

Previsão de casos de COVID19 usando o modelo SEIRD e técnicas de Aprendizado de Máquina

Felipe Passarella

Agosto, 2020

Resumo

Foram estudados os modelos epidemiológicos baseados no SIR. Foi modelada uma variação que acrescenta compartimentos para indivíduos mortos e que leva em conta período de incubação do vírus e taxa de hospitalização. Os parâmetros desse modelo foram obtidos com heurísticas e o número de reprodução R_0 foi considerado como variante ao longo do tempo e portanto foi ajustado usando o método dos mínimos quadrados. Também é proposta uma abordagem para avaliar a efetividade de medidas de intervenção. Por fim, prevemos o número de casos e óbitos para a cidade de São Paulo.

Palavras-chaves: (COVID19, SARS-COV-2, SEIRD).

Introdução

A COVID-19 (do inglês Coronavirus Disease 2019) é uma doença viral infecciosa causada pelo coronavírus da síndrome respiratória aguda grave 2 (SARS-COV-2). O primeiro caso da doença surgiu na cidade de Wuhan, China em dezembro de 2019.

No dia 30 de janeiro de 2020 a Organização Mundial da Saúde reconheceu estado de emergência internacional e em 11 de março de 2020 foi declarada a pandemia com mais de 125,000 casos confirmados.

Em 02 de abril de 2020, mais de um milhão de casos haviam sido reportados como positivos e mais de 50,000 pessoas haviam morrido da doença.

Governos do mundo todo tomaram diversas decisões para conter o avanço da pandemia, como por exemplo isolamento de cidades, imposição de quarentenas, fechamento de fronteiras, etc.

Este trabalho tem o objetivo de estudar a progressão da doença em diversos cenários, os modelos necessários para tal, e por consequência a efetividade das medidas de contenção tomadas ao redor do mundo.

1 Modelo SIR

Doenças infecciosas podem ser classificadas em duas categorias, as causadas por microparasitas (vírus e bactérias) e por macroparasitas (vermes). A COVID-19 é causada por um microparásita, o coronavírus. Doenças causadas por esse tipo de agente são transmitidas quando há contato entre um hospedeiro carregando o agente e um indivíduo suscetível.

De forma simples, poderíamos usar um modelo populacional de presa-predador, como o Lotka-Volterra, para descrevermos a infecção. Teríamos a população viral como predador e indivíduos suscetíveis como presa. Porém, a população viral não nos interessa, uma vez que a carga viral que um indivíduo infectado carrega não influencia em quantos outros irá infectar. Além disso, esse tipo de modelo supõe que o vírus circula livremente pelo ambiente, o que não é verdade, já que ele está atrelado ao hospedeiro ¹.

Podemos dividir a população hospedeira em três grupos: **Suscetíveis** (S), **Infectantes** (I) e **Removidos** (R). Os infectantes são aqueles que possuem a doença e carregam consigo o vírus, os suscetíveis são aqueles que ainda estão saudáveis porém não possuem imunidade a doença e os removidos são aqueles que já contraíram a doença ou por algum motivo possuem imunidade. Além disso, os removidos podem também ser indivíduos postos em isolamento ou que morreram. Em doenças que há reinfecção, podemos acrescentar aos removidos aqueles que possuem imunidade temporária.

Partindo dessa divisão, é proposto o modelo SIR. A figura 1(a) ilustra a transição entre os grupos. Por exemplo, indivíduos suscetíveis são convertidos para infectantes, ou seja, são contaminados com a doença de acordo com uma constante β , e transicionam para o grupo de removidos de acordo com uma constante ν .

Já a figura 1(b) ilustra o modelo SIRS, que considera reinfecções, causadas de acordo com uma constante γ .

