Modelagem Dinâmica Aplicada à COVID-19

Sebastião Cícero Pinheiro Gomes

Carlos Rodrigues Rocha

Igor Oliveira Monteiro

Prefácio

A pandemia da COVID-19 surpreendeu o planeta em razão da rapidez do seu espalhamento. Com foco inicial em Hubei (China), identificado em dezembro de 2019, precisou apenas de dois meses para se tornar uma pandemia. O vírus SAR-CoV-2 (também conhecido como o novo coronavírus) já matou milhares de pessoas no planeta e, no Brasil, no dia em que escrevemos esse prefácio, estamos infelizmente contabilizando oitenta mil óbitos.

Movidos pelo desejo de ajudar a combater essa pandemia, começamos avidamente a estudá-la, principalmente a partir de fevereiro de 2020. Muitos artigos científicos, teses e dissertações foram estudados, sempre com o foco na COVID-19. Diversos modelos dinâmicos foram implementados e testados, principalmente a partir da confrontação entre simulações e dados reais de cidades do Brasil e do exterior. Uma nítida dicotomia foi se formando à medida que o estudo avançava: ampliar a complexidade estrutural do modelo pode ampliar os erros de origem paramétrica (modelos com estruturas complexas possuem mais parâmetros); diminuir a complexidade do modelo pode ampliar os erros de origem estrutural. Erros estruturais e paramétricos são inerentes aos modelos dinâmicos, sem distinção, aplicados a qualquer ramo da ciência. No caso da COVID-19, os dados reais disponíveis não são muito confiáveis em razão da subnotificação e também de retardos no tempo, correlatos a problemas na detecção/identificação de novos casos. Portanto, nesta pandemia, a identificação paramétrica passou a ser um problema crítico. Este fato foi preponderante na opção que fizemos por um modelo de estrutura simples, com menos parâmetros a estimar. Porém, houve a necessidade de realizarmos adaptações neste modelo de estrutura simples, a fim de torná-lo mais preditivo e realista, com respostas que pudessem reproduzir a evolução epidêmica na cidade ou região em estudo. Desta forma, o compromisso assumido nesta opção privilegiou um meio termo entre simplicidade e complexidade estrutural. Surgiu assim a proposta de modelo que denominamos MdSIR, significando modelo SIR modificado. A estrutura SIR (Suscetíveis, Infectados, Recuperados) é a mais simples estrutura dentre os chamados modelos compartimentais. O MdSIR tem complexidade estrutural um pouco maior comparado ao SIR e os seus parâmetros variam no tempo. Um dos capítulos do presente livro é dedicado a explicar as variações paramétricas e a sugerir como estes parâmetros podem ser identificados ao longo do tempo.

Um breve resumo histórico sobre as pandemias que assolaram a humanidade é feito no primeiro capítulo. Neste momento de angústia à espera de uma vacina, quando o isolamento social tem se tornado muito difícil e mesmo impossível de se realizar para algumas pessoas, saber que a Peste Negra durou duzentos anos e matou milhões de

pessoas na idade média nos dá uma ideia sobre o sofrimento que representou essa que foi a maior das tragédias em termos de pandemia na história da humanidade. Precisamos esperar mais alguns meses por uma vacina, mas nossos antepassados sapiens esperaram dois séculos para terem algum alívio. Se assumirmos que o tempo de uma geração seja em média equivalente a vinte e cinco anos, eles conviveram com a grande tragédia por oito gerações consecutivas. Talvez a simples lembrança desse fato histórico ajude na conscientização e traga um pouco de tranquilidade, ancorada na certeza de que tudo isso passará brevemente e fará parte da história de luta da humanidade para sobreviver na Terra.

Este livro resume os principais resultados desse nosso estudo recente sobre a pandemia da COVID-19. Tentamos ser didáticos ao gerar um texto repleto de resultados e ideias expostos de forma gráfica. Esperamos que leitores oriundos de diversas áreas do conhecimento científico possam ler e entender o seu conteúdo, muito embora algum conhecimento prévio sobre equações diferenciais seja importante para a sua compreensão. O último capítulo contém explicações sobre o aplicativo Simcovid 2.0, desenvolvido justamente para facilitar a utilização prática do conteúdo teórico do presente livro. Mesmo o usuário não especialista em matemática ou computação poderá realizar suas simulações e ainda, gerar e exportar os resultados numéricos e gráficos.

Finalmente, entendemos que este livro pretende ser uma pequena contribuição que vem a se somar a tantas outras de colegas do mundo acadêmico, nessa luta que é de todos nós.

Rio Grande, RS, 20/07/2020.

Os autores.

Sobre os Autores

Sebastião Cícero Pinheiro Gomes

Doutorado em Robótica e Automação, obtido no Institut Supérieur de l'Aéronautique et de l'Espace (ISAE – SUPAERO), Toulouse, França, 1992.

Professor do Instituto de Matemática, Estatística e Física (IMEF), da Universidade Federal do Rio Grande (FURG), Rio Grande, RS, Brasil.

Currículo Lattes: http://lattes.cnpq.br/5895086145314795

Carlos Rodrigues Rocha

Doutorado em Engenharia Mecânica, obtido na Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil, 2012.

Professor do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS), Rio Grande, RS, Brasil.

Currículo Lattes: http://lattes.cnpq.br/8918020972394590

Igor Oliveira Monteiro

Doutorado em Matemática Aplicada, obtido na Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil, 2015.

Professor do Instituto de Matemática, Estatística e Física (IMEF), da Universidade Federal do Rio Grande (FURG), Rio Grande, RS, Brasil.

Currículo Lattes: http://lattes.cnpq.br/3884751552646684

Capítulo I

Introdução

1.1 Histórico

A história da humanidade está repleta de epidemias que dizimaram muitas vidas ao longo do tempo. A peste Antonina atingiu a cidade de Roma em 166 d.C. e há relatos indicando que esta epidemia, no seu auge, matava duas mil pessoas por dia na cidade. Porém, os problemas com epidemias surgiram desde os primórdios da civilização humana, há dez mil anos, quando se iniciou a chamada revolução agrícola. Estes nossos antepassados passaram de caçadores-coletores em vidas nômades a agricultores-pastores em assentamentos, formando assim as primeiras aglomerações, ainda no período neolítico. Esses primeiros assentamentos vieram a formar as cidades reinos, presentes, por exemplo, na Mesopotâmia, região que fazia parte do que veio no século XX a ser batizado como Crescente Fértil. A vida humana organizada em cidades cobrou o seu preço, facilitando o surgimento de doenças que se tornaram epidemias.

Certamente, nos primórdios da nossa espécie sobre o planeta Terra, os deslocamentos entre regiões mais afastadas impediam que epidemias se tornassem pandemias. Em Romero (2019) há uma importante narrativa sobre as doenças na humanidade, as quais sempre estiveram correlacionadas aos eventos políticos que transformaram e ditaram o rumo da nossa história. Neste livro há uma vasta revisão bibliográfica sobre o tema que aborda e, em uma das passagens, há o seguinte texto:

> “Ao nos transformarmos de caçadores-coletores em agricultores-pastores, alteramos o ambiente natural, trouxemos uma série de animais – aves, cavalo, porco, ovelha, cabra, gado, camelo, cão, rato – para o nosso convívio doméstico e, principalmente, aumentamos nossa densidade populacional. Por fim, por meio do comércio e da guerra, estabelecemos mecanismos de comunicação entre os diferentes assentamentos humanos. Com isso, criamos oportunidades para uma miríade de microparasitas – vírus, bactérias, protozoários, vermes – prosperarem entre nós (McNeil, 1976, p.69).”

Conforme explicado em Romero (2019) e em McNeill (1976), o homem adentrou o ciclo de doenças que eram restritas aos animais silvestres e que em alguns casos foram transferidas para animais de criação, doenças estas que também atingiram os humanos. Essas transferências de parasitas entre hospedeiros são frequentes até os dias atuais, mas com o agravamento de convivermos com maiores concentrações humanas nas grandes cidades e mais facilidades de deslocamentos entre regiões do planeta. Portanto, todo esse contexto faz com que, de tempos em tempos, tenhamos que encarar uma pandemia.

