



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA
MESTRADO ACADÊMICO EM FÍSICA

RAFAEL DE LIMA PEIXOTO

INFERÊNCIA BAYESIANA APLICADA A MODELOS EPIDEMIÓLOGICOS
COMPARTIMENTAIS

FORTALEZA

2022

RAFAEL DE LIMA PEIXOTO

INFERÊNCIA BAYESIANA APLICADA A MODELOS EPIDEMIÓLOGICOS
COMPARTIMENTAIS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Física do Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Física. Área de Concentração: Física da Matéria Condensada.

Orientador: Prof. Dr. Saulo Davi Soares e Reis.

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- P431i Peixoto, Rafael de Lima.
Inferência Bayesiana Aplicada a Modelos Epidemiológicos Compartmentais / Rafael de Lima Peixoto. –
2022.
57 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação
em Física, Fortaleza, 2022.
Orientação: Prof. Dr. Saulo Davi Soares e Reis.
1. SARS-CoV-2. 2. Inferência Bayesiana. 3. Filtro de Kalman. 4. SEIIR. I. Título.

CDD 530

RAFAEL DE LIMA PEIXOTO

INFERÊNCIA BAYESIANA APLICADA A MODELOS EPIDEMIÓLOGICOS
COMPARTIMENTAIS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Física do Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Física. Área de Concentração: Física da Matéria Condensada.

Aprovada em: 26/09/2022.

BANCA EXAMINADORA

Prof. Dr. Saulo Davi Soares e Reis (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Erneson Alves de Oliveira
Universidade de Fortaleza (UNIFOR)

Prof. Dr. Humberto de Andrade Carmona
Universidade Federal do Ceará (UFC)

À minha mãe Flaviane por todo o cuidado e apoio em minhas jornadas.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter conservado eu e minha família com saúde.

A Alanna Bezerra por me acompanhar nessa jornada, não me deixar fraquejar em nenhum momento na realização deste trabalho e por sempre estar ao meu lado em meio a grandes decisões.

A minha família, em especial à minha mãe Flaviane por sempre estar ao meu lado e se preocupar em sempre garantir o meu bem-estar e minha felicidade.

Ao Prof. Dr. Saulo Davi Soares e Reis por me orientar em mais um trabalho e pelos conselhos sobre sempre buscar melhorar.

Ao Prof. Dr. Humberto de Andrade Carmona pelas conversas produtivas em momentos cruciais da realização deste trabalho.

Aos meus amigos e colegas de pós-graduação, em especial Wagner Pereira, Wallace Vale e João Victor pelos momentos de estudo, troca de ideias e companherismo.

Aos meus colegas do Laboratório de Sistemas Complexos, em especial Marciel, Edson, Lara, Germano, André e Laísa pelos inúmeros momentos compartilhados.

Agradeço a todos os professores do Departamento de Física e de Matemática por me proporcionarem grandes conhecimentos, lições pessoais e profissionais, em especial aos professores: Saulo, Wandemberg, Ascânio, Cláudio Lucas, Carlos William, José Ramos, Lenz, Nildo, Fernando Pimentel e Renivaldo Sena.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Ao longo dos últimos anos, intensificou-se o estudo de modelos epidemiológicos compartimentais com o objetivo de estudar a dinâmica de transmissão do vírus SARS-CoV-2. Neste trabalho buscamos através de técnicas de Inferência Bayesiana, mais precisamente o framework do Filtro de Kalman Iterativo por Amostragem (IEKF), investigar a evolução temporal do número de reprodução efetivo R_t da pandemia de COVID-19 na cidade de Fortaleza. O método de IEKF assume uma distribuição gaussiana para as probabilidades a priori e a likelihood, ajustando a distribuição a priori para a posteriori utilizando a regra de Bayes. O cálculo do número de reprodução é deste modo feito utilizando o modelo epidemiológico SEIIR juntamente com o IEKF, o que nos permitiu estudar a evolução da dinâmica de propagação do SARS-CoV-2 na cidade de Fortaleza a partir de dados referentes ao número de óbitos diários confirmados disponibilizados pela Secretaria de Saúde do Governo do Estado do Ceará (SESA). Dentro dessas análises, inferimos a evolução temporal de todos os parâmetros do modelo SEIIR bem como seu respectivo número efetivo de reprodução, o qual indica períodos de maior propagação da doença e, portanto, é um importante indicador a ser considerado por autoridades governamentais durante sua tomada de decisão no combate a crises epidemiológicas.

Palavras-chave: sars-cov-2; inferência bayesiana; filtro de Kalman; seiiir.

ABSTRACT

Over the past few years, the study of compartmental epidemiological models has intensified to study the transmission dynamics of the SARS-CoV-2 virus. In this work, we seek through Bayesian Inference techniques, more precisely the Iterative Sampling Kalman Filter (IEKF) framework, to investigate the temporal evolution of the effective reproduction number R_t of the COVID-19 pandemic in the city of Fortaleza. The IEKF method assumes a Gaussian distribution for the a priori and likelihood probabilities, adjusting the a priori to a posteriori distribution using Bayes' rule. The calculation of the reproduction number is therefore done using the SEIIR epidemiological model together with the IEKF, which allowed us to study the evolution of the dynamics of the spread of SARS-CoV-2 in the city of Fortaleza from data referring to the number of deaths. confirmed diaries made available by the Secretary of Health of the Government of the State of Ceará (SESA). Within these analyses, we inferred the temporal evolution of all the parameters of the SEIIR model as well as their respective effective reproduction number, which indicates periods of greater spread of the disease and, therefore, is an important indicator to be considered by government authorities during its analysis and decision-making in combating epidemiological crises.

Keywords: sars-cov-2; bayesian inference; Kalman filter; seiir.

LISTA DE FIGURAS

Figura 1 – Representações pictóricas de patógenos: Corona Vírus que que pode causar síndromes respiratórias mais graves, o vírus da dengue que é transmitida pelo mosquito <i>Aedes aegypti</i> e o vírus HIV responsável pela AIDS.	15
Figura 2 – Diagrama do modelo epidemiológico SI. Cada indivíduo suscetível torna-se com o passar do tempo infeccioso a uma taxa β	17
Figura 3 – Fração do número de infectados em função do tempo de propagação da doença para $\beta = 0.50$, $\beta = 1.00$, $\beta = 0.50$ e $i_0 = 0.01$	18
Figura 4 – Diagrama do modelo epidemiológico SIS. Um indivíduo torna-se infeccioso a uma taxa β e novamente suscetível a uma taxa μ	19
Figura 5 – Fração do número de indivíduos infectados em função do tempo de propagação da doença para $i_0 = 0.01$ e os parâmetros $[(\beta = 2.00, \mu = 1.00), (\beta = 2.80, \mu = 0.80), (\beta = 3.00, \mu = 0.10)]$	19
Figura 6 – Diagrama do modelo epidemiológico SIR. No primeiro estágio, temos que um indivíduo suscetível torna-se infeccioso a uma taxa β , no estágio seguinte pode evoluir para um estado de imunização a uma taxa μ ou infelizmente, evoluir para óbito.	22
Figura 7 – Solução numérica do modelo SIR obtido através do método RK4. Em (a) onde $\mu = 0.9$ temos que a curva de indivíduos infectados é baixíssima tendo em vista que a taxa de recuperação é elevada, em (b) $\mu = 0.8$ começamos a observar uma tendência de crescimento dos infectados, subsequentemente em (c) e (d) onde $\mu = 0.6$ e $\mu = 0.3$ respectivamente. Podemos constatar que limitando o valor de μ mantendo β fixo, a curva de infectados torna-se menos achatada adquirindo uma forma característica onde observamos um ponto de máximo em movimento de translação.	23
Figura 8 – Fração do número de indivíduos infectados para os modelos SI, SIS e SIR. Para a mesma taxa de transmissão β , o modelo SI indica um crescimento acelerado. Os modelos SIS e SIR, crescem mais lentamente do que o modelo SI devido a taxa de recuperação. Observa-se que para uma alta taxa de recuperação μ , os modelos SIS e SIR, sinalizam o estado livre da doença, em que o número de infectados decai exponencialmente com o tempo.	24

- Figura 9 – Número de Casos Novos diários (a esquerda) e o número de óbitos diários (a direita) ao longo do ano de 2020 até março de 2021. Após o mês de maio de 2020 observa-se a diminuição dos casos diários sendo assim caracterizada o fim da primeira onda de contaminação. A medida em que foi-se aproximando do final do ano, o número volta a crescer dando início a segunda onda de contaminação. A curva de óbitos apresentou um crescimento mais acentuado próximo a maio de 2020, já a curva de casos o crescimento foi maior em meados de novembro e dezembro voltando a cair próximo ao mês de março de 2021. 29
- Figura 10 – Média móvel semanal do número de Casos Novos diários notificados pelo sistema de saúde (a esquerda) e a média móvel do número de óbitos diários notificados (a direita). Observa-se a formação de duas "ondas" com altas taxas de contaminação. A primeira sendo anterior ao decreto de *Lockdown*, que foi estabelecido em 05/05/2020, após esta data o número de casos novos foi diminuindo ao longo do tempo assim como o número de óbitos, até voltar a um crescimento exponencial iniciando no final do ano de 2020 e no início de 2021. A curva de casos aumentou de forma exponencial sendo maior que o primeiro pico e a de óbitos também cresceu, felizmente, a uma taxa menor em relação a primeira onda. 30
- Figura 11 – O modelo epidemiológico SEIIR. A população total N nutre a população de indivíduos suscetíveis S . Os indivíduos suscetíveis tornam-se expostos E a uma taxa $\lambda(t)$. Os indivíduos expostos tornam-se infectados contagiosos a uma taxa σ . Uma fração α da população infectada torna-se infectada relatada, I_r , e uma fração $(1 - \alpha)$ se torna infectados não relatados I_u . Os indivíduos infectados relatados e não relatados são removidos a uma taxa γ . Uma fração ϕ dos infectados relatados evoluem para óbito. 31

- Figura 12 – Framework Filtro de Kalman Iterativo por Amostragem. Munido do vetor de estados e de parâmetros iniciais gera-se um *ensemble* de partículas. As partículas geradas são aplicadas no modelo SEIIR e integradas numericamente, após tal integração obtêm-se uma nova configuração dos parâmetros e comparamos esses resultados com os dados. Para cada comparação é atribuída um peso, o valor deste peso é utilizado para selecionar as partículas que melhores representam o conjunto de dados. O processo continua a cada iteração m até que tenhamos obtido a melhor configuração para aquela janela de dados. 45
- Figura 13 – Inferência para a janela 12/04/2020 a 02/05/2020, sendo destacada em cinza e a data de publicação do Decreto 05/05/2020 que instituiu o *Lock-down*. A esquerda o conjunto total de dados analisados e direita em destaque a região de Inferência. O ajuste da máxima verossimilhança do modelo SEIIR esta destacada na linha vermelha e o número cumulativo de óbitos de acordo com o período analisado nos círculos brancos. Pode-se observar claramente que a inferência gerou resultados bastantes satisfatórios com os dados observados, logo, o modelo consegue descrever bem este conjunto de dados. 49
- Figura 14 – Inferência para a janela 07/07/2020 a 27/07/2020, sendo destacada em cinza e a data de publicação do Decreto 05/05/2020 que instituiu o *Lock-down*. A esquerda o conjunto total de óbitos acumulados notificados e direita em destaque a região de Inferência. 49
- Figura 15 – Inferência para a janela 13/02/2021 a 05/03/2021, sendo destacada em cinza. A esquerda o conjunto total de óbitos acumulados notificados e direita em destaque a região de Inferência. 50

- Figura 16 – Evolução temporal do parâmetro R_t ao longo do período de 19/03/2020 até 21/03/2021. A janela utilizada foi de 20 dias semanais com exceção entre outubro e novembro na qual uma janela de 50 dias foi aplicada. A partir do mês de Maio observa-se a tendência de queda entre os três meses seguintes, tal fator sendo explicado devido a publicação do decreto em 05/05/2020 que instituiu o *Lockdown* na cidade de Fortaleza. Após este período têm-se o crescimento novamente, fator explicado pelo relaxamento nos cuidados e festas de final de ano onde as pessoas tendem a confraternizar de maneira mais conjunta com um número maior de pessoas. 50
- Figura 17 – Evolução dos parâmetros do modelo obtidas através do filtro de Kalman. Inicialmente os parâmetros começam com um grande flutuação em torno da região dos parâmetros iniciais atribuídos e a medida que vamos selecionando os parâmetros que maximizam a *likelihood* com maior peso de probabilidade observa-se uma convergência em cada um dos parâmetros gerados. 51
- Figura 18 – Evolução Temporal dos parâmetros ao longo do período analisado. Cada janela possui o tamanho de 20 dias com um exceção entre outubro e novembro no qual uma janela de 50 dias foi utilizada. Observa-se em certas regiões uma estabilidade perante de seus valores, comparado esses resultados com a Fig. 16, têm-se que esta região está de acordo com as janelas nas quais os valores de R_t apresentavam valores próximos. 52

SUMÁRIO

1	INTRODUÇÃO	12
2	MODELOS EPIDEMOLÓGICOS	14
2.1	O modelo SI	16
2.2	O modelo SIS	18
2.3	O modelo SIR	21
2.3.1	<i>Solução numérica para o modelo SIR</i>	21
2.4	Número Básico de Reprodução R_0	24
2.4.1	<i>Construção da matriz de próxima geração (NGM)</i>	26
2.4.2	<i>Método da Função de Sobrevivência</i>	27
2.5	Dados Utilizados	28
2.6	O modelo SEIIR	29
3	INFERÊNCIA BAYESIANA	32
3.1	Abordagem clássica ou frequentista	32
3.2	Abordagem Bayesiana	34
3.2.1	<i>Regra de Bayes</i>	34
3.2.2	<i>Monte Carlo Cadeias de Markov</i>	37
3.3	Filtro de Kalman (KF)	40
3.4	Filtro de Kalman Iterativo por Amostragem (IEKF)	43
4	RESULTADOS DAS INFERÊNCIAS	48
5	CONCLUSÃO E PERSPECTIVAS FUTURAS	53
	REFERÊNCIAS	54

1 INTRODUÇÃO

A saga do coronavírus teve seu início ao final do ano de 2019 na cidade de Wuhan, província de Hubei na China, no qual tais casos foram inicialmente reportados como casos de pneumonia de origem desconhecida e que rapidamente espalharam-se pelas demais províncias chinesas (ORGANIZATION, 2020). Em março de 2020, casos confirmados em regiões fora da China em países asiáticos, europeus, africanos e americanos aumentaram de forma exponencial alertando assim a alta taxa de transmissão do vírus (DU *et al.*, 2020). Neste período, foram iniciados esforços para conter a transmissão do patógeno, como emissões de orientações sobre o monitoramento do paciente, coleta de amostras e tratamento, além do fornecimento de informações atualizadas sobre esse surto, entretanto, devido a incertezas sobre a transmissibilidade e virulência, não foi possível verificar a eficácia de tais esforços (GIBBONS *et al.*, 2014; MUNSTER *et al.*, 2020). Em 11 de março de 2020, a Organização Mundial de Saúde declarou a doença de coronavírus (COVID-19), como uma pandemia, levando assim a bloqueios em todo o mundo, afetando a economia global, o turismo, a questão da saúde pública e ameaçando uma possível recessão global (LIM; TO, 2022; YANG *et al.*, 2020; FORUM, 2020).

Uma característica epidemiológica crítica a ser analisada para avaliarmos o potencial pandêmico de um vírus respiratório em evolução, é a fração de casos não documentados porém infecciosos. Tais infecções geralmente apresentam sintomas leves ou nenhum e, portanto, acabam passando despercebidas, podendo assim expor uma grande parcela da população ao vírus (LI *et al.*, 2020). Para estudar a evoluções de epidemias, é possível utilizar modelos matemáticos para sua descrição, além de que deve-se levar em consideração as particularidades de cada doença. É importante ressaltar que um modelo pode ser extremamente eficiente para um tipo de doença enquanto outro não, dependendo das características epidemiológicas da doença. Os modelos de epidemia mais simples são: SI (Suscetível - Infeccioso) que apresenta dois compartimentos e parte da máxima que conforme se dá a evolução temporal todos os indivíduos serão infectados. O Modelo SIS (Suscetível - Infeccioso - Suscetível, apresenta três compartimentos, e leva em consideração que um indivíduo que contraiu o patógeno *a priori*, pode se recuperar e tornar-se suscetível novamente em um outro momento. O modelo SIR (Suscetível - Infeccioso - Removido), que consiste em um modelo no qual apresenta três compartimentos e diferentemente do SIS, um indivíduo pode se recuperar do patógeno, adquirindo imunidade ou vindo a óbito, sendo removido da população, portanto, não torna-se novamente suscetível. Os três modelos acima, apenas o SI e o SIS, apresentam solução analítica fechada. O modelo SIR, por sua vez,

requer uma solução numérica rápida e eficiente.