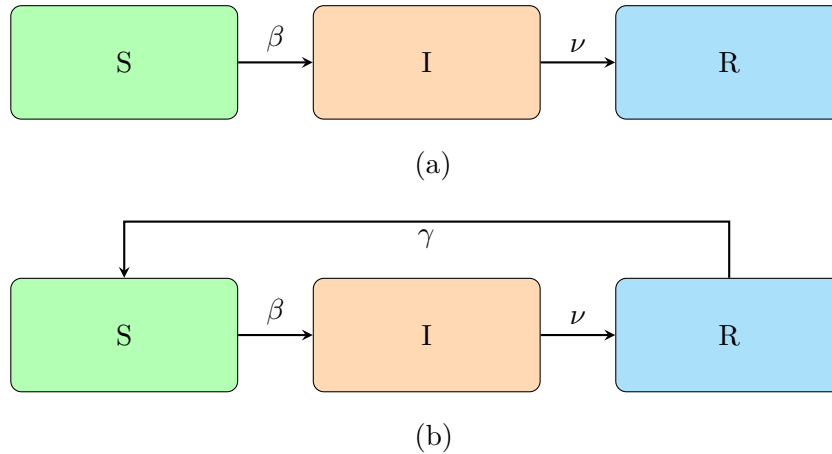


Figura 1 – (a) Modelo SIR, transição de suscetíveis para infectantes e destes para removidos. (b) Modelo SIRS, os removidos podem transicionar para suscetíveis.

Se assumirmos que a taxa de transmissão da doença é proporcional a quantidade de encontros entre suscetíveis e infectantes e que a efetividade da transmissão dependa da

¹ Após ser liberado pelo hospedeiro, o vírus pode permanecer em algumas superfícies por certo tempo [4].

constante β podemos escrever as equações do modelo SIRS como:

$$\frac{dS}{dt} = -\beta IS + \gamma R \quad (1)$$

$$\frac{dI}{dt} = \beta IS - \nu I \quad (2)$$

$$\frac{dR}{dt} = \nu I - \gamma R \quad (3)$$

Note que o modelo SIR é um caso particular do SIRS, considerando $\gamma = 0$, ou seja, sem taxa de reinfeção.

A razão β/ν tem uma interpretação interessante, ν é a taxa de remoção da classe infectante, portanto é dada em unidade de 1/tempo. Logo o tempo médio de infecção é dado por $1/\nu$ e β/ν indica a fração da população que entra em contato com a doença.

Seja $N = S + I + R$ a população total, podemos definir $R_0 = N\beta/\nu$, o número de contato infeccioso.

R_0 representa a média de infecções secundárias causadas ao ser introduzido um indivíduo infeccioso em uma população com N suscetíveis. Portanto, se $R_0 > 1$ a doença será mantida naquela população, caso contrário, a doença tende a ser erradicada.

2 Modelo SEIR

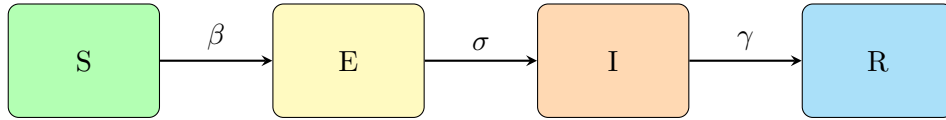


Figura 2 – Modelo SEIR

A maioria das doenças infecciosas não afetam imediatamente o indivíduo exposto a elas, é necessário um tempo de ativação até que o infectado comece a ter sintomas ou a propagar a doença.

Fazendo uma extensão natural do modelo SIR, o SEIR adiciona uma categoria, a dos **Expostos** (E), que são os indivíduos que estão infectados, porém não apresentam sintomas ou não infectam outros, por o vírus ainda estar em seu período de incubação.

Portanto, neste modelo indivíduos **Suscetíveis** (S) são convertidos para **Expostos** (E) e estes após certo tempo são convertidos para **Infectados** (I).

O sistema de equações que descreve o modelo é o seguinte:

$$\frac{dS}{dt} = (1 - S)\mu - \frac{\beta IS}{N} \quad (4)$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - (\sigma + \mu)E \quad (5)$$

$$\frac{dI}{dt} = \sigma E - (\gamma + \mu)I \quad (6)$$

$$\frac{dR}{dt} = \gamma I - \mu R \quad (7)$$

Onde μ indica taxa de nascimento da população, neste modelo a taxa de nascimento é considerada igual a de mortalidade.

O σ está relacionado com o período de incubação.

Já o γ está relacionado com o tempo em que o indivíduo permanece doente.

E o parâmetro β indica a taxa de conversão pela qual indivíduos suscetíveis tornam-se expostos.

Ainda, o β é governado por outros dois parâmetros: p que indica a probabilidade um indivíduo infectar o outro caso haja contato entre os dois. E n que é uma média de quantos indivíduos cada um encontra por dia. Portanto:

$$\beta = pn \quad (8)$$

É interessante notar que a probabilidade de infecção pode ser afetada pelo uso de máscaras, higienização correta das mãos, etc. Já o número médio de contatos interpessoais pode ser impactado pelo distanciamento social, ou no caso prejudicial, por aglomerações.