Algumas doenças eram comuns nas cidades reinos do Crescente Fértil e podiam migrar entre cidades, afetando principalmente as que ficavam às margens de rios. Um mesmo rio comunicava diversas cidades e servia também de transporte para estas doenças. As principais doenças que acometiam as pessoas no terceiro milénio a.C. eram as de transmissão aérea (sarampo, varíola e gripe) e as de transmissão hídrica (disenteria e febre tifoide). Porém, antes de meados do primeiro milênio a.C. não há indícios de que tenha havido uma pandemia, principalmente por razões vinculadas às dificuldades de locomoção entre as diversas regiões povoadas de então.

Segundo Romero (2019), uma das primeiras pandemias que se tem notícias ocorreu entre os anos de 436 e 426 a.C. Teve o seu início provável em regiões próximas à Etiópia e se propagou pelo norte da África, passando por Egito e Líbia, chegando a atingir em cheio a Grécia antiga. Ficou conhecida como a Peste de Atenas, caracterizada por grande letalidade e sintomas cruéis. Porém, não há consenso sobre a sua etiologia. Alguns pesquisadores já imputaram a causa a algumas doenças, tais como varíola, peste bubônica, antraz, sarampo e febre tifoide. Estava assim oficialmente inaugurada a temporada de pandemias na história da humanidade nesse planetinha por enquanto azul.

Voltemos então à grande pandemia da história antiga: a Peste dos Antoninos, que em meados do século II d.C. atingiu todo o império romano. Teve origem no oriente e as candidatas mais prováveis foram identificadas como varíola ou sarampo. Foi trazida à península itálica por soldados do exército romano que voltavam de campanhas na Pérsia. Inicialmente atingiu as regiões do Danúbio e do Reno, tendo chegado a Roma em 166 d.C. Esta pandemia se espalhou por todos os cantos do império romano e depois por toda a Europa, tendo durado ao todo quatorze anos. Estima-se que tenha matado entre um quarto e um terço da população da Itália e, em toda a Europa, entre quatro e sete milhões de pessoas, com uma mortalidade estimada entre 25% e 30% (Romero, 2019).

Outra importante pandemia da história antiga ficou conhecida como a Peste de Justiniano, nome do imperador romano da época. Estamos nos referindo ao ano do seu início, em 540 d.C., quando eclodiu no Egito e espalhou-se em seguida por toda a costa do

Mediterrâneo. Hoje é fato conhecido que se tratou da primeira aparição da peste bubônica, que já chegou com grandes agressividade e letalidade. Das cidades costeiras do Mediterrâneo migrou para o interior do continente europeu e há relatos históricos indicando que foi grande a devastação que causou. Alcançou ainda todo o norte da África e chegou também ao Oriente Médio e à Pérsia. Depois seguiu devastando vidas na Índia, China e Japão. Foi sem dúvida uma pandemia, mas com velocidade de disseminação lenta entre as diversas regiões nas quais chegou, pois esteve ativa até o ano de 750 d.C. (MacNeill, 1975), ou seja, por duzentos e dez anos aniquilou vidas humanas nas regiões afetadas.

A peste bubônica voltou a assolar a Eurásia por diversas vezes ao longo dos séculos. O caso mais grave passou para história com o nome de Peste Negra, cujo pico de contaminação na Europa aconteceu por volta do ano de 1351. Acredita-se que tenha sido a mais letal de todas as pandemias da história humana. Estima-se que tenham morrido entre 75 e 200 milhões de pessoas em toa a Eurásia. A Figura 1 mostra o avanço da Peste Negra na Europa, na qual podemos perceber a rapidez, para os padrões da época, da disseminação. Atribui-se o seu início como tendo acontecido na Mongólia, China, em 1334, onde morreram cinco milhões de pessoas (Rezende, 2009). Conforme ilustra a Figura 1, sua chegada à Europa ocorreu em 1347, pela Itália e nos portos de Marselha (França). Com taxa de mortalidade alta, a disseminação da bactéria *Yersinia pestis* foi rápida e trouxe consigo o pânico e o caos. Transmitida inicialmente de ratos para humanos a partir de pulgas, posteriormente passou a ser transmitida também por contato com doentes e corpos, que de tantos, não havia como sepultá-los. A falta de compreensão sobre as formas de contágio contribuiu para ampliar o crescimento da pandemia. O relato de um cirurgião francês que sobreviveu à peste é uma preciosidade, que nos auxilia a compreender como a doença acometia as pessoas (Rezende, 2009):

> “A grande mortandade teve início em Avignon em janeiro de 1348. A epidemia se apresentou de duas maneiras. Nos primeiros dois meses manifestava-se com febre e expectoração sanguinolenta e os doentes morriam em três dias; decorrido esse tempo manifestou-se com febre contínua e inchação nas axilas e nas virilhas e os doentes morriam em cinco dias. Era tão contagiosa que se propagava rapidamente de uma pessoa a outra; o pai não ia ver seu filho nem o filho a seu pai; a caridade desaparecera por completo. Não se sabia qual a causa desta grande mortandade. Em alguns lugares pensava-se que os judeus haviam envenenado o mundo e por isso os mataram”.

Ao final do ciclo deste surto pandêmico específico, a doença ainda permaneceu de forma endêmica, durante mais de dois séculos, voltando a atacar de tempos em tempos.

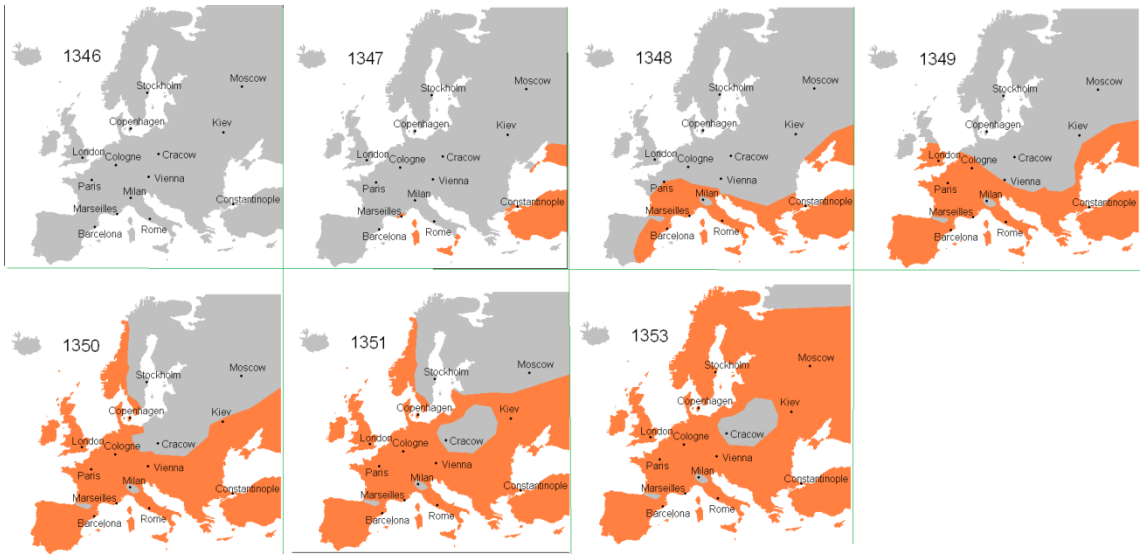


Figura 1. As fases da Peste Negra na Europa, adaptada de Black Death strikes Europe: 1347-1353, extraída do wikipedia e que está em domínio público, de autoria de Andrei Nacu, no link: (https://pt.wikipedia.org/wiki/Peste_Negra#/media/Ficheiro:Blackdeath2.gif).

