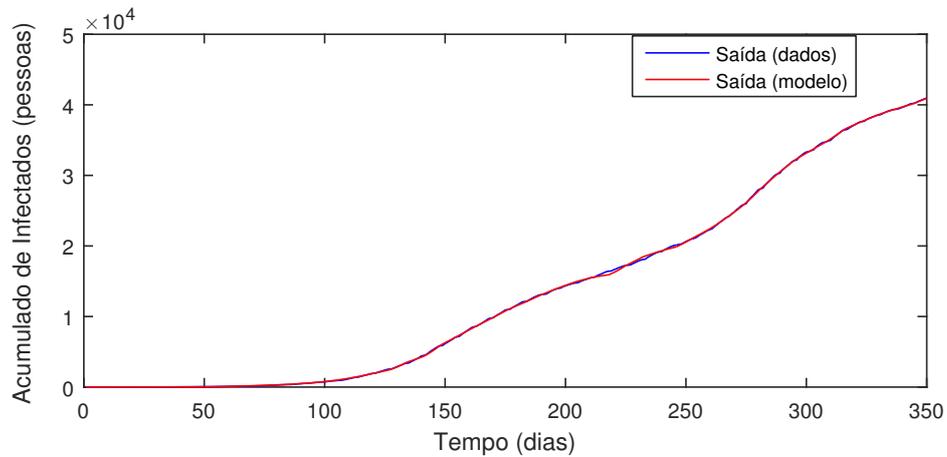


Figura 4.8: Modelo SIR para R08 - Canoas

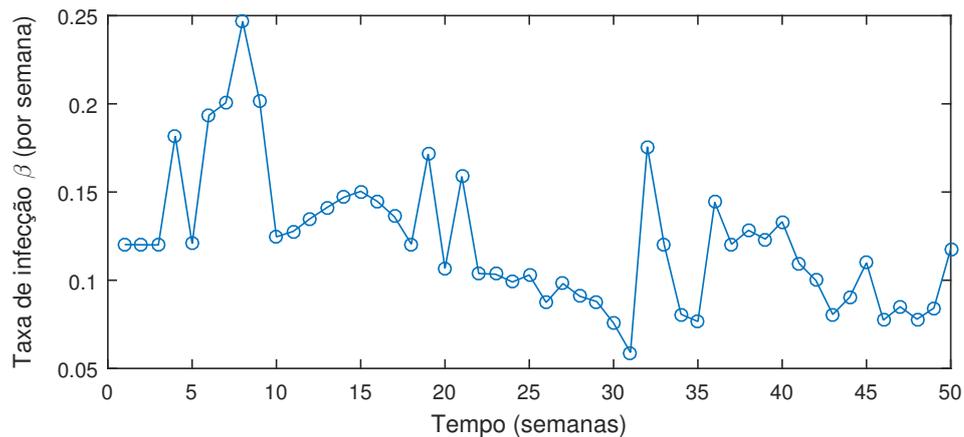


Figura 4.9: Taxas de Infecção para R08

A análise dessas taxas revela que houve um pico de infecção por volta da 8ª semana. Após esse pico, a taxa apresentou uma queda, mantendo-se praticamente estável por algumas semanas. Posteriormente, verificou-se um período de quedas quase consecutivas entre a 22ª e a 31ª semana. No entanto, em seguida, as taxas voltaram a crescer de forma acentuada.

4.3.4 Identificação para R04,05 - Capão da Canoa

A região analisada é composta por 23 municípios localizados no litoral norte do Estado, com Capão da Canoa como cidade-sede, onde foi estabelecido o centro de saúde responsável pelo registro e organização dos dados epidemiológicos. A partir desses dados, realizou-se a identificação do modelo, que demonstrou ser capaz de reproduzir de forma confiável o comportamento do coronavírus nesse território, como ilustrado na Figura 4.10.

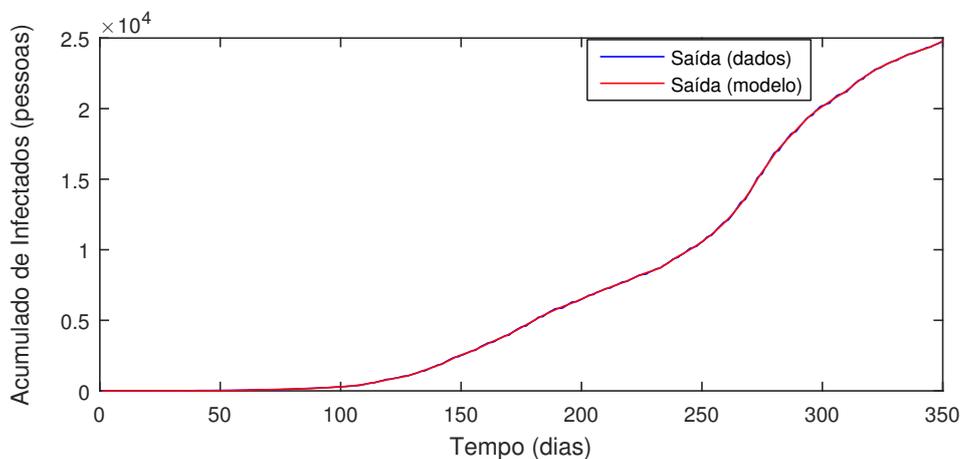


Figura 4.10: Modelo SIR para R04 e R05 - Capão da Canoa

Os resultados obtidos mostram uma boa aderência entre as curvas geradas pelo modelo e os dados observados, evidenciando a eficácia da abordagem adotada. Essa precisão é fundamental para a aplicação do modelo em cenários semelhantes, uma vez que reforça sua capacidade de representação das dinâmicas de infecção em diferentes contextos regionais.

Na Figura 4.11, apresenta-se a evolução das taxas de infecção ao longo do período analisado. Observa-se que o momento de maior probabilidade de contrair a doença ocorreu na 8ª semana, caracterizando o pico de infecção para essa região. Esse dado é consistente com padrões observados em outras regiões, indicando

que fatores contextuais, como densidade populacional e interação social, podem ter influenciado no aumento das taxas nesse intervalo de tempo.

Além disso, o gráfico destaca a importância de medidas de controle eficazes para mitigar a propagação da doença. Após o pico da 8ª semana, é possível notar uma leve tendência de queda nas taxas de infecção, sugerindo que as ações adotadas pelas autoridades de saúde podem ter contribuído para conter a transmissão do vírus.

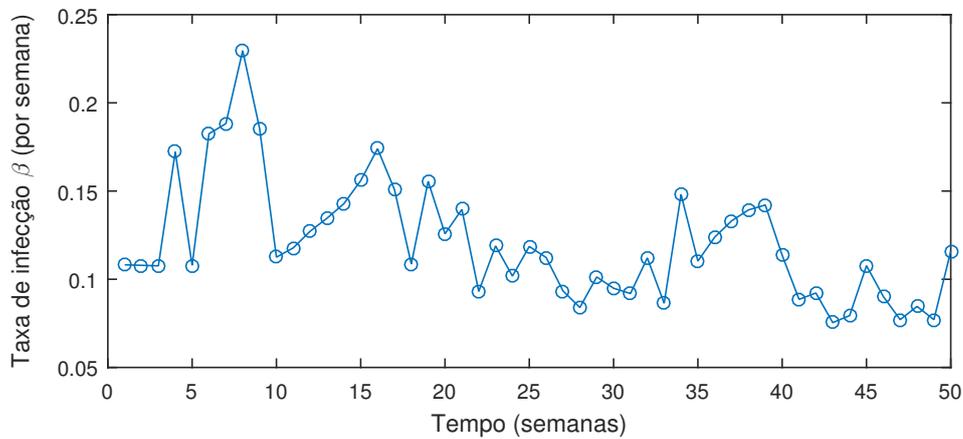


Figura 4.11: Taxas de Infecção para R04 e R05

4.3.5 Identificação para R23, 24, 25, 26 - Caxias do Sul

A região R23, 24, 25, 26 é composta por 50 cidades e tem, em Caxias do Sul, a base dos centros de saúde. A partir dos registros obtidos para essa zona de agrupamento, identificaram-se as taxas semanais de infecção para simular, por meio do modelo, a situação da pandemia nesses locais. A Figura 4.12 apresenta a comparação entre o resultado obtido pela identificação e o comportamento real gerado pelos dados.

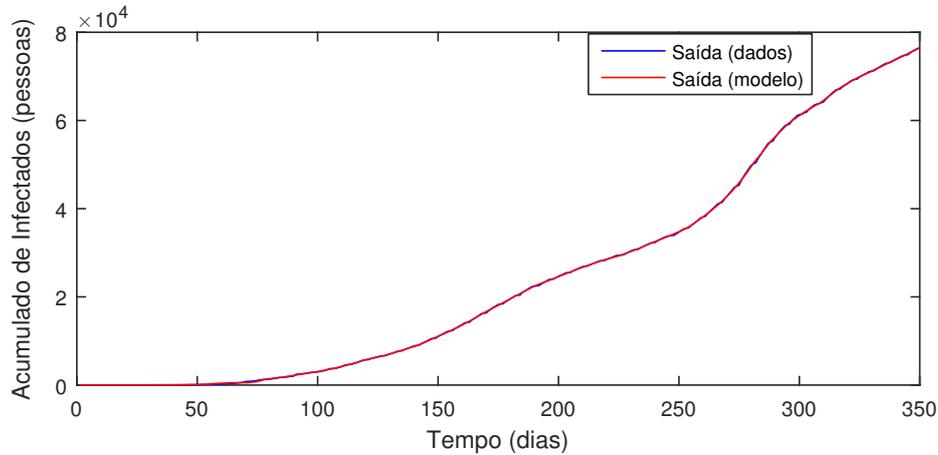


Figura 4.12: Modelo SIR para R23, R24, R25 e R26 - Caxias do Sul

Mais uma vez, observa-se a semelhança entre os dois gráficos sobrepostos. Assim, conclui-se que o modelo demonstrou precisão também nesse caso.

Outra análise possível é a de que o número acumulado de infectados começou a crescer mais precocemente do que nas regiões anteriores. Nota-se que a inclinação da curva aumenta significativamente antes do centésimo dia. Em contrapartida, nas regiões previamente analisadas, o crescimento mais acelerado foi observado apenas próximo ao dia 100. Dessa forma, esperava-se que as taxas identificadas indicassem valores mais elevados em semanas anteriores, em comparação aos casos observados anteriormente.

De fato, verifica-se, na Figura 4.13, que apresenta as taxas de infecção identificadas, chances consideráveis de infecção já na quarta semana. Além disso, percebe-se que as taxas permaneceram elevadas por um período significativo, atingindo um pico na décima primeira semana. Contudo, observaram-se quedas nas semanas seguintes, com estabilização dos valores, sem grandes oscilações a partir desse ponto.

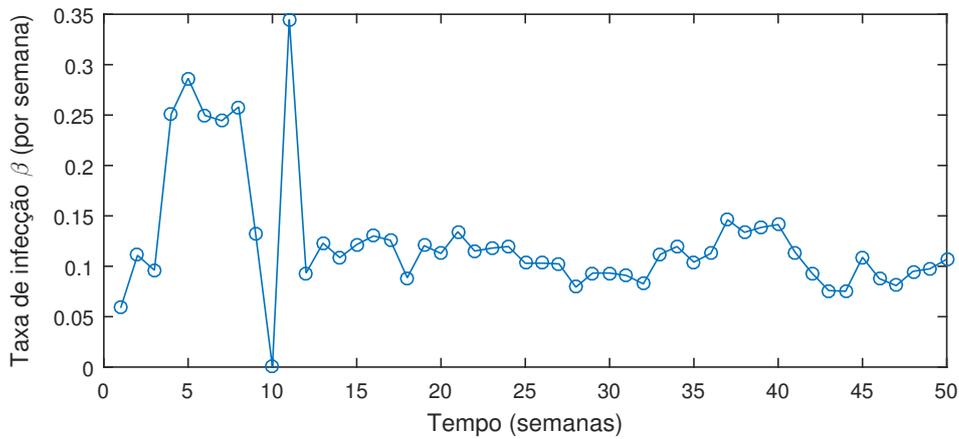


Figura 4.13: Taxas de Infecção para R23, R24, R25 e R26

4.3.6 Identificação para R12 - Cruz Alta

Esta região agrupa 13 municípios, tendo Cruz Alta como base do centro de saúde. A identificação realizada a partir dos dados desse centro reproduziu uma atividade semelhante à observada nos registros, conforme ilustrado na Figura 4.14.

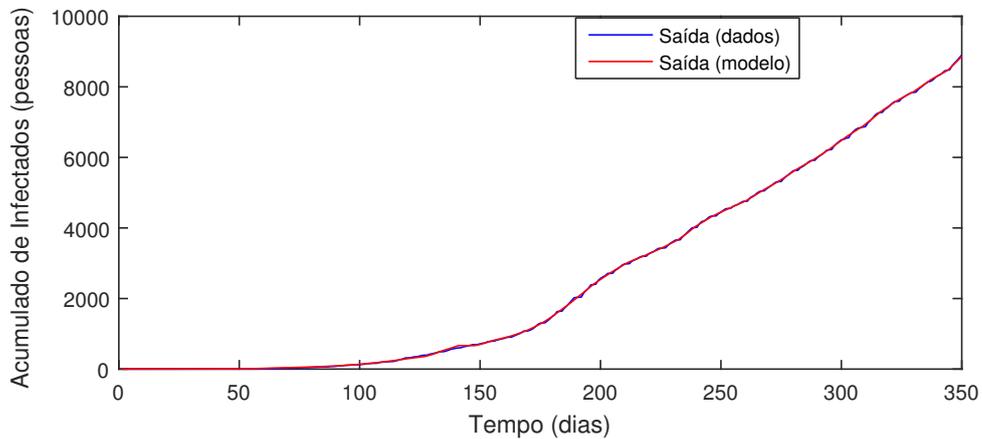


Figura 4.14: Modelo SIR para R12 - Cruz Alta

Ao analisar a forma como a taxa de infecção se alterou semanalmente, observa-se que houve um aumento nas primeiras semanas, atingindo um pico na oitava semana. Em seguida, possivelmente devido às medidas adotadas pelas auto-

ridades para conter o contágio, registrou-se uma queda nos valores. Uma nova alta acentuada foi observada nas 19^a e 22^a semanas. Posteriormente, ocorreu outra queda na 25^a semana, que se manteve até a 31^a, quando os casos voltaram a crescer. A partir desse ponto, a amplitude do gráfico manteve-se baixa, indicando pouca variação nos valores da taxa nas semanas seguintes.