Assim como no modelo SIR, o SEIR também possui a medida R_0 , que no caso é dada por:

$$R_0 = \frac{\sigma}{\sigma + \mu} \frac{\beta}{\gamma + \mu} \quad (9)$$

Todos os parâmetros citados acima, menos os que compõem o β , são fixos. Portanto, para que se tenha noção de diferentes cenários, é preciso explorar diversos valores de p e n . E para fazer previsões acuradas, queremos boas estimativas para eles.

Como queremos $R_0 < 1$ para que a doença seja erradicada, temos $\mu \approx 0$, então precisamos que $\beta < \gamma$.

3 Modelo SEIRD

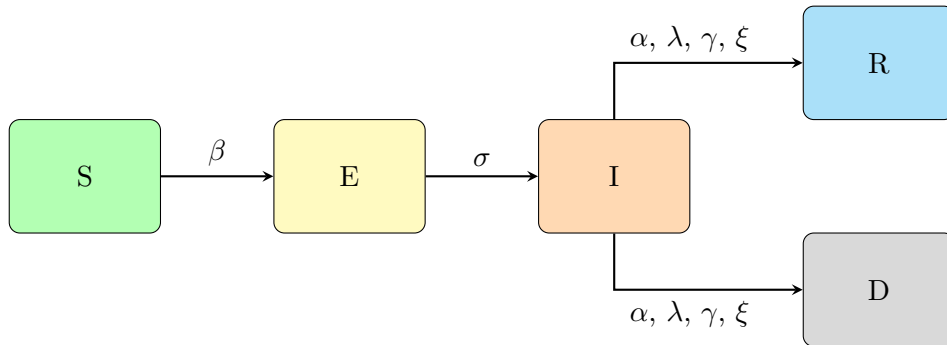


Figura 3 – Modelo SEIRD

Quando uma doença provoca uma quantidade significativa de óbitos, é importante que estes também sejam avaliados e incorporados ao modelo.

Fizemos então uma extensão do SEIR, com algumas mudanças. Não consideramos as taxas de mortalidade geral e natatalidade. Adicionamos a variável D , que indica o número de mortos, e adicionamos outros três parâmetros: α que indica a taxa de internação dos infectados, λ que indica o tempo de permanência médio de um indivíduo no hospital, e ξ que indica a taxa de mortalidade de pacientes hospitalizados.

Dessa forma, temos as seguintes equações:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \quad (10)$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - \sigma E \quad (11)$$

$$\frac{dI}{dt} = \sigma E - (1 - \alpha)\gamma I - \alpha\lambda I \quad (12)$$

$$\frac{dR}{dt} = (1 - \alpha)\gamma I + \alpha(1 - \xi)\lambda I \quad (13)$$

$$\frac{dD}{dt} = \alpha\xi\lambda I \quad (14)$$

Ainda,

$$N = S + E + I + R \quad (15)$$

e

$$\frac{dN}{dt} = 0 \quad (16)$$

Portanto, a medida R_0 aqui é dada por:

$$R_0 = \frac{\beta}{\gamma} \quad (17)$$

Logo, para que a doença seja erradicada, precisamos ainda que $\beta < \gamma$.

4 Obtenção dos parâmetros

Em fevereiro de 2020 a Organização Mundial da Saúde (OMS) enviou uma missão para Wuhan, até então epicentro da doença. Foi então publicado um relatório [5] com diversas informações sobre demografia, período de incubação, etc. Daí podemos extrair que a média do período de incubação (relacionado com σ) está entre 5-6 dias. Já de acordo com [6] essa média é de 5.2 dias.

Se assumirmos que o período de incubação pode ser aproximado por uma distribuição exponencial com média σ , isto é, $T_{inc} \approx \text{Exp}(\sigma)$, então $\sigma = \frac{1}{T_{inc}} \approx 0,192$

A taxa de hospitalização (α) reportada em [5] é de 20%. Consideramos que todos os casos severos e críticos devem ser hospitalizados.

Já de acordo com [7] a taxa de mortalidade de pacientes hospitalizados (ξ) está por volta de 30%.