Mais recentemente a humanidade sofreu com outra grande pandemia que veio na esteira da primeira guerra mundial. Trata-se da gripe espanhola, que teve início em 1918 e cujo agente causador é o vírus H1N1, que viria a causar outra pandemia em 2009. Neste caso de 1918, apesar de já haver muita documentação nesta época, os dados não são muito precisos. Estima-se que a gripe matou entre 20 e 40 milhões de pessoas, mas há pesquisadores que admitem a possibilidade do número de mortos ter chegado a 100 milhões. Todos os mortos na primeira grande guerra somam aproximadamente 15 milhões. Apesar de ter recebido o nome de gripe espanhola, sabe-se que esta pandemia não se originou na Espanha, embora ainda seja controversa a sua origem. Há estudos que apontam o seu surgimento no Kansas, USA, assim como outros estudos asseguram o seu surgimento em meio a tropas britânicas sediadas na França, ainda em 1917. Estima-se que 30% da população do planeta tenha sido infectada. A Figura 2 é histórica e mostra curvas de mortes por todas as causas, normalizadas por 100 mil habitantes, referentes a Nova York, Londres, Paris e Berlim. Esta figura mostra que os picos epidêmicos aproximadamente coincidem no tempo, ou seja, os instantes de pico nestas quatro cidades foram próximos ao dia 25/10/1918. No Brasil, aconteceram aproximadamente 35 mil mortes, localizadas principalmente no Rio de Janeiro e em São Paulo. O pico da epidemia no Rio de Janeiro parece ter sido por volta do dia 22/10/1918, quando foram computados 930 óbitos (Goulart, 2005). Portanto, o instante de pico no Rio de Janeiro foi aproximadamente o mesmo

verificado nas cidades da Figura 2. Só no Rio de Janeiro, dois terços da população contraiu a virose e aproximadamente 15 mil morreram. Em Romero (2019) há uma descrição detalhada dessa pandemia e da sua correlação com eventos políticos e econômicos.

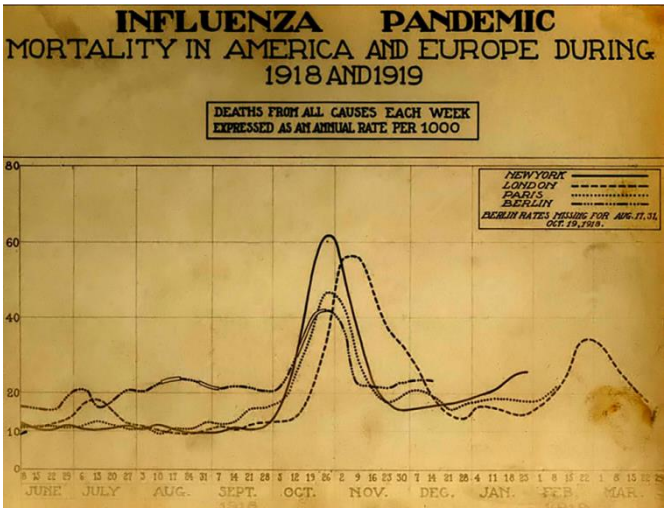


Figura 2. Gráficos mostrando o número de mortes (todas as causas) por 1000 habitantes (fonte: NICHOLLS (2006), imagem do acervo do National Museum of Health and Medicine, Armed Forces Institute of Pathology, Washington, D.C., United States).

1.2 A COVID-19

Depois desse mergulho histórico sobre pandemias que assolaram a humanidade, voltemos à realidade dos dias atuais. Estamos em meio a uma grande pandemia, fruto de uma virose que ficou conhecida como COVID-19. O vírus causador pertence à família SARS, mais precisamente o SARS-CoV-2. Os sintomas no infectado são semelhantes ao da gripe comum, mas com uma maior taxa de mortalidade. Há muita subnotificação, até porque muitos infectados não desenvolvem sintomas. Esta subnotificação mascara a taxa de mortalidade. Países que testam mais a população para a detecção da virose apresentam menores taxas de mortalidade, fato já esperado. Na Figura 3 há um gráfico em barras que mostra a evolução do número de óbitos no Brasil até o dia 05/07/2020. Chegamos a 64900 óbitos e com uma curva ainda em crescimento. Já estamos próximos a termos o dobro das mortes ocorridas na gripe espanhola de 1918. Portanto, para o Brasil, essa já é a maior das tragédias em termos de epidemias, de acordo com os números absolutos. No dia 03/07/2020 o Brasil tinha 1543341 casos confirmados, com 63254 óbitos, indicando uma taxa de mortalidade equivalente 4.1%. Porém, em razão da subnotificação, essa taxa de mortalidade deve ser menor. O portal stastista.com estima a taxa de mortalidade da COVID-

19 em 2.2%. Como a taxa no Brasil está quase o dobro dessa média mundial, é provável que tenhamos um alto índice de subnotificação. Esta taxa de 2.2% de mortalidade da COVID-19 é alta comparada ao H1N1 de 2009 (0.02%), mas é bem menor do que a mortalidade causada por um vírus da mesma família, o SARS em 2002 (9.6%) e, principalmente, muito menor do que a mortalidade com a gripe aviária de 2013 (39.3%), esta última causada pelo vírus H7N9. Felizmente, o risco de transmissão sustentada entre pessoas foi baixo, no caso da epidemia por H7N9.

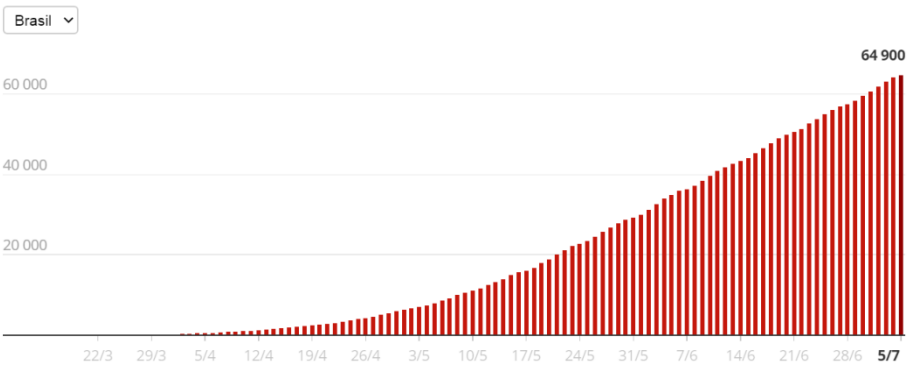


Figura 3. Número de óbitos no Brasil até o dia 05/07/2020 (fonte: portal G1).

Os primeiros casos de COVID-19 surgiram na cidade de Wuhan, capital da província de Hubei (China), no mês de Dezembro de 2019. Em fevereiro de 2020 já era uma pandemia, com impressionante velocidade de contágio. O governo Chinês resolveu adotar uma estratégia de *lockdown* e a epidemia em Hubei foi interrompida com sucesso, conforme pode ser observado nas Figuras 4a (número acumulado de casos) e 4b (número de óbitos).

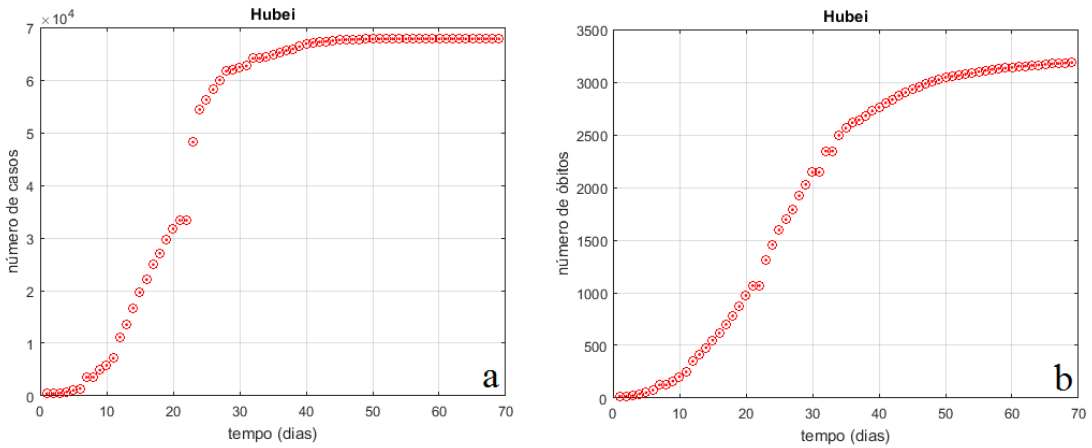


Figura 4. Número acumulado de casos e número de óbitos em Hubei (China), com dados coletados entre os dias 22/01/2020 e 30/03/2020.