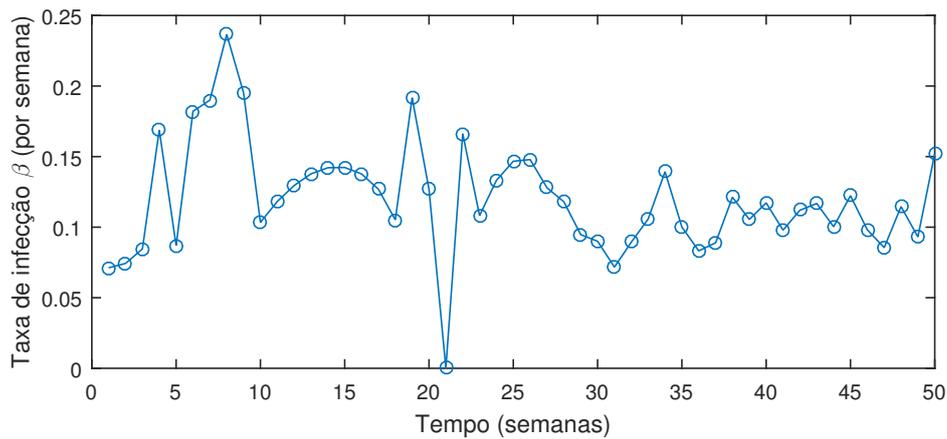


Figura 4.15: Taxas de Infecção para R12

4.3.7 Identificação para R16 - Erechim

Localizada no norte do estado, a região R16 abrange 34 cidades, tendo em Erechim sua base de saúde. Com as informações obtidas nesse centro, realizou-se a identificação do vetor das taxas de infecção, permitindo a plotagem dos gráficos do acumulado de infectados. Ao confrontar as curvas geradas a partir do modelo e dos dados oficiais, como ilustrado na Figura 4.16, observa-se uma grande semelhança nos comportamentos, o que confirma a eficácia da identificação.

Desse modo, também é possível analisar o desenvolvimento das taxas identificadas ao longo das semanas consideradas, como apresentado na Figura 4.17. No gráfico, nota-se a ocorrência de períodos de aumento significativo na taxa de infecção, que atingiu seu pico na décima quarta semana. Esse aumento pode ser

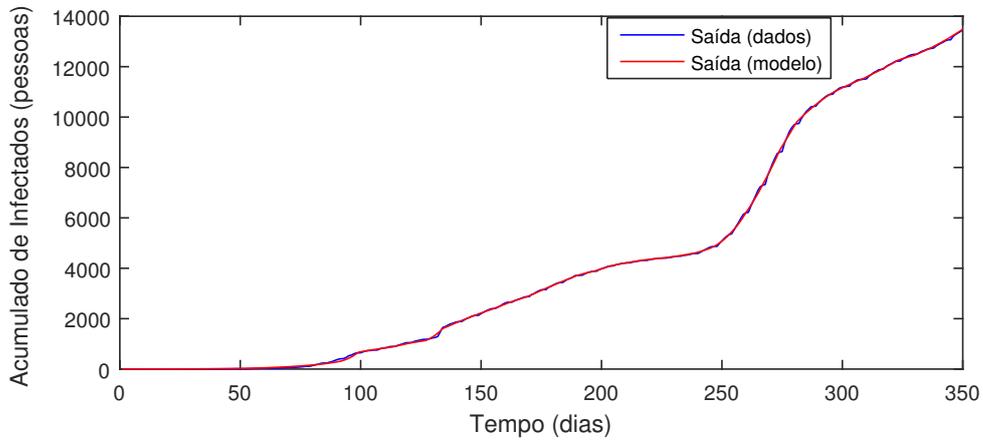


Figura 4.16: Modelo SIR para R16 - Erechim

corroborado pela aceleração observada nos casos acumulados no gráfico anterior. Por exemplo, observa-se que o crescimento acentuado na 32^a semana resultou em uma curva com inclinação mais acentuada no total de acumulados algumas semanas depois, por volta do 250^o dia.

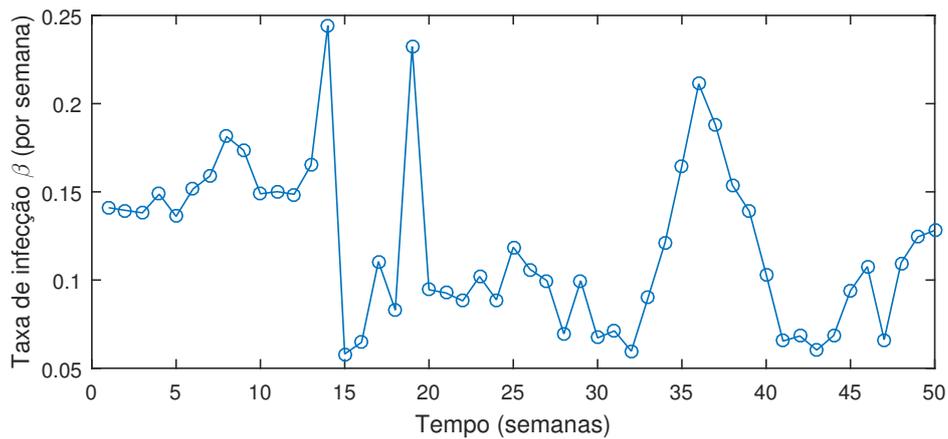


Figura 4.17: Taxas de Infecção para R16

4.3.8 Identificação para R09 - Guaíba

Nesta região, estão agrupadas 19 cidades, incluindo Guaíba, onde se localiza o centro de saúde responsável pelo registro dos dados epidemiológicos. As taxas de infecção semanais identificadas a partir dessas informações geraram uma saída para o modelo, cujo comportamento se aproximou consideravelmente dos dados reais. Essa semelhança pode ser observada na Figura 4.18, que apresenta uma comparação entre os dois gráficos.

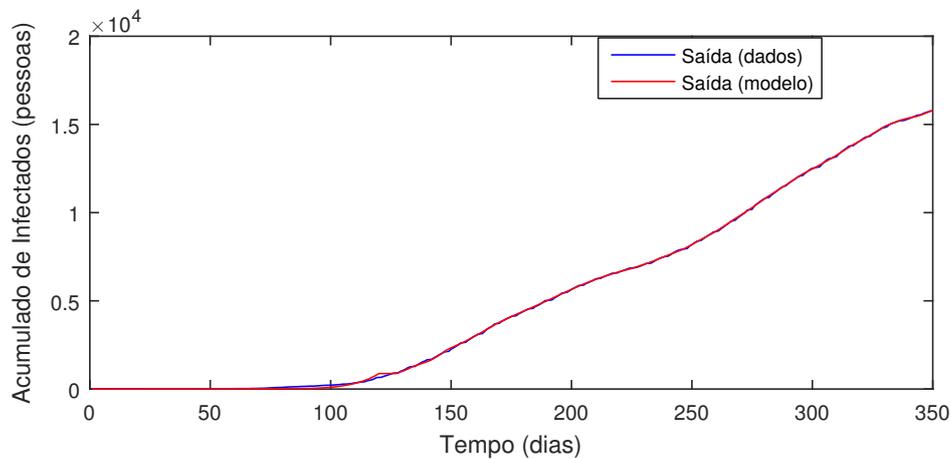


Figura 4.18: Modelo SIR para R09 - Guaíba

Os valores identificados permitem, ainda, uma análise detalhada das taxas semanais e de suas variações ao longo do tempo. A Figura 4.19 ilustra essa dinâmica, evidenciando a evolução da pandemia na região. Percebe-se que o período de maior probabilidade de infecção ocorreu por volta da 14^a semana. Entre a 10^a e a 20^a semana, observou-se uma maior oscilação nas taxas de infecção. Após esse intervalo, os valores tornaram-se mais estáveis, indicando uma diminuição nas flutuações da propagação do vírus.

A estabilidade observada nas semanas seguintes pode refletir tanto o impacto de intervenções de saúde pública quanto a adaptação da população às medidas preventivas.

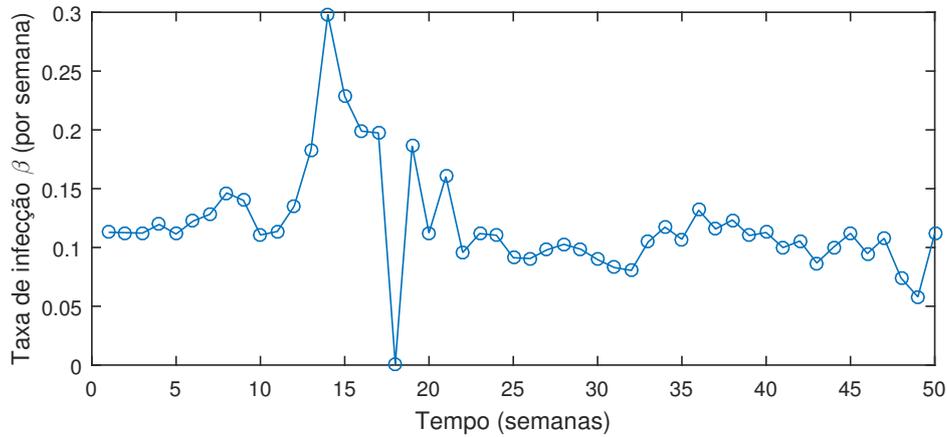


Figura 4.19: Taxas de Infecção para R09

4.3.9 Identificação para R13 - Ijuí

Com o centro de saúde base localizado em Ijuí, a região R13 abrange 20 municípios do noroeste gaúcho. Os dados coletados nesse centro permitiram a identificação do modelo, que conseguiu representar com boa precisão a dinâmica da pandemia nessa zona de agrupamento. A comparação entre o resultado do modelo e o comportamento real do número total de infectados pode ser conferida na Figura 4.20.

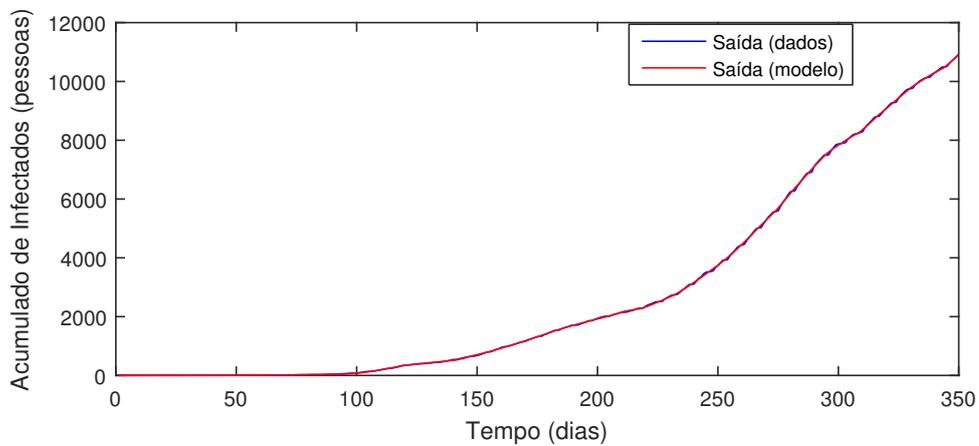


Figura 4.20: Modelo SIR para R13 - Ijuí

O gráfico que apresenta a evolução das taxas de infecção identificadas está ilustrado na Figura 4.21. Observa-se que a taxa começou a apresentar uma ascensão significativa a partir da décima semana, atingindo o pico de infecção na décima quarta semana, quando então passou a declinar. Posteriormente, houve um novo crescimento próximo à vigésima semana, embora a amplitude das variações tenha se tornado menor nos períodos subsequentes.

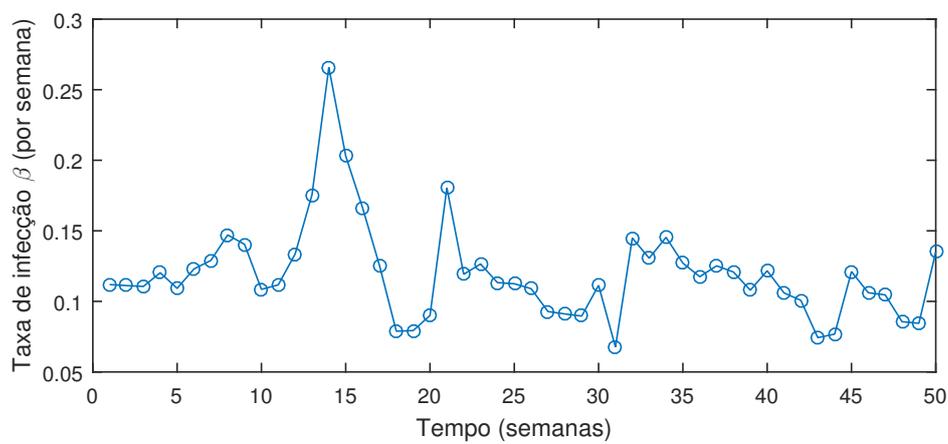


Figura 4.21: Taxas de Infecção para R13

4.3.10 Identificação para R29,30 - Lajeado

A região é composta por 37 cidades, dentre as quais Lajeado, onde se localiza a base dos centros de saúde desses municípios. Com base nos dados registrados nessa unidade, realizou-se a identificação do modelo, obtendo-se um resultado compatível com a situação real da pandemia nesse território, conforme apresentado na Figura 4.22.

Para compreender a dinâmica do contágio nessas cidades, verifica-se a evolução das taxas de infecção identificadas, ilustrada na Figura 4.23. Observa-se que, nessa região, os valores das taxas já eram elevados nas primeiras semanas, o que indica um possível surto de COVID-19 nesse período. Esse fato pode explicar

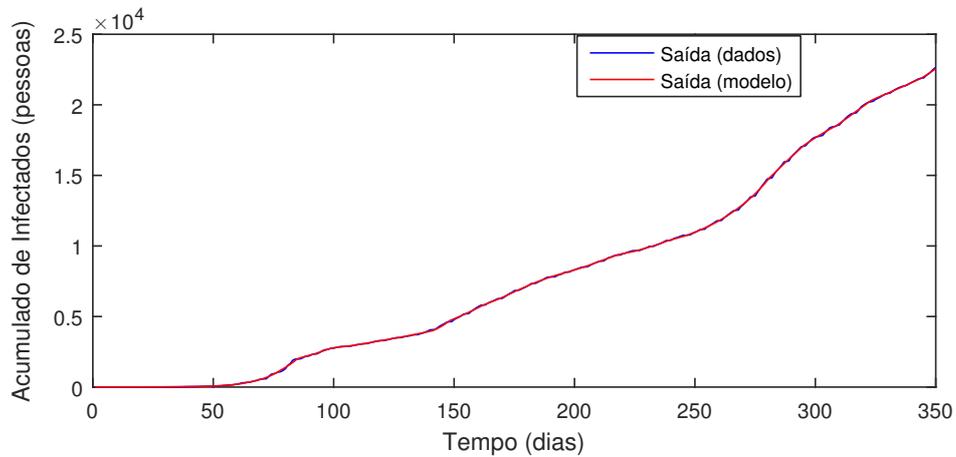


Figura 4.22: Modelo SIR para R29 e R30 - Lajeado

o crescimento precoce do número acumulado de infecções, por volta do 50^o dia, conforme destacado na Figura 4.22. A taxa apresentou uma queda significativa a partir da 12^a semana, mas voltou a crescer de forma acentuada na 20^a semana. A partir de então, as taxas demonstraram um comportamento mais estável, com menor amplitude entre os valores.