Segundo [8] os pacientes internados ficam cerca de 5-8 dias nos hospitais. Para obter λ , assim como com o tempo de incubação, supomos que o período de permanência nos hospitais pode ser aproximado por uma distribuição exponencial, ou seja, $T_h \approx \text{Exp}(\lambda)$. Escolhemos $T_h = 7$, portanto $\lambda = \frac{1}{T_h} \approx 0,143$

Para γ seguimos a metodologia [9]. Dessa forma obtivemos que um indivíduo permanece infectante por cerca de 3.2 dias. Portanto novamente assumimos que esse período pode ser aproximado por uma distribuição exponencial tal que $T_i = \text{Exp}(\gamma)$. Logo, $\gamma = \frac{1}{3.2} \approx 0,313$.

Já β foi o parâmetro que variamos em nossas simulações e obtivemos seus diversos valores baseado em valores reportados de R_0 .

<i>Parâmetro</i>	<i>Valor</i>
σ	0,192
α	0,2
ξ	0,3
λ	0,143
γ	0,313

Tabela 1 – Valores obtidos dos parâmetros.

5 SIR vs SEIRD

A princípio a diferença mais visível entre os modelos está no nome. Nosso modelo possui dois compartimentos a mais do que o SIR, que é o modelo que tomamos como base. Temos o **E** referente ao grupo de indivíduos expostos e o **D** referente ao grupo de indivíduos que faleceram devido a doença.

Ao adicionarmos a categoria de expostos, estamos levando em conta que a doença possui um período de incubação, ou seja, existe uma fase de transição entre o estado suscetível (indivíduos saudáveis) e o estado infectado, que é o estado que consideramos que haja propagação efetiva da doença². Portanto, a propagação da doença nesse modelo é mais lenta do que no SIR, como ilustrado nas figuras 4 e 5. Também é possível notar que sob as mesmas condições, o pico da doença no modelo SIR ocorre com uma porcentagem maior da população infectada simultaneamente.

Já a categoria dos mortos nos dá uma noção quantitativa da severidade da doença. Neste compartimento utilizamos parâmetros relacionados a taxa de hospitalização e a mortalidade de pacientes hospitalizados³. Também usamos essa categoria para regular o R_0 do modelo, uma vez que a subnotificação de óbitos é significativamente menor do que a de casos confirmados.

² Poderíamos considerar que indivíduos passando pelo período de incubação também propagam a doença, porém precisaríamos de um novo parâmetro que seria a constante de proporcionalidade referente a taxa de conversão entre indivíduos suscetíveis e expostos.

³ Outra aproximação que fazemos, sem levar em conta eventual colapso do sistema hospitalar.

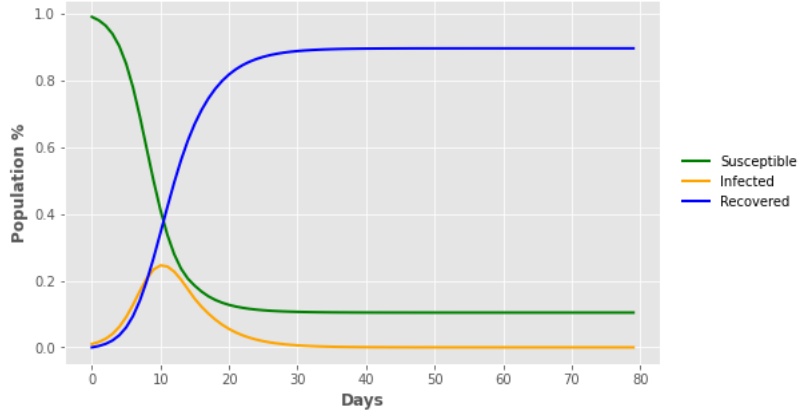


Figura 4 – Simulação do modelo SIR com $p = 5\%$, $n = 16$, $\nu = 0,313$ e $R_0 = 2,56$.

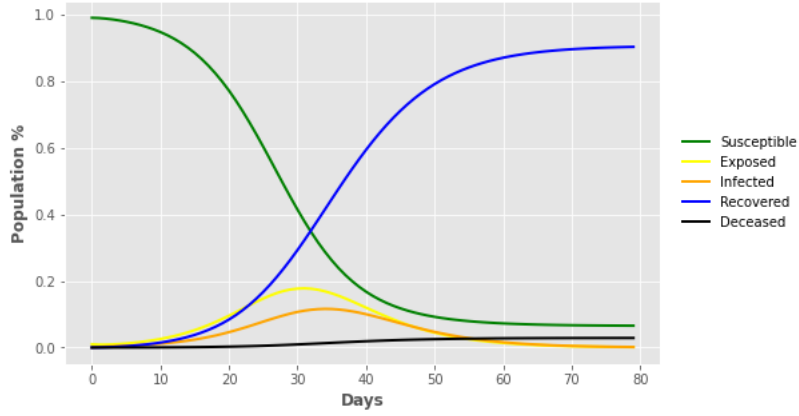


Figura 5 – Simulação do modelo SEIRD com $p = 5\%$, $n = 16$ e $R_0 = 2,56$.