No presente trabalho utilizaremos o modelo SEIIR que trata-se de uma variação do modelo SEIR (Suscetível - Exposto - Infectado - Removido) (NEWMAN, 2018). Neste modelo, dividimos os infecciosos em duas classes de infecciosos, nomeadamente, I_r , infectados relatados, confirmados pelo sistema de saúde, e I_u os indivíduos infectados que não foram confirmados, abrangendo também os assintomáticos. Aqui estamos interessados em inferir o número de reprodução efetiva R_e sendo definido como o número médio de novos casos de infecção causados por um indivíduo infectado (DIEKMANN *et al.*, 2010).

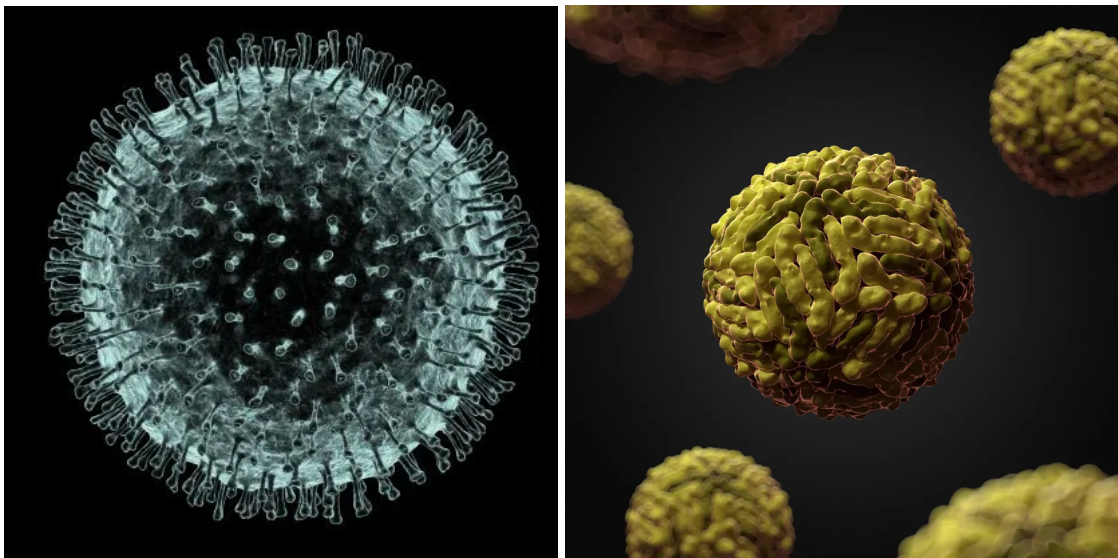
Para isto, utilizaremos o framework Filtro de Kalman Iterativo por Amostragem (IEFK) (LI *et al.*, 2020; IONIDES *et al.*, 2006; KING *et al.*, 2008) no modelo SEIIR para inferir os parâmetros do modelo, tomando como passo inicial uma distribuição normal multivariada dos parâmetros e estados. Além disto utilizamos dados públicos disponibilizados pela secretaria de Saúde do Governo do Estado do Ceará (SESA) e assim fazmos previsões sobre o número óbitos de maneira mais eficiente com menos imprecisões. A escolha da cidade de Fortaleza para o estudo deu-se devido a sua importância no desenvolvimento da epidemia estadual, sendo uma das maiores cidades de estado além de conter um grande fluxo de pessoas diariamente.

2 MODELOS EPIDEMOLÓGICOS

Doenças se propagam através da rede de contato entre os indivíduos: doenças transmitidas pelo ar, como a gripe e tuberculose, são transmitidas quando duas pessoas (uma contaminada e outra não contaminada) respiram o ar na mesma sala; doenças sexualmente transmissíveis (DST's) como o HIV, transmitidas através do ato sexual, entre outras. Podemos representar esse padrões de contatos como redes e assim utilizarmos técnicas matemáticas para modelarmos a dinâmica de propagação (NEWMAN, 2018).

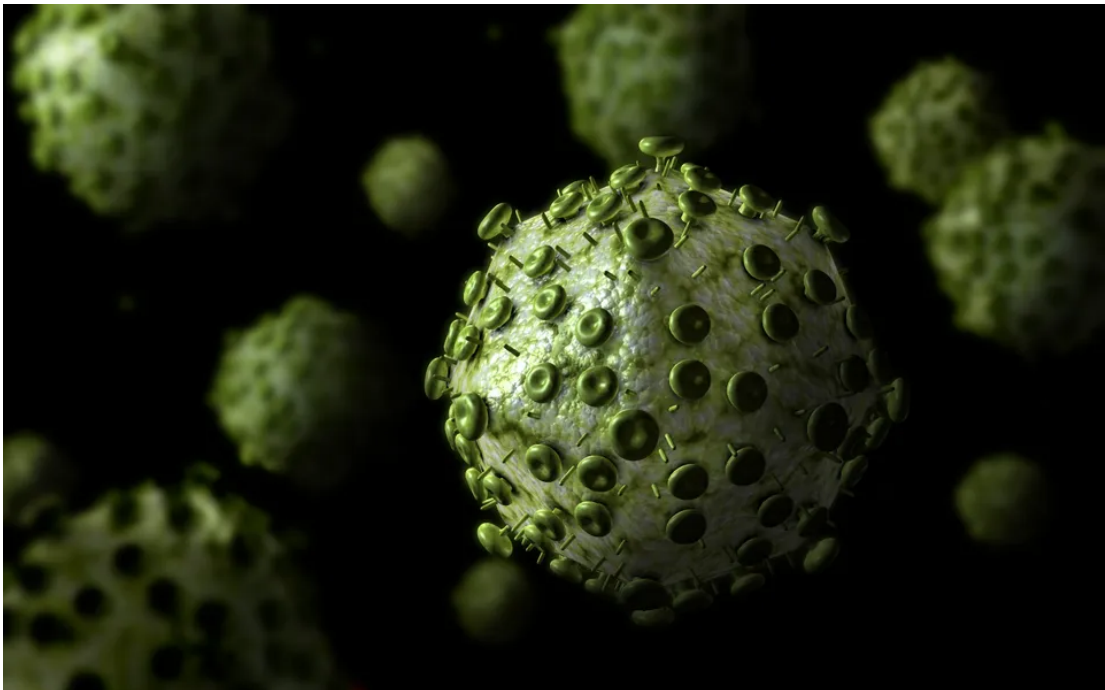
Existem outros processos utilizando redes que apresentam semelhanças como a propagação de doenças, dentre as quais podemos citar a disseminação de notícias, rumores ou fofocas por meio de uma população onde ambas apresentam similaridade com a propagação. Quando uma pessoa toma conhecimento de uma notícia ou uma informação de um conhecido, ela se torna capaz de espalhar essa notícia para outras pessoas, podemos afirmar então que elas foram “infectadas”. Este processo configura-se sendo a notícia verdadeira ou não, como no caso da propagação de *Fake News* (PAPANASTASIOU, 2020). Quando tratamos com propagação de notícias, esse processo é mais complicado do que a de doenças, uma vez que a informação pode ser disseminada por vários meios, como a internet, a mídia, livros, entre outros. Apesar disso, as ideias e os modelos de propagação de doenças nos ajudam a entender melhor a disseminação de informações (NEWMAN, 2018).

Os modelos epidemiológicos para o estudo da propagação de doenças foram desenvolvidos pela Epidemiologia como uma estrutura analítica e numérica robusta para modelar a disseminação de patógenos. Sabe-se da biologia que um patógeno é definido como um organismo que causa doença em seu hospedeiro. Os patógenos são taxonomicamente diversos, compreendendo vírus e bactérias, além de seres eucariotos unicelulares e multicelulares. Todo organismo vivo é afetado por patógenos, incluindo bactérias, que são alvejados por vírus especializados chamados fagos (BALLOUX; DORP, 2017). Na Fig. 1 têm-se algumas representações de patógenos estudados.



(a) Corona Vírus - SARS

(b) Vírus da Dengue



(c) Vírus da Imunodeficiência Humana(HIV)

Figura 1 – Representações pictóricas de patógenos: Corona Vírus que pode causar síndromes respiratórias mais graves, o vírus da dengue que é transmitida pelo mosquito *Aedes aegypti* e o vírus HIV responsável pela AIDS.

Fonte: (HOWARD, 2014)

Os modelos de epidemia podem ser caracterizados por duas hipóteses:

i. **Compartimentalização**

Esta hipótese classifica cada indivíduo baseado no estágio em que a doença os afeta.

Estes modelos apresentam estados diversos, como indivíduos saudáveis, infecciosos, imunes ao patógeno, isto é, não podem ser infectados, ou indivíduos latentes, indivíduos expostos

a doença mas ainda não são contagiosos. Os indivíduos podem transitar pelos compartimentos. Por exemplo, no caso de uma nova influenza, inicialmente todos os indivíduos encontram-se no estado suscetível (S), desde que esse indivíduo entre em contato com uma pessoa infecciosa ela pode tornar-se infectada. Com o tempo ela se recuperará e desenvolverá imunidade, perdendo sua suscetibilidade à cepa específica da influenza (BARABÁSI, 2016).

ii. Mistura Homogênea

Esta hipótese, também chamado de aproximação totalmente mista ou de ação em massa, assume que cada indivíduo tem a mesma chance de entrar em contato com um indivíduo infectado. Essa hipótese elimina a necessidade de sabermos toda a rede de contatos no qual a doença se espalha, substituindo-a pelo presuposto que qualquer pessoa pode infectar outra.

2.1 O modelo SI

Neste modelo vamos considerar que uma certa doença propaga-se em uma população de N indivíduos. Seja $S(t)$ o número de indivíduos suscetíveis, isto é, saudáveis, no tempo t e $I(t)$ o número de indivíduos infecciosos (os quais podem infectar um indivíduo suscetível). No instante de tempo inicial $t = 0$, toda população é suscetível ($S(0) = N$) e ninguém está infectado ($I(0) = 0$). A pergunta que podemos fazer é: “Dado um indivíduo infectado no tempo inicial $t = 0$, isto é, $I(0) = 1$, quantos indivíduos serão infectados a um tempo posterior t ?”

De acordo com a hipótese da mistura homogênea, a probabilidade com a qual uma pessoa infectada encontre um indivíduo suscetível é dada por $S(t)/N$. Desde que os indivíduos infecciosos $I(t)$ são transmissores do patógeno, cada um deles a uma taxa β , o número médio de novas infecções $dI(t)$ variando no tempo é dada por:

$$\frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N}. \quad (2.1)$$

Vamos utilizar as seguintes notações:

$$s(t) = \frac{S(t)}{N}, \quad e$$

$$i(t) = \frac{I(t)}{N}.$$

para a fração de suscetíveis e de infecciosas no tempo t . Podemos fazer as seguintes mudanças para removermos a dependência direta de s :

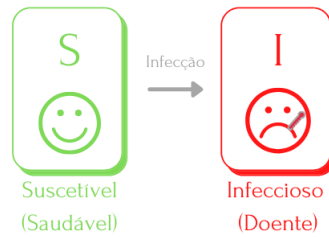


Figura 2 – Diagrama do modelo epidemiológico SI. Cada indivíduo suscetível torna-se com o passar do tempo infeccioso a uma taxa β .

Fonte: Elaborado pelo autor.

$$\frac{di(t)}{dt} = \beta si = \beta i(1 - i). \quad (2.2)$$

onde β é chamado de taxa de transmissão ou transmissibilidade e utilizamos o fato de $S = 1 - i$. Pode-se resolver a equação acima utilizando o método de separação de variáveis (BOYCE *et al.*, 2021) e usando que $i_0 = i(t = 0)$, obtemos então que a fração de indivíduos infecciosos aumenta no tempo na forma:

$$i = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}. \quad (2.3)$$

Na Fig. 2 temos a representação em compartimentos do modelo SI. Com base na equação 2.3 e na Figura 3 podemos concluir que no início da infecção, a fração de indivíduos infectados aumenta exponencialmente. Na verdade, no início da propagação o indivíduo infeccioso encontra apenas indivíduos suscetíveis, facilitando ainda mais a transmissão do patógeno.

O tempo característico τ necessário para alcançar a fração $1/e$ de todos os indivíduos suscetíveis é:

$$\tau = \frac{1}{\beta}. \quad (2.4)$$

A Eq.2.4 nos indica que se aumentarmos β , aumenta-se a velocidade de transmissão do patógeno e reduz-se o tempo característico conforme apresentado na Figura 3.

A medida em que avançamos no tempo os indivíduos infecciosos encontram cada vez menos indivíduos suscetíveis. Então o crescimento de i cai com a evolução temporal. O

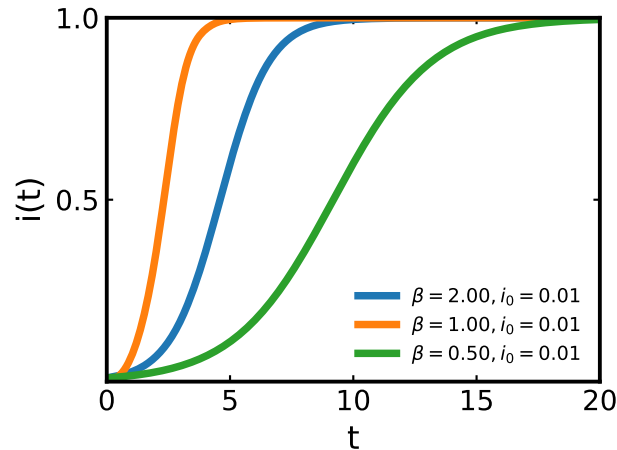


Figura 3 – Fração do número de infectados em função do tempo de propagação da doença para $\beta = 0.50$, $\beta = 1.00$, $\beta = 0.50$ e $i_0 = 0.01$.

Fonte: Elaborado pelo autor.

comportamento do modelo sai do regime exponencial e entra no chamado regime de saturação, no qual todos os indivíduos da população estão infectados. A dinâmica encerra-se quando todos os indivíduos estão infectados, $i(t \rightarrow \infty) = 1$ e $s(t \rightarrow \infty) = 0$.

2.2 O modelo SIS

O modelo SIS (Suscetível-Infecioso-Suscetível) é uma extensão do modelo SI, permitindo a reinfeção por parte do indivíduo infeccioso. Esta modificação permite que o modelo SIS seja aplicável a um maior número de doenças, uma vez que é sabido que a maioria dos patógenos não garante a imunidade às vítimas após tratamento, ou confere apenas uma imunidade limitada, sendo assim possível a reinfeção mais de uma vez, como por exemplo a gripe.

Define-se μ como sendo a taxa de recuperação na qual o indivíduo torna-se novamente suscetível. Pela perspectiva matemática, o desenvolvimento da equação que rege esse modelo é semelhante a do SI, pode-se apenas acrescentar o termo $-\mu i$ que nos dá a taxa na qual a população recupera-se da doença na equação do modelo SI, portanto:

$$\frac{di(t)}{dt} = \beta i(1-i) - \mu i. \quad (2.5)$$

A equação 2.5 possui a seguinte solução:

$$i = \left(1 - \frac{\mu}{\beta}\right) \frac{C e^{(\beta-\mu)t}}{1 + C e^{(\beta-\mu)t}}. \quad (2.6)$$

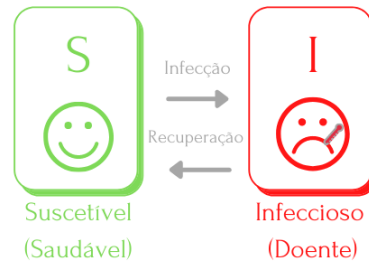


Figura 4 – Diagrama do modelo epidemiológico SIS. Um indivíduo torna-se infeccioso a uma taxa β e novamente suscetível a uma taxa μ .

Fonte: Elaborado pelo autor.

Valendo-se das condições iniciais do nosso modelo $i_0 = i(t = 0)$ temos que a constante C é dada por:

$$C = \frac{i_0}{\left(1 - i_0 - \frac{\mu}{\beta}\right)}. \quad (2.7)$$

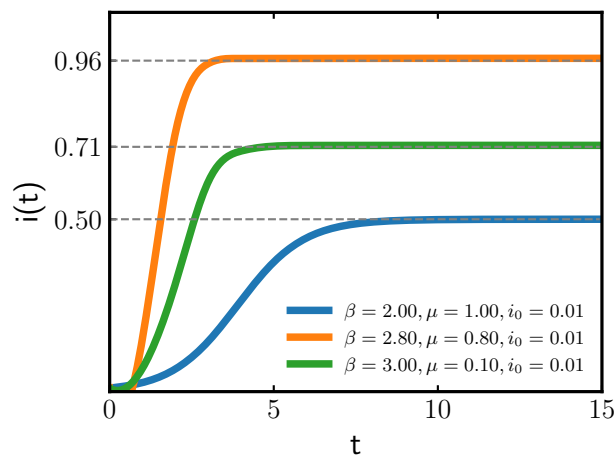


Figura 5 – Fração do número de indivíduos infectados em função do tempo de propagação da doença para $i_0 = 0.01$ e os parâmetros $[(\beta = 2.00, \mu = 1.00), (\beta = 2.80, \mu = 0.80), (\beta = 3.00, \mu = 0.10)]$.