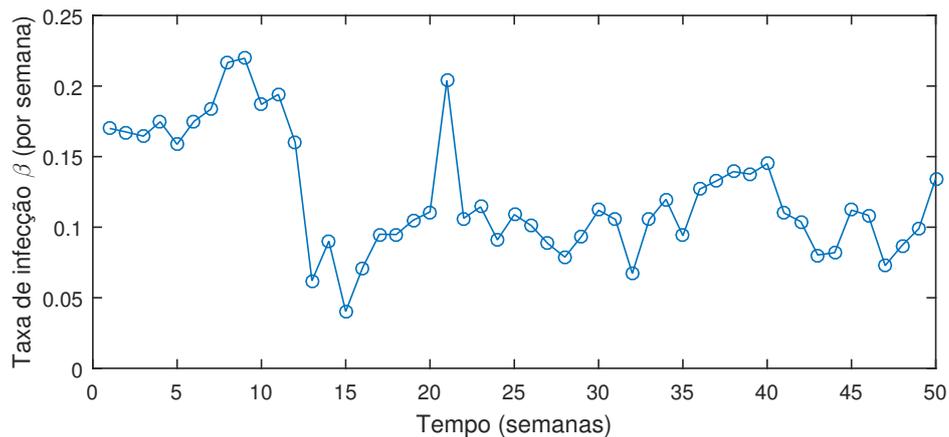


Figura 4.23: Taxas de Infecção para R29 e R30

4.3.11 Identificação para R07 - Novo Hamburgo

A zona de agrupamento R07, cujo centro de saúde base está localizado em Novo Hamburgo, abrange 15 cidades. Os casos registrados nessa região possibilitaram a identificação dos parâmetros do modelo, de forma que o comportamento simulado correspondesse de maneira precisa ao observado na realidade. O gráfico comparativo entre os dois comportamentos está apresentado na Figura 4.24.

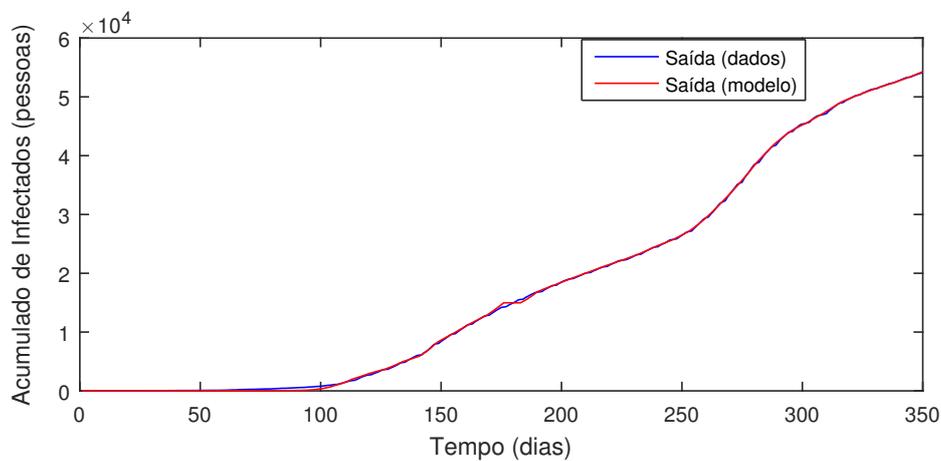


Figura 4.24: Modelo SIR para R07 - Novo Hamburgo

A evolução das taxas de infecção identificadas ao longo do tempo é exibida na Figura 4.25. Observa-se no gráfico que o maior crescimento dessa taxa ocorreu a partir da 10^a semana, quando houve um aumento expressivo nos valores, até atingir um pico na 14^a semana. Após esse período, observou-se uma redução nas taxas, que permaneceram relativamente estáveis até o final do período analisado.

Essa estabilidade posterior pode indicar uma desaceleração na propagação do vírus, possivelmente em função da adoção de medidas de controle mais eficazes ou de uma mudança no comportamento da população.

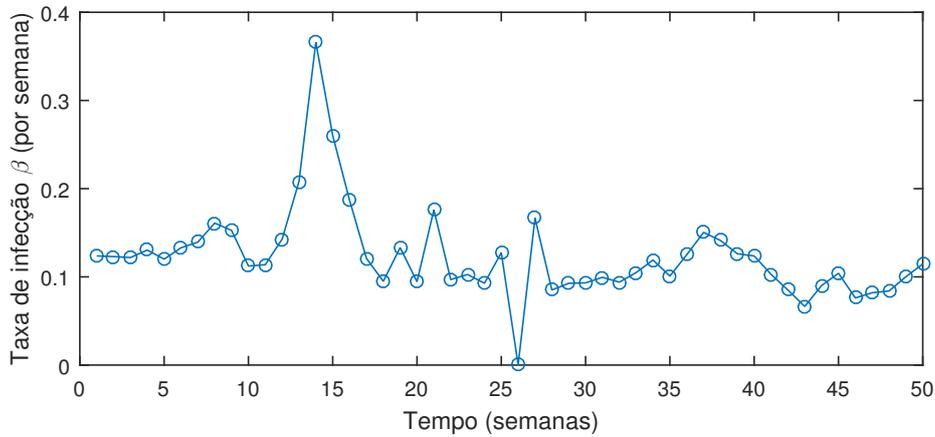


Figura 4.25: Taxas de Infecção para R07

4.3.12 Identificação para R15, 20 - Palmeira das Missões

Formada por 52 cidades do norte do estado, esta região tem em Palmeira das Missões o centro de saúde base, no qual estão registrados os dados sobre as infecções nesses municípios. Com base nessas informações públicas, determinou-se o vetor β , que contém as taxas de infecção semanais, com o objetivo de simular a atividade da COVID-19 nessa zona de agrupamento. O resultado obtido foi satisfatório, pois foi possível observar um comportamento bastante semelhante entre a simulação e os dados reais, conforme ilustrado na Figura 4.26.

Na Figura 4.27, observa-se a variação da taxa de infecção ao longo das semanas analisadas. Percebe-se que o ponto de maior probabilidade de contágio ocorreu na 14^a semana. Após esse pico, os níveis de contágio permaneceram substancialmente mais baixos até o final do período estudado.

A redução nos níveis de infecção após o pico pode sugerir a eficácia de medidas de controle implementadas ou uma diminuição natural na propagação do vírus. Essa diminuição gradual é importante para entender a evolução da pandemia e planejar futuras intervenções de saúde pública.

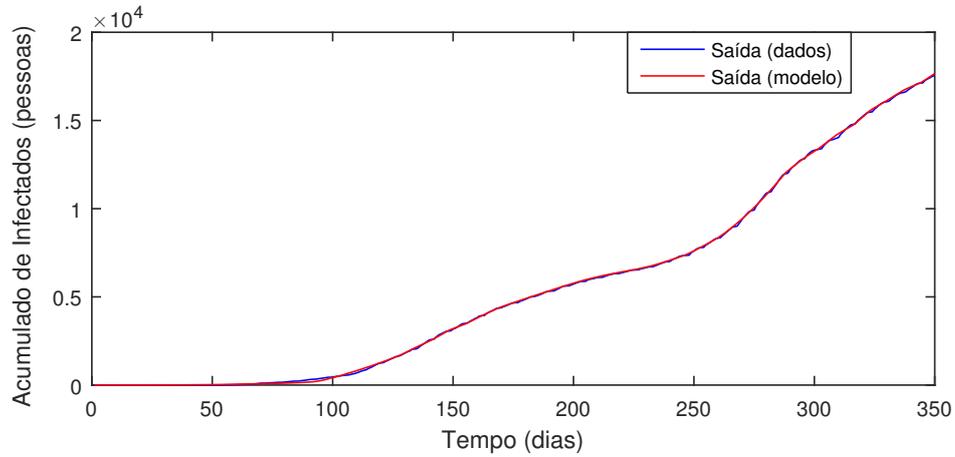


Figura 4.26: Modelo SIR para R15 e R20 - Palmeira das Missões

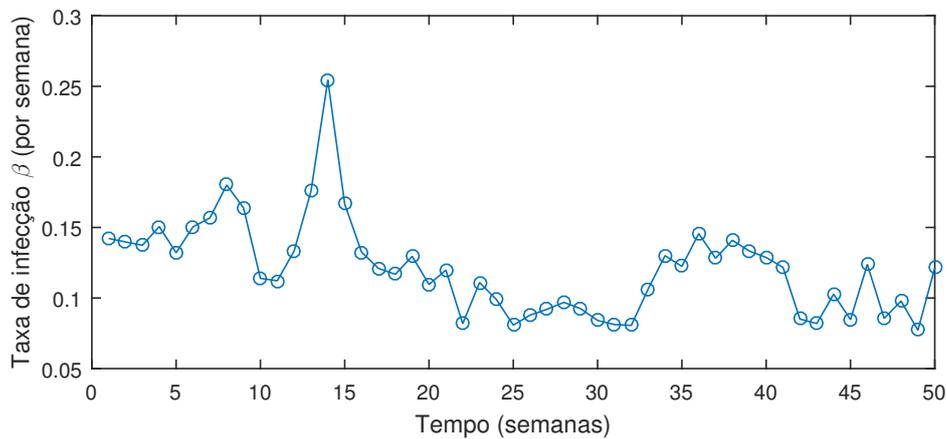


Figura 4.27: Taxas de Infecção para R15 e R20

4.3.13 Identificação para R17, 18, 19 - Passo Fundo

Esta região abrange informações de 63 cidades, cujos dados são centralizados e registrados no centro de saúde base localizado em Passo Fundo. A análise conduzida para essa área permitiu a identificação de parâmetros do modelo epidemiológico, possibilitando a simulação da evolução da pandemia. Os resultados obtidos demonstraram grande aderência à realidade, o que pode ser constatado na Figura 4.28, onde as curvas geradas pelo modelo e pelos dados reais apresentam sobreposição quase completa. Esse nível de precisão reforça a confiabilidade do método

de identificação utilizado, validando-o como ferramenta robusta para o estudo de cenários epidemiológicos.

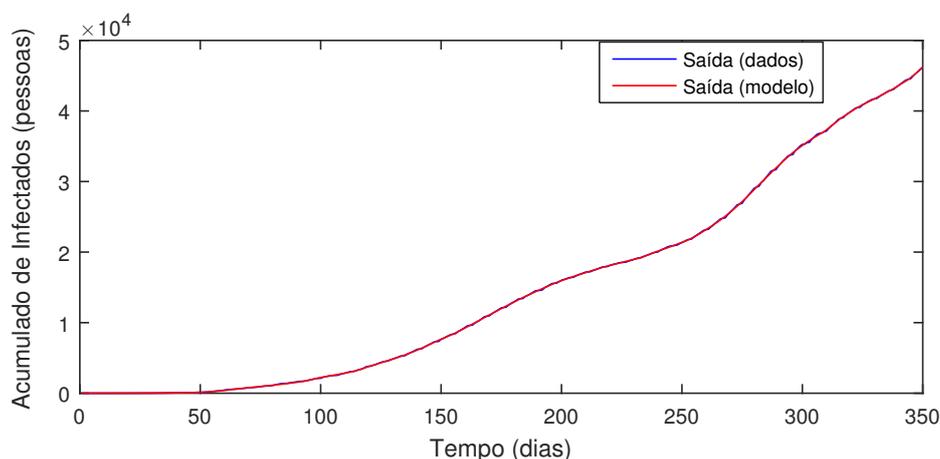


Figura 4.28: Modelo SIR para R17, R18 e R19 - Passo Fundo

No que diz respeito à dinâmica das taxas de infecção identificadas, representadas na Figura 4.29, observa-se que a maior probabilidade de contágio concentrou-se nas primeiras semanas analisadas, alcançando o ápice na 8^a semana. Esse comportamento sugere que a região enfrentou um período inicial de disseminação intensa, possivelmente influenciado por fatores como a ausência de medidas de contenção efetivas ou elevada taxa de transmissibilidade inicial do vírus.

A partir da 10^a semana, os dados indicam uma estabilização das taxas de infecção, com a ausência de picos expressivos nos valores subsequentes. Tal estabilidade pode estar relacionada à implementação de intervenções sanitárias, como distanciamento social, uso de máscaras e campanhas de conscientização, além do aumento da imunidade na população após o pico inicial. No entanto, mesmo com essa redução, é importante ressaltar a necessidade de monitoramento constante, dado que oscilações, ainda que menores, podem ocorrer em resposta a novos surtos ou relaxamento das medidas preventivas.

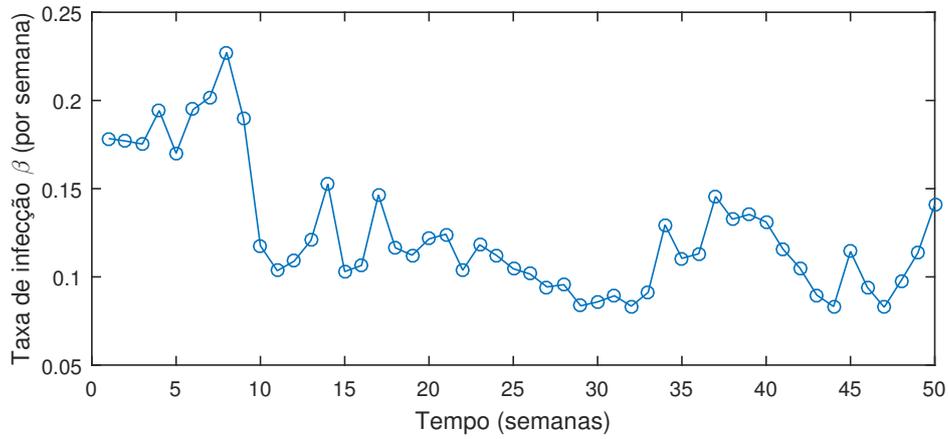


Figura 4.29: Taxas de Infecção para R17, R18 e R19

4.3.14 Identificação para R21 - Pelotas

Contando com 22 municípios localizados no sul do estado, a região R21 possui em Pelotas a base para o registro das informações relacionadas à pandemia nesses locais. A partir desses dados, foi possível calcular as taxas semanais de infecção e modelar o comportamento da doença. A Figura 4.30 apresenta a comparação entre os resultados da simulação e os dados reais observados.

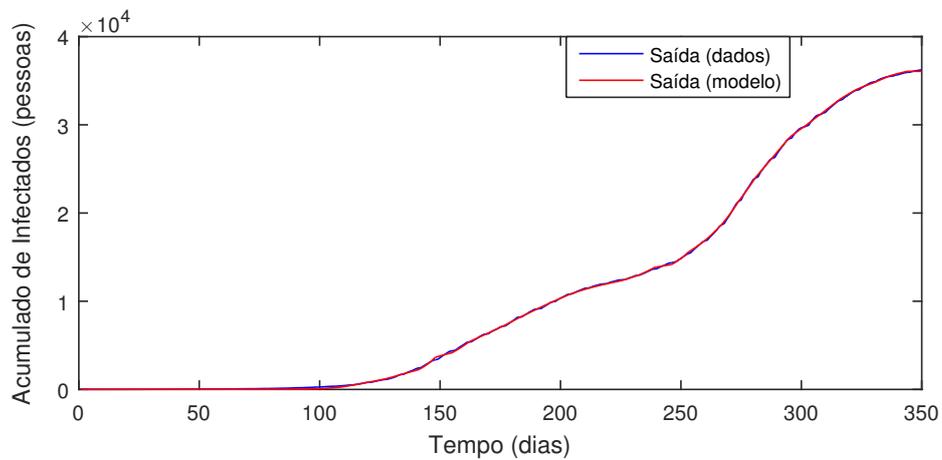


Figura 4.30: Modelo SIR para R21 - Pelotas

Conforme se verifica na figura, a precisão do modelo é evidenciada pela proximidade entre as curvas simuladas e reais. Assim, passa-se a analisar a variação das taxas ao longo dos 350 dias, conforme ilustrado na Figura 4.31. O gráfico demonstra que, apesar de um início mais estável, houve um elevado índice de contágio por volta da 14^a semana, momento em que a pandemia atingiu seu pico na região. Após esse período, ainda que tenham ocorrido outros episódios de alta, a taxa manteve-se significativamente abaixo do valor máximo registrado.