6 Metodologia para análise das medidas sanitárias

As medidas sanitárias afetam diretamente o parâmetro R_0 , ou seja, a capacidade de propagação da doença.

O uso de máscaras e o distanciamento social influenciam o R_0 através do parâmetro p , ou seja, a probabilidade de infecção. De acordo com [10] a Covid-19 pode ser espalhada por meio de gotículas, partículas aéreas e aerossóis emitidos por pessoas tossindo, espirrando e até falando. Esse estudo e outros citados em [11] comprovam a eficácia do uso de máscaras na redução da transmissibilidade da doença.

Segundo [12], o distanciamento social é eficaz, porém a distância necessária varia conforme condições ambientais como intensidade e direção de ventos, umidade relativa do ar e também com o uso de máscaras.

Já o isolamento social impacta o R_0 por intermédio do parâmetro n , ou seja, na média de contatos entre os indivíduos.

Portanto, iremos assumir $\beta = \beta(t) = \rho(t)\eta(t)$, onde $\rho(t) = \delta(t)p$ e $\eta(t) = \tau(t)n$, sendo $\delta(t)$ um parâmetro que indica a deformação do p ao longo do tempo e $\tau(t)$ um que

indica a deforção do n ao longo do tempo.

A estimativa desses parâmetros separadamente é uma tarefa difícil, porém podemos estimar $\beta(t)$ direto e daí calcular R_t dependente do tempo em vez de R_0 . Com isso podemos avaliar as medidas sanitárias já adotadas e também prevermos os casos futuros.

Ainda, para avaliarmos a eficiência de uma política pública como medida sanitária, iremos considerar:

$$\beta(t) = \beta_0 \prod_{i \in \mathbb{N}} \beta_i^{p_i}, p_i \in \{0, 1\}$$

Onde β_0 indica a taxa de infecção natural da doença e β_i a contribuição da i -ésima política a ser avaliada. Com o valor da potência p_i podemos escolher ativar ou não a contribuição de uma medida.

6.1 O Plano São Paulo [13]

No dia 1º de julho de 2020 entrou em vigor no estado de São Paulo o chamado "Plano São Paulo". O governo estadual propôs uma estratégia para a retomada da economia que consiste em dividir o estado em 17 departamentos regionais de saúde e categorizar cada um em uma escala de cinco níveis de abertura.

A classificação desses níveis depende de fatores como a taxa de ocupação de leitos de UTIs, números de novas internações, óbitos, etc.

Essa classificação é feita semanalmente e uma região pode passar de uma escala mais relaxada para uma mais restritiva e vice-versa.

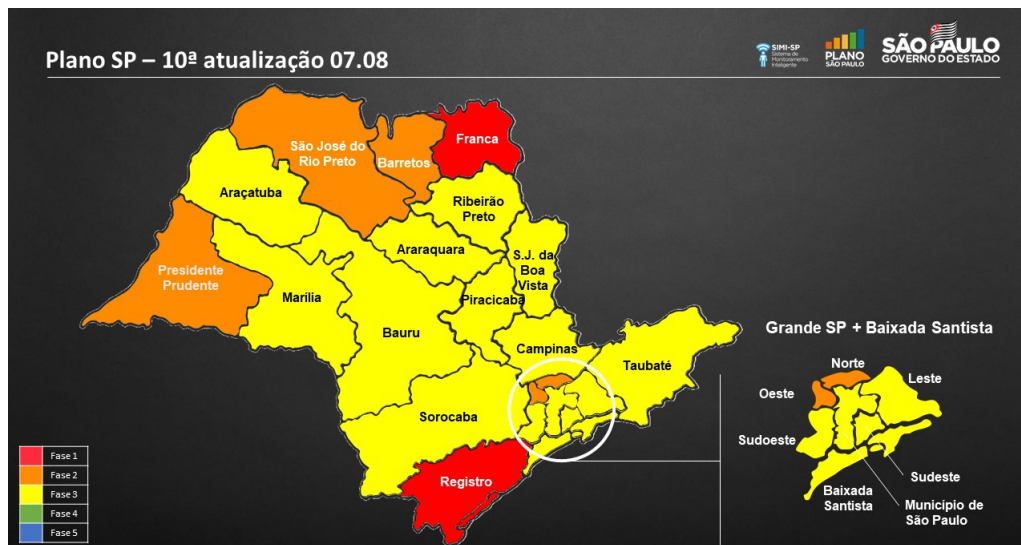


Figura 6 – Mapa da classificação das regiões do Estado de São Paulo conforme o plano São Paulo na semana do dia 07/08. disponível em <https://www.saopaulo.sp.gov.br/planosp/>