Em todo o planeta, até o dia 06/07/2020, 11563004 pessoas já foram infectadas com o novo coronavírus e desses, 536843 foram a óbito, dados do portal statista.com. Considerando que a taxa de mortalidade de 2.2% do SAR-COV-2 seja real e ainda, que o número de óbitos seja próximo do real, isto significaria que o número de casos em todo o planeta, em 06/07/2020, deveria ser próximo a 2440195, ou seja, haveria no mínimo duas vezes mais casos do que o número de notificados, indicando uma subnotificação superior a 100%. Uma característica dessa pandemia é que uma significativa parcela dos infectados não desenvolve sintomas e, portanto, estes não são identificados na maioria dos casos, principalmente nos países que estão efetuando poucos testes na população em geral. Esta característica talvez seja a responsável por uma parte desta subnotificação e só uma estratégia de aplicação em massa de testes poderia diminuir esse problema que é, sem dúvida, significativo, principalmente porque há relatos indicando que os assintomáticos também podem transmitir o vírus.

1.3 Comentários Finais

Com mais de meio milhão de óbitos nesse início de julho de 2020, a COVID-19 ainda está longe de ser uma das mais mortais pandemias da história da humanidade. A Tabela I contém comparações entre as principais pandemias que flagelaram a humanidade, ao longo de dois milênios e meio. A Peste Negra foi realmente a mais mortal, encontrando a humanidade na escuridão de uma idade média que chegava lentamente ao seu fim. Vale ressaltar que a Peste de Justiniano durou quase 200 anos e ainda, há historiadores que afirmam ela continuou ativa por séculos, embora em muito menor escala e de forma endêmica, até o eclodir da Peste Negra.

Tabela I. Comparações entre as principais pandemias da história da humanidade.

Pandemia	Doença	Ano do início	Número de óbitos
Peste de Atenas	Febre Tifoide	436 a.C.	60 a 70 mil
Peste dos Antoninos	Varíola ou sarampo	166	4 a 7 milhões
Peste de Justiniano	Peste bubônica	540	25 a 50 milhões
Peste Negra	Peste bubônica	1346	75 a 200 milhões
Gripe Espanhola	Influenza H1N1	1918	20 a 40 milhões
COVID-19	Influenza SARS-CoV-2	2019	600 mil, até Jul/2020

O Homo Sapiens é muito novo na Terra, sendo a única das espécies do gênero Homo que sobreviveu e comprovadamente se adaptou com êxito às condições do planeta. Nascida com berço em uma região africana onde hoje se encontra a Etiópia, essa espécie surgiu há apenas 200 mil anos (HARARI, 2015). Esse é um intervalo de tempo ínfimo na escala cósmica. Exemplificando, adotando uma escala reduzida no tempo, consideremos que 4 bilhões e 500 milhões de anos (quando o planeta Terra surgiu) correspondem a um ano na escala cósmica, ou seja, nesta nova escala, o planeta surgiu na primeira hora do dia 01/janeiro. Nesta nova escala, o Homo Sapiens teria surgido no dia 31/dezembro, às 23 horas, 36 minutos e 38 segundos, ou seja, surgimos no planeta há apenas 23 minutos e 22 segundos. Caso consideremos o tempo médio de uma geração equivalente a 25 anos, o Homo Sapiens surgiu há apenas 8000 gerações. A primeira grande pandemia (Peste de Atenas) surgiu há insignificantes 100 gerações. Em síntese, somos muito novos na Terra, de forma que ainda desconhecemos muito sobre a natureza em que vivemos, suas leis e seus segredos, a ponto de uma gripe diferente conseguir matar meio milhão de pessoas em cinco meses, sob os olhares perplexos dos cientistas do século XXI.

Por enquanto, a única forma eficiente de controle da COVID-19 tem sido o isolamento social, praticado tanto no Brasil quanto no resto do planeta. Há outras ações de prevenção que também auxiliam a conter o ímpeto da infecção viral, como são os casos do uso obrigatório de máscaras e do distanciamento social em lugares públicos. Porém, a mais eficiente forma de controle epidêmico tem sido de fato o isolamento social que, no Brasil, tem sido praticado parcialmente. Segundo dados do site inloco (https://mapabrasileirodacovid.inloco.com.br/pt/), no dia 04/07/2020 havia 41% da população brasileira em isolamento social. Conforme será visto nos próximos capítulos, este isolamento social parcial ajuda, pois diminui a velocidade da contaminação, desloca o instante de pico para a direita e diminui a amplitude de infectados neste instante de pico. Porém, não resolve o problema, a infecção continua a se expandir entre os suscetíveis a contraí-la, restando-nos torcer para que a comunidade científica consiga desenvolver rapidamente uma vacina que seja eficaz no controle epidêmico.

Referências

GOULART, A. da C. Revisitando a espanhola: a gripe pandêmica de 1918 no Rio de Janeiro. História, Ciências, Saúde – Manguinhos, v. 12, n. 1, p. 101-42, jan.-abr. 2005.

HARARI, Y. N. Sapiens - Uma breve história da humanidade. L&PM POCKET, vol. 1288, 2018, do original: Sapiens, a brief history of humankind. Harper, 2015.

McNEILL, W. H. Plagues and peoples. Nova York, Anchor Press, 1976.

NICHOLLS H. Pandemic Influenza: The Inside Story. PLoS Biol 4(2): e50, 2006. https://doi.org/10.1371/journal.pbio.0040050.

REZENDE, J. M. À sombra do plátano: crônicas de história da medicina [online]. São Paulo: Editora Unifesp, 2009. As grandes epidemias da história. pp. 73-82.

ROMERO L. C. P. Saúde & Política, a doença como protagonista da história. Outubro edições, 2019.

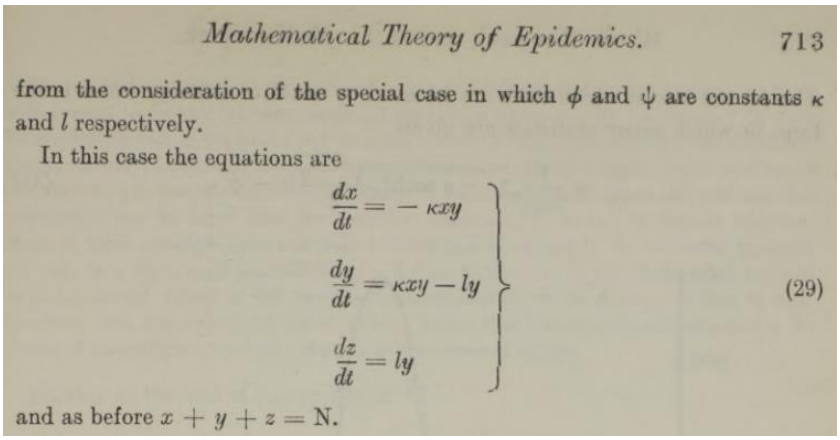
Capítulo II

Modelos Epidemiológicos

2.1 Introdução

Conforme visto no capítulo anterior, a humanidade tem sofrido com muitas pandemias ao longo da história. A moderna facilidade em termos de viagens intercontinentais facilita a que epidemias se tornem pandemias, como foi o caso da gripe espanhola em 1918. Evidentemente, maiores são as facilidades para essas viagens nos dias atuais, de forma que a COVID-19 surgiu na China no final de 2019 e em março de 2020 já era uma pandemia. Apesar de muito recente, já há um vasto material bibliográfico sobre a modelagem da dinâmica de espalhamento do coronavírus ou SARS-CoV-2. Entretanto, antes de discorrer sobre a bibliografia recente que trata da COVID-19, faremos uma breve explanação sobre os dois principais modelos dinâmicos epidemiológicos.

Os cientistas ingleses Kermarck e McKendrick postularam em 1927 o primeiro modelo matemático para representar a dinâmica de uma epidemia. Este artigo histórico, publicado em um congresso da *Royal Society A, Mathematical, Physical and Engineering Science*, deu início a toda uma área de pesquisa, hoje conhecida como modelos matemáticos para a dinâmica de crescimento populacional. Na Figura 4 podemos ver um extrato do artigo no qual são postuladas as equações diferenciais do modelo por eles proposto. Esta proposição recebeu o nome de modelo SIR, uma alusão aos compartimentos Suscetíveis, Infectados e Recuperados. Kermarck e McKendrick inauguraram assim a era dos chamados modelos compartimentais. Neste caso, o compartimento dos Suscetíveis (*x*, na Figura 4) perde indivíduos para o compartimento dos Infectados (*y*, na Figura 4), que por sua vez perde indivíduos para o compartimento dos Recuperados (*z*, na Figura 4). As duas taxas de conversão de indivíduos entre os compartimentos foram denominadas k e l.