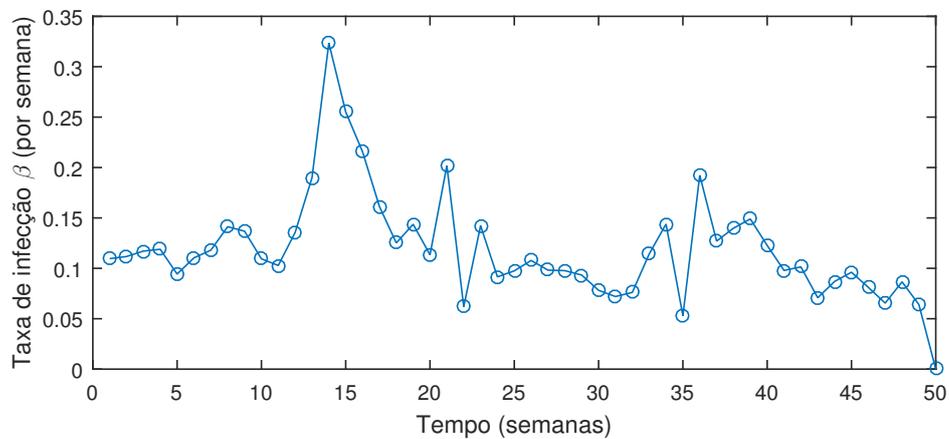


Figura 4.31: Taxas de Infecção para R21

4.3.15 Identificação para R10 - Porto Alegre

Porto Alegre, juntamente com outras 5 cidades de sua região metropolitana, compõe a região R10, cujos dados sobre a pandemia são sistematicamente catalogados na capital gaúcha. Esses dados permitiram identificar o vetor das taxas de infecção semanais, fornecendo subsídios para a análise da precisão das simulações em comparação com o comportamento real do número total de casos. A Figura 4.32 apresenta a correspondência entre as curvas simuladas e os dados observados, evidenciando a eficácia do modelo na representação da realidade. Essa proximidade

reforça a confiabilidade do método empregado para projetar a evolução da pandemia na região.

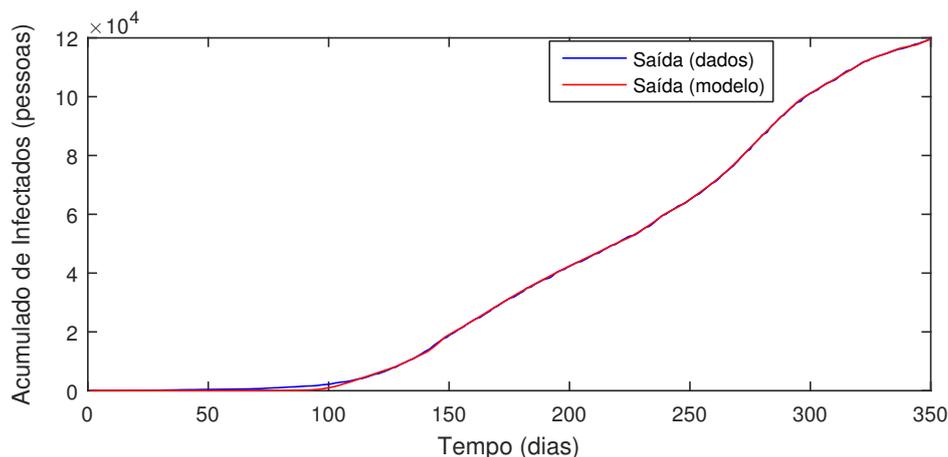


Figura 4.32: Modelo SIR para R10 - Porto Alegre

No intuito de compreender a dinâmica das taxas semanais de infecção, é analisado o gráfico da Figura 4.33. Observa-se que, nas semanas iniciais, as taxas permaneceram relativamente baixas, refletindo uma fase de controle inicial da disseminação. No entanto, a partir da 10^a semana, verifica-se um crescimento acentuado na probabilidade de contaminação, que culmina no ápice registrado na 14^a semana. Após esse pico, a taxa de infecção apresentou uma queda significativa, estabilizando-se em patamares relativamente baixos até o final do período avaliado.

Esses padrões sugerem a possível influência de intervenções sanitárias e comportamentais no controle da disseminação do vírus, ainda que outros fatores possam ter contribuído para essa dinâmica.

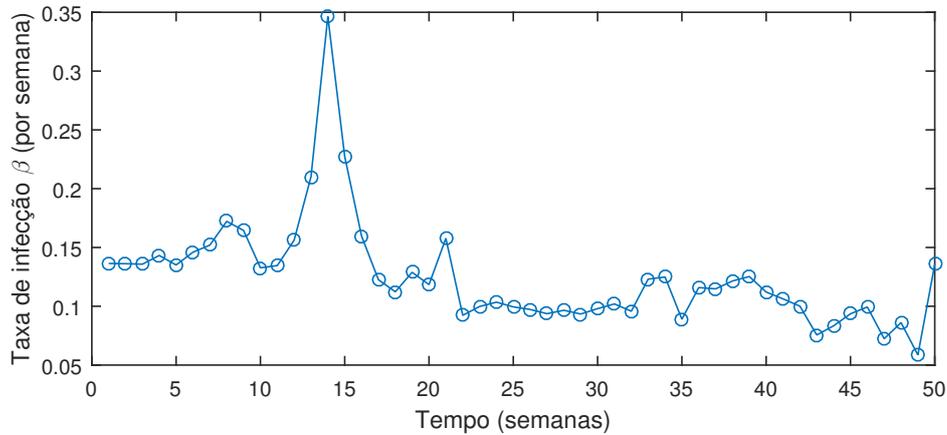


Figura 4.33: Taxas de Infecção para R10

4.3.16 Identificação para R28 - Santa Cruz

Esta zona de agrupamento abrange 13 municípios localizados no Vale do Rio Pardo, tendo em Santa Cruz do Sul a sede dos centros de saúde responsáveis pelo gerenciamento e catalogação das informações relacionadas à pandemia. As taxas semanais de infecção identificadas para esses territórios permitiram realizar simulações da atividade da doença e comparar os resultados com o desenvolvimento real dos casos na região. A Figura 4.34 ilustra a correspondência entre os dados simulados e os observados, destacando a precisão do modelo empregado para representar a evolução da pandemia nesse agrupamento territorial.

A partir das taxas identificadas, foi possível analisar como ocorreu o processo de infecção nessa região. O cenário apresentado é caracterizado por uma considerável oscilação, indicando pouca estabilidade na propagação da doença ao longo do período avaliado. Nas semanas iniciais, observou-se uma maior probabilidade de contaminação, com o pico das taxas registrado na 8ª semana. Apesar das flutuações semanais, a amplitude das variações manteve-se relativamente limitada, sugerindo que, embora os índices tenham oscilado, o comportamento geral da doença não foi marcado por mudanças extremas. Esse padrão pode refletir a interação

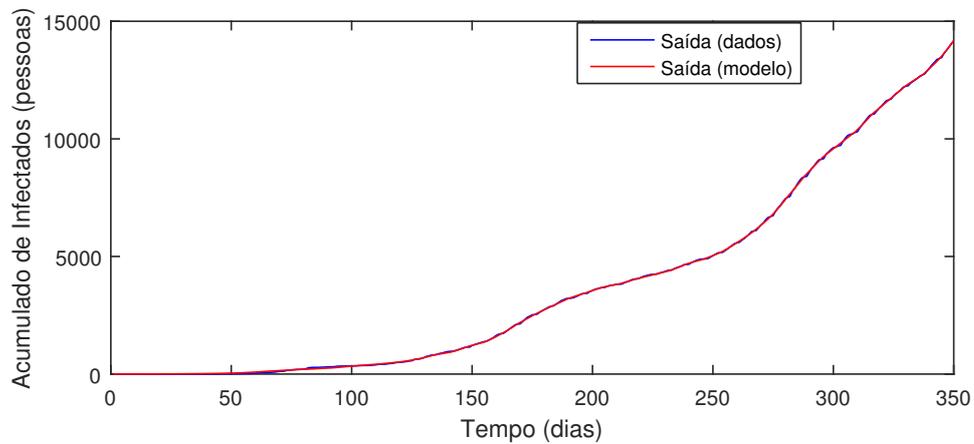


Figura 4.34: Modelo SIR para R28 - Santa Cruz

entre fatores locais, como medidas de contenção, comportamento populacional e características epidemiológicas específicas.

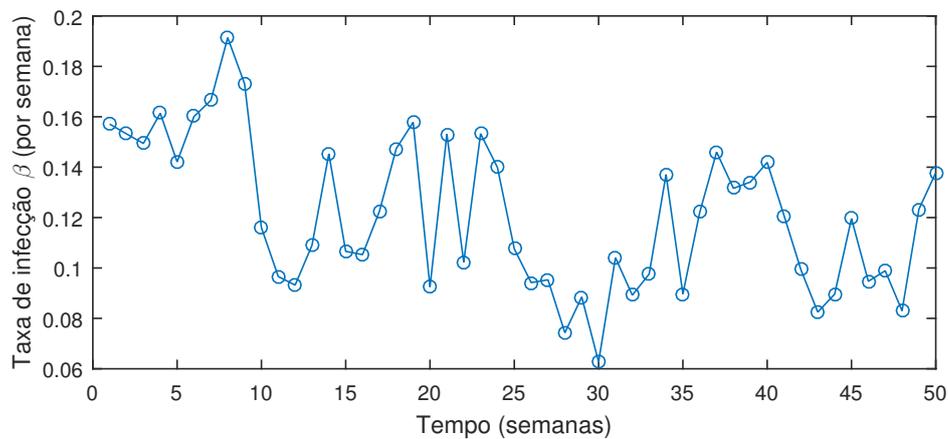


Figura 4.35: Taxas de Infecção para R28

4.3.17 Identificação para R01,02 - Santa Maria

Na área central do estado, encontram-se as 32 cidades que integram a zona de agrupamento de Santa Maria, onde está situada a sede base dos centros de saúde responsáveis pelo gerenciamento das ações de saúde pública nesses municípios.

Os dados coletados nessas localidades foram utilizados para calcular as taxas de infecção semanais da região, possibilitando a simulação do comportamento do novo coronavírus. A Figura 4.36 ilustra a correspondência entre os resultados da simulação e os dados reais, evidenciando a precisão do modelo na representação da dinâmica da doença.

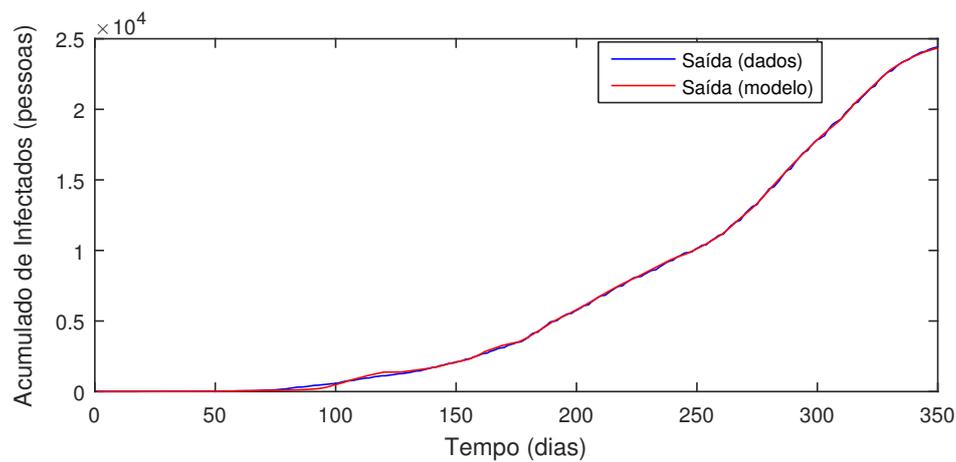


Figura 4.36: Modelo SIR para R01 e R02 - Santa Maria

No que se refere às taxas identificadas, verifica-se um crescimento mais acentuado a partir da 11^a semana, culminando no pico registrado na 14^a semana. Posteriormente, observa-se uma queda abrupta nesses valores, o que pode indicar a implementação de medidas de contenção por parte das autoridades de saúde. Após esse declínio, a taxa volta a apresentar crescimento na 18^a semana, estabilizando-se em um comportamento mais regular a partir desse ponto. A Figura 4.37 apresenta graficamente essas mudanças nos valores, oferecendo uma visão clara das flutuações ocorridas ao longo do período analisado.

Esses padrões sugerem que a adoção de estratégias de contenção pode ter exercido papel significativo na redução inicial das taxas de infecção, enquanto o crescimento posterior pode refletir possíveis relaxamentos nas medidas ou alterações no comportamento da população. A análise contínua desse tipo de dado é funda-

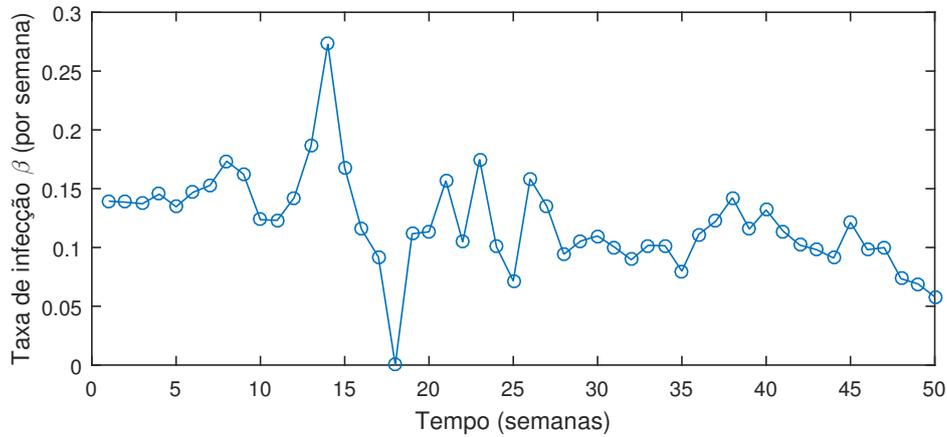


Figura 4.37: Taxas de Infecção para R01 e R02 - Santa Maria

mental para compreender a efetividade das políticas públicas implementadas e para a definição de estratégias futuras no enfrentamento de pandemias.

4.3.18 Identificação para R14 - Santa Rosa

No noroeste do estado está localizada a região R14, composta por 22 municípios cujos dados sobre a população infectada foram coletados e registrados. A sede dos Centros de Saúde responsáveis pelo monitoramento e catalogação das informações encontra-se em Santa Rosa. A partir desses dados, e com o uso de um algoritmo de otimização, foi possível determinar as taxas de infecção semanais, que serviram de base para a simulação do comportamento da doença nesse território. A Figura 4.38 apresenta a correspondência entre as curvas geradas pela simulação e pelos dados reais, evidenciando que o modelo foi capaz de reproduzir, com eficiência, a dinâmica do número total de infectados.

Além de reproduzir a evolução dos casos totais, o vetor identificado permite analisar a variação da taxa de infecção ao longo das semanas avaliadas, conforme ilustrado na Figura 4.39. O gráfico evidencia que o período mais crítico da pandemia na região ocorreu na 14^a semana, quando as taxas atingiram seu pico.