Da forma que definimos o β , podemos ter a intuição do impacto de cada decisão. Por exemplo, uma fase que permita o funcionamento dos estabelecimentos com maior capacidade e em mais horas fará com que o n aumente, bem como o β .

Portanto, basta avaliar a intensidade em que o β é, ou foi, perturbado por cada medida.

7 Ajuste do R_0

Todos os dias o Ministério da Saúde divulga os dados da pandemia e disponibiliza em [14] uma planilha contendo estes, tornando fácil o acesso. Nesta planilha é possível checar a situação de cada estado e de cada município.

Os dados são coletados pelas secretarias estaduais de saúde e curados pelo ministério. Essas secretarias trabalham em regime reduzido aos fins de semana, portanto estas datas sofrem com subnotificação que é compensada ao longo da semana. Este é o primeiro problema que deve ser contornado, esta variação afeta o valor diário do β .

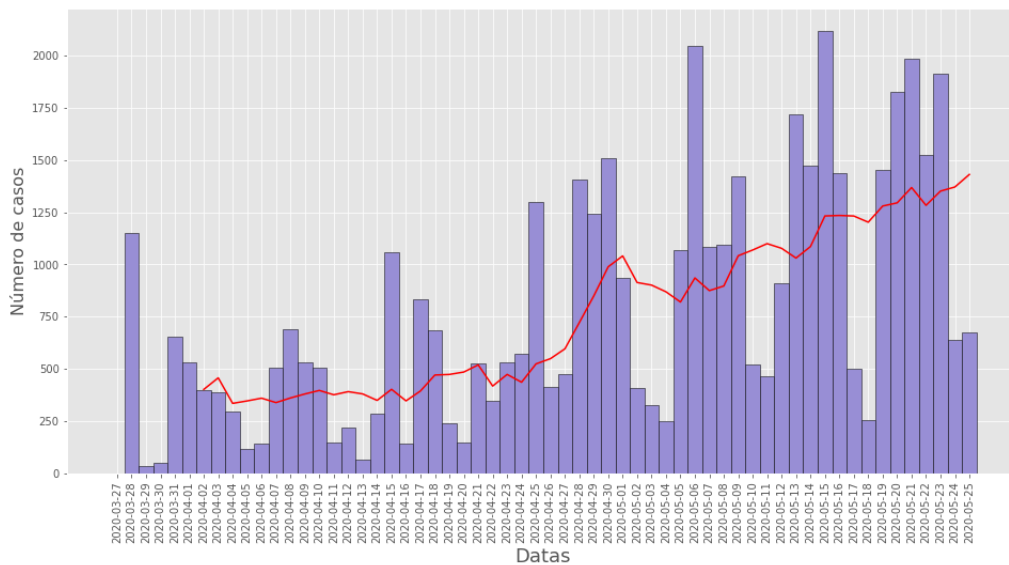


Figura 7 – Número de casos diários na cidade de São Paulo conforme divulgado pelo Ministério da Saúde. Em vermelho a média móvel a cada 7 dias.

O segundo problema que é encontrado está relacionado com a quantidade de testes realizados. A maioria dos indivíduos testados apresentou algum sintoma e procurou assistência médica. Como discutido na seção 4, a taxa de hospitalização é cerca de 20%, portanto muitos casos deixam de ser reportados. Porém todos os óbitos suspeitos são testados, logo é mais interessante regular o modelo de acordo com o número de óbitos, pois estes apresentam menor grau de subnotificação.

O terceiro problema provém da velocidade com que os resultados dos testes são obtidos. As entradas no sistema do Ministério da Saúde correspondem ao dia em que o resultado do exame foi divulgado e não efetivamente ao dia que algum indivíduo supostamente começou a apresentar os sintomas da doença. Portanto há um fator de atraso entre as predições do modelo e a realidade.

Para contornar esses problemas, foi decidido ajustar o modelo ao número de óbitos acumulados.