Mathematical Theory of Epidemics. 713

from the consideration of the special case in which ϕ and ψ are constants κ and l respectively.

In this case the equations are

$$\left.\begin{aligned} \frac{dx}{dt} &= -\kappa xy \\ \frac{dy}{dt} &= \kappa xy - ly \\ \frac{dz}{dt} &= ly \end{aligned}\right\} \qquad (29)$$

and as before $x + y + z = \mathrm{N}$.

Figura 4. Um trecho do artigo proposto por Kermarck e McKendrick, em 1927.

2.2 O Modelo SIR

Conforme citado anteriormente, o modelo SIR foi idealizado originalmente por Kermarck e McKendrick, em 1927. A Figura 5 contém o diagrama de blocos com os compartimentos do modelo.

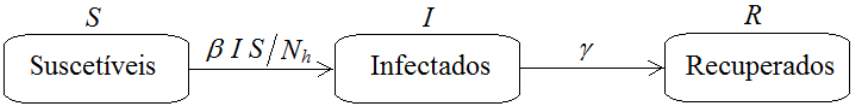


Figura 5 - Diagrama de blocos representando o modelo SIR.

Neste modelo, S é número de habitantes suscetíveis que podem entrar em contato com o vírus. I é o número de infectados e R o número de recuperados (ou removidos). N_h corresponde ao número de habitantes e, no presente artigo, considera-se que $N_h = S + I + R$, de forma que não estão sendo considerados nascimentos ou mortes de causas naturais modificando a quantidade de suscetíveis S. Esta estrutura simples de modelo necessita que sejam identificados apenas dois parâmetros: β (taxa de crescimento) e γ (taxa de recuperação). As seguintes equações diferenciais descrevem este modelo SIR original:

$$\left.\begin{aligned} \dot{S} &= \frac{-\beta SI}{N_h} \\ \dot{I} &= \frac{\beta SI}{N_h} - \gamma I \\ \dot{R} &= \gamma I \end{aligned}\right\} \qquad (1)$$

O leitor pode observar que estas equações diferenciais são as mesmas mostradas no artigo original (Figura 4). A diferença é que, no artigo original, a variável *x* relativa aos Suscetíveis está normalizada, ou seja, dividida pelo número de habitantes.

O parâmetro γ é equivalente ao inverso do tempo de infecção, ou seja, $\gamma = 1/T_{inf}$. Entende-se o T_{inf} como o intervalo de tempo entre um indivíduo contrair a infecção e infectar outro indivíduo. Este parâmetro também é chamado de intervalo de geração (GANYANI *et al*. 2020) e será abordado em maiores detalhes no Capítulo IV. Um parâmetro importante em qualquer epidemia é o que se convencionou chamar de índice de reprodução basal, definido como sendo a taxa de crescimento vezes o tempo de infecção: $R_0 = \beta T_{inf} = \beta/\gamma$. Para $S = S_0 = N_h$ e $I = I_0$ no tempo inicial t_0, demonstra-se facilmente que $I = e^{(\beta-\gamma)t} I_0$ no início da epidemia. Portanto, se $\beta > \gamma \rightarrow R_0 > 1$, haverá crescimento exponencial da epidemia no início do processo contagioso. Se $R_0 < 1$, haverá atenuação também exponencial, até acabar a epidemia. Portanto, o R_0 é um parâmetro fundamental para o controle da epidemia.

Imaginemos inicialmente uma cidade com 100 mil habitantes $(N_h = 100000)$. Vamos considerar ainda que $R_0 = 2.25$ e que $T_{inf} = 5.5\, dias$, valores identificados para Hubei, no início da epidemia naquela província chinesa. Neste caso, a taxa de recuperação será $\gamma = 1/5.5 = 0.1818$. A taxa de crescimento será $\beta = R_0\gamma = 0.4091$. Estes parâmetros têm a mesma unidade física, equivalente a $(1/dias)$. A Figura 6 contém os resultados de uma simulação com o modelo dinâmico SIR (equação (1)), considerando os parâmetros estipulados acima e ainda, um vetor condição inicial equivalente a: $[S_0;\ \ I_0;\ \ R_0] = [N_h - 1;\ \ 1;\ \ 0]$. O ";" separa linhas, de forma que trabalhamos com vetores coluna. Nas simulações deste livro, sempre que a condição inicial for omitida é porque foi esta a condição inicial adotada. Na simulação da Figura 6, bem como nas demais simulações apresentadas neste livro, utilizamos a rotina ODE45 do MATLAB para a integração numérica das equações diferenciais do modelo dinâmico. Os dados reais de cidades que serão utilizados nos próximos capítulos são disponibilizados com atualizações diárias, de forma que utilizamos um passo de integração equivalente a $1\, dia$. Entretanto, é importante ressaltar que a rotina de integração ODE45 é chamada n_p (número de pontos do domínio discreto) vezes, ou seja, atualizamos a condição inicial a cada novo passo de integração, sendo que esta rotina trabalha com um ajuste automático do passo interno de integração, agregando assim maior estabilidade numérica.

Algumas considerações são importantes que sejam feitas, correlatas à simulação da Figura 6.

a) O instante de pico seria no tempo $t = 51\ dias$. Neste tempo, o número de infectados atinge o seu valor máximo, equivalente a 19513. Considerando que 5% destes precisarão de cuidados de terapia intensiva (UTI), seriam necessários 976 leitos de UTI, número irreal para uma cidade de 100 mil habitantes. Conforme será visto nos próximos capítulos, o isolamento social tem a função de mudar esse quadro catastrófico de colapso do sistema de saúde, deslocando o instante de pico para a direita e diminuindo a amplitude de infectados no instante de pico.

b) O número de recuperados (R) engloba os indivíduos que se recuperaram clinicamente e os que vieram a óbito.

c) O número acumulado de casos corresponde à soma dos infectados e dos recuperados $(I + R)$ e, no final do ciclo epidêmico, temos $I = 0$. O número acumulado de casos se torna igual ao número de recuperados no final do ciclo que, neste caso, foi $R = 85342$. A conclusão imediata é que, caso nenhuma medida de proteção seja tomada, incluindo a ausência de isolamento social, aproximadamente 85% da população seria contaminada com a virose.

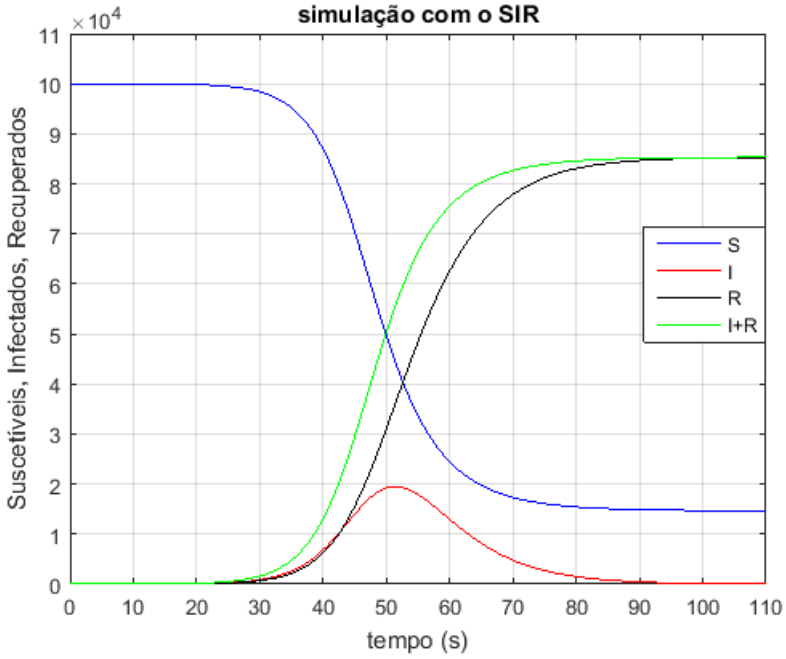


Figura 6. Simulação com o modelo SIR para uma cidade com 100000 habitantes, considerando os seus parâmetros constantes.