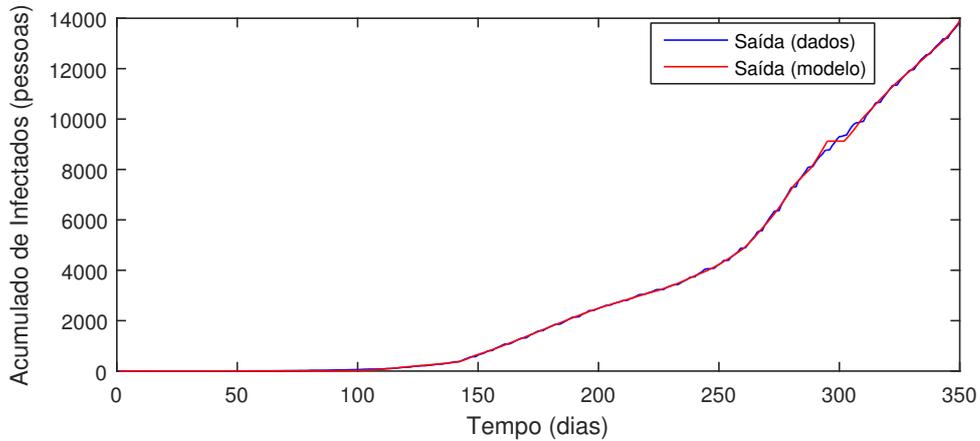


Figura 4.38: Modelo SIR para R14 - Santa Rosa

Um segundo aumento significativo foi observado na 22^a semana, embora de menor intensidade em comparação ao anterior. Após esse ponto, a propagação da doença apresentou um comportamento mais estável, com oscilações moderadas no restante do período analisado.

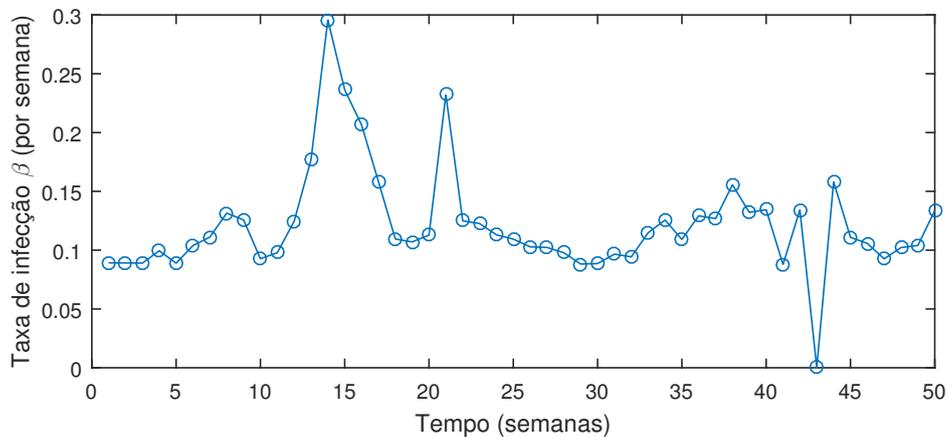


Figura 4.39: Taxas de Infecção para R14

Esses padrões sugerem que a região enfrentou duas ondas distintas de aumento na taxa de infecção, provavelmente relacionadas a eventos específicos ou alterações nas medidas de controle. A estabilidade registrada no período final pode

indicar o impacto positivo de intervenções sanitárias mais consistentes ou mudanças no comportamento populacional.

4.3.19 Identificação para R11 - Santo Ângelo

A região R11 compreende 24 municípios localizados nas Missões e no oeste do estado do Rio Grande do Sul. A quantidade total de infectados nessas localidades, ao longo das semanas analisadas, foi informada pela base dos Centros de Saúde, situada em Santo Ângelo. A partir desses dados, foi calculado o vetor β , que representa a taxa de infecção semanal da região, com o objetivo de descrever a evolução dos casos totais registrados. A Figura 4.40 apresenta a comparação entre os resultados obtidos pela identificação dos parâmetros e o gráfico gerado com base nos dados reais. Essa comparação evidencia que o modelo conseguiu reproduzir, com eficácia, o comportamento da pandemia na região.

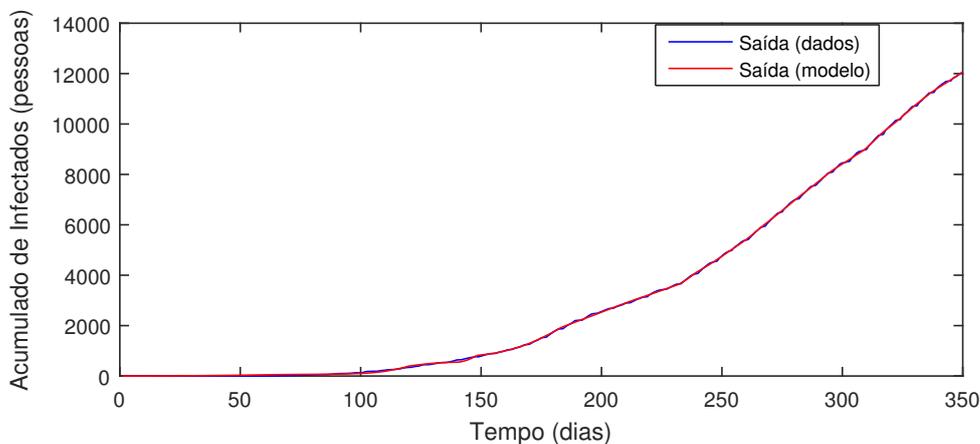


Figura 4.40: Modelo SIR para R11 - Santo Ângelo

Em relação às taxas identificadas, detalhadas na Figura 4.41, verifica-se que a região enfrentou dois períodos de crescimento mais acentuado: o primeiro ocorreu entre a 11^a e a 16^a semanas; o segundo, entre as semanas 20 e 21, sendo neste último momento registrado o ápice da probabilidade de infecção. Após esse período,

o comportamento da pandemia apresentou maior regularidade, com oscilações menos intensas nas taxas de infecção.

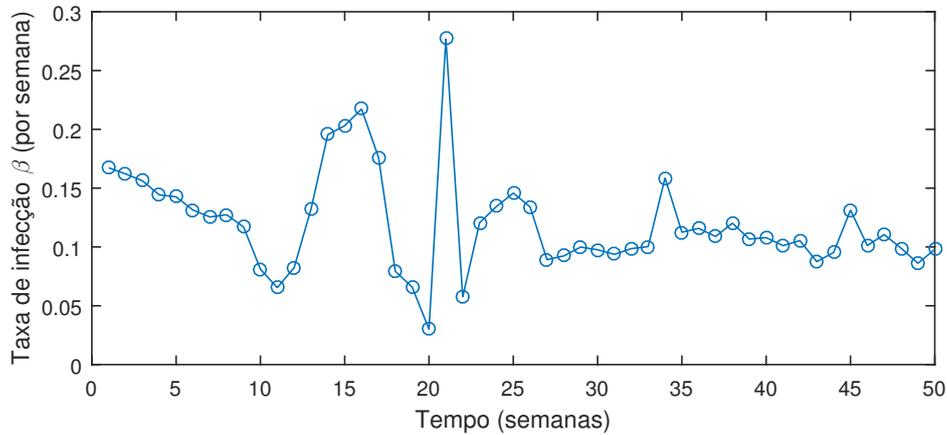


Figura 4.41: Taxas de Infecção para R11

Esses padrões sugerem que a região experimentou uma fase inicial de aceleração na transmissão do vírus, seguida de uma segunda onda de crescimento, de menor duração, antes de atingir maior estabilidade. As diferenças nas curvas podem refletir fatores como alterações no comportamento da população, mudanças nas políticas de controle sanitário ou a introdução de intervenções específicas.

4.3.20 Identificação para R06 - Taquara

Na zona de agrupamento R06 estão localizadas 8 cidades, sendo Taquara a sede dos Centros de Saúde responsáveis por esses municípios. Os casos registrados na população dessa região foram utilizados para determinar o vetor β , representando as taxas de infecção semanais. Com base no vetor identificado, foi possível simular o modelo epidemiológico e comparar o número total de infectados estimado pelo modelo com os dados reais registrados. A Figura 4.42 ilustra essa comparação, evidenciando a correspondência entre os valores simulados e os observados.

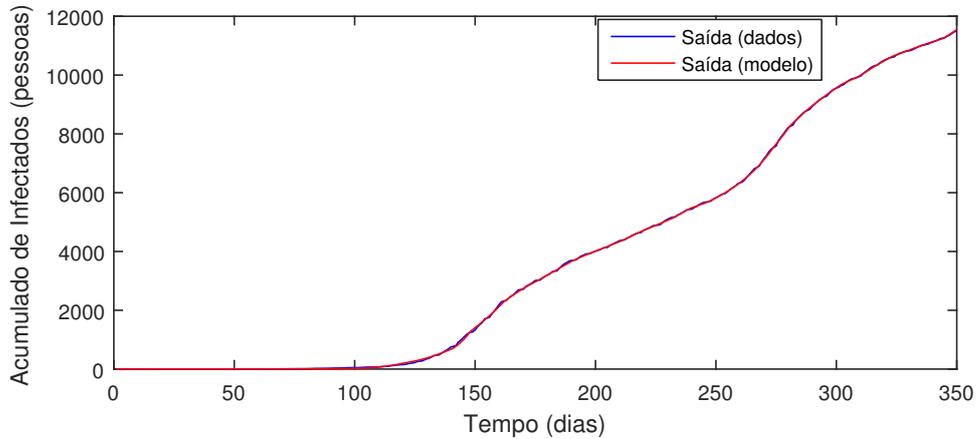


Figura 4.42: Modelo SIR para R06 - Taquara

A semelhança entre os comportamentos observados confirma a eficiência do modelo em reproduzir a dinâmica real da pandemia na região. Assim, procede-se à análise das taxas calculadas, conforme demonstrado no gráfico da Figura 4.43. Observa-se que, após um início com pouca variação, a região experimentou um aumento significativo na taxa de infecção a partir da 10^a semana, atingindo o pico da pandemia na 14^a semana. Posteriormente, a taxa de infecção apresentou uma tendência de estabilização, indicando um controle mais eficaz da disseminação da doença.

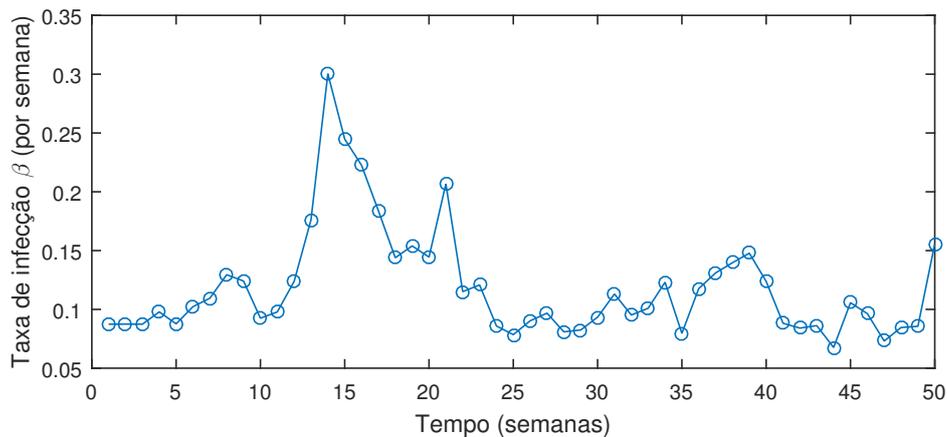


Figura 4.43: Taxas de Infecção para R06

4.3.21 Identificação para R03 - Uruguaiana

A última região analisada corresponde a 11 cidades localizadas na fronteira oeste do Rio Grande do Sul. Uruguaiana é a sede dos Centros de Saúde responsáveis pelo registro e catalogação dos dados epidemiológicos desses municípios, pertencentes à zona de agrupamento R03. Com base nas informações da população infectada, foi determinado o vetor β , que representa as taxas de infecção semanais. Essas taxas permitiram a simulação da atividade da COVID-19 nessa região e sua comparação com os dados reais registrados. A Figura 4.44 demonstra que o modelo reproduziu de maneira eficiente a dinâmica da doença, evidenciando sua precisão e confiabilidade.

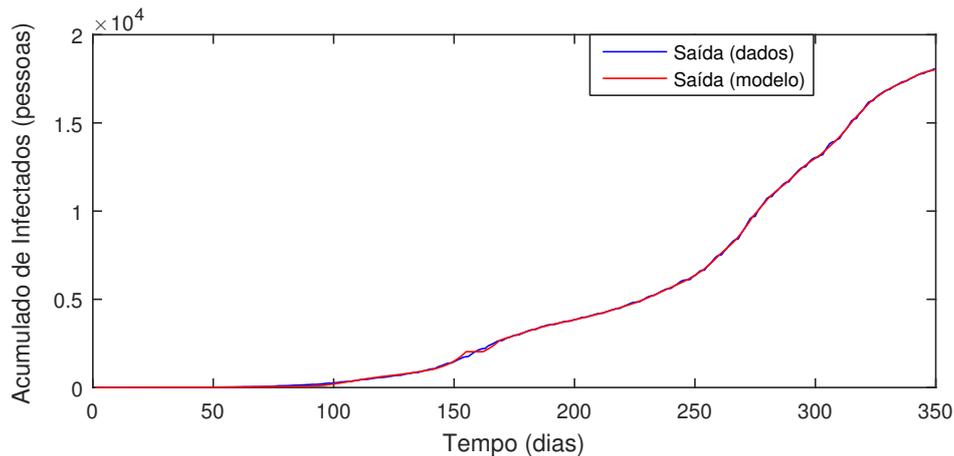


Figura 4.44: Modelo SIR para R03 - Uruguaiana

O estudo da dinâmica das taxas de infecção semanal está ilustrado no gráfico da Figura 4.45. A análise do vetor β revela que a pandemia atingiu seu ponto mais crítico na 14^a semana, com o maior índice de contaminação registrado. Além disso, entre a 20^a e a 25^a semana, foram observados outros períodos de elevação, possivelmente relacionados a eventos ou fatores que impulsionaram a transmissão do vírus. Após esse intervalo, as taxas de infecção apresentaram pouca variação, mantendo-se relativamente estáveis até o término do período analisado.

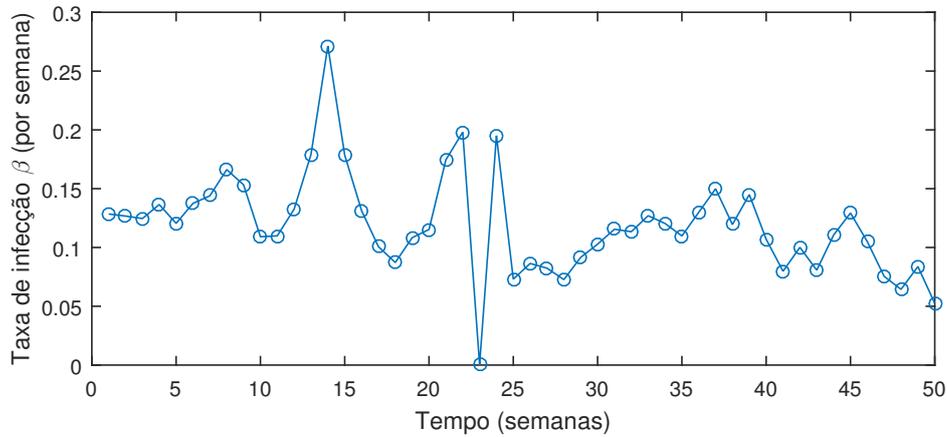


Figura 4.45: Taxas de Infecção para R03

Essa estabilização nas taxas pode indicar o impacto de medidas de controle adotadas na região ou uma possível imunização parcial da população em decorrência de infecções prévias. A análise contínua das taxas é essencial para avaliar a eficácia das intervenções e identificar tendências que possam orientar futuras estratégias de saúde pública.