Para fazer o ajuste consideramos $R_0 = R_0(t)$ como a função logística:

$$R_0(t) = \frac{R_{0s} - R_{0e}}{1 + e^{-k(-t+t_0)}} + R_{0e}$$

,

onde R_{0s} indica o valor inicial do R_0 , R_{0e} seu valor final, k e t_0 parâmetros a serem ajustados.

Daí o ajuste foi feito usando o método dos mínimos quadrados com as condições iniciais do modelo indicadas na tabela 2. Os parâmetros obtidos encontram-se na tabela 3.

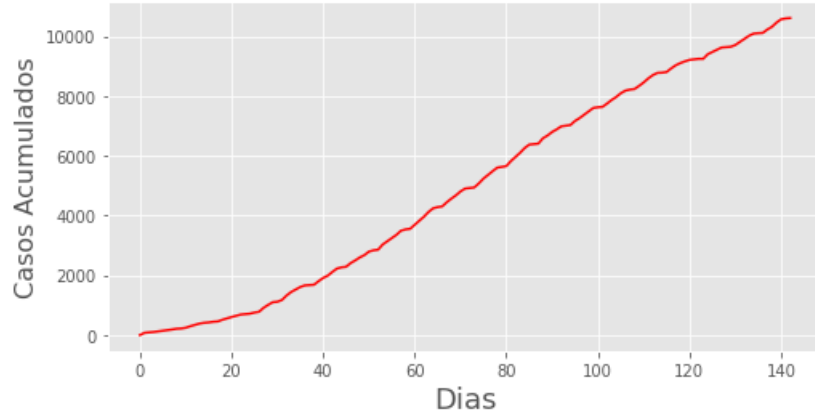


Figura 8 – Óbitos acumulados na cidade de São Paulo conforme divulgado pelo Ministério da Saúde.

<i>Compartimento</i>	<i>Valor</i>
N	$12e6$
S	$N-1$
E	1
I	0
R	0
D	0

Tabela 2 – Condições iniciais do modelo, $N = 12e6$ indica a população do município de São Paulo.

<i>Parâmetro</i>	<i>Melhor Valor</i>
R_{0s}	$6,0$
k	$5,0$
t_0	$29,7$
R_{0e}	$0,85$

Tabela 3 – Melhores valores obtidos para os parâmetros com as condições iniciais da tabela 2.

<i>Data</i>	<i>Dado Real</i>	<i>Dado do Modelo</i>
12/08/2020	10311	10295
13/08/2020	10453	10353
14/08/2020	10563	10412
15/08/2020	10592	10469

Tabela 4 – Dados Reais vs Dados fornecidos pelo modelo de acordo com os parâmetros da tabela 3 e condições iniciais da tabela 2.

Data	Óbitos Previstos
17/08/2020	10583
18/08/2020	10640
19/08/2020	10695
20/08/2020	10750
21/08/2020	10805

Tabela 5 – Número de óbitos acumulados, no município de São Paulo, previstos pelo modelo para cinco dias além dos dados de treino.

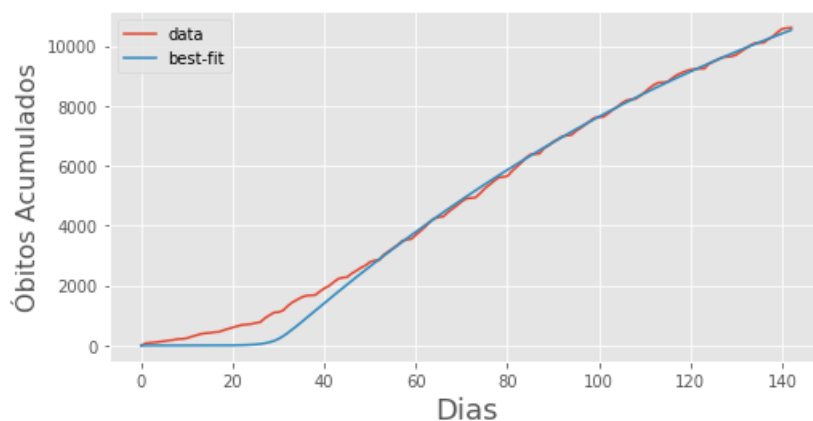


Figura 9 – Modelo ajustado. Em vermelho o número de óbitos acumulados no município de São Paulo, em azul o número previsto pelo modelo.