Fonte: Elaborado pelo autor.

De acordo com a Equação 2.6 e a Fig. 5 podemos perceber que para uma baixa taxa de recuperação, a fração de indivíduos infectados segue uma curva logística de crescimento similar ao que foi observado no modelo SI, tal estado é chamado na epidemiologia de estado endêmico ($\mu < \beta$). No entanto, vale ressaltar um aspecto muito importante, no modelo SIS, cujo diagrama está representado na 4, nunca ocorrerá uma situação na qual teremos toda a população

infectada com a doença, pois i atinge um valor constante ao longo de um tempo suficientemente longo $i(\infty) < 1$. Conclui-se que em qualquer momento apenas uma fração finita da população está infectada, diferentemente do modelo SI que prediz a contaminação de toda a população. Neste estado, o número de pessoas infectadas é igual ao número de indivíduos que se recuperaram da doença, têm-se que a fração da população que foi infectada não muda com tempo. Podemos obter o valor de $i(\infty)$ apenas tomando $di/dt = 0$ na Eq. 2.5, onde obtemos:

$$i(\infty) = 1 - \frac{\mu}{\beta}. \quad (2.8)$$

A medida que a taxa de recuperação μ aumenta de maneira significativa, i decai exponencialmente com tempo, indicando que o processo de infecção cessa. Tal fato ocorre devido ao número de indivíduos recuperados por unidade de tempo exceder o número de novas infecções individuais. Portanto o patógeno desaparece da população, este estado recebe nome de estado livre da doença ($\mu > \beta$).

O modelo SIS prevê que alguns patógenos irão persistir na população enquanto outros não, pois depende da taxa de recuperação dos indivíduos. Esta diferença está em possíveis resultados, escrevendo o tempo característico de um patógeno como:

$$\tau = \frac{1}{\mu (R_0 - 1)}, \quad (2.9)$$

onde

$$R_0 = \frac{\beta}{\mu}. \quad (2.10)$$

é definido como o número básico de reprodução. Este número representa o número médio de indivíduos suscetíveis infectados por um indivíduo infeccioso em uma população totalmente suscetível (DIEKMANN *et al.*, 2010). Podemos então definir R_0 como o número de novas infecções que cada indivíduo infeccioso causa em circunstâncias ideais.

Para $R_0 > 1$, a epidemia encontra-se no estado endêmico. Na prática, se cada indivíduo infeccioso infecta mais de uma pessoa saudável, o patógeno tende a se espalhar e a perdurar na população. Quanto maior o valor de R_0 mais rápido se dá o processo de contaminação. Se $R_0 < 1$, a epidemia cessa, tendo em vista que se cada indivíduo infeccioso infecta menos do que um indivíduo, o patógeno não consegue perdurar na população.

2.3 O modelo SIR

Sabe-se que em certos tipos de doenças os indivíduos tornam-se imunes, seja temporariamente ou vindo a óbito. Tendo em vista tal observação, precisamos de um modelo que seja capaz de levar em consideração os casos acima descritos. Este é o caso do modelo SIR (Suscetível - Infeccioso - Removido) desenvolvido por Ronald Ross, William Hamer e outros no início do século XX (ANDERSON, 1991). O fato de estarmos adicionando mais um compartimento a esse modelo, estamos tornando o estudo mais geral. Vale ressaltar que do ponto de vista epidemiológico, diferentemente de termos humanos, faz pouca diferença para o estudo da doença se o indivíduo está vivo ou não. De qualquer forma, ele é removido do conjunto de hospedeiros em potencial para doença. Assim, o compartimento R representa a população recuperada ou morta. Neste modelo, temos dois momentos: no primeiro momento o indivíduo suscetível entra em contato com um indivíduo infectado a uma taxa β , tornando-se igualmente infectado. Em um segundo momento ele pode se recuperar com uma taxa μ (NEWMAN, 2018).

Com base nestas características, temos que as equações diferenciais que regem esse modelo são:

$$\frac{ds}{dt} = -\beta si, \quad (2.11a)$$

$$\frac{di}{dt} = -\mu i + \beta si, \quad (2.11b)$$

$$\frac{dr}{dt} = \mu i, \quad (2.11c)$$

onde temos o vínculo que representa toda a população escrito como:

$$s + i + r = 1. \quad (2.12)$$

2.3.1 Solução numérica para o modelo SIR

Na literatura, existem alguns trabalhos que utilizam técnicas como pontos de equilíbrio (SILVA, 2018) (ALMEIDA, 2014), estimativa de parâmetros (COSTA *et al.*, 2021) e até uma solução analítica (HARKO *et al.*, 2014) para o modelo em questão, entretanto, neste trabalho utilizaremos o método de Runge-Kutta de quarta ordem (RK4) o qual consiste em uma série de procedimentos utilizados para resolver equações diferenciais não-lineares de forma numérica

com precisão (NEWMAN, 2013). O método consiste em expandir uma determinada função $f(x)$ em um dado intervalo h usando a expansão de Taylor. No método de quarta ordem em especial, a acurácia é da ordem de h^4 e o erro da ordem de h^5 , sendo h suficientemente pequeno têm-se uma boa visão sobre como a solução da equação parece. O método de Euler para solução de equações diferenciais consiste em um dos métodos de Runge-Kutta, o de primeira ordem.

Seguindo o método, ao invés de escrevermos nossa equação diferencial ordinária (EDO) da maneira como na equação 2.11, vamos reescrever na forma:

$$\frac{ds}{dt} = f(s, i), \quad (2.13a)$$

$$\frac{di}{dt} = g(s, i), \quad (2.13b)$$

$$\frac{dr}{dt} = v(i). \quad (2.13c)$$

e expandir essas funções em torno de uma série de pequenos passos h

$$h = \frac{(b-a)}{N}. \quad (2.14)$$

sendo N o número total de passos.

No modelo SIR, possuímos mais de uma variável a ser analisada, conforme descrito em (NEWMAN, 2013), ao invés de usarmos Runge-Kutta para cada uma das funções, podemos criar uma função vetorial que armazene cada uma dessas funções e resolvendo esta função, obtemos a solução de cada uma das variáveis tornando a solução mais didática. De maneira mais sucinta, podemos discretizar o método completo em um único passo onde calculamos os

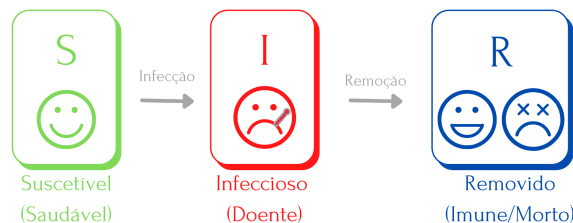


Figura 6 – Diagrama do modelo epidemiológico SIR. No primeiro estágio, temos que um indivíduo suscetível torna-se infeccioso a uma taxa β , no estágio seguinte pode evoluir para um estado de imunização a uma taxa μ ou infelizmente, evoluir para óbito.

Fonte: Elaborado pelo autor.

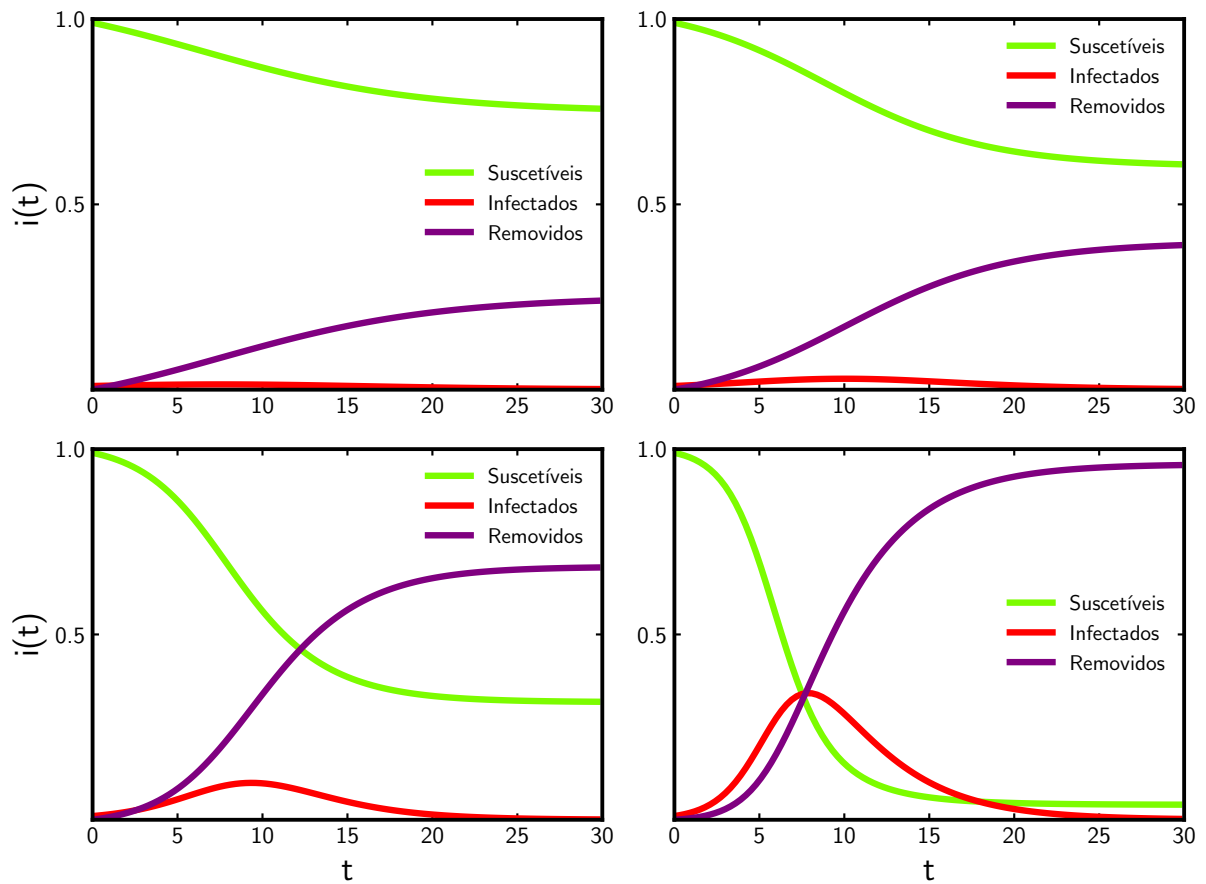


Figura 7 – Solução numérica do modelo SIR obtido através do método RK4. Em (a) onde $\mu = 0.9$ temos que a curva de indivíduos infectados é baixíssima tendo em vista que a taxa de recuperação é elevada, em (b) $\mu = 0.8$ começamos a observar uma tendência de crescimento dos infectados, subsequentemente em (c) e (d) onde $\mu = 0.6$ e $\mu = 0.3$ respectivamente. Podemos constatar que limitando o valor de μ mantendo β fixo, a curva de infectados torna-se menos achatada adquirindo uma forma característica onde observamos um ponto de máximo em movimento de translação.

Fonte: Elaborado pelo autor.

parâmetros descritos na Eq. 2.15 generalizando-os para o caso multivariável, para apenas uma variável temos:

$$k_1 = hf(s, t) \quad (2.15a)$$

$$k_2 = hf\left(s + \frac{1}{2}k_1, t + \frac{1}{2}h\right) \quad (2.15b)$$

$$k_3 = hf\left(s + \frac{1}{2}k_2, t + \frac{1}{2}h\right) \quad (2.15c)$$

$$k_4 = hf\left(s + \frac{1}{2}k_3, t + \frac{1}{2}h\right) \quad (2.15d)$$

$$s(t+h) = s(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (2.15e)$$

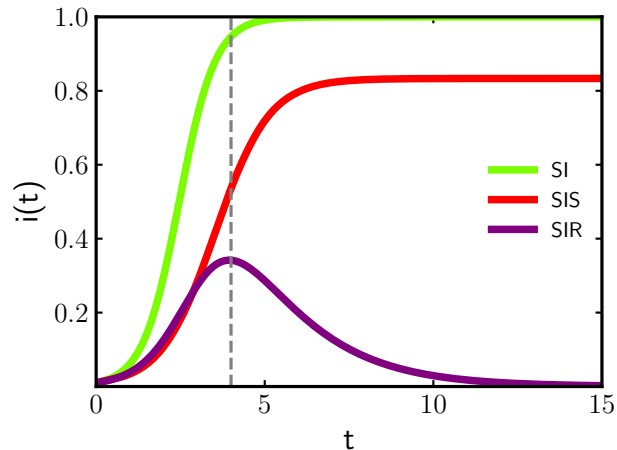


Figura 8 – Fração do número de indivíduos infectados para os modelos SI, SIS e SIR. Para a mesma taxa de transmissão β , o modelo SI indica um crescimento acelerado. Os modelos SIS e SIR, crescem mais lentamente do que o modelo SI devido a taxa de recuperação. Observa-se que para uma alta taxa de recuperação μ , os modelos SIS e SIR, sinalizam o estado livre da doença, em que o número de infectados decai exponencialmente com o tempo.

Fonte: Elaborado pelo autor.

Provido do algoritmo RK4, vamos tomar o passo $h = 0.3$ obtido por 2.14, obtendo assim a solução apresentada na Fig.7, onde as condições iniciais consideradas foram $s(0) = 0.99$, $i(0) = 0.01$ e $r(0) = 0$, o parâmetro $\beta = 1$ e tomamos uma coleção de valores de μ , a considerar $[0.3, 0.6, 0.8, 0.9]$, para que possamos ter uma visão global sobre os diferentes casos de variação da taxa de recuperação.

As predições do modelos SI, SIS e SIR concordam entre si nos primeiros estágios da epidemia, em que temos o número de infectados individuais pequeno, a doença propaga-se livremente e a curva de infectados aumenta exponencialmente. A medida em que a epidemia vai se alastrando, o número de infectados aumenta e percebe-se uma mudança no comportamento de cada curva. No modelo SI todos tornam-se infectados; no SIS a epidemia entra estado endêmico, no qual observa-se que apenas uma fração finita dos indivíduos é infectada, ou a infecção cessou; no SIR, todos acabam sendo removidos ao final. Em termos do número básico de reprodução R_0 , se $R_0 < 1$ o patógeno persiste na população, para $R_0 > 1$ ele é removido naturalmente (BARABÁSI, 2016).

2.4 Número Básico de Reprodução R_0

O número básico de reprodução R_0 é, sem dúvidas, umas das quantidades epidemiológicas mais importantes para quantificar doenças infecciosas. Seu valor nos permite projetar e entender os estágios em que a contaminação está ocorrendo e assim tomar medidas de controle.

É definido como sendo o número médio de novos casos de infecções oriundas a partir de um indivíduo infectado em uma população composta de indivíduos suscetíveis. No contexto do estudo da dinâmica de modelos epidemiológicos, determina regimes importantes. Se $R_0 > 1$, cada indivíduo infeccioso infecta em média mais de um indivíduo suscetível, logo, tal resultado representa que a taxa de infecção está maior do que a taxa de recuperação, implicando assim, na persistência do patógeno na população. No caso de $R_0 = 1$, a epidemia persistirá, entretanto, com o número de infectados mantendo-se constante, tornando-se uma endemia. Por fim, se $R_0 < 1$, a doença está controlada. Vale ressaltar que o valor de R_0 é uma constante característica do tipo de doença, temos na Tabela 1 alguns valores de R_0 para diversos tipos de doenças.

Para o cálculo de R_0 utilizaremos a abordagem da matriz de próxima geração, do inglês, *next-generation matrix* (NGM). Neste método utiliza-se o número de novos indivíduos infectados em vários compartimentos da população de interesse (DIEKMANN *et al.*, 1990). O valor de R_0 é obtido como o auto-valor dominante da matriz de próxima geração (DIEKMANN *et al.*, 2010). A abordagem da matriz de próxima geração tem sido amplamente utilizada em modelos epidemiológicos de tempo contínuo (DRIESSCHE; WATMOUGH, 2002; DIEKMANN *et al.*, 1990), no modelos de tempo discreto tal abordagem é usada com menos frequência (JONG *et al.*, 1994; LEWIS *et al.*, 2006; WESLEY *et al.*, 2009).