A análise de todas as regiões ressalta a relevância de se considerar a dinâmica local da pandemia na formulação de estratégias mais direcionadas e eficazes de controle. Observa-se que o comportamento das taxas de infecção apresenta variações significativas entre as regiões, mesmo dentro de um único estado. Esses resultados indicam que abordagens homogêneas podem não ser suficientes para lidar com as especificidades epidemiológicas de cada localidade.

A heterogeneidade observada reflete fatores diversos, como densidade populacional, infraestrutura de saúde, adesão às medidas preventivas e características socioeconômicas. Assim, políticas públicas fundamentadas em dados regionais permitem intervenções mais assertivas, que consideram tanto as particularidades de cada comunidade quanto a evolução temporal da pandemia.

Além disso, a análise regional destaca a importância do monitoramento contínuo e da aplicação de modelos matemáticos para prever cenários futuros. Essa abordagem não apenas auxilia no controle da disseminação do vírus, mas também otimiza o uso de recursos e aumenta a eficácia das medidas de mitigação.

4.4 Resultado da Identificação para o Estado

Logo após verificar a precisão do modelo em descrever a atividade da doença nas regiões a cada semana, avaliou-se sua eficácia em retratar o comportamento da COVID-19 no estado como um todo. Para esse propósito, considerou-se um período abrangente de 770 dias, equivalentes a 110 semanas. A análise, nesse caso, foi refinada para a escala diária, permitindo uma observação mais detalhada da propagação da doença ao longo do tempo. Para isso, calculou-se as taxas de infecção diárias, possibilitando o acompanhamento das variações na transmissão em cada dia específico. Esse processo resultou na determinação de um vetor β com 770 entradas, cujos valores iniciais foram atribuídos utilizando o processo iterativo previamente descrito.

Com a identificação desses parâmetros, verificou-se que, mesmo nesse nível mais detalhado de análise, o modelo demonstrou alta capacidade de reproduzir fielmente a evolução real do número de casos totais registrados no estado. As curvas simuladas e observadas mostraram-se praticamente sobrepostas, corroborando a eficácia do modelo. Tal resultado pode ser visualizado na Figura 4.46, que evidencia a acurácia do algoritmo ao descrever o comportamento epidemiológico da doença.

Além disso, a dinâmica de variação das taxas de infecção diárias, ilustrada na Figura 4.47, revelou oscilações ao longo do período estudado. Entre os padrões observados, destaca-se um aumento significativo entre o 650^o e o 700^o dia, indicando uma aceleração no ritmo de novos casos. Uma possível explicação para esse crescimento acentuado pode ser atribuída ao surgimento de novas variantes

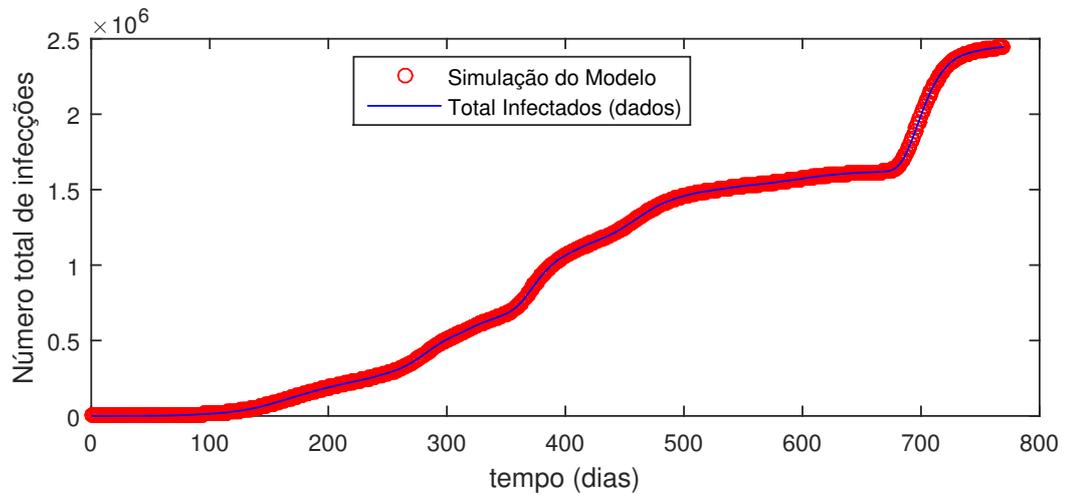


Figura 4.46: Comportamento Real x Simulado

do SARS-CoV-2, que, nesse intervalo, apresentaram maiores índices de transmissibilidade. Essas variantes, identificadas em diversos estudos realizados no período, foram amplamente reconhecidas por modificar a dinâmica da pandemia, reforçando a importância de estratégias adaptativas de controle.

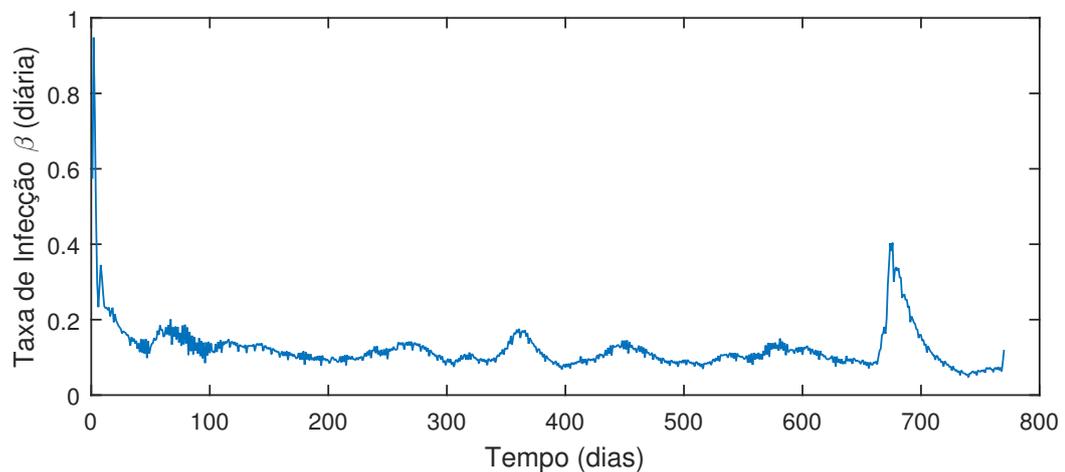


Figura 4.47: Taxas de Infecção diárias para o RS

Com a identificação dos parâmetros do modelo, também foi realizada uma simulação numérica do sistema de equações que o configura. Essa abordagem

permitiu analisar detalhadamente a evolução das diferentes classes epidemiológicas ao longo dos 770 dias analisados. A Figura 4.48 ilustra esse comportamento, evidenciando as dinâmicas entre as categorias da população (suscetíveis, infectados e recuperados). Essa análise fornece uma visão abrangente sobre o impacto temporal da pandemia, permitindo a validação do modelo frente aos dados observados.

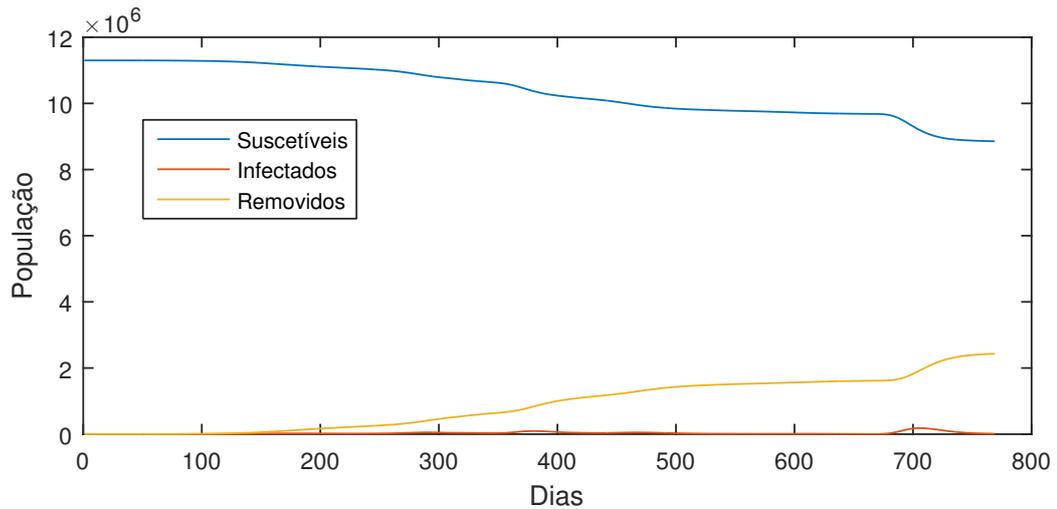


Figura 4.48: Simulação do Modelo com β variante no tempo

A partir dos dados disponíveis, foi possível calcular as taxas de infecção diárias, possibilitando um acompanhamento detalhado do comportamento e da evolução das infecções ao longo do período. O modelo demonstrou boa precisão em retratar a dinâmica da pandemia nos dias passados, fornecendo informações importantes para compreender os padrões históricos da disseminação da COVID-19. Entretanto, cabe questionar: essa técnica, além de explicar os eventos passados, pode ser utilizada para prever o comportamento da doença no futuro?

Essa indagação motivou a investigação sobre a capacidade preditiva do modelo. Considerou-se a possibilidade de utilizar os parâmetros identificados e as tendências observadas para projetar a atividade da COVID-19 em períodos subsequentes aos já registrados. Tal análise poderia oferecer subsídios para a formulação de estratégias de controle e mitigação. No próximo capítulo, será apresentado um

estudo aprofundado sobre essa possibilidade, explorando as potencialidades e as limitações do modelo na previsão de cenários futuros.

5 PREDIÇÃO DO MODELO

A predição por meio de modelos matemáticos constitui uma área fundamental em diversas disciplinas, como física, engenharia, economia e biologia. Por meio de equações matemáticas, busca-se descrever e prever o comportamento de sistemas com base em dados conhecidos ou em suposições sobre o seu funcionamento [10]. No contexto da epidemiologia, prever a dinâmica de uma doença infecciosa pode auxiliar as autoridades na antecipação de cenários e na adoção de medidas preventivas necessárias.

Para obter resultados precisos, é essencial compreender que a modelagem preditiva exige a observância de etapas indispensáveis. As condições para a predição incluem o desenvolvimento de um modelo matemático que represente com precisão o sistema estudado. Nesse processo, torna-se imprescindível definir as variáveis relevantes para o problema e realizar adaptações ou simplificações, quando necessário. Além disso, é fundamental validar a precisão do modelo, comparando as simulações geradas com dados reais. Essa validação permite avaliar se o modelo é capaz de representar adequadamente o comportamento do sistema [20].

Neste estudo, empregou-se um modelo matemático com parâmetro variante no tempo para descrever a dinâmica de uma epidemia em um período passado, no qual os dados observacionais estavam disponíveis para a identificação dos parâmetros do modelo. Com base na análise retrospectiva, foi determinado o comportamento temporal do parâmetro ao longo da fase histórica da epidemia.

Para a projeção futura da epidemia, utilizou-se o último valor calculado desse parâmetro como aproximação para simular o comportamento da doença nos períodos subsequentes, para os quais não há dados disponíveis. Dessa forma, realizaram-se previsões para horizontes de 7, 14, 30, 60 e 180 dias, avaliando a ro-

bustez e a adequação dos resultados frente à ausência de dados futuros que permitam uma estimativa contínua do parâmetro.

5.1 Resultados

Um dos objetivos deste trabalho é avaliar a eficácia do modelo desenvolvido na previsão do comportamento de uma doença infecciosa. Para isso, foram utilizados dados observacionais referentes à pandemia de COVID-19 no estado do Rio Grande do Sul, com o propósito de validar a capacidade do modelo de representar, com precisão, a dinâmica da doença ao longo de um período de 770 dias. A partir dessa validação, busca-se, também, examinar o potencial do modelo para prever o comportamento da epidemia em intervalos temporais futuros.

O processo de verificação da eficácia do modelo em prever o comportamento da COVID-19 no estado ocorreu da seguinte forma: inicialmente, assumiu-se o 100^o dia da pandemia como ponto de partida, considerando que, a partir desse momento, não havia informações adicionais sobre a evolução da doença. Com base nisso, foi simulada a progressão da epidemia utilizando um parâmetro constante, identificado no modelo para o 100^o dia, até o final dos 770 dias. Os resultados simulados foram comparados aos dados reais observados para intervalos de 7, 14, 30, 60 e 180 dias à frente. Para cada intervalo, calculou-se o erro percentual e_T entre as previsões do modelo e os valores reais observados, dado por:

$$e_T = \frac{Y(t+T) - \hat{Y}(t+T, \beta)}{Y(t+T)},$$

onde $T = 7, 14, 30, 60, 180$ é o intervalo de tempo em que será feita a predição.

O mesmo procedimento foi repetido sucessivamente a partir do 101^o dia, seguido pelo 102^o dia, e assim por diante, até o 500^o dia da pandemia. Para cada horizonte de previsão futuro (7, 14, 30, 60 e 180 dias), foram realizadas 400 simulações, correspondentes aos diferentes pontos de partida. Em cada simulação,

calculou-se o erro percentual entre os valores projetados pelo modelo e os dados reais observados, permitindo uma avaliação detalhada da precisão do modelo ao longo dos diversos períodos.

Ao término de todas as simulações, os erros percentuais calculados para cada um dos períodos futuros foram organizados em histogramas. Para cada horizonte de previsão, determinaram-se o erro percentual médio e o desvio padrão dos valores obtidos. Ademais, comparou-se a distribuição dos erros, representada nos histogramas, com a curva gaussiana, o que possibilitou a análise da aderência dos erros à distribuição normal.

5.1.1 Predição para 7 dias ($T = 7$)

Nesta subseção, foi aplicado o método descrito anteriormente para avaliar a capacidade de predição do modelo em relação ao horizonte de 7 dias à frente. Essa avaliação consistiu na análise da distribuição dos erros percentuais entre os valores projetados pelo modelo e os dados observados, conforme ilustrado na Figura 5.1.

Os resultados obtidos indicaram que a distribuição dos erros percentuais apresentou uma semelhança notável com a distribuição normal, representada pela curva Gaussiana sobreposta ao histograma dos dados. Especificamente, a média dos erros percentuais foi de 0,0013, com um desvio padrão de 0,0108. Esses valores sugerem que a maior parte dos erros está concentrada em torno da média, que, por sua vez, é muito próxima de zero. Tal comportamento denota baixa variabilidade nos erros percentuais e reforça a adequação do modelo para prever, com elevada precisão, o cenário pandêmico uma semana à frente.