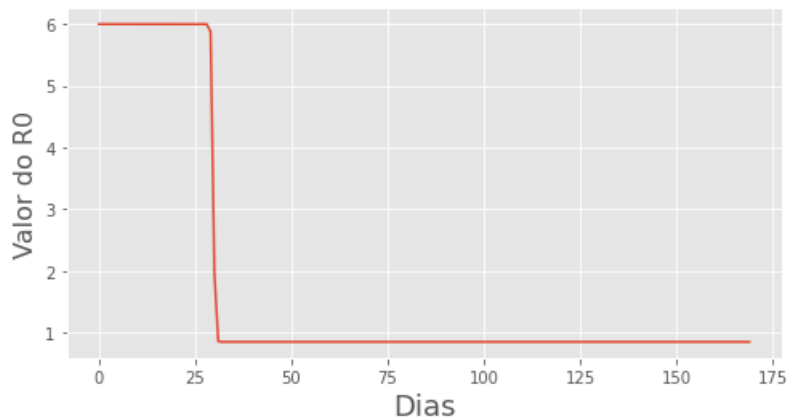


Figura 10 – Valor de $R_0(t)$ ao longo do tempo.

8 Discussão

Primeiramente, o valor do R_0 como observado na figura 10 começa muito alto e sofre uma queda brusca. Portanto, é notável que ele não reflete o valor real. É possível que ele tenha apresentado esse comportamento por causa das condições iniciais.

Por exemplo, foram considerados como suscetíveis todos os moradores do município de São Paulo, e fora introduzido apenas um indivíduo infectado. Na realidade não é assim, como a cidade é uma metrópole, pessoas de diversas cidades circulam por lá diariamente. Além disso, não houve controle da entrada de pessoas no país, portanto um único vôo provindo de algum epicentro pode ter trazido mais de um indivíduo infectado para a cidade. Logo, a divergência nos primeiros dias que pode ser observada na figura 9.

Por outro lado, o modelo se baseia em condições de teste perfeitas, *i.e.*, com resultados instantâneos e todos os indivíduos infectados sendo identificados. Como mencionado na seção 6, isto não é possível. Portanto, vai haver de fato certo atraso por parte das previsões do modelo.

Outro fator que pode afetar a capacidade de generalização do modelo está relacionado com as heurísticas que foram usadas para estimar os parâmetros que deixamos fixos. As distribuições usadas, os tempos médios escolhidos, etc. Por exemplo, conforme a doença vai progredindo diversos tratamentos são testados, estes afetam a letalidade da doença que mantivemos fixa.

9 Conclusão

O modelo é capaz de generalizar bem a progressão da doença, com isso é possível gerar diferentes cenários e analisar a efetividade de medidas a serem tomadas.

Para que a progressão da doença seja acompanhada é necessário testar a população com frequência.

Usar máscaras, manter o isolamento e o distanciamento é essencial.

Referências

1. Organização Mundial da Saúde <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>>
2. Ministério da Saúde do Brasil <<https://coronavirus.saude.gov.br/>>
3. BNO News - Rastreado o Coronavírus <<https://bnonews.com/index.php/2020/04/the-latest-coronavirus-cases/>>
4. Kampf, Günter, et al. "Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents." *Journal of Hospital Infection* (2020).
5. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) <<https://www.who.int/Docs/Default-Source/Coronaviruse/Who-China-Joint-Mission-on-Covid-19-Final-Report.pdf>>
6. Li, Qun, et al. "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia." *New England Journal of Medicine* (2020).
7. Docherty, Annemarie B., et al. "Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol." *medRxiv* (2020).
8. Rees, Eleanor M., et al. "COVID-19 length of hospital stay: a systematic review and data synthesis." (2020).
9. Wu, Joseph T., Kathy Leung, and Gabriel M. Leung. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study." *The Lancet* 395.10225 (2020): 689-697.
10. Aydin, Onur, Md Abul Bashar Emon, and M. Taher A. Saif. "Performance of fabrics for home-made masks against spread of respiratory infection through droplets: a quantitative mechanistic study." *medRxiv* (2020).
11. Centers for Disease Control and Prevention <<https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html#recent-studies>>
12. Feng, Yu, et al. "Influence of wind and relative humidity on the social distancing effectiveness to prevent COVID-19 airborne transmission: A numerical study." *Journal of aerosol science* (2020): 105585.
13. Plano SP (governo do estado) <<https://www.saopaulo.sp.gov.br/planosp/>>
14. Painel Coronavírus <<https://covid.saude.gov.br/>>