Tabela 1 – R_0 para diversos patógenos estudados

Doença	Transmissão	R_0
Sarampo	Pelo ar	12-18
Coqueluche	Pelo ar	12-17
Difteria	Saliva	6-7
Varíola	Contato social	5-7
Poliomelite	Via fecal-oral	5-7
Rubéola	Pelo ar	5-7
Caxumba	Pelo ar	4-7
HIV/AIDS	Contato sexual	2-5
SARS	Pelo ar	2-5
Gripe (1918)	Pelo ar	2-3

Fonte:(NEWMAN, 2018)

2.4.1 Construção da matriz de próxima geração (NGM)

Façamos $X(t) = (x_0, x_1, \dots, x_n)^T$, onde n é o número de compartimentos da população no instante t . Tais compartimentos podem ser de indivíduos suscetíveis, infecciosos, removidos, etc. Tomando

$$\frac{dX(t)}{dt} = G(X(t)), \quad (2.16)$$

onde $G : \mathbf{R}_+^n \rightarrow \mathbf{R}_+^n$ e $G \in C^1(\mathbf{R}_+^n)$ para $\mathbf{R}_+^n = \{Y(y_1, y_2, \dots, y_n) | y_j \geq 0, j = 1, 2, \dots, n\}$, e y_j representa as taxas de variação do estado j . Pode-se separar $X(t)$ em dois conjuntos distintos, $X_1(t)$, representando os indivíduos infectados, indivíduos contidos nos primeiros m estados (com $m < n$) e $X_2(t)$ representando os indivíduos não infectados, contidos nos estados $(n - m)$, logo:

$$X(t) = \begin{cases} X_1(t) = (x_1, x_2, \dots, x_m)^T \\ X_2(t) = (x_{m+1}, x_{m+2}, \dots, x_n)^T. \end{cases} \quad (2.17)$$

A Equação 2.16 pode então ser escrita na forma:

$$\left(\frac{dX_1(t)}{dt} \quad \frac{dX_2(t)}{dt} \right) = \left(G_1(t) \quad G_2(t) \right). \quad (2.18)$$

Para determinar a Equação 2.18 assume-se que existe um único ponto de equilíbrio livre de doenças, *DFE* (*Disease-free equilibrium*), no qual $X_1(t) = 0$ e $X_2(t) > 0$, com 0 representando a matriz nula, e que ao linearizar a Equação 2.16, valendo-se do *DFE*, obtêm-se (ALLEN; DRIESSCHE, 2008):

$$\frac{dY(t)}{dt} = JY(t), \quad (2.19)$$

no qual J é a matriz jacobiano quadrada de dimensão n , das variáveis G com as variáveis Y , tal matriz é representada:

$$J = \begin{pmatrix} F + T & 0 \\ A & C \end{pmatrix}. \quad (2.20)$$

O elemento $F + T$, são as submatrizes quadradas não negativas irredutíveis. Tais matrizes são obtidas a partir da derivação em termo de G_0 , uma forma de obtê-los é utilizando as matrizes \mathcal{F} e \mathcal{T} , que representam o vetor de novas infecções e o vetor de transferência entre os estados, respectivamente. Fazemos então $G_1(X(t)) = F(t) + \mathcal{T}(t)$. F e T são obtidos a partir das relações:

$$F = \frac{d\mathcal{F}_i(X(t^*))}{dx_j} \quad e \quad T = -\frac{d\mathcal{T}_i(X(t^*))}{dx_j}. \quad (2.21)$$

onde cada x_j representa os j -ésimos estados de X_1 , \mathcal{F}_i os i -ésimos estados da matriz \mathcal{F} . O número básico de reprodução será calculado através de (DRIESSCHE; WATMOUGH, 2002):

$$R_0 = \rho(FT^{-1}). \quad (2.22)$$

$X_1(t^*)$ é um *DFE*, um ponto de extremo do sistema considerado em Eq 2.16, T^{-1} é a matriz inversa de T e $\rho(FT^{-1})$ é o raio espectral da matriz FT^{-1} . O raio espectral é definido como o autovalor dominante da matriz em questão

$$\rho = \max_{(i=1,2,\dots,n)}[\lambda_i(FT^{-1})], \quad (2.23)$$

no qual $\lambda_i(FT^{-1})$ são os i -ésimos autovalores da matriz. Na literatura encontramos alternativas para o cálculo de R_0 utilizando uma abordagem determinística conhecida como método da “função de sobrevivência”. Tal método é utilizado quando as probabilidades de transmissão não constantes entre classes são assumidas. Para modelos que possuem várias classes de indivíduos infecciosos, o método da Matriz de próxima geração é um extensão desta abordagem (HEFFERNAN, 2005).

2.4.2 Método da Função de Sobrevivência

O método da função de sobrevivência é uma definição dos primeiros princípios de R_0 , sendo abordada em detalhes em (HEESTERBEEK; DIETZ, 1996). Consideremos uma grande população, tomando $F(a)$ como a probabilidade de um dado indivíduo recém infectado permaneça infectado por um tempo a e $b(a)$ como o número médio de indivíduos recém-infectados no qual um indivíduo infeccioso contaminará por unidade de tempo, quando o tempo de infecção total for a . Desta maneira, R_0 é definido:

$$R_0 = \int_0^{\infty} b(a)F(a)da. \quad (2.24)$$

A Eq 2.24 nos permite obter R_0 , sendo apropriada para qualquer modelo cujas expressões dadas em formas fechadas, possam ser expressas por probabilidade de sobrevivência subjacente, $F(a)$ e a infecciosidade como função do tempo, $b(a)$. Esta maneira de se obter R_0 não se restringe a apenas sistemas descritos por equações diferenciais (HEFFERNAN, 2005).

2.5 Dados Utilizados

Antes de iniciarmos a exemplificação do modelo que utilizamos neste trabalho, faz-se necessário salientarmos previamente sobre a natureza dos dados utilizados. Inicialmente, os dados em que estavamos trabalhando eram oriundos da Secretaria de Saúde em uma tabela diária dos casos notificados no Estado do Ceará no formato csv. Após a implementação do portal Integra SUS, esses dados foram coletados diretamente do portal de maneira mais eficiente. Os dados (no formato de tabelas) contém informações de forma anônima de todos os pacientes, a exemplo, idade e município de residência, bem como os resultados dos exames e evolução da doença oriunda de diversos outros sistemas de informação, como:

- e-SUS-VE- Plataforma de Notificação Saúde do Ministério da Saúde
- SIVEP - Sistema de Informação de Vigilância Epidemiológica - Ministério da Saúde
- COVEP - Coordenadoria de Vigilância Epidemiológica e Prevenção em Saúde.
- CESAP - Célula de Saúde e Promoção Social
- SESAI - Secretaria Especial de Saúde Indígena

Na Fig. 9 têm-se a evolução do número de Casos Novos diários notificados pelo Sistema de Saúde (a esquerda) a partir da Data dos Primeiros Sintomas, é importante destacar este fator, tendo em vista as imprecisões no momento da coleta desta informação, um paciente pode enganar-se quanto ao real dia em que teve o primeiro sintoma trazendo assim uma incerteza. A direita ainda na Fig.9 temos a evolução no número de casos diários a partir da Data de Óbito, esta, apresenta uma incerteza menor do que a anterior, tornando-se uma medida mais precisa, sendo assim um dos motivos da escolha de se fazer inferências com base na data de óbito.

De acordo com a Fig. 9 observa-se que nos estados iniciais da dinâmica de propagação do vírus em Fortaleza apresenta um crescimento exponencial bem expressivo caracterizando a primeira onda da contaminação. Após a Decreto que instaurou o *Lockdown*, temos a diminuição

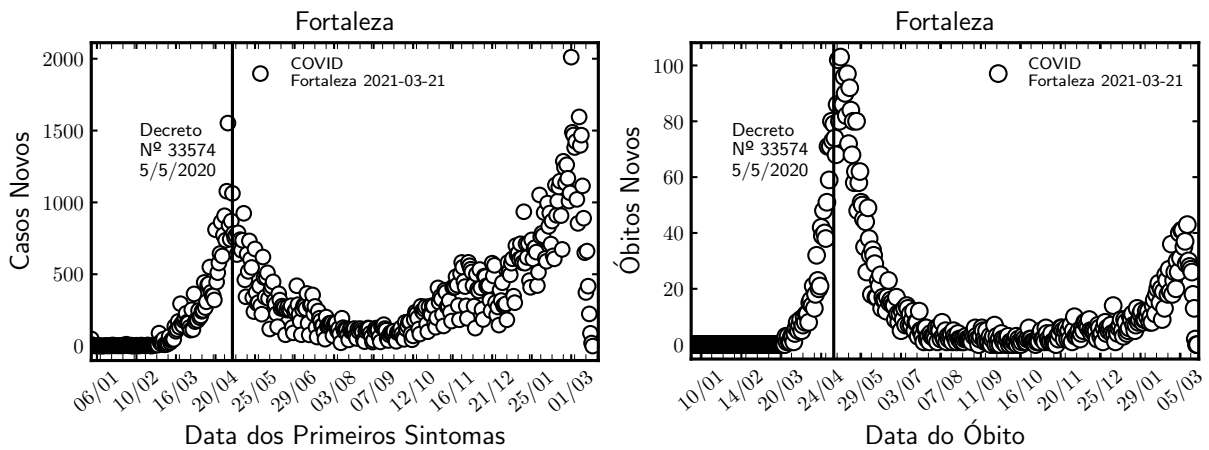


Figura 9 – Número de Casos Novos diários (a esquerda) e o número de óbitos diários (a direita) ao longo do ano de 2020 até março de 2021. Após o mês de maio de 2020 observa-se a diminuição dos casos diários sendo assim caracterizada o fim da primeira onda de contaminação. A medida em que foi-se aproximando do final do ano, o número volta a crescer dando início a segunda onda de contaminação. A curva de óbitos apresentou um crescimento mais acentuado próximo a maio de 2020, já a curva de casos o crescimento foi maior em meados de novembro e dezembro voltando a cair próximo ao mês de março de 2021.

Fonte: Elaborado pelo autor.

do número de casos e de óbitos de maneira significativa evidenciando a eficácia de tal medida de contenção. A medida em que aproxima-se do final do ano de 2020 e início do ano de 2021 têm-se uma tendência novamente de crescimento, sendo o número de casos de forma mais agressiva diferentemente do número de óbitos que apresentou um crescimento com menor atenuação.

Na Fig. 10 temos uma maneira alternativa de mostrar os mesmos dados apresentados na Fig. 9, nela temos em azul-claro a representação na forma de histograma estes casos diários e óbitos, na linha azul-escuro temos a média móvel semanal dos casos e óbitos diários. Esta visualização dos dados nos permite observar de mais maneira mais direta a formação das duas ondas de contaminação no período estudado, sendo a primeira no início do ano período anterior ao *Lockdown* e a segunda mais próxima do final do ano. Evidenciamos ainda que a curva de casos na segunda onda foi superior a primeira e a de número de óbitos, felizmente, caiu de maneira bastante representativa na segunda onda.

2.6 O modelo SEIIR

O modelo que utilizamos para estudar os dados referentes à infecção por COVID-19 no Ceará, SEIIR, é uma variação do modelo SEIR (Susceptível, Exposto, Infectado, Removido) onde levamos em consideração os infectados não relatados e cujo objetivo é inferir o número

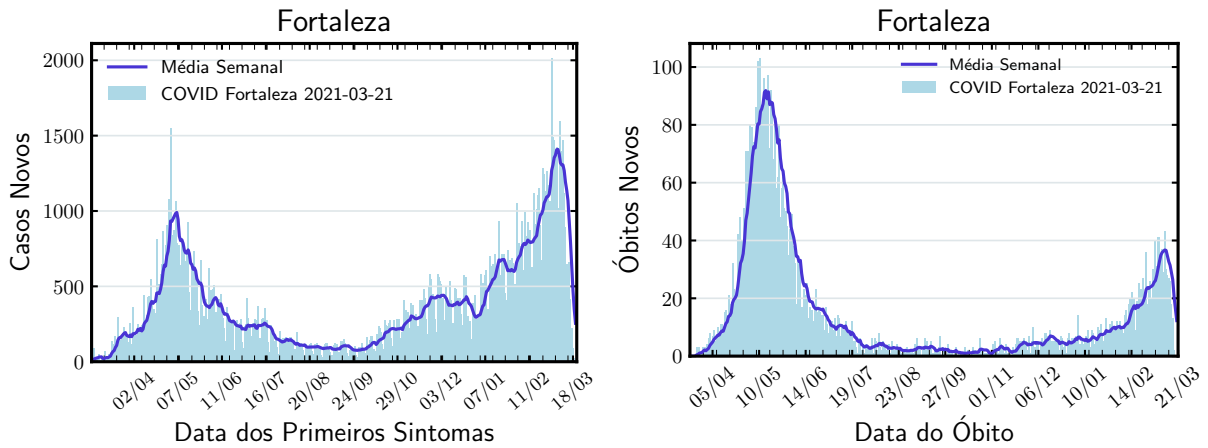


Figura 10 – Média móvel semanal do número de Casos Novos diários notificados pelo sistema de saúde (a esquerda) e a média móvel do número de óbitos diários notificados (a direita). Observa-se a formação de duas "ondas" com altas taxas de contaminação. A primeira sendo anterior ao decreto de *Lockdown*, que foi estabelecido em 05/05/2020, após esta data o número de casos novos foi diminuindo ao longo do tempo assim como o número de óbitos, até voltar a um crescimento exponencial iniciando no final do ano de 2020 e no início de 2021. A curva de casos aumentou de forma exponencial sendo maior que o primeiro pico e a de óbitos também cresceu, felizmente, a uma taxa menor em relação a primeira onda.

Fonte: Elaborado pelo autor.

de reprodução efetiva R_e que é número de reprodução R_0 no tempo (LI *et al.*, 2020). Para tal, dividimos a população infecciosa em duas classes: a primeira corresponde à população que desenvolveu os sintomas e teve a infecção por SARS-CoV-2 confirmada pelo sistema, tal população recebeu a nomenclatura de infectada infecciosa I_r ; a segunda classe corresponde a população infecciosa assintomática, ou que não tenha a infecção confirmada pelo sistema de saúde, rotulamos como I_u (população não relatada). Definimos β como sendo a taxa de transmissão para a população infecciosa relatada I_r .

Neste modelo assume-se que a grande maioria dos indivíduos infecciosos relatados são casos sintomáticos, em contraparte, o número de indivíduos não relatados - dos quais são supostamente em sua maioria são assintomáticos. Esta suposição é fundamental, sendo ainda reforçada com a descoberta de que as pessoas assintomáticas são 42% menos propensas a transmitir o SARS-CoV-2 do que as sintomáticas (BYAMBASUREN *et al.*, 2020). Diante destas informações a população infecciosa relatada I_r , é reduzida por um fator μ . Deste modo, temos que a taxa dependente do tempo que os indivíduos ficaram expostos ao vírus é dada nesse modelo por

$$\lambda(t) = \beta \frac{(I_r + \mu I_u)}{N}, \quad (2.25)$$

onde N é a população total que assumimos como constante.

Definimos que uma certa fração α da população é considerada infecciosa a uma taxa σ , e fração complementar $(1 - \alpha)$ evolui para os infectados não relatados no mesmo período. É comum assumirmos que ambas as populações infecciosas tornam-se parte da população removida a mesma taxa γ . Para fins de comparação entre o modelo utilizado e os dados disponíveis sobre o número diário de óbitos, fez-se necessário o monitoramento também da fração ϕ da população infecciosa relatada removida que evolui para óbito, dessa maneira, a população falecida relatada D_r aumenta a uma taxa $(\phi \gamma I_r)$.

O número de reprodução efetivo para este modelo é dado por (LI *et al.*, 2020):

$$R_e = \left(\alpha \frac{\beta}{\gamma} + (1 - \alpha) \frac{\mu \beta}{\gamma} \right) \frac{S}{N}. \quad (2.26)$$

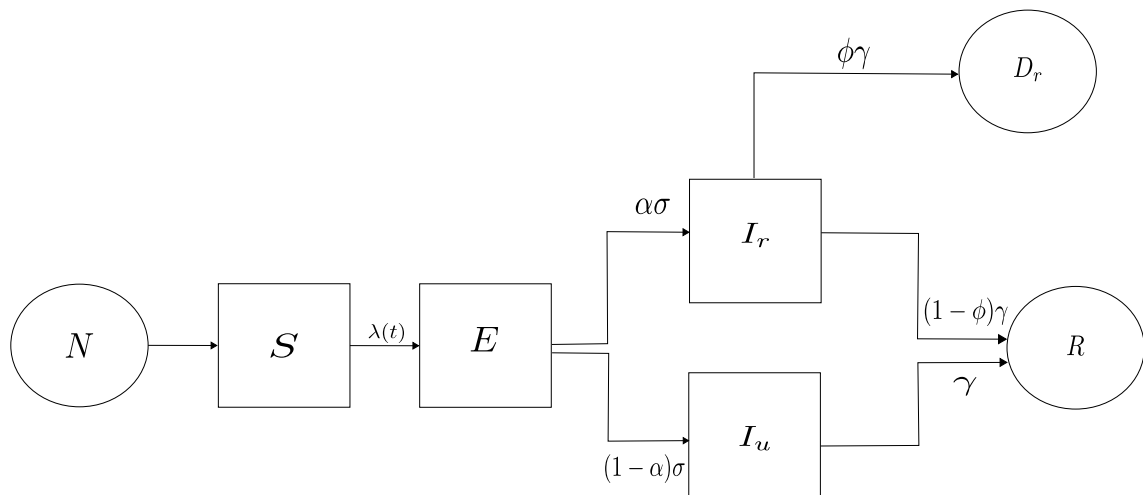


Figura 11 – O modelo epidemiológico SEIIR. A população total N nutre a população de indivíduos suscetíveis S . Os indivíduos suscetíveis tornam-se expostos E uma taxa $\lambda(t)$. Os indivíduos expostos tornam-se infectados contagiosos a uma taxa σ . Uma fração α da população infectada torna-se infectada relatada, I_r , e uma fração $(1 - \alpha)$ se torna infectados não relatados I_u . Os indivíduos infectados relatos e não relatados são removidos a uma taxa γ . Uma fração ϕ dos infectados relatados evoluem para óbito.