Vimos que o R_0 é o principal parâmetro de uma epidemia e assim, passaremos a analisar como modificações nesse parâmetro alteram os resultados das simulações. Na Figura 7 plotamos curvas relativas ao número de infectados (I) considerando três diferentes valores para o índice de reprodução basal: $R_0 = 2.25$; $R_0 = 1.2 * 2.25$; $R_0 = 0.8 * 2.25$. Esta

figura põe em evidência um fator crucial em todas as epidemias: o controle epidêmico consiste sempre na tentativa de diminuir o R_0. Este é um tópico muito explorado a partir do Capítulo III. Veremos que o principal fator que diminui o R_0 é o isolamento social. Quando parte da população adere ao isolamento social, a densidade populacional de suscetíveis a contraírem a virose diminui, diminuindo assim o R_0 e, consequentemente, a velocidade da contaminação. Veremos também que, para que haja um efetivo controle de forma a interromper o processo epidêmico, é necessário que o R_0 se torne menor do que 1.

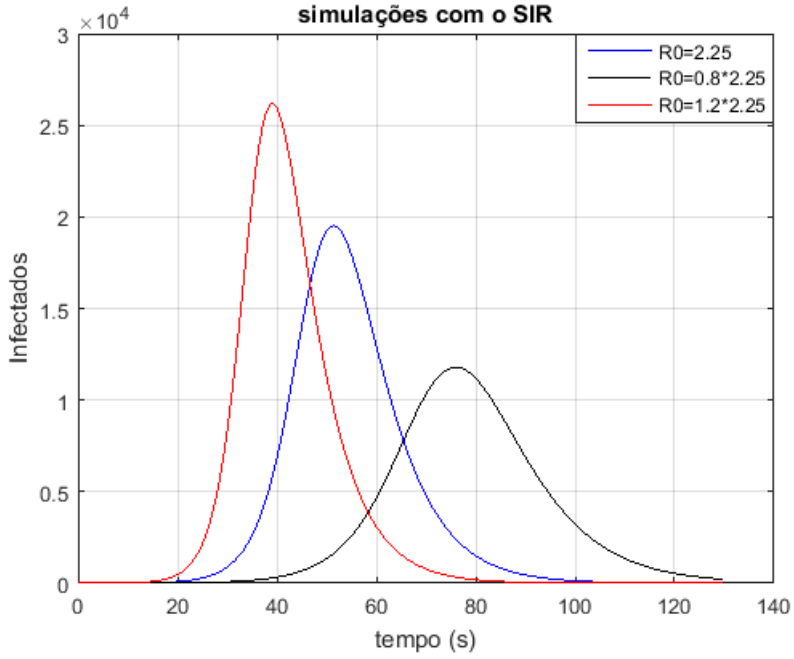


Figura 7. Simulações considerando diferentes valores para o R_0.

2.3 Outros Modelos Derivados do SIR

Dentre os modelos que são variações do SIR, o principal é o que se convencionou chamar de SEIR, cujos compartimentos são constituídos de Suscetíveis, Expostos, Infectados e Removidos. A novidade é o compartimento dos Expostos, formado pelos indivíduos que já tiveram contato com o vírus, mas que não desenvolveram ainda sintomas da infecção porque o vírus ainda está na fase de incubação. Torna-se necessário, portanto, incluir mais uma equação diferencial à dinâmica, bem como mais uma variável de estado para representar os Expostos (E) e mais um parâmetro, correspondente ao tempo de incubação (T_{inc}). Definimos a taxa de conversão de E para I (de Expostos para Infectados) equivalente a $\alpha = (1/T_{inc})$, de forma que o SEIR pode ser representado a partir do diagrama de blocos visto na Figura 8.

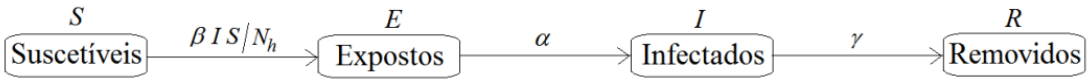


Figura 8. Diagrama de blocos representando o modelo SEIR.

O modelo SEIR pode ser descrito a partir do seguinte sistema de equações diferenciais:

$$\dot{S} = -\frac{\beta SI}{N_h}$$
$$\dot{E} = \frac{\beta SI}{N_h} - \alpha E \quad (2)$$
$$\dot{I} = \alpha E - \gamma I$$
$$\dot{R} = \gamma I$$

Conforme definido anteriormente, os parâmetros α e γ dependem das constantes tempo de incubação e tempo de infecção, nas formas $\alpha = 1/T_{inc}$ e $\gamma = 1/T_{inf}$. Veremos a seguir que o SEIR inclui um retardo no tempo, com relação ao SIR, que será tanto diminuído quanto o T_{inc} se aproximar de zero, ou seja, para $T_{inc} \to 0$, $\text{SEIR} \to \text{SIR}$.

Na Figura 9 são apresentadas três simulações, considerando diferentes valores de T_{inc}. Utilizamos os mesmos valores $\gamma = 1/T_{inf} = 0.1818$ e $\beta = R_0\gamma = 0.4091$ da simulação com o SIR (Figura 6). Os valores considerados para o T_{inc} foram 0.005, 2 e 4. Podemos perceber que à medida que o T_{inc} cresce, aumenta o retardo no tempo, mas mantém-se o mesmo valor final para o R, que se estabiliza em 85341 após o I voltar a ser zero, quando se encerra o surto epidêmico. Observamos ainda que, para $T_{inc} = 0.005$, as curvas em vermelho da Figura 9 são praticamente idênticas ao R e ao I da Figura 6, ou seja, o SEIR produz respostas iguais ao SIR. Como as amplitudes do número de infectados nos instantes de pico são menores para T_{inc} maiores, as curvas do número de infectados são mais abertas à medida que os seus instantes de pico se deslocam para a direita, caracterizando assim retardos no tempo.

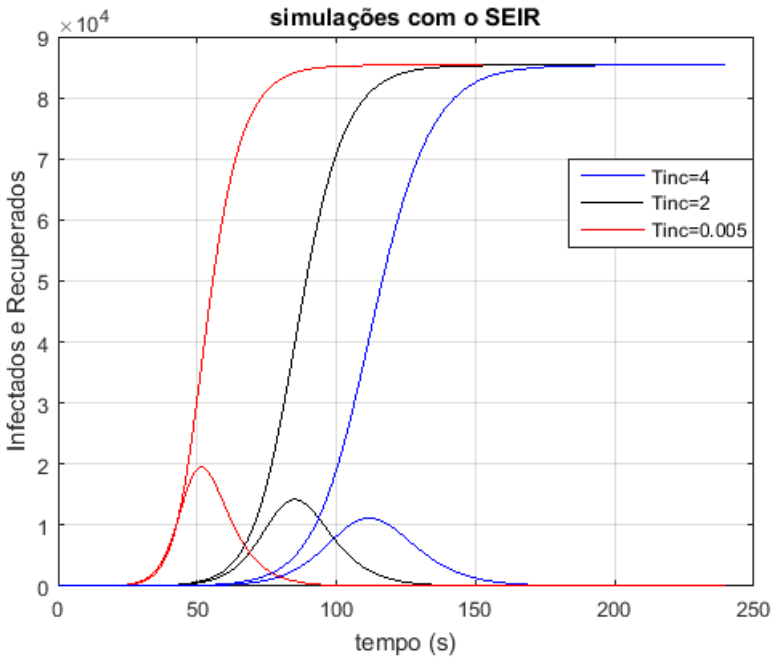


Figura 9. Simulações com o modelo SEIR, para diferentes valores de T_{inc}.

A partir dos anos setenta, quando os computadores digitais permitiram facilidades para a realização de simulações, muitos trabalhos foram publicados propondo modelos derivados do SIR e ainda, surgiu todo um grupo de modelos derivados do SEIR. De fato, o SEIR ganhou tal importância que gerou um rastro na história, de modelos que já são modificações desta mesma sua estrutura. Um exemplo típico foi apresentado em Prem *et al*. (2020), cujo diagrama de blocos pode ser visto na Figura 10. Neste artigo os autores dividem o compartimento dos Infectados em duas categorias, os sintomáticos ou clínicos e os assintomáticos ou subclínicos, de forma que a probabilidade de um infectado se tornar sintomático é definida pelo parâmetro τ. Neste SEIR modificado, as variáveis de estado I_s e I_a representam os infectados sintomáticos e os assintomáticos, respectivamente.