Adicionalmente, a proximidade da distribuição dos erros à curva gaussiana sugere que as incertezas no modelo podem ser atribuídas, em grande parte,

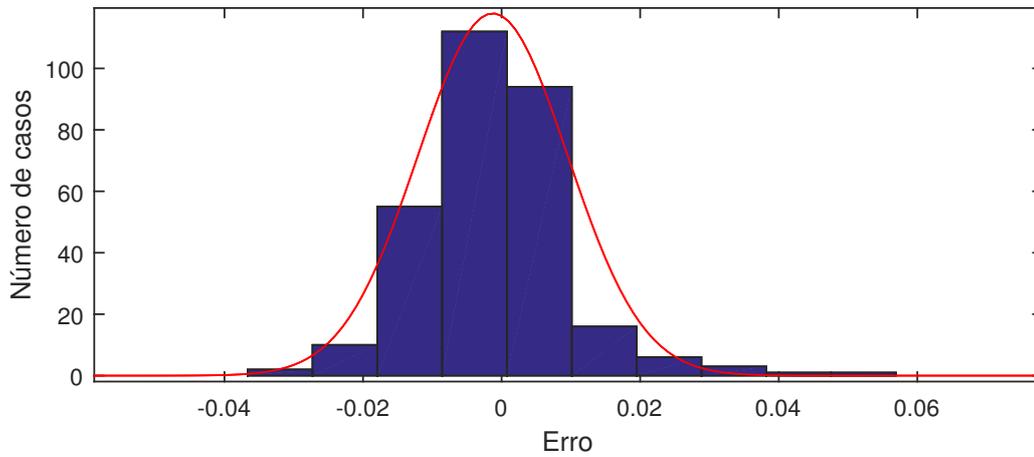


Figura 5.1: Erros percentuais para 7 dias no futuro

a fatores aleatórios e não a problemas estruturais na formulação do modelo matemático. Essa característica é desejável em modelos preditivos, pois implica que os desvios observados são previsíveis dentro de um padrão estatístico bem compreendido.

Desse modo, pode-se concluir que o modelo desenvolvido apresenta boa capacidade de predição para um intervalo de 7 dias, fornecendo informações úteis para o planejamento de ações e a alocação de recursos no contexto da pandemia.

5.1.2 Predição para 14 dias ($T = 14$)

Após verificar a eficácia do modelo em prever a evolução da pandemia para um horizonte de uma semana, buscou-se avaliar sua habilidade em prever a situação duas semanas no futuro. Esse cenário é ilustrado no gráfico da Figura 5.2, que apresenta a distribuição dos erros percentuais obtidos nas simulações realizadas.

Os resultados indicaram que o erro percentual médio foi de 0,006, enquanto o desvio padrão atingiu 0,031. Esses valores sugerem uma maior dispersão

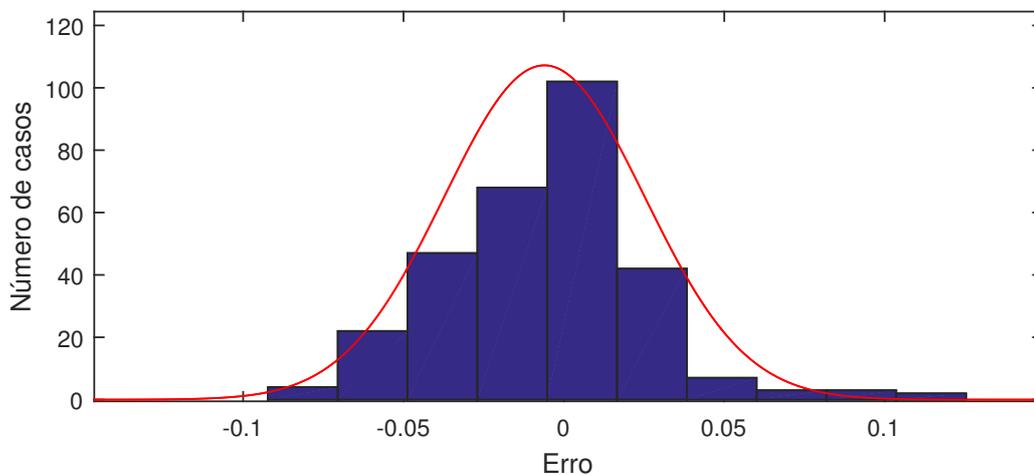


Figura 5.2: Erros percentuais para 14 dias no futuro

em relação ao horizonte de 7 dias, evidenciada por uma quantidade mais significativa de simulações cujo erro ultrapassou o intervalo de um desvio padrão da média. Ainda assim, a distribuição dos erros manteve alguma semelhança com a curva gaussiana, conforme observado no histograma.

A análise desses resultados revela que, embora o modelo tenha apresentado uma precisão ligeiramente inferior no horizonte de 14 dias, ele ainda demonstrou capacidade de capturar tendências gerais da evolução da pandemia nesse intervalo. Essa leve redução na eficiência preditiva pode ser atribuída à maior incerteza inerente a horizontes temporais mais amplos, que ampliam a possibilidade de eventos externos interferirem na dinâmica epidemiológica.

Portanto, conclui-se que, para a previsão de duas semanas, o modelo manteve desempenho satisfatório, mas sua eficácia foi moderadamente reduzida em comparação ao horizonte de uma semana. Esse comportamento é esperado em modelos preditivos devido ao acúmulo de incertezas à medida que se estende o intervalo de projeção.

5.1.3 Predição para 30 dias ($T = 30$)

O objetivo desta subseção foi avaliar a resposta do modelo ao simular a evolução da pandemia em um horizonte de 30 dias no futuro. Para isso, aplicou-se o procedimento descrito anteriormente, gerando os resultados apresentados na Figura 5.3, que ilustra a distribuição dos erros percentuais obtidos nas simulações e sua comparação com a curva normal.

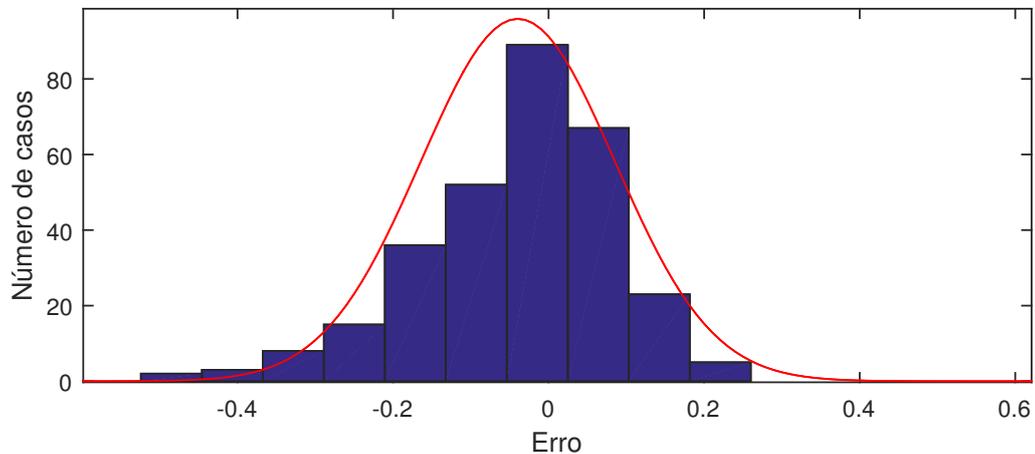


Figura 5.3: Erros percentuais para 30 dias no futuro

Os resultados revelaram um erro percentual médio de 0,039 e um desvio padrão de 0,125. Esses valores indicam que, embora a média dos erros tenha permanecido relativamente baixa, o desvio padrão elevado evidencia uma considerável dispersão nos erros. Isso significa que muitos valores simulados pelo modelo ficaram distantes da média, caracterizando uma variabilidade significativa nas previsões.

A análise da distribuição dos erros percentuais reforça essa conclusão, uma vez que foram observados numerosos casos com desvios substanciais em relação ao valor médio, indicando uma menor aderência à distribuição normal. Tal comportamento sugere que o modelo, ao tentar projetar a atividade da pandemia para um

horizonte de 30 dias, é sensível a fatores não capturados adequadamente por sua formulação.

Portanto, conclui-se que, nesse cenário, o modelo não apresenta um desempenho confiável para prever a dinâmica da pandemia em períodos de um mês. Essa limitação reflete a complexidade crescente em previsões de longo prazo, que estão mais suscetíveis a variáveis externas e eventos inesperados, destacando a necessidade de abordagens complementares ou refinamentos no modelo para melhorar sua precisão nesse horizonte temporal.

5.1.4 Predição para 60 dias ($T = 60$)

Nesta subseção, apresentam-se os resultados da simulação do modelo desenvolvido para prever a evolução da pandemia em um horizonte de 60 dias a partir do último ponto conhecido. O procedimento consistiu em calcular os erros percentuais de cada estimativa, comparando os valores simulados com os dados reais observados, conforme detalhado anteriormente.

Os resultados indicaram um erro percentual médio de aproximadamente 0,23 evidenciando um elevado grau de imprecisão nas projeções. Além disso, o desvio padrão calculado foi de 0,5332, um valor significativo que reflete a ampla variabilidade nos erros observados. Esses números demonstram que, ao considerar um horizonte temporal mais longo, o modelo apresenta maior dificuldade em representar adequadamente a dinâmica da pandemia.

A Figura 5.4 apresenta a distribuição dos erros percentuais obtidos, acompanhada de uma curva que representa a distribuição normal para comparação. Observa-se claramente que a distribuição dos erros percentuais não possui similaridade com a curva normal, caracterizando uma dispersão acentuada e um comportamento assimétrico em relação à média.

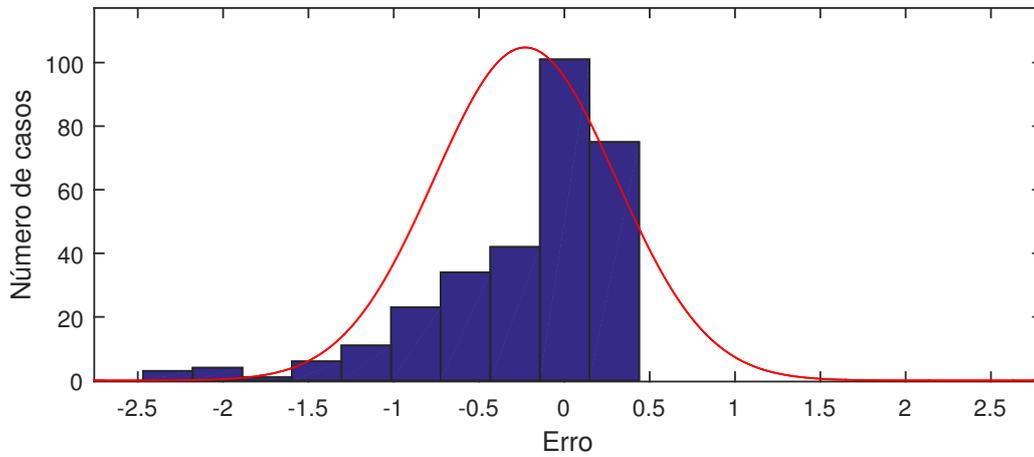


Figura 5.4: Erros percentuais para 60 dias no futuro

Esses resultados indicam que, para previsões de 60 dias, o modelo proposto não é eficiente em capturar a complexidade da evolução da pandemia. A alta imprecisão pode ser atribuída a fatores como mudanças inesperadas na dinâmica epidemiológica, intervenções externas não consideradas no modelo, ou limitações na abordagem matemática utilizada. Isso ressalta a necessidade de aprimorar os parâmetros que possam aumentar a capacidade preditiva do modelo para horizontes temporais mais longos.

5.1.5 Predição para 180 dias ($T = 180$)

Após constatar que o modelo não apresentou precisão suficiente para simular o comportamento da pandemia em horizontes de 1 ou 2 meses no futuro, realizou-se uma nova análise para avaliar sua capacidade de prever a dinâmica da doença em um período de 6 meses (180 dias) à frente. Os resultados dessa simulação estão apresentados na Figura 5.5, que exhibe a distribuição dos erros percentuais obtidos, acompanhada da curva gaussiana para comparação.

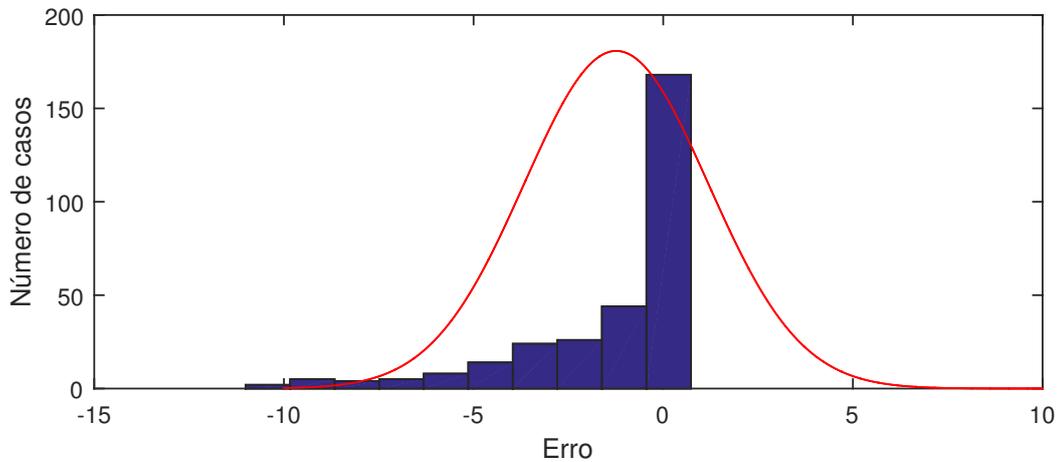


Figura 5.5: Erros percentuais para 180 dias no futuro

Os dados revelaram um erro percentual médio superior a 1,23, ou seja, 123%, indicando uma discrepância extremamente elevada entre as previsões do modelo e os dados reais. Além disso, o desvio padrão calculado foi próximo a 2,42 (243%), refletindo uma variação extremamente ampla nos erros. Esses valores deixam claro que o modelo, nesse horizonte temporal, não é capaz de representar com confiabilidade a evolução da pandemia.

A análise da Figura 5.5 reforça essa conclusão, uma vez que a distribuição dos erros percentuais não apresenta qualquer semelhança com a curva normal, evidenciando uma dispersão significativa e um comportamento altamente irregular. Tais resultados sugerem que a capacidade preditiva do modelo se degrada substancialmente em horizontes temporais mais extensos, provavelmente devido à amplificação de erros acumulados e à ausência de fatores externos relevantes no modelo, como novas intervenções de controle, mudanças de comportamento populacional ou mutações no agente infeccioso.

Portanto, conclui-se que o modelo proposto é inviável para previsões de 6 meses no futuro. Essa limitação decorre da incapacidade do parâmetro constante de capturar as mudanças dinâmicas que caracterizam a evolução de uma epidemia,

especialmente em contextos influenciados por intervenções das autoridades, mutações da doença ou alterações comportamentais na população.

6 CONCLUSÃO

O uso de modelos compartimentais simples, como o modelo *SIR*, é amplamente discutido na literatura científica. No entanto, os parâmetros desses modelos, geralmente representados por constantes positivas nas equações diferenciais que os definem, restringem sua aplicação a cenários específicos. Neste trabalho, foi aplicado um modelo compartimental *SIR* para descrever a dinâmica do novo coronavírus em um longo intervalo de tempo. Para tanto, modificou-se a formulação clássica do sistema de equações, ajustando o parâmetro da taxa de infecção. Em vez de assumir um valor constante, esse parâmetro foi definido como variável ao longo do tempo.

A modelagem adotou uma função constante por partes para a taxa de infecção, representada por um vetor em que cada elemento é uma constante específica para uma unidade de tempo analisada. Essa abordagem permitiu desenvolver um modelo capaz de descrever a evolução da COVID-19 em períodos prolongados para os quais existem dados observacionais.