Fonte: Elaborado pelo autor.

Na Eq. 2.26 o número de reprodução efetiva da população infectada relatada reside no primeiro termo, ao mesmo tempo que o segundo termo é a contribuição da população infectada não relatada. Na Fig. 12 temos a representação do modelo dividida nos compartimentos.

3 INFERÊNCIA BAYESIANA

Atualmente duas abordagens estatísticas são utilizadas em análise de dados. A primeira, é a abordagem frequentista, a qual avalia os procedimentos com base em amostras repetidas, partindo de um modelo particular (a *likelihood*), no qual se define a distribuição condicional de probabilidade com parâmetros desconhecidos. As propriedades deste procedimento são avaliadas em um *framework* repetido para trabalhar com parâmetros fixos desconhecidos. A segunda, a abordagem bayesiana requer um modelo de amostras e uma distribuição *a priori* de todas as quantidades desconhecidas do modelo (parâmetros e dados ausentes). A distribuição *a priori* e a *likelihood* são utilizadas para a obtenção da distribuição *a posteriori*.

3.1 Abordagem clássica ou frequentista

A abordagem clássica ou frequentista cujo os “fundadores”: Karl Pearson, Ronald A. Fisher e Jerzy Neyman, tem como principal objetivo determinar, sempre que possível, generalizações sobre a população estudada a partir de uma amostra da mesma recolhida. A amostra é definida como um conjunto de observações ou dados estatísticos resultantes de experimentos repetidos em condições semelhantes (*ensembles*), e a população é o conjunto de todas as observações possíveis em condições semelhantes. Definindo $x = (x_1, x_2, x_3, \dots, x_n)$ o vetor que representa o conjunto de uma determinada amostra, onde n é a dimensão da amostra ou conjunto de dados, podemos então estabelecer um conjunto \mathcal{X} de todas as amostras possíveis e assim defini-lo por espaço amostral (PAULINO *et al.*, 2018).

Vale ressaltar que a inferência frequentista consiste em reconhecer a variabilidade de uma amostra para uma amostra. Para o estabelecimento de inferências é importante ressaltar que dados observados constituem apenas um dos muitos conjuntos que poderiam ter sido obtidos nas mesmas condições. A interpretação dos dados depende não apenas do conjunto observado mas também das hipóteses que foram adotadas nos possíveis conjuntos alternativos de dados. Sendo \mathcal{F}_0 uma função de distribuição que representa a variabilidade ou incerteza na observação de \mathcal{X} , definida como uma variável aleatória que representa as observações das amostras. Quando trabalhamos com um fenômeno e buscamos modelá-lo, têm-se o conhecimento sobre o fenômeno ou processo aleatório que leva a propor uma família de distribuições \mathcal{F} no qual \mathcal{F}_0 pertence, designando assim um modelo estatístico para \mathcal{X} (PAULINO *et al.*, 2018).

As distribuições \mathcal{F} são representadas pela função de densidade de probabilidade

e se estas são rotuladas por um dado parâmetro θ com domínio no conjunto Θ , atribuído por espaço de parâmetros, sendo assim, um modelo estatístico segue a forma:

$$\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}, x \in \mathcal{X}. \quad (3.1)$$

Neste modelo de variáveis aleatórias (X_1, X_2, \dots, X_n) supostas independentes condicionalmente em θ , pode-se representá-las em termos das densidades marginais das variáveis X_i , onde $i = 1, 2, \dots, n$,

$$\mathcal{F} = \left\{ f(x|\theta) = \prod_{i=1}^n f_i(x_i|\theta) : \theta \in \Theta \right\}, x \in \mathcal{X}. \quad (3.2)$$

Podemos ter o caso mais simples no qual as variáveis são independentes e identicamente distribuídas (i.i.d), tem-se a chamada amostragem casual e o modelo pode então ser escrito:

$$\mathcal{F} = \left\{ f(x|\theta) = \prod_{i=1}^n f(x_i|\theta) : \theta \in \Theta \right\}, x \in \mathcal{X}. \quad (3.3)$$

A escolha de \mathcal{F} resulta de vários fatores:

- experimentos anteriores em condições semelhantes;
- considerações teóricas sobre o objetivo do estudo e a natureza dos fenômenos;
- considerações teóricas sobre a natureza das técnicas experimentais utilizadas.

De modo mais sucinto, a inferência frequentista busca responder se os dados x são compatíveis com a família \mathcal{F} e se forem, quais conclusões podem ser obtidas. O princípio da amostragem repetida nos diz que os métodos estatísticos devem ser avaliados através de um número indefinido de repetições, ditas hipotéticas, efetuadas em condições semelhantes. A inferência frequentista baseia-se na interpretação frequentista da probabilidade, no qual a utilização de frequências como medidas de incertezas além da avaliação dos procedimentos em termos da frequência com que produzem respostas corretas ou bons resultados (PAULINO *et al.*, 2018).

Esta abordagem apresenta uma interpretação da probabilidade baseando-se que um dado acontecimento pode ser medido ao observar a frequência relativa deste mesmo acontecimento em uma sucessão de experimentos em condições idênticas e semelhantes, representando

assim uma das suas limitações, a exemplo o lançamento de uma moeda viciada ou irregular (sendo que as faces “cara” e “coroa” não são igualmente prováveis) ao fim de 1000 sendo obtidos 600 “coroas” a probabilidade empírica seria de $600/1000 = 0,6$ (PAULINO *et al.*, 2018).

3.2 Abordagem Bayesiana

O método de se utilizar os dados observados na determinação da distribuição a *posteriori* recebe o nome de abordagem de Bayes.

No modelo Bayesiano inicia-se com uma quantidade $\mathbf{y} = (y_1, \dots, y_n)$ de amostras (dados observados) e um vetor de parâmetros desconhecidos θ . As amostras geralmente são dadas na forma de uma distribuição de probabilidade $f(\mathbf{y}|\theta)$. Quando observamos a relação inversa, $f(\theta|\mathbf{y})$, esta função é chamada de *likelihood* e escrita como $\mathcal{L}(\theta; \mathbf{y})$. Para valores em particular de \mathbf{y} , pode-se obter o valor $\hat{\theta}$ que maximiza a *likelihood* (CARLIN; LOUIS, 2008)

$$\hat{\theta} = \operatorname{argmax} \mathcal{L}(\theta; \mathbf{y}). \quad (3.4)$$

Este valor é chamado de *Maximum Likelihood Estimation* (MLE) de θ , proposto inicialmente nos trabalhos de Fisher (FISHER, 1922) e é um dos métodos estatísticos mais utilizados atualmente.

3.2.1 Regra de Bayes

Ao invés de supor que θ é um parâmetro fixo, pode-se pensar nele como uma quantidade aleatória. Adotando-se uma distribuição de probabilidade de θ , tal distribuição resume toda e qualquer informação que possuímos de θ . Vale ressaltar que esta informação não está relacionada com aquela fornecida pelo dado \mathbf{y} , define-se esta distribuição como a *priori*.

Assim como a *likelihood* apresenta parâmetros θ , a distribuição a *priori* também apresenta parâmetros η , parâmetros comumente chamados de hiperparâmetros. Por ora, assume-se que estes hiperparâmetros são conhecidos e escrevemos a distribuição a *priori* como $\pi(\theta) \equiv \pi(\theta|\eta)$. A inferência em relação a θ é, então baseada, na distribuição a *posteriori*:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}, \theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}, \theta)}{\int p(\mathbf{y}, \theta) d\theta} = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int p(\mathbf{y}, \theta)\pi(\theta) d\theta}. \quad (3.5)$$

A Eq. 3.5 é conhecida como Teorema de Bayes. A distribuição a *posteriori* é o produto da *likelihood* e a distribuição a *priori* normalizada. A Eq. 3.5 é uma versão contínua

da fórmula de Bayes (BAYES, 1763), em geral, a versão discreta desta equação é mais familiar sendo introduzida em cursos introdutórios de estatística:

$$P(B_j|A) = \frac{P(A, B_j)}{P(A)} = \frac{P(A, B_j)}{\sum_j P(A, B_j)} = \frac{P(A|B_j)P(B_j)}{\sum_j P(A|B_j)P(B_j)}. \quad (3.6)$$

Um dos grandes desafios na obtenção da *posteriori* é o cálculo da integral no denominador da Eq. 3.5. Em alguns casos, a integral pode ser calculada de forma fechada onde obtêm-se uma distribuição *posteriori* também fechada (CARLIN; LOUIS, 2008).

Um exemplo bem simples de aplicação do Teorema de Bayes no qual utiliza-se o mercado de ações (MORETTIN; BUSSAB, 2017).

Exemplo: Seja y o rendimento do IBOVESPA (Índice da Bolsa de Valores de São Paulo), em porcentagem, por período. Nosso objetivo é estimar se o rendimento é positivo ($y > 0$) ou negativo ($y < 0$). Façamos θ o “estado do mercado”, sendo (θ_1) mercado em alta e (θ_2) mercado em baixa. Supondo que toda informação que temos *a priori* sobre θ_1 e θ_2 esteja resumida em:

$$\begin{aligned} p(\theta_1) &= p(\theta = \theta_1) = 3/5, \\ p(\theta_2) &= p(\theta = \theta_2) = 2/5. \end{aligned} \quad (3.7)$$

As verossimilhanças são dadas por:

$$\begin{aligned} p(y > 0|\theta), \\ p(y < 0|\theta). \end{aligned} \quad (3.8)$$

Tais verossimilhanças são conhecidas no Teorema de Bayes e vamos supor que seus valores expressos:

$$\begin{aligned} p(y > 0|\theta_1) &= 2/3, p(y > 0|\theta_2) = 1/3, \\ p(y < 0|\theta_1) &= 1/3, p(y < 0|\theta_2) = 2/3. \end{aligned} \quad (3.9)$$

Calcula-se as probabilidades conjuntas por meio de $p(y, \theta) = p(\theta)p(y|\theta)$. Assim,

$$\begin{aligned}
p(y > 0 | \theta = \theta_1) &= p(\theta = \theta_1) \cdot p(y > 0 | \theta = \theta_1) = 3/5 \times 2/3 = 6/15, \\
p(y > 0 | \theta = \theta_2) &= p(\theta = \theta_2) \cdot p(y > 0 | \theta = \theta_2) = 2/5 \times 1/3 = 2/15, \\
p(y < 0 | \theta = \theta_1) &= p(\theta = \theta_1) \cdot p(y < 0 | \theta = \theta_1) = 3/5 \times 1/3 = 3/15, \\
p(y < 0 | \theta = \theta_2) &= p(\theta = \theta_2) \cdot p(y < 0 | \theta = \theta_2) = 2/5 \times 2/3 = 4/15.
\end{aligned} \tag{3.10}$$

O Teorema de Bayes, fornece as probabilidades *a posteriori* de θ_1 e θ_2 , de acordo com o valor observado y :

$$p(\theta | y) = \frac{p(\theta)p(y|\theta)}{p(y)}. \tag{3.11}$$

O termo $p(y)$ é chamado de probabilidades marginais preditoras, sendo calculadas de acordo com:

$$p(y) = \sum_{\theta} p(\theta)p(y|\theta). \tag{3.12}$$

Valendo-se da Eq. 3.12 em nosso caso, obtêm-se:

$$\begin{aligned}
p(y > 0) &= p(\theta)p(y > 0 | \theta_1) + p(\theta)p(y > 0 | \theta_2) = 8/15, \\
p(y < 0) &= p(\theta)p(y < 0 | \theta_1) + p(\theta)p(y < 0 | \theta_2) = 7/15,
\end{aligned} \tag{3.13}$$

A distribuição *a posteriori* é então obtida por:

$$\begin{aligned}
p(\theta = \theta_1 | y > 0) &= \frac{p(\theta_1)p(y > 0 | \theta_1)}{p(y > 0)} = 3/4, \\
p(\theta = \theta_2 | y > 0) &= \frac{p(\theta_2)p(y > 0 | \theta_2)}{p(y > 0)} = 1/4, \\
p(\theta = \theta_1 | y < 0) &= \frac{p(\theta_1)p(y < 0 | \theta_1)}{p(y < 0)} = 3/7, \\
p(\theta = \theta_2 | y < 0) &= \frac{p(\theta_2)p(y < 0 | \theta_2)}{p(y < 0)} = 4/7.
\end{aligned} \tag{3.14}$$

Como os resultados das Eq. 3.14 pode-se “estimar” a alta ou baixa do mercado tomando o valor máximo da probabilidade *a posteriori* dada a informação sobre o rendimento. Este modelo é estático, em modelos dinâmicos esses valores são atualizados de período para período (diário, mensal etc.) (MORETTIN; BUSSAB, 2017).

3.2.2 Monte Carlo Cadeias de Markov

O método Bayesiano Monte Carlo cadeias de Markov, ou simplesmente MCMC é utilizado com frequência em trabalhos relacionados à epidemia para estimar parâmetros desconhecidos em modelos epidemiológicos. Sua abordagem é feita utilizando a Inferência Bayesiana a relacionando a informação *a priori* dos parâmetros do modelo, seja de estudos anteriores ou a própria análise inicial da coleta de informações, combinada com os novos dados para gerar estimativas na forma de uma distribuição de parâmetros (REIS *et al.*, 2022) (PANDEY *et al.*, 2013).

Valendo-se das distribuições *a priori* dos parâmetros $p(\theta)$ e da *likelihood* $p(D|\theta)$ a distribuição *a posteriori* é obtida pela regra de Bayes Eq. 3.5. A *likelihood* precisa ser maximizada para obter-se o melhor conjunto de parâmetros ajustados (PANDEY *et al.*, 2013). Assume-se que os erros são do tipo gaussiano com média zero e variância σ^2 sendo o erro dado por:

$$E^2 = \sum_i (D_i - y_i(\theta))^2, \quad (3.15)$$

no qual D são os dados analisados e $y_i(\theta)$ são os valores estimados. A *likelihood* é expressa por:

$$\mathcal{L}(\theta) = p(D|\theta) \propto \exp\left(\frac{-E^2}{2\sigma^2}\right). \quad (3.16)$$

Dada distribuição *a priori* dos parâmetros, a *posteriori* é calculada utilizando a regra de Bayes, para o caso em questão:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}. \quad (3.17)$$

Em geral existe uma dificuldade no cálculo da integral no denominador da Eq. 3.17, tal dificuldade pode ser contornada valendo-se da relação de proporcionalidade:

$$p(\theta|D) \propto p(D|\theta)p(\theta). \quad (3.18)$$

Usando a Eq. 3.18, forma-se uma cadeia de Markov que assintoticamente converge para a distribuição *a posteriori* dos parâmetros através do algoritmo de Metropolis. O algoritmo

de Metropolis é um procedimento iterativo que utiliza uma regra de aceitação-rejeição para a convergência da distribuição requerida, sendo necessário apenas uma função proporcional da distribuição a ser feita a amostragem (CARLIN; LOUIS, 2008)(PANDEY *et al.*, 2013).

Após a inicialização do vetor de parâmetros iniciais $\theta^{(0)}$, vetor este obtido a partir de $p(\theta)$, a i -ésima iteração do algoritmo é definida como:

1. Para cada iteração $i = 1, 2, 3 \dots$. O novo conjunto de parâmetros θ^* é definido a partir da distribuição proposta $J(\theta^*|\theta^i)$, onde $J(\theta^*|\theta^i)$ deve ser simétrica.
2. A probabilidade de aceitação de θ^* é calculada

$$r = \min\left(\frac{p(\theta^*|D)}{p(\theta^i|D)}, 1\right) = \min\left(\frac{p(D|\theta^*)p(\theta^*)}{p(D|\theta^i)p(\theta^i)}, 1\right); \quad (3.19)$$

3. Gerar um número aleatório $\delta \sim U(0, 1)$ e fazer

$$\theta^{i+1} = \begin{cases} \theta^*, & \text{se } \delta < r; \\ \theta^i, & \text{caso contrário.} \end{cases} \quad (3.20)$$

Este algoritmo deve ter um número suficiente de iterações para que haja uma convergência dos parâmetros para a distribuição a *posteriori*. Existem vários testes que podem ser utilizados para testar se a cadeia convergiu, um deles é o teste Gelman-Rubin no qual a geração de múltiplas cadeias semelhantes inicializadas com um diferente conjunto de parâmetros obtidos através de um distribuição uniforme (GELMAN; RUBIN, 1992). As variâncias de cada cadeia é comparada com a variância entre as cadeias e grandes desvios entre essas variâncias indicam não-convergência (PANDEY *et al.*, 2013).