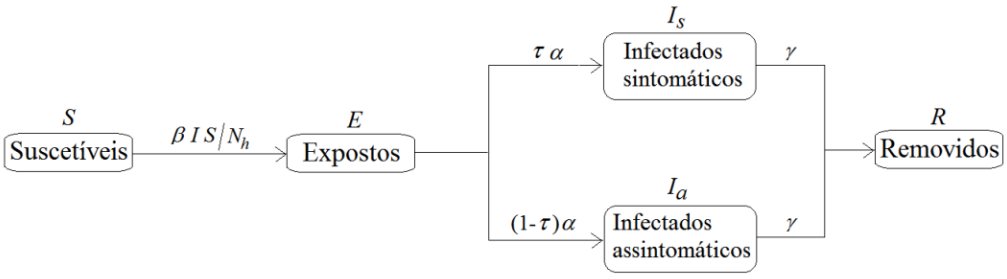


Figura 10. Diagrama de blocos do modelo SEIR modificado para a inclusão de Infectados sintomáticos e assintomáticos.

Alguns autores consideram estruturas mais complexas para a modelagem dinâmica de epidemias. Este é o caso do artigo publicado por Giordano *et al*. (2020), cujo diagrama de blocos do modelo pode ser visto na Figura 11. Os autores trabalharam com uma base de dados reais coletados na Itália, correlatos à COVID-19. Puseram o nome de SIDARTHE no modelo que propuseram. Quando observamos detalhadamente esse modelo, percebemos que se trata de uma modificação da estrutura SIR. No diagrama de blocos da Figura 11, excetuando-se os compartimentos S e I, todos os demais formam o compartimento dos removidos (R) na estrutura SIR. Chama a atenção o grande número de parâmetros que essa estrutura de modelo necessita, igual a quinze, no total.

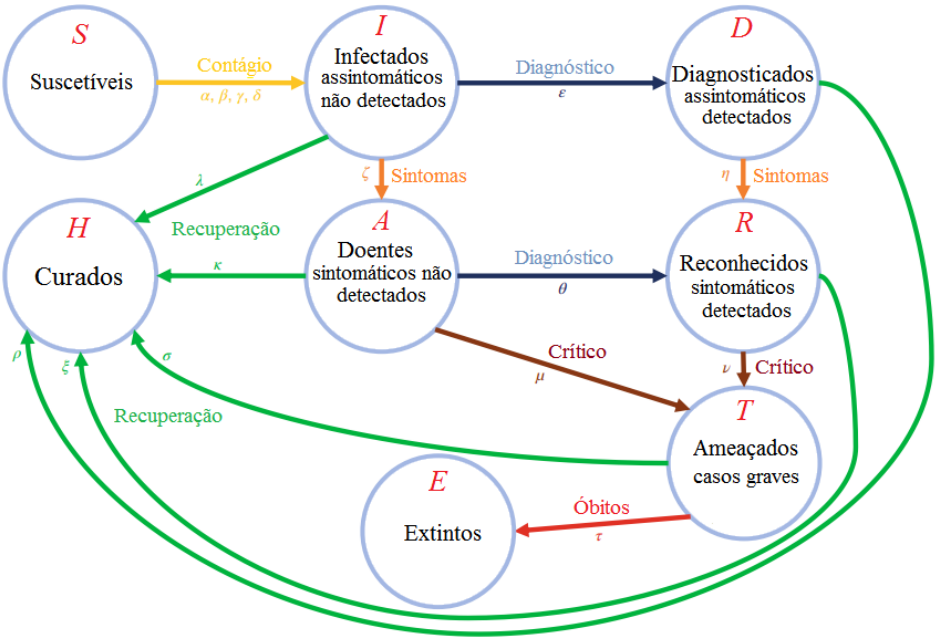


Figura 11. Diagrama de blocos do modelo SIDARTHE (adaptado de GIORDANO *et al*. 2020).

Existe um conflito em termos de custo-benefício na escolha de um modelo que possa representar bem a dinâmica de uma epidemia, localizada em uma determinada cidade ou região: se a escolha for por um modelo de estrutura mais simples, este terá menos parâmetros e, portanto, menos erros de identificação paramétrica, mas poderemos incorrer em maiores erros estruturais; se a escolha for por um modelo de estrutura mais complexa, este terá mais parâmetros e, portanto, mais erros de identificação paramétrica, embora possa haver menos erros estruturais. Após a realização de diversos estudos com dados reais de cidades da Europa, da China, dos USA e do Brasil, chegamos à conclusão que um

modelo de estrutura mais simples pode fornecer bons resultados em termos de reproduzir a dinâmica real epidêmica da COVID-19, desde que este modelo receba algumas adaptações. Em geral, os dados reais sobre a COVID-19 não se mostram muito confiáveis por diversos fatores, destacando-se, dentre estes, a subnotificação. Dados menos confiáveis significam amplificação de erros na identificação paramétrica e, quanto maior for o número de parâmetros a estimar, maiores serão os problemas com erros paramétricos introduzidos no modelo dinâmico. Nos capítulos seguintes mostramos que, de fato, um modelo de estrutura simples, com algumas adaptações, pode gerar bons resultados para o caso específico da epidemia por COVID-19, localizada em uma determinada cidade ou região.

Referências

GANYANI T., Kremer C., Chen D., Torneri A., Faes C., Wallinga J. and Hens N.. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data. Euro Surveill. 2020 Apr 30; 25(17): 2000257.

GIORDANO G., Blanchini F., Bruno R., Colaneri P., Di Filippo A., Di Matteo A. and Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 2020, https://doi.org/10.1038/s41591-020-0883-7.

KERMACK, W.; MCKENDRICK, A. Contributions to the Mathematical Theory of epidemics. Proceedings of the Royal Society A, Mathematical, Physical and Engineering Science, 1927, https://doi.org/10.1098/rspa.1927.0118.

PREM K. *et al*. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. www.thelancet.com/public-health Published online March 25, 2020 https://doi.org/10.1016/S2468-2667(20)30073-6.

Capítulo III

Primeiro Estudo – O Modelo MdSIR

3.1 Introdução

O presente capítulo tem como principal objetivo explicar detalhes sobre o modelo dinâmico proposto, que passa a fazer parte do grupo de modelos também derivados do SIR. Esta adaptação proposta recebeu o nome de MdSIR, por se tratar de modificações impostas à estrutura SIR, de forma a representar bem a dinâmica, aplicada especificamente à COVID-19 em uma determinada cidade ou região. Dados reais de diversas cidades são utilizados nesse primeiro estudo, sendo o principal deles o caso da província de Hubei (China). Este caso é particularmente interessante porque em Hubei, houve um ciclo epidêmico completo em razão do eficiente *lockdown* decretado. Dados reais da cidade de Nova York e de algumas cidades espanholas também são utilizados.

Uma vez que os principais modelos dinâmicos de epidemias já foram apresentados ao leitor no capítulo anterior, apresentamos a seguir uma breve revisão bibliográfica sobre modelos epidemiológicos, principalmente enfocando trabalhos voltados à COVID-19. Evidentemente, é fundamental uma revisão bibliográfica antes de se iniciar qualquer pesquisa, por motivos óbvios.

Conforme mencionado anteriormente, a maioria dos trabalhos utilizam os chamados modelos compartimentais, cujos principais são o SIR (Suscetíveis, Infectados, Removidos) e o SEIR (Suscetíveis, Expostos, Infectados, Removidos). As equações diferenciais destes modelos descrevem como ocorrem as mudanças de indivíduos entre os compartimentos. No caso do SIR tradicional, indivíduos que contraem a doença deixam o compartimento dos Suscetíveis e passam a integrar o compartimento dos Infectados e deste, passam a Removidos após o término da infecção. É importante ressaltar que no compartimento dos Removidos estão os que se recuperaram clinicamente e os que vieram a óbito. Conforme também informado anteriormente, o modelo SIR tradicional foi proposto originalmente por KERMARCK e MCKENDRICK em 1927. Desde então muitas variações deste modelo têm sido propostas na literatura, sendo o modelo SEIR a mais citada. O presente trabalho também propõe uma variação do modelo SIR, especificamente adaptada ao estudo dinâmico da COVID-19.