A validação do modelo proposto foi realizada mediante a identificação do vetor de parâmetros. Para isso, utilizaram-se dados da população infectada pelo novo coronavírus no estado do Rio Grande do Sul. Foi aplicado um algoritmo de otimização baseado no método dos mínimos quadrados, que comparou a quantidade total de infectados registrada nos dados observacionais com as estimativas de infectados geradas pelo modelo desenvolvido.

Inicialmente, foi verificada a capacidade do modelo de descrever o comportamento semanal da doença, identificando-se um vetor de parâmetros que representa as taxas de infecção semanais. Para essa análise, compararam-se os resultados da otimização com os dados provenientes de 21 regiões de agrupamento, nas quais o estado do Rio Grande do Sul foi subdividido. A análise gráfica revelou que, em

todas as regiões, o modelo reproduziu com precisão a evolução do número total de casos, demonstrando sua eficácia. Adicionalmente, as taxas semanais em cada zona de agrupamento foram apresentadas, permitindo identificar períodos de maior ou menor risco de contágio.

Em seguida, avaliou-se a capacidade do modelo de representar a atividade diária da COVID-19 em todo o estado. Essa análise baseou-se na comparação entre o gráfico gerado a partir dos dados oficiais e aquele simulado pelo modelo ajustado por otimização. Os resultados evidenciaram que o modelo reproduziu com eficiência o comportamento diário da doença, uma vez que as curvas resultantes permaneceram praticamente sobrepostas, comprovando a precisão da modelagem.

Neste trabalho, desenvolveu-se também um algoritmo iterativo inovador, projetado para inicializar o vetor de parâmetros a ser identificado com valores previamente ajustados, aproximando-os da solução desejada. Essa estratégia teve como objetivo mitigar o risco de convergência para mínimos locais, superando as dificuldades típicas de problemas de otimização não convexos. O algoritmo opera de maneira progressiva, aumentando gradualmente a dimensão do vetor de parâmetros ao longo das iterações. Em cada etapa, otimiza-se os parâmetros adicionados até alcançar a quantidade total previamente estabelecida. Essa abordagem assegurou elevada precisão na simulação, permitindo a identificação eficaz de um grande número de parâmetros.

Este trabalho também propôs o uso de um modelo simples e determinístico para prever o comportamento de uma epidemia. Para tal, adotou-se uma taxa de infecção constante, correspondente ao último valor identificado pelo modelo com taxa variável no tempo, com o objetivo de simular a atividade da epidemia em diferentes períodos futuros. Realizou-se uma análise estatística, cujos resultados indicaram que o modelo apresenta boa precisão até duas semanas à frente. Simulando a atividade da doença para 7 dias no futuro, obteve-se um erro médio inferior a 0,13%, o que evidencia sua capacidade preditiva. Para 14 dias, o erro percen-

tual gerado foi inferior a 0,6%, indicando que o modelo também é adequado para compreender o comportamento da pandemia nesse intervalo. Ressalta-se, contudo, que, para períodos mais longos, não se esperava boa precisão, motivo pelo qual foi proposto um modelo com taxa variável no tempo para tais casos.

Conclui-se, adicionalmente, que os procedimentos utilizados na identificação de parâmetros do modelo podem ser particularmente úteis para sistemas cuja dinâmica seja menos influenciada por fatores ambientais.

6.1 Trabalhos Futuros

A obtenção dos resultados deste trabalho abriu possibilidades para investigações mais aprofundadas na adaptação de modelos simples para o estudo de doenças epidemiológicas. Algumas direções para estudos futuros incluem:

6.1.1 Identificação de uma Taxa de Remoção Variante no Tempo

Nos resultados apresentados, fixou-se o valor de γ , parâmetro que mede a taxa de remoção da população infectada, com base no período médio em que um indivíduo permanece portador da doença. No entanto, é plausível que, em determinados períodos, esse tempo varie, podendo ser maior ou menor. Assim, pode-se adotar a mesma abordagem de considerar esse parâmetro como variante no tempo, permitindo a análise de condições nas quais a doença persista por mais tempo em indivíduos.

6.1.2 Modelagem com Vacinação

No final de 2020, diferentes vacinas começaram a ser desenvolvidas para conter a propagação da COVID-19. Com a implementação de políticas de vacinação específicas em cada localidade, surge a possibilidade de incluir um fator de vacinação nos modelos epidemiológicos. Por exemplo, no modelo *SIR* utilizado para descrever a dinâmica do novo coronavírus, pode-se estabelecer uma função $v(t)$ para representar a atividade de imunização no cenário a ser simulado.

A função $v(t)$ é responsável por levar uma parcela da população suscetível diretamente para a classe dos removidos. Nesse contexto, o modelo se caracteriza como:

$$\begin{cases} \frac{dS(t)}{dt} = -\frac{\beta(t)I(t)S(t)}{N} - v(t) \\ \frac{dI(t)}{dt} = \frac{\beta(t)I(t)S(t)}{N} - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) + v(t) \end{cases} \quad (6.1)$$

A função $v(t)$ pode ser definida de diversas formas. Uma abordagem é assumir que a relação entre a vacinação e a população suscetível seja proporcional, representada por δS , em que δ é a taxa de vacinação. Além disso, outros fatores, como a efetividade da vacina, podem ser incorporados ao modelo. Essa extensão permite determinar as taxas de infecção em populações parcialmente vacinadas e comparar os resultados com aqueles obtidos no presente estudo, no qual a vacinação não foi considerada.

6.1.3 Técnicas de Controle

Além da identificação de parâmetros, técnicas de controle podem ser ferramentas valiosas para aplicação em modelos epidemiológicos. Essas técnicas en-

volvem o estudo de estratégias que permitam determinar valores dos parâmetros de modo a projetar um cenário desejado. Por exemplo, é possível desenvolver um método que defina o valor da taxa de infecção β , de forma a garantir que a demanda sobre o sistema de saúde local permaneça abaixo de um limite preestabelecido.

Para essas aplicações, diferentes técnicas podem ser comparadas. Entre os métodos mais tradicionais, destaca-se o controle *PID* (Proporcional Integral Derivativo), amplamente utilizado em sistemas dinâmicos para ajuste contínuo. Métodos mais avançados, como o controle preditivo, também podem ser empregados. Esses métodos utilizam a otimização em tempo real para prever o comportamento do sistema e ajustar os parâmetros com maior precisão, considerando restrições e metas específicas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Center for disease control and prevention. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>, 2020.
- [2] European centre for disease prevention and control. Disponível em: <https://www.ecdc.europa.eu/en/\textsc{COVID}-19/latest-evidence/transmission>, 2020.
- [3] Painel coronavírus rs. Disponível em: <https://ti.saude.rs.gov.br/\textsc{COVID}19/>, 2020,2021.
- [4] Números do coronavírus. Disponível em: <https://especiais.gazetadopovo.com.br/coronavirus/numeros/>, 2021.
- [5] Protocolo de prevenção obrigatório dos modelos de distanciamento controlado do estado do rio grande do sul. disponível em: <https://distanciamentocontrolado.rs.gov.br>, 2021.
- [6] World health organization - who. clinical management of COVID-19: Interim guidance. Disponível em: <https://www.who.int/publications-detail/clinical-management-of-\textsc{COVID}-19>, 2021.
- [7] ANTON, H., BIVENS, I. C., AND DAVIS, S. L. *Cálculo: Um novo horizonte - Vol.2*, 8th ed. Bookman Companhia Ed., 2007.
- [8] ARROYO-MARIOLI, F., BULLANO, F., KUCINSKAS, S., AND RONDON-MORENO, C. Tracking r of COVID-19: A new real-time estimation using the kalman filter. *Plos One*, v. 16, nr. 0244474 (2021).
- [9] BAUER, F., AND CASTILLO-CHAVEZ, C. *Mathematical Models in Population Biology and Epidemiology*, second ed. Springer, 2010.
- [10] BENDER, E. A. *An Introduction to Mathematical Modelling*. Wiley, 1978.

- [11] BERNARD, O., HADJ-SADOK, Z., AND DOCHAIN, D. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology and Bioengineering* (2001).
- [12] BUCHTER, J. C. *Numerical Methods for Ordinary Differential Equations*, 2nd ed. John Wiley and Sons, 2008.
- [13] BURDEN, R. L., AND FAIRES, J. D. *Numerical Analysis*, 9th ed. Brooks/Cole Cengage Learning, 2010.
- [14] CALAFIORE, C. G., NOVARA, C., AND POSSIERI, C. A time-varying SIRD model for the COVID-19 contagion in italy. *Annual Reviews in Control*, v. 50, p. 361-372 (2020).
- [15] CERVELIN, B. H., AND EHRHARDT, M. A. D. Um método de minimização irrestrita baseado em derivadas simplex. *XXXIV Congresso Nacional de Matemática Aplicada Computacional* (2012).
- [16] CHENG, Y., LU, P., CHANG, C., AND LIU, T. A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Transactions on Network Science and Engineering*, vol. 7, p.3279-3294 (2020).
- [17] COOPER, I., MONDAL, A., AND ANTONOPOULOS, C. G. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons and Fractals, Volume 139* (2020).
- [18] D'ARENZIO, M., AND CONIGLIO, A. Assessment of the sars-cov-2 basic reproduction number, r_0 , based on the early phase of COVID-19 outbreak in italy. *Biosafety and Health*, v. 2, p. 57-59 (2020).
- [19] DOCHAIN, D. *Automatic Control of Bioprocesses*, first ed. John Wiley and Sons Inc, 2008.
- [20] DYM, C. *Principles Of Mathematical Modelling*, 2nd ed. Elsevier Academic Press, 2004.

- [21] GANDHI, R. T., LYNCH, J. B., AND DEL RIO, C. Mild or moderate COVID-19. *The New England Journal of Medicine*, 383 (2020), 1757–1766.
- [22] GODIO, A., PACE, F., AND VERGNANO, A. SEIR modeling of the italian epidemic of sars-cov-2 using computational swarm intelligence. *International Journal of Environmental Research and Public Health*, 3555 (2020).
- [23] HASELEIN, W. M., AND ECKHARD, D. Identification of a time-varying sir model for covid-19, 2024.
- [24] HASELEIN, W. M., AND ECKHARD, D. Identificação de um modelo SIR variante no tempo para COVID-19. *XXV Congresso Brasileiro de Automática. Rio de Janeiro, RJ - Brasil* (2024).
- [25] HETHCOTE, H. W. *Three Basic Epidemiological Models*. 1999.
- [26] IOANNIDIS, J. P. A., CRIPPS, S., AND TANNER, M. A. Forecasting for COVID-19 has failed. *International Journal of Forecasting*, v. 38, p. 423-438 (2022).
- [27] IRONS, J. N., AND RAFTERY, E. A. Estimating sars-cov-2 infections from deaths, confirmed cases, tests, and random surveys. *Proceedings of the National Academy of Sciences of the United States of America*, v. 118, n.2103272118 (2021).
- [28] ISERLES, A. *A first course in the Numerical Analysis of Differential Equations*, 2nd ed. Cambridge University Press, 2009.
- [29] IVORRA, B., R., F. M., VELA-PEREZ, M., AND RAMOS, A. M. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. the case of china. *Communications in Nonlinear Science and Numerical Simulation*, Volume 88, nr. 105303 (2020).
- [30] IZMAILOV, A. F., AND SOLODOV, M. V. *Newton Type Methods for optimization and Variational problems*, 1st ed. Springer, 2014.

- [31] JARDIM, C. L. T. F., FERREIRA, L. A. F., AND PRATES, D. B. Estudo da influenza A (H1N1) aplicada ao modelo SIRS. *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics* (2014).
- [32] JR, J. D., AND WOODS, D. E. Optimization on microcomputers: The nelder-mead simplex algorithm. *Technical Report, Workshop on Microcomputers in Delaware* (1985).
- [33] JUNIOR, M. L. D. P., DOS SANTOS, S. R., AND PERIÇARO, G. A. Metodos de newton e quase-newton para otimizacao irrestrita. *VIII Encontro de Produção Científica e Tecnológica* (2013).
- [34] JUSTO, D. A. R., SAUER, E., DE AZEVEDO, F. S., GUIDI, L. F., DOS SANTOS, M. C., AND DE ALMEIDA KONZEN, P. H. *Cálculo Numérico - Um livro colaborativo*, 1st ed. Universidade Federal do Rio Grande do Sul, 2016.
- [35] KALACHEV, L., GRAHAM, J., AND LANDGUTH, E. L. A simple modification to the classical sir model to estimate the proportion of under-reported infections using case studies in flu and COVID-19. *Infectious Disease Modelling*, v. 9, p.1147-1162 (2024).
- [36] LAGARIAS, J. C., REEDS, J. A., WRIGHT, M. H., AND WRIGHT, P. E. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM - Society of Industrial and Applied Mathematics*, Vol.9 pp.112-147 (1998).
- [37] LUIZ, M. H. R. *Modelos Matemáticos em Epidemiologia*. 2012.
- [38] MARTINEZ, J. M., AND RAMOS, S. A. *Métodos Computacionais de Otimização*, 1st ed. UNICAMP, 1998.
- [39] MOHAMADOU, Y., HALIDOU, A., AND KAPEN, P. T. A review of mathematical modeling, artificial intelligence and datasets used in the study, predic-

- tion and management of COVID-19. *Applied Intelligence*, v. 50, p. 3913-3925 (2020).
- [40] MOLTER, A., QUADROS, R. S., RAFIKOV, M., BUSKE, D., AND GONÇALVES, G. A. Mathematical modelling to perform an analysis of social isolation periods in the dissemination of COVID-19. *Trends in Computational and Applied Mathematics*, v. 22, p.1-14 (2021).
- [41] MURRAY, J. D. *Mathematical Biology I: An Introduction*, third ed. Springer, 2001.
- [42] NEUMAIER, A. *Introduction to Numerical Analysis*, 1st ed. Cambridge University Press, 2001.
- [43] POSTNIKOV, E. B. Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons and Fractals*, v. 135, nr. 109841 (2020).
- [44] RODA, W. C., VARUGHESE, M. B., HAN, D., AND LEE, M. Y. Why is it difficult to accurately predict the COVID-19 epidemic?, journal = Infectious Disease Modelling, Vol. 5, p.271-281, year = 2020,.
- [45] WANGPING, J., KE, H., YANG, S., WENZHE, C., SHENGSHU, W., SHANSHAN, Y., JIANWEI, W., FUYIN, K., T, P., JING, L., MIAO, L., AND YAO, H. Extended SIR prediction of the epidemics trend of COVID-19 in italy and compared with hunan, china. *Frontiers and Medicine*, v. 7, nr. 169 (2020).
- [46] WARNATZ, J., MAAS, U., AND DIBBLE, R. W. *Combustion: Physical and Chemical Fundamentals, Modeling and Simulation, Experiments, Pollutant Formation*, 4th ed. Springer-Verlag Berlin Heidelberg, 2006.
- [47] WEISSMAN, G. E., CRANE-DROESCH, A., CHIVERS, C., MIKKELSEN, M. E., AND HALPERN, S. D. Locally informed simulation to predict hospital capacity

needs during the COVID-19 pandemic. *Annals of Internal Medicine*, v. 173, p. 21-28 (2020).

[48] WOLBERG, J. *Data Analysis Using the Method of Least Squares*, first ed. Springer, 2006.

[49] YOU, C., DENG, Y., HU, W., SUN, J., QIUSHI, L., ZHOU, F., PANG, C. H., CHEN, Z., AND ZHOU, X. H. Estimation of the time-varying reproduction number of COVID-19 outbreak in china. *International Journal of Hygiene and Environmental Health*, v. 228, nr. 113555 (2020).