Na literatura encontram-se trabalhos que utilizam o método MCMC, no qual para $J(\theta^*|\theta^i)$ utilizaram uma distribuição normal multivariada (REIS *et al.*, 2022)(PANDEY *et al.*, 2013):

$$J(\theta^*|\theta^i) \sim N(\theta^i, \lambda^2 \Sigma) \quad (3.21)$$

sendo (θ^i) a média e $(\lambda^2 \Sigma)$ a variância. Para a distribuição normal multivariada proposta a taxa de aceitação é (ROBBINS; MONRO, 1951)

$$\hat{\alpha} = \begin{cases} 0.44 & \text{se } d = 1 \\ 0.23 & \text{se } d > 1 \end{cases} \quad (3.22)$$

onde d é o número de parâmetros a serem estimados.

3.3 Filtro de Kalman (KF)

Proposto em 1960 por Rudolf E. Kalman, o filtro de Kalman consiste em um algoritmo utilizado em sistemas dinâmicos para estimar variáveis desconhecidas em dados observados que contenham algum tipo de ruído ou outros tipos de imprecisões. Ao longo dos últimos anos, este algoritmo vem sendo utilizado em diversos tipos de trabalhos a saber, cálculos de órbita de espaçonaves, rastreamento de alvo de manobra e posicionamento de GPS devido sua eficiência e rapidez na execução (LI *et al.*, 2015). Contudo, o filtro de Kalman também apresenta limitações. A principal delas está no fato de que sua aplicação se dá apenas em sistemas lineares requerindo que o tipo de ruído seja ruído branco Gaussiano, ou seja, cada amostra possui uma distribuição normal com média nula. Conclui-se que a aplicação na forma padrão do filtro de Kalman é limitado. Tais limitações são corrigidas por variações deste método tais como filtro de Kalman Extendido (EKF) e o filtro de Kalman Sem Rasto (UKF) que abrangem sistemas não lineares. Neste trabalho utilizamos o Filtro de Kalman Iterativo por Amostragem (IEFK).

KF é um dos métodos recursivos de estimativa linear mais famosos das teorias de filtros Bayesianos e é composto por um conjunto de equações que estima o estado de um sistema linear partindo de um sucessão de medições de grandezas realizadas ao longo de uma evolução temporal. O fato do filtro ser recursivo reside no fato de que ao fazer a estimativa do estado atual, precisa-se apenas da estimativa no passo anterior e a medição atual (SANTOS, 2017). As equações do filtro de Kalman são:

$$x_k = \mathbb{A}x_{k-1} + \mathbb{B}u_{k-1} + w_k, \quad (3.23)$$

$$z_k = Hx_k + v_k, \quad (3.24)$$

em que x_k é o vetor de estado, z_k o vetor de observação, \mathbb{A} é a matriz de transição de estados, H a matriz de observação, w_k é o vetor de ruído do sistema, u_{k-1} é o vetor de controle do sistema e v_k é o vetor ruído de observação. Os ruídos do processo são independentes entre si, e conforme descrito acima são ruídos brancos com distribuição de probabilidade normal. Os vetores w_k e v_k são simétricos, não correlacionados e satisfazem (LI *et al.*, 2015):

$$E(w) = 0, \quad (3.25a)$$

$$\text{cov}(w) = E(w w^T) = \mathbb{Q}, \quad (3.25b)$$

$$E(v) = 0, \quad (3.26a)$$

$$\text{cov}(v) = E(vv^T) = \mathbb{R}, \quad (3.26b)$$

$$E(wv^T) = 0. \quad (3.27)$$

Associado ao ruído do sistema é definida a matriz \mathbb{Q} , como sendo a covariância de w ; temos a matriz \mathbb{R} referente a covariância do ruído das medidas v . A matriz \mathbb{Q} , reflete a medida da confiança no modelo dinâmico utilizado pelo filtro de Kalman, posto que quanto maior a confiabilidade no modelo, menor os valores das componentes da matriz. De modo semelhante, a matriz \mathbb{R} representa a confiança nas observações do processo (SPINGARN, 1987).

O objetivo principal do filtro de Kalman é minimizar a variância dos estados estimados do sistema dinâmico linear. Para isto, utiliza-se a função custo da média dos quadrados dos erros de estimação (Minimu Mean Square Error- MMSE), utilizada por possuir um mínimo global que representa a covariância dos estados estimados, além de possuir uma convergência rápida (LEI *et al.*, 2017):

$$E \left[(x_k - \hat{x}_k)^2 \right] = E[e_k^2], \quad (3.28)$$

onde temos E representando a média, x_k o vetor de estado do sistema no tempo k , \hat{x}_k o vetor de estado estimado e e_k é o erro entre o estado medido e o estado observado. Valendo-se de $k - 1$ medidas, conseguimos uma estimativa *a priori* do estado, sendo representado por \hat{x}_k^- . O valor obtido é a estimativa *a posteriori* do estado \hat{x}_k no instante k .

Os desvios nas estimativas *a priori* e *a posteriori* são expressas por:

$$e_k^- = x_k - \hat{x}_k^-, \quad (3.29)$$

$$e_k = x_k - \hat{x}_k, \quad (3.30)$$

As equações equivalentes das estimativas dos erros nas covariações *a priori* e *a posteriori* são dadas por:

$$P_k^- = E[e_k^- e_k^-]^T, \quad (3.31)$$

$$P_k = E[e_k e_k^T]. \quad (3.32)$$

O objetivo do algoritmo reside em obter \hat{x}_k que calcule a estimativa de um estado posterior para conseguir as equações do filtro de Kalman. Para isso, precisamos tomar \hat{x}_k como uma combinação linear de estimativa *a priori* e a diferença ponderada entre as medições reais e o valor da previsão medida (LI *et al.*, 2015). O filtro é dividido nas Fases de Predição, onde faz-se a estimativa *a priori* do próximo estado do sistema, valendo-se das medidas disponíveis até o instante anterior e na Fase de Atualização temos que a estimativa *a priori* é combinada com a observação atual para refinar a estimativa do estado (SANTOS, 2017). As equações de previsão são definidas como:

$$\hat{x}_k^- = A\hat{x}_{k-1}^- + Bu_{k-1}, \quad (3.33)$$

$$P_k^- = AP_{k-1}^- A^T + Q, \quad (3.34)$$

Enquanto que as equações de atualização seguem:

$$K_k = \hat{P}_k^- H^T (H\hat{P}_k^- H^T + R)^{-1}, \quad (3.35)$$

$$\hat{x}_{k-1}^- = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-), \quad (3.36)$$

$$P_k = (I - K_k H)P_k^-. \quad (3.37)$$

Em que K_k é a matriz Kalman, \hat{x}_k o valor de filtro ideal, P_k a matriz de desvio do filtro e I a matriz unidade. Na literatura podemos encontrar alguns trabalhos que utilizam esta

forma padrão do Filtro de Kalman tais como nos trabalhos de (MAHFOUZ *et al.*, 2014) que propôs um método utilizando a combinação de Machine Learning com filtro de Kalman para estimar posições instantâneas de um alvo em movimento, (PATEL; THAKORE, 2013) que aplicou o filtro de Kalman no rastreamento de um único objeto em movimento no campo da segurança de sistemas de vigilância automatizados.

3.4 Filtro de Kalman Iterativo por Amostragem (IEKF)

Para estudarmos como se dá a evolução na dinâmica de propagação do SARS-CoV-2 na cidade de Fortaleza, fez-se necessário a utilização do modelo SEIIR concomitantemente com o framework de Filtro de Kalman Iterativo por amostragem descrito no Cap. 2. Esta estrutura permite-nos investigar a evolução temporal do número de reprodução eficaz, inferindo os parâmetros descrito do modelo SEIIR e as populações valendo-se dos dados de observação do número cumulativo diário de mortes confirmadas pelo Sistema de Saúde oriundas de infecções por SARS-Cov-2 e o número de casos disponíveis no Governo do Ceará COVID-19 dados públicos (CEARÁ, 2022b).

Utilizamos o framework IEKF para uma janela de 20 dias semanalmente, (em alguns períodos varia-se a janela para testar a eficácia do modelo) permitindo assim estimar o parâmetro R_e , em cada janela foi utilizado os dados da janela anterior bem como os novos dados do Governo relatados para definir os parâmetros do modelo e o valor inicial no primeiro dia da janela.

O conjunto de equações diferenciais acopladas que regem o modelo SEIIR são expressas por:

$$\frac{dS}{dt} = -\lambda S, \quad (3.38a)$$

$$\frac{dE}{dt} = \lambda S - \sigma E, \quad (3.38b)$$

$$\frac{dI_r}{dt} = \alpha \sigma E - \gamma I_r, \quad (3.38c)$$

$$\frac{dI_u}{dt} = (1 - \alpha) \sigma E - \nu I_u, \quad (3.38d)$$

$$\frac{dR_r}{dt} = (1 - \phi) \gamma I_r, \quad (3.38e)$$

$$\frac{dD_r}{dt} = \phi \gamma I_r, \quad (3.38f)$$

$$\frac{dR_u}{dt} = \nu I_u. \quad (3.38g)$$

Sujeitas ao vínculo de que a população total $N = S + E + I_r + I_u + R_r + R_u + D_r$, no nosso caso a população de Fortaleza, uma das maiores cidades do Brasil, de aproximadamente de 2,67 milhões de pessoas, é conservada.

Conforme mencionado anteriormente, existem diversos filtros que são utilizados para inferir os parâmetros de modelos epidemiológicos, a exemplo o *framework Iterated Filtering* (IF) (IONIDES *et al.*, 2006) (KING *et al.*, 2008). Tal *framework* é utilizado para inferir pelo método MLE e tem sido aplicado no estudo de doenças como cólera (KING *et al.*, 2008) e sarampo (HE *et al.*, 2010). No IF é gerado um ensemble de estados, no qual cada ensemble representa a distribuição dos parâmetros e das variáveis de interesse, através de uma série de iterações aplica-se técnicas de filtragem buscando assim a diminuição da variância. Neste processo, a distribuição dos parâmetros iterável é otimizada e converge para valores que maximizam a *likelihood* (LI *et al.*, 2020).

O modelo SEIIR apresenta uma alta dimensionalidade, logo utilizar *frameworks* como *Iterated Filtering* para inferir os parâmetros epidemiológicos requer um grande número de de partículas (SNYDER *et al.*, 2008); enquanto que o IEKF produz resultados similares com um número menor de *ensembles*, desta forma reduz o custo computacional (PEI *et al.*, 2018). Na literatura encontram-se trabalhos que comparam os diversos filtros existentes testando sua acurácia, custo computacional e grau de otimização. Pode-se citar o trabalho (YANG *et al.*, 2014) no qual são testados diversos filtros para estimar a incidência de gripe semanalmente em cidades dos Estados Unidos.

Utilizamos o *framework* Filtro de Kalman Iterativo por Amostragem (IEKF) para inferir os parâmetros do modelo, em que as populações iniciais são baseadas nas comparações

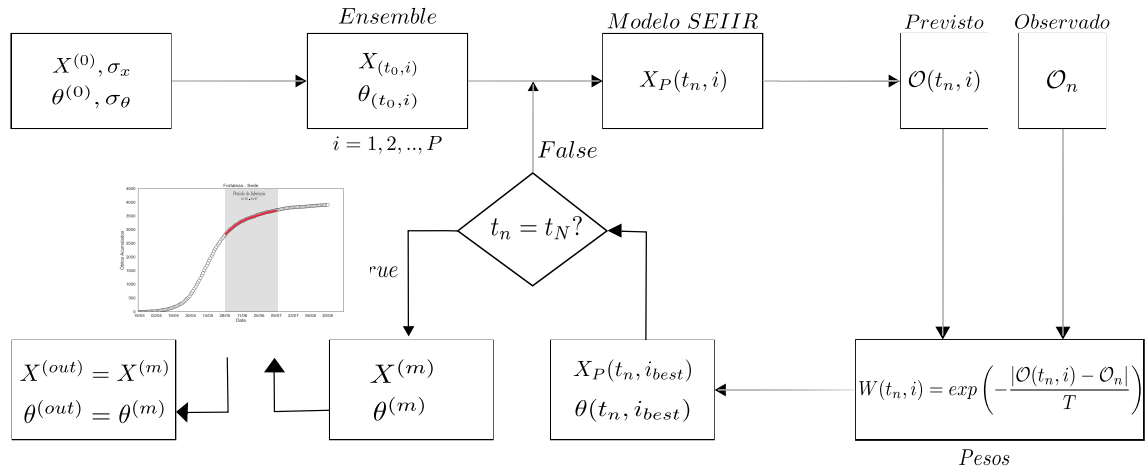


Figura 12 – Framework Filtro de Kalman Iterativo por Amostragem. Munido do vetor de estados e de parâmetros iniciais gera-se um *ensemble* de partículas. As partículas geradas são aplicadas no modelo SEIIR e integradas numericamente, após tal integração obtêm-se uma nova configuração dos parâmetros e comparamos esses resultados com os dados. Para cada comparação é atribuída um peso, o valor deste peso é utilizado para selecionar as partículas que melhores representam o conjunto de dados. O processo continua a cada iteração m até que tenhamos obtido a melhor configuração para aquela janela de dados.

Fonte: Elaborado pelo autor.

de previsões do modelo definino pela integração das equações 3.38 com um conjunto de T observações obtidas em tempos discretos em uma janela de observações. Esta mesma abordagem utilizada aqui foi utilizada em (PONTE *et al.*, 2021) no qual buscava-se investigar as taxas de transmissão da COVID-19 em ônibus em diferentes fases da pandemia comparando-os com os resultados de R_e na cidade de Fortaleza, nos ônibus e entre os profissionais da saúde que utilizam o transporte público. Este filtro assume que a distribuição *a priori* e a *likelihood* são gaussianas e que o ajuste feito da distribuição *a priori* para a *posteriori* segue a regra de Bayes (ANDERSON, 2001).

O framework é inicializado com um vetor de estados inicial definido por $X^{(0)} = \{S, E, I_r, I_u, R_r, R_u, D_r\}^{(0)}$, e um vetor de parâmetros iniciais $\theta^{(0)} = \{\beta, \mu, \alpha, \sigma, \gamma, \nu, \phi\}^{(0)}$. Tais vetores possuem incertezas conhecidas pelas matrizes de variância σ_X e σ_θ respectivamente. O processo de cada iteração m inicia-se com a geração de um *ensemble* de P "partículas" de tal modo que o estado inicial no tempo t_0 é sorteado de uma distribuição normal multivariada com média $X^{(m-1)}$ e variância $a^{(m-1)}\sigma_X$, onde a que varia $0 < a < 1$ é um fator de resfriamento dos estados do modelo. Cada partícula i possui o estado inicial definido por $X(t_0, i) = \text{normal}(X^{(m-1)}, a^{(m-1)}\sigma_X)$. Esses estados são utilizados para fixar o estado o vetor $X_f(t_0, i)$ que definem a distribuição *a posteriori* no instante t_0 . Vale ressaltar que cada partícula i possui seu vetor de parâmetros iniciais $\theta(t_0, i) = \text{normal}(\theta^{(m-1)}, a^{(m-1)}\sigma_\theta)$, no qual $0 < b < 1$ é um fator de

resfriamento para os parâmetros.

A inferência se dá integrando numericamente o modelo valendo-se das condições iniciais, de tal maneira que o vetor de estados previsto para uma dada partícula i no tempo t_n é dado por: $X_P(t_n, i) = f(X_F(t_{n-1}, i), \theta(t_{n-1}, i))$, isto é, a distribuição a *posteriori* é obtida a partir dos resultados obtidas da distribuição a *posteriori*. Munido dessas previsões, atribui-se um peso $W(t_n, i)$ para cada partícula i definido como:

$$W(t_n, i) = \exp\left(\frac{|\mathcal{O}(t_n, i) - \mathcal{O}_n|}{T}\right). \quad (3.39)$$

$\mathcal{O}(t_n, i)$ é o resultado previsto para a quantidade que é observada no tempo t_n pela partícula i . A escolha da função 3.39 deu-se devido a rápida convergência, sendo assim um fator crucial na obtenção de resultados a fim de monitorar propagação de doenças com altas taxas de transmissão. Para nosso estudo, $\mathcal{O}(t_n, i)$ é a previsão do número de óbitos acumulados relatados $D_r(t_n, i)$.