No capítulo anterior analisamos o modelo proposto por GIORDANO *et al*. (2020), que utilizam uma variação do SIR para modelar a epidemia na Itália. São muitos parâmetros a serem estimados e eles concluem que uma infinidade de combinações com diferentes parâmetros consegue um mesmo resultado bom em termos de reproduzir o fenômeno real.

Interessantes estimativas de contágio foram publicadas por LI R. *et al*. (2020), enfocando particularmente a epidemia de COVID-19 na china.

HAMZAH *et al*. (2020) utilizaram o modelo SEIR, sem, no entanto, considerar quarentena ou limitações de circulação. Os autores não mostraram confrontações entre dados reais e resultados de simulações, mas apresentaram uma boa revisão da bibliografia recente sobre o assunto.

KRAMER *et al*. (2020) utilizam dados de mobilidade em tempo real a fim de monitorar a eficácia e o impacto das medidas de controle adotadas nas diversas províncias chinesas.

RACHAH e TORRES (2017) utilizam o modelo SEIR para analisar a epidemia pelo vírus ebola na Libéria (África), considerando a vacinação como variável de controle da epidemia.

ALMEIDA *et al.* (2020) aplicam o SEIR para a previsão da evolução dinâmica da epidemia de COVID-19 na cidade de Belo Horizonte.

Alguns trabalhos específicos da área de medicina trazem importantes levantamentos de dados sobre a atual epidemia (LI Q. *et al*. 2020).

PREM *et al*. (2020) apresentam um estudo em simulação para a província de Wuhan, China, utilizando o modelo SEIR, incluindo análises das curvas de crescimento para diferentes escolhas de parâmetros do modelo.

HUBBS (2020) utiliza o modelo SEIR para analisar mudanças nos picos das curvas de respostas do modelo, em razão de modificações em alguns parâmetros.

QUINTANS e SILVA (2018) utilizam o SIR para a dinâmica de epidemias de HIV e gripe, incluindo análises de estabilidade do modelo.

Há uma vasta documentação especifica sobre modelos dinâmicos de epidemias, na forma de artigos e teses. KABIR *et al*. (2019) estudam efeitos da conscientização da população sobre a dinâmica com o modelo SIR. Seguindo o mesmo tema, o fator medo da doença na população foi proposto como adaptação ao modelo SIR por EPSTEIN *et al*. (2008).

ZHOU *et al*. (2017) desenvolveram estudos sobre análise de estabilidade de um modelo SIR discreto de vários grupos. SONG *et al*. (2016) também introduziram a análise de Lyapunov de estabilidade em um modelo SIR com atrasos distribuídos e taxa de incidência não linear.

LONDOÑO (2014) apresenta métodos de estimação de parâmetros dos modelos SIR e SEIR, focando na estimação do índice de reprodução basal, que se constitui no principal parâmetro, relacionado com a velocidade de propagação do contágio.

O modelo SIR em sua forma discreta é amplamente discutido por SABETI (2016), incluindo estrutura etária da população.

ZHEN *et al*. (2018) propõem modificações em um modelo SIR para incorporar a difusão do Laplaciano e o atraso espaço-temporal para modelar a transmissão de doenças transmissíveis.

GAI *et al*. (2020) investigam o modelo epidêmico SIRS com difusão espacial e taxas de incidência não lineares. Este modelo é uma adaptação do SIR tradicional para representar a o espalhamento da infecção considerando a sua difusão espacial. A ideia que transporta esse tipo de modelo é que, em uma mesma cidade, pode haver focos distribuídos espacialmente com maiores ou menores velocidades de contaminação.

CHOWELL *et al*. (2020) mostraram estudos usando o Modelo de Crescimento Generalizado (GMG) e o Modelo Logístico Generalizado (GLM) aplicado a uma epidemia de Ebola.

DORN (2016) utiliza o SIR para fazer análises comparativas entre diferentes epidemias de dengue, utilizando estimativas do índice de reprodução basal. Nas dissertações e teses encontradas há extensa bibliografia sobre modelos dinâmicos de epidemias. Especificamente sobre a COVID-19, há em WHO-China (2019) importantes informações sobre dados clínicos identificados para evolução epidêmica na China.

Um dos principais objetivos do presente capítulo é mostrar que um modelo de estrutura simples pode reproduzir bem resultados reais da evolução dinâmica da COVID-19. A motivação para perseguir esse objetivo está vinculada à grande dificuldade de se estimar parâmetros para a dinâmica localizada em cidades ou regiões. Modelos com estruturas mais complexas têm muitos parâmetros a estimar e isso tende a amplificar os erros paramétricos. A opção foi utilizar a estrutura mais simples de todas: o modelo SIR. Porém, este modelo em sua forma original com parâmetros constantes não consegue reproduzir a dinâmica epidêmica, principalmente porque existem variações na densidade populacional de suscetíveis, por imposição de governos locais, variações estas comumente conhecidas como isolamento social. Conforme será visto na sequência do estudo, o isolamento social

provoca diminuição na velocidade da contaminação. Com o isolamento social, apenas uma parcela da população pode circular livremente. Este percentual de circulação foi introduzido no modelo modificado MdSIR, de forma a torná-lo mais realista. Esta foi a principal motivação do presente estudo: adaptar uma estrutura de modelo simples, de forma a introduzir na dinâmica o percentual de circulação, o qual se constitui na principal variável de controle epidêmico, principalmente porque ainda não há vacina contra a COVID-19.

Outras duas modificações também foram necessárias: a consideração no modelo do fenômeno aqui denominado de aderência; a introdução na dinâmica de mais uma variável de controle, correspondente a uma taxa de quarentena de infectados. Esta última modificação se deve ao fato de pessoas já identificadas com a infecção viral entrarem em quarentena, seja em suas próprias residências ou em hospitais e assim, não contribuem mais com a contaminação.

Finalmente, ressaltamos que foram encontrados na literatura vários trabalhos que utilizam também variações do modelo SIR, na modelagem da COVID-19: além de GIORDANO *et al*. (2020) e GAI *et al*. (2020) já referenciados, há os trabalhos do grupo GDISPEN (2020), FRANCO (2020), REIS *et al*. (2020), SCHULZ *et al*. (2020), WU *et al*. (2020) e VILLAVERDE e JONES (2020).

3.2 Modificações na Estrutura do SIR

3.2.1 O Modelo MdSIR

Foram propostas três modificações no modelo SIR, a fim de torná-lo mais preditivo e mais adaptado à epidemia com o novo coronavírus. A primeira delas é a consideração de um efeito dinâmico que intitulamos nesse estudo de zona ou região de aderência. Conforme será explicado posteriormente, a dinâmica epidêmica pode ser “capturada” caso algumas condições a façam entrar nesta zona de aderência: a contaminação é interrompida e o número de infectados busca o zero.

Outra modificação originalmente proposta no presente estudo consiste na determinação da variação da taxa de crescimento (β) em função do percentual de circulação (p). Definimos o percentual de circulação como sendo a parcela da população que pode circular livremente, dividida pelo total de habitantes, de forma que: $0 \leq p \leq 1$. Outra modificação consiste em considerar que alguns infectados são identificados e entram em quarentena, em suas próprias casas ou em tratamento nos hospitais. Essa parcela de infectados identificados

(sintomáticos ou não) não contribui mais com a contaminação. O modelo MdSIR (GOMES *et al*. 2020a) passa então a ter a seguinte forma:

$$\dot{S} = \frac{-\beta(p)SI}{N_h}$$

$$\dot{I} = \frac{\beta(p)SI}{N_h} - (\gamma + k)I \tag{3}$$

$$\dot{R} = (\gamma + k)I$$

O parâmetro k representa a taxa de infectados (sintomáticos ou não) que são detectados e que entram em quarentena em hospitais ou nas suas próprias residências, passando a não mais contribuírem com a contaminação. Com isso, a nova taxa de recuperação será: $\bar{\