O processo de filtragem se dá mantendo-se no modelo apenas as partículas com maior peso de probabilidade $\mathcal{P} = W(t_n, i) / \sum_j W(t_n, j)$. Os estados das partículas filtradas $X_F(t_n, i) = X_P(t_n, i_{best})$ irão definir a distribuição a *posteriori* no tempo t_n , i_{best} são os índices das partículas filtradas (KING *et al.*, 2008). O vetor que armazena os parâmetros é constantemente atualizado no tempo t_n no qual sorteia-se um novo valor com base nos índices i_{best} a partir de uma distribuição normal multivariada $\theta(t_n, i) = \text{normal}(\theta(t_{n-1}, i_{best}), b^{(m-1)} \sigma_\theta)$. As iterações continuam até serem estabelecidos o vetor de estados $X^{(m)}$ e o vetor de parâmetros $\theta^{(m)}$ para a próxima iteração. O vetor de parâmetros seguinte é obtido por (IONIDES *et al.*, 2006):

$$\theta^{(m)} = \theta^{(m-1)} + V(t_1) \sum_{n=1}^N V^{-1}(t_n) (\bar{\theta}(t_n) - \bar{\theta}(t_{n-1})). \quad (3.40)$$

no qual têm-se $\bar{\theta}(t_n)$ é a média amostral dos melhores parâmetros no tempo t_n , $\bar{\theta}(t_n, i_{best})$ e $V(t_n)$ é a variância (IONIDES *et al.*, 2006) (KING *et al.*, 2008). O próximo vetor de estados é calculado pela média amostral dos estados filtrados.

$$X^{(m)} = \frac{1}{P} \sum_{j_{best}=1}^P X(t_0, j_{best}). \quad (3.41)$$

O vetor de estados $X^{(m)}$ e o vetor de parâmetros $\theta^{(m)}$ a cada iteração m são utilizados para computar a evolução da obtenção dos estados e parâmetros para a janela de observações

$1, \dots, t_T$. Para se verificar a precisão das inferências obtidas é calculado o erro com base nos resultados obtidos e dos dados disponíveis

$$\varepsilon^{(m)} = \frac{1}{T} \sum_{n=1}^T \left| \hat{\theta}_n^{(m)} - \theta_n \right|^2. \quad (3.42)$$

Continuam-se as iterações até que o módulo da diferença entre os erros $\left| \varepsilon^{(m)} - \varepsilon^{(m-1)} \right|$ seja menor que 0.01.

4 RESULTADOS DAS INFERÊNCIAS

A partir da integração das equações do modelo SEIIR, conseguimos obter resultados interessantes para a região de Fortaleza. As Figs.13,14 e 15 nos mostram o quão preciso foram as inferências, mostrando assim que o modelo responde bem ao conjunto de dados analisados. Vale ressaltar que um dos problemas fundamentais neste conjunto de dados é a presença de muitos mínimos locais, dificultando assim em alguns momentos resultados mais precisos, uma vez que o modelo pode se fixar em um mínimo, acertar nos primeiros dados e distanciar-se dos demais ao longo da janela. Ao todo foram analisadas 13 janelas de 20 dias cada, com uma exceção no qual testamos para uma janela maior de 50 entre os meses de outubro e novembro de 2020. A escolha do tamanho do tamanho da janela, deu-se devido a necessidade de testar o modelo como um auxílio no estudo da pandemia para tomada de decisões relativas às políticas públicas sanitárias.

Optou-se por se fazer as inferências no número de óbitos, devido, infelizmente, ser o dado no qual têm-se uma precisão maior da data do ocorrido, uma vez que cada paciente ao relatar a data dos primeiros sintomas da doença, pode enganar-se no dia, de fato, em que apresentou os sintomas iniciais da doença, levando assim uma imprecisão na data de contaminação real.

Na Fig. 16 temos a evolução temporal ao longo de 2020 até março de 2021 de um dos parâmetros epidemiológicos mais importantes, o número de reprodução efetivo R_t . Em consequência da variação do R_t ser ao longo do tempo, a análise deste número permite analisar a eficácia de políticas públicas para conter o avanço da contaminação de doenças, em especial a COVID-19, ou contribuição para a tomada de novas medidas, sejam elas de restrição ou relaxação.

Observa-se nos estágios iniciais da pandemia o número que o R_t apresenta valores altíssimos, sendo necessário a intensificação de medidas mais restrições para conter o avanço do patógeno. Na Fig. 9 temos a comparação da evolução no número de casos e óbitos diários notificados, inicialmente seguindo um comportamento semelhante de crescimento, sendo que o número de óbitos vinha crescendo de forma mais expressiva nos estágios iniciais da pandemia. Após a caracterização da primeira onda de contaminação os papéis se invertem e observa-se que o número de casos aumenta de forma mais expressiva ao final do ano de 2020 e início do ano de 2021, esse fato se deve ao avanço na vacinação, as pessoas continuavam a pegar o vírus porém com sintomas mais leves. A medida em que a pandemia vai evoluindo vemos que o R_t vai diminuindo, fator evidenciado pelos decretos (CEARÁ, 2022a) publicados que foram restringindo o fluxo de pessoas com o isolamento social entre outras medidas.

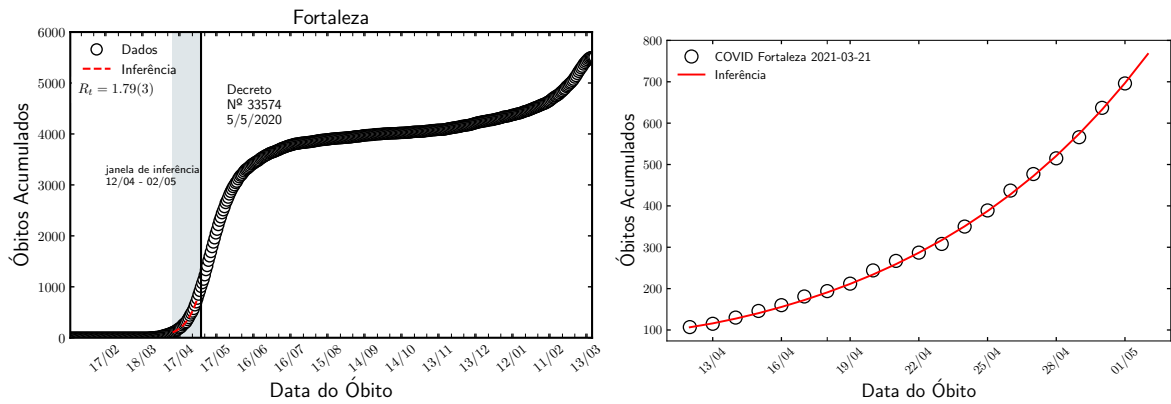


Figura 13 – Inferência para a janela 12/04/2020 a 02/05/2020, sendo destacada em cinza e a data de publicação do Decreto 05/05/2020 que instituiu o *Lockdown*. A esquerda o conjunto total de dados analisados e direita em destaque a região de Inferência. O ajuste da máxima verossimilhança do modelo SEIIR esta destacada na linha vermelha e o número cumulativo de óbitos de acordo com o período analisado nos círculos brancos. Pode-se observar claramente que a inferência gerou resultados bastante satisfatórios com os dados observados, logo, o modelo consegue descrever bem este conjunto de dados.

Fonte: Elaborado pelo autor.

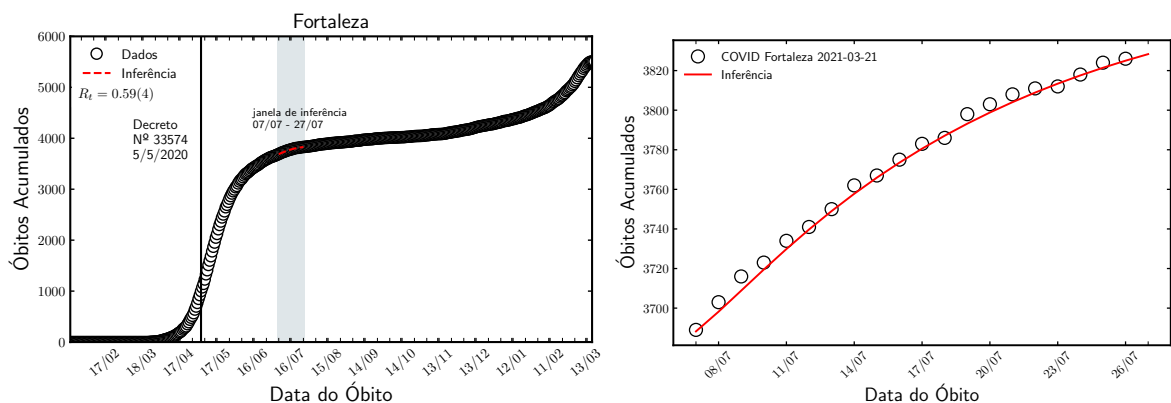


Figura 14 – Inferência para a janela 07/07/2020 a 27/07/2020, sendo destacada em cinza e a data de publicação do Decreto 05/05/2020 que instituiu o *Lockdown*. A esquerda o conjunto total de óbitos acumulados notificados e direita em destaque a região de Inferência.

Fonte: Elaborado pelo autor.

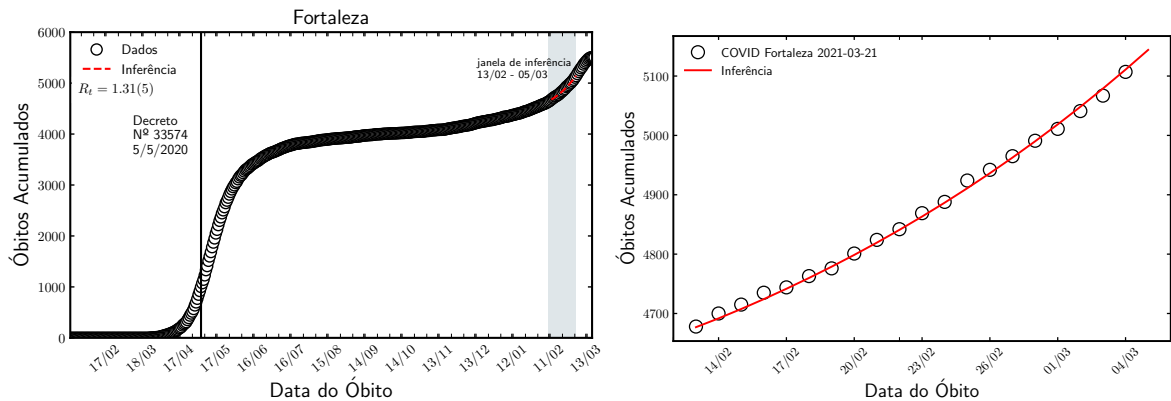


Figura 15 – Inferência para a janela 13/02/2021 a 05/03/2021, sendo destacada em cinza. A esquerda o conjunto total de óbitos acumulados notificados e direita em destaque a região de Inferência.

Fonte: Elaborado pelo autor.

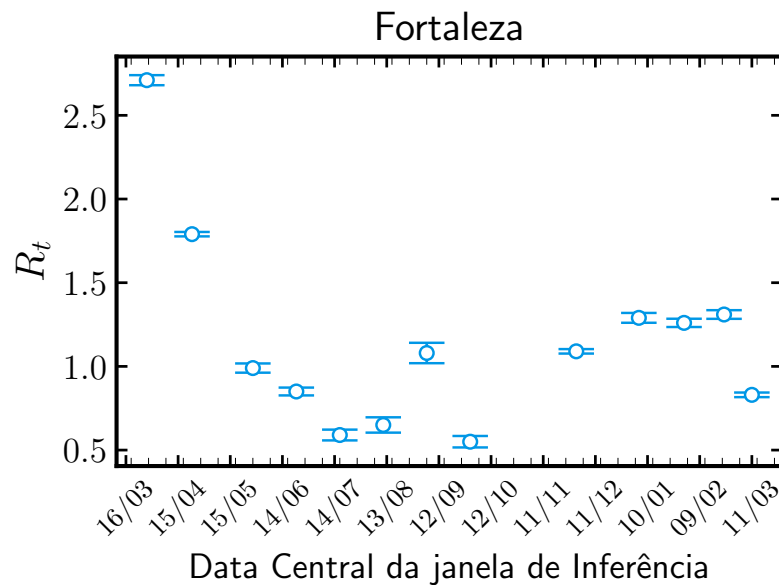


Figura 16 – Evolução temporal do parâmetro R_t ao longo do período de 19/03/2020 até 21/03/2021. A janela utilizada foi de 20 dias semanais com exceção entre outubro e novembro na qual uma janela de 50 dias foi aplicada. A partir do mês de Maio observa-se a tendência de queda entre os três meses seguintes, tal fator sendo explicado devido a publicação do decreto em 05/05/2020 que instituiu o *Lockdown* na cidade de Fortaleza. Após este período têm-se o crescimento novamente, fator explicado pelo relaxamento nos cuidados e festas de final de ano onde as pessoas tendem a confraternizar de maneira mais conjunta com um número maior de pessoas.

Fonte: Elaborado pelo autor.

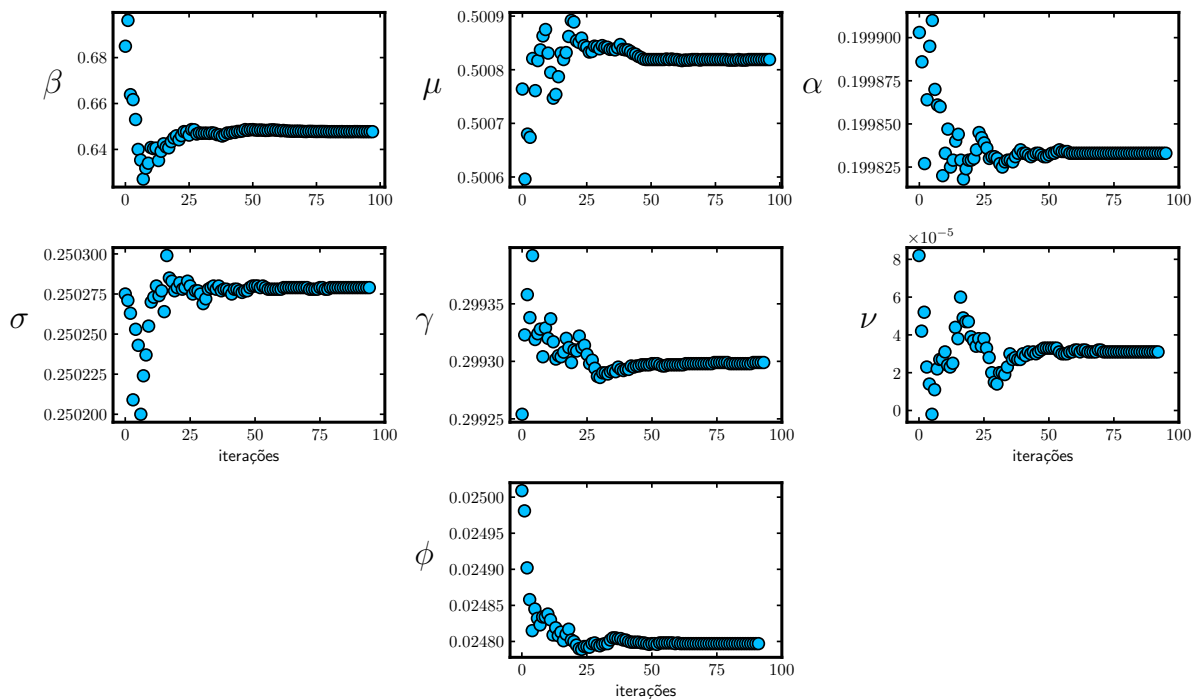


Figura 17 –Evolução dos parâmetros do modelo obtidas através do filtro de Kalman. Inicialmente os parâmetros começam com um grande flutuação em torno da região dos parâmetros iniciais atribuídos e a medida que vamos selecionando os parâmetros que maximizam a *likelihood* com maior peso de probabilidade observa-se uma convergência em cada um dos parâmetros gerados.

Fonte: Elaborado pelo autor.

A Fig. 17 nos mostra a evolução da convergência dos parâmetros para um valor que melhor representa o conjunto de dados estudados, verificando-se assim a eficácia do processo de filtragem a cada iteração. Alguns parâmetros levam mais tempo para convergirem do que outros, entretanto, em média a partir da quadragésima iteração começam a se estabilizar. Vale ressaltar que o parâmetro ν apresenta uma variação da ordem de 10^{-5} , sendo juntamente com o parâmetro σ os parâmetros que apresentam as maiores precisões no modelo. Na Fig. 18 têm-se a evolução temporal dos parâmetros do modelo, o parâmetro β que representa a taxa de transmissão do patógeno foi um dos parâmetros que mais variou ao longo do período analisado tendo um peso maior no cálculo do R_t . No estágio inicial vemos que o parâmetro μ que representa o fator redutor de β apresenta um comportamento constante entre maio e agosto de 2020 período em que o R_t manteve-se aproximadamente em queda.

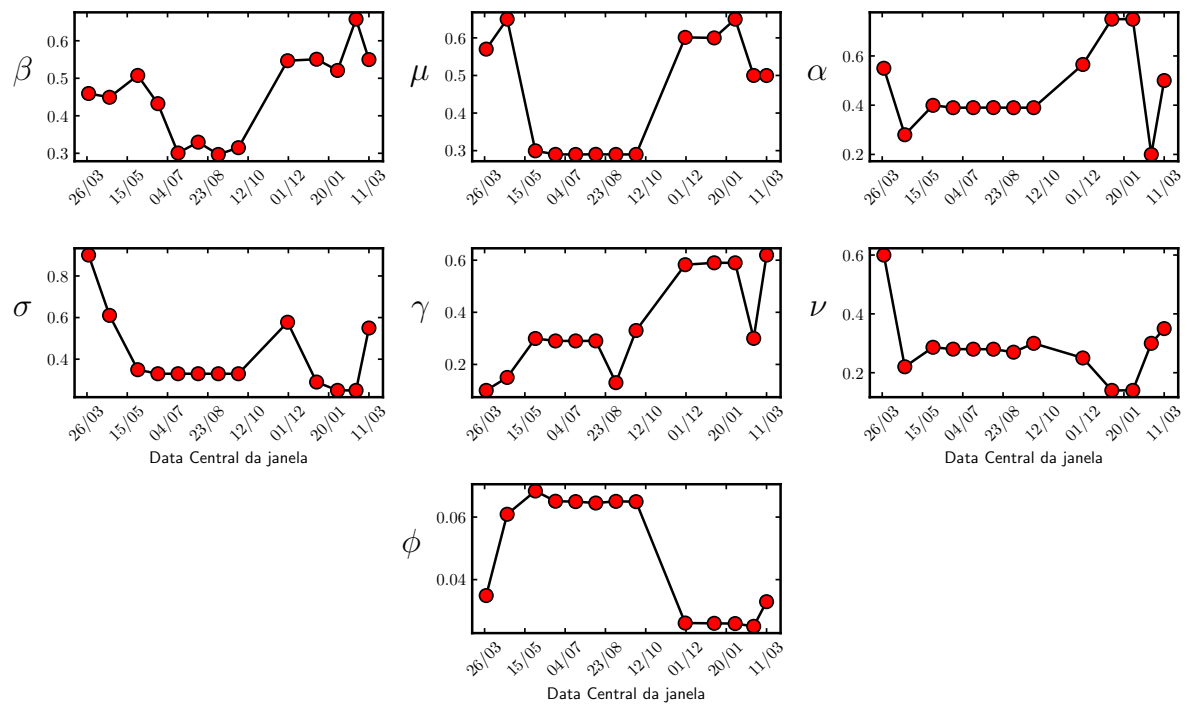


Figura 18 – Evolução Temporal dos parâmetros ao longo do período analisado. Cada janela possui o tamanho de 20 dias com um exceção entre outubro e novembro no qual uma janela de 50 dias foi utilizada. Observa-se em certas regiões uma estabilidade aparente de seus valores, comparado esses resultados com a Fig. 16, têm-se que esta região está de acordo com as janelas nas quais os valores de R_t apresentavam valores próximos.

Fonte: Elaborado pelo autor.

5 CONCLUSÃO E PERSPECTIVAS FUTURAS

Diante dos resultados apresentados, pode-se concluir que a escolha do modelo SEIIR para o tratamento do dados da pandemia da COVID-19 na cidade de Fortaleza apresentou resultados satisfatórios nos quais as inferências representaram de maneira conveniente os dados coletados. O *framework* (IEFK), mostrou-se uma ferramenta apropriada para a filtragem das partículas geradas pelo modelo, sendo necessário apenas algumas realizações para que fossem obtidos resultados que concordassem com o número cumulativo de óbitos notificados. Vale ressaltar que o *framework* aqui utilizado para modelos epidemiológicos, apresenta uma utilização mais ampla, podendo ser aplicado a sistemas dinâmicos que evolua no tempo com o objetivo de otimizar seus resultados.

A inferência Bayesiana se manteve presente em todas as etapas deste trabalho, sua utilização também não se restringe a modelos epidemiológicos, sendo aplicada em trabalhos como (LUAN *et al.*, 2022) (ALMEIDA *et al.*, 2022) para o estudo de processos de propagação do congestionamento de estradas e estimar parâmetros relevantes sobre buracos negros em acreção respectivamente.

O período estudado neste trabalho corresponde ao primeiro ano de pandemia em Fortaleza. Pretende-se dar continuidade neste trabalho analisando os dados até o ano de 2022 das cinco Superintendências Regionais de Saúde do estado Ceará, inserindo informações referentes ao IDH de cada município e verificar a relação com o número de óbitos, além de testar o modelo com os dados de grandes cidades mundiais como Nova York, Londres e Paris.

REFERÊNCIAS

- ALLEN, L. J.; DRIESSCHE, P. Van den. The basic reproduction number in some discrete-time epidemic models. **Journal of difference equations and applications**, Taylor & Francis, v. 14, n. 10-11, p. 1127–1147, 2008.
- ALMEIDA, I.; DUARTE, R.; NEMMEN, R. Deep learning bayesian inference for low-luminosity active galactic nuclei spectra. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 509, n. 4, p. 5657–5668, 2022.
- ALMEIDA, P. R. d. Modelos epidêmicos sir, contínuos e discretos, e estratégias de vacinação. Universidade Federal de Viçosa, 2014.
- ANDERSON, J. L. An ensemble adjustment kalman filter for data assimilation. **Monthly weather review**, American Meteorological Society, v. 129, n. 12, p. 2884–2903, 2001.
- ANDERSON, R. M. Discussion: the kermack-mckendrick epidemic threshold theorem. **Bulletin of mathematical biology**, Springer, v. 53, n. 1, p. 1–32, 1991.
- BALLOUX, F.; DORP, L. van. Q&a: What are pathogens, and what have they done to and for us? **BMC biology**, BioMed Central, v. 15, n. 1, p. 1–6, 2017.
- BARABÁSI, A.-L. **Network Science**. [S. l.]: Cambridge University Press, 2016.
- BAYES, T. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. **Philosophical transactions of the Royal Society of London**, The Royal Society London, n. 53, p. 370–418, 1763.
- BOYCE, W. E.; DIPRIMA, R. C.; MEADE, D. B. **Elementary differential equations and boundary value problems**. [S. l.]: John Wiley & Sons, 2021.
- BYAMBASUREN, O.; CARDONA, M.; BELL, K.; CLARK, J.; MCLAWS, M.-L.; GLASZIOU, P. Estimating the extent of asymptomatic covid-19 and its potential for community transmission: systematic review and meta-analysis. **Official Journal of the Association of Medical Microbiology and Infectious Disease Canada**, University of Toronto Press, v. 5, n. 4, p. 223–234, 2020.
- CARLIN, B. P.; LOUIS, T. A. **Bayesian methods for data analysis**. [S. l.]: CRC Press, 2008.
- CEARÁ, G. do Estado do. **Decretos do Governo do Ceará com ações contra o coronavírus**. 2022. <https://www.ceara.gov.br/decretos-do-governo-do-ceara-com-acoes-contr-o-coronavirus/>. Acessado em : 03/07/2022.
- CEARÁ, G. do Estado do. **IntegraSUS- Transparência da Saúde do Estado do Ceará**. 2022. <https://integrasus.saude.ce.gov.br/#/home>. Acessado em : 06/02/2022.
- COSTA, F.; SOUSA, Í. J. de; SANTOS, J. A.; SILVA, L. Modelo sir aplicado na dinâmica da covid-19 no estado do maranhão, brasil. **Revista de Matemática**, v. 1, n. 1, p. 18–34, 2021.
- DIEKMANN, O.; HEESTERBEEK, J.; ROBERTS, M. G. The construction of next-generation matrices for compartmental epidemic models. **Journal of the royal society interface**, The Royal Society, v. 7, n. 47, p. 873–885, 2010.

DIEKMANN, O.; HEESTERBEEK, J. A. P.; METZ, J. A. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. **Journal of mathematical biology**, Springer, v. 28, n. 4, p. 365–382, 1990.

DRIESSCHE, P. Van den; WATMOUGH, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. **Mathematical biosciences**, Elsevier, v. 180, n. 1-2, p. 29–48, 2002.

DU, Z.; WANG, L.; CAUCHEMEZ, S.; XU, X.; WANG, X.; COWLING, B. J.; MEYERS, L. A. Risk for transportation of coronavirus disease from wuhan to other cities in china. **Emerging infectious diseases**, Centers for Disease Control and Prevention, v. 26, n. 5, p. 1049, 2020.

FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character**, The Royal Society London, v. 222, n. 594-604, p. 309–368, 1922.

FORUM, W. E. **The economic toll of the coronavirus – from iPhones to solar panels to tourism**. 2020. <https://www.weforum.org/agenda/2020/02/economic-toll-coronavirus-manufacturing-tourism-china-asia/>. Acessado em : 29/06/2022.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, JSTOR, p. 457–472, 1992.

GIBBONS, C. L.; MANGEN, M.-J. J.; PLASS, D.; HAVELAAR, A. H.; BROOKE, R. J.; KRAMARZ, P.; PETERSON, K. L.; STUURMAN, A. L.; CASSINI, A.; FÈVRE, E. M. *et al.* Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. **BMC public health**, BioMed Central, v. 14, n. 1, p. 1–17, 2014.

HARKO, T.; LOBO, F. S.; MAK, M. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. **Applied Mathematics and Computation**, Elsevier, v. 236, p. 184–194, 2014.

HE, D.; IONIDES, E. L.; KING, A. A. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. **Journal of the Royal Society Interface**, The Royal Society, v. 7, n. 43, p. 271–283, 2010.

HEESTERBEEK, J.; DIETZ, K. The concept of r_0 in epidemic theory. **Statistica neerlandica**, Wiley Online Library, v. 50, n. 1, p. 89–110, 1996.

HEFFERNAN, J. J., sr, and m., wl (2005). **Perspectives on the basic reproductive ratio**. **Journal of Royal Society Interface**, v. 2, n. 4, p. 281–293, 2005.

HOWARD, J. **These 12 Viruses Look Beautiful Up Close But Would Kill You If They Could (PHOTOS)**. 2014. https://www.huffpost.com/entry/deadly-viruses-beautiful-photos_n_4545309. Acessado em: 01/12/2021.

IONIDES, E. L.; BRETÓ, C.; KING, A. A. Inference for nonlinear dynamical systems. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 103, n. 49, p. 18438–18443, 2006.

JONG, M. D.; DIEKMANN, O.; HEESTERBEEK, J. The computation of r_0 for discrete-time epidemic models with dynamic heterogeneity. **Mathematical biosciences**, Elsevier, v. 119, n. 1, p. 97–114, 1994.

- KING, A. A.; IONIDES, E. L.; PASCUAL, M.; BOUMA, M. J. Inapparent infections and cholera dynamics. **Nature**, Nature Publishing Group, v. 454, n. 7206, p. 877–880, 2008.
- LEI, B.; XU, G.; FENG, M.; ZOU, Y.; HEIJDEN, F. Van der; RIDDER, D. D.; TAX, D. M. **Classification, parameter estimation and state estimation: an engineering approach using MATLAB**. [S. l.]: John Wiley & Sons, 2017.
- LEWIS, M. A.; RENCLAWOWICZ, J.; DRIESSCHE, P. v. d.; WONHAM, M. A comparison of continuous and discrete-time west nile virus models. **Bulletin of Mathematical Biology**, Springer, v. 68, n. 3, p. 491–509, 2006.
- LI, Q.; LI, R.; JI, K.; DAI, W. Kalman filter and its application. In: IEEE. **2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)**. [S. l.], 2015. p. 74–77.
- LI, R.; PEI, S.; CHEN, B.; SONG, Y.; ZHANG, T.; YANG, W.; SHAMAN, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). **Science**, American Association for the Advancement of Science, v. 368, n. 6490, p. 489–493, 2020.
- LIM, W. M.; TO, W.-M. The economic impact of a global pandemic on the tourism economy: The case of covid-19 and macao’s destination-and gambling-dependent economy. **Current Issues in Tourism**, Taylor & Francis, v. 25, n. 8, p. 1258–1269, 2022.
- LUAN, S.; KE, R.; HUANG, Z.; MA, X. Traffic congestion propagation inference using dynamic bayesian graph convolution network. **Transportation research part C: emerging technologies**, Elsevier, v. 135, p. 103526, 2022.
- MAHFOUZ, S.; MOURAD-CHEHADE, F.; HONEINE, P.; FARAH, J.; SNOUSSI, H. Target tracking using machine learning and kalman filter in wireless sensor networks. **IEEE Sensors Journal**, IEEE, v. 14, n. 10, p. 3715–3725, 2014.
- MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. [S. l.]: Saraiva Educação SA, 2017.
- MUNSTER, V. J.; KOOPMANS, M.; DOREMALEN, N. van; RIEL, D. van; WIT, E. de. A novel coronavirus emerging in china—key questions for impact assessment. **New England Journal of Medicine**, Mass Medical Soc, v. 382, n. 8, p. 692–694, 2020.
- NEWMAN, M. **Networks: Second Edition**. [S. l.]: Oxford University Press, 2018.
- NEWMAN, M. E. **Computational physics**. [S. l.]: Createspace, 2013.
- ORGANIZATION, W. H. **Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update**. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. Acessado em : 29/06/2022.
- PANDEY, A.; MUBAYI, A.; MEDLOCK, J. Comparing vector–host and sir models for dengue transmission. **Mathematical biosciences**, Elsevier, v. 246, n. 2, p. 252–259, 2013.
- PAPANASTASIOU, Y. Fake news propagation and detection: A sequential model. **Management Science**, INFORMS, v. 66, n. 5, p. 1826–1846, 2020.
- PATEL, H. A.; THAKORE, D. G. Moving object tracking using kalman filter. **International Journal of Computer Science and Mobile Computing**, v. 2, n. 4, p. 326–332, 2013.

- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B.; SILVA, G. L. **Estatística Bayesiana**. 2a. ed. Lisboa: Fundação Calouste Gulbenkian, 2018.
- PEI, S.; KANDULA, S.; YANG, W.; SHAMAN, J. Forecasting the spatial transmission of influenza in the united states. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 115, n. 11, p. 2752–2757, 2018.
- PONTE, C.; CARMONA, H. A.; OLIVEIRA, E. A.; CAMINHA, C.; LIMA, A. S.; ANDRADE, J. S.; FURTADO, V. Tracing contacts to evaluate the transmission of covid-19 from highly exposed individuals in public transportation. **Scientific reports**, Nature Publishing Group, v. 11, n. 1, p. 1–11, 2021.
- REIS, S. D. S.; BÖTTCHER, L.; NOGUEIRA, J. P. d. C.; SOUSA, G. S.; NETO, A. S. L.; HERRMANN, H. J.; ANDRADE, J. S. Spatio-temporal characteristics of dengue outbreaks. **Frontiers in Physics**, Frontiers, p. 579, 2022.
- ROBBINS, H.; MONRO, S. A stochastic approximation method. **The annals of mathematical statistics**, JSTOR, p. 400–407, 1951.
- SANTOS, E. C. d. Avaliação de desempenho de algoritmos de localização de emissores de rf baseados em medidas de marcação em um cenário naval. 2017.
- SILVA, N. R. Modelos epidemiológicos sir, continuo com estratégias de vacinação como método de controle epidêmico. 2018.
- SNYDER, C.; BENGTSSON, T.; BICKEL, P.; ANDERSON, J. Obstacles to high-dimensional particle filtering. **Monthly Weather Review**, v. 136, n. 12, p. 4629–4640, 2008.
- SPINGARN, K. Passive position location estimation using the extended kalman filter. **IEEE Transactions on Aerospace and Electronic Systems**, IEEE, n. 4, p. 558–567, 1987.
- WESLEY, C. L.; ALLEN, L. J.; JONSSON, C. B.; CHU, Y.-K.; OWEN, R. D. A discrete-time rodent-hantavirus model structured by infection and developmental stages. **Advanced Studies in Pure Mathematics**, v. 53, p. 387–398, 2009.
- YANG, W.; KARSPECK, A.; SHAMAN, J. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. **PLoS computational biology**, Public Library of Science San Francisco, USA, v. 10, n. 4, p. e1003583, 2014.
- YANG, Y.; ZHANG, H.; CHEN, X. Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. **Annals of tourism research**, Elsevier, v. 83, p. 102913, 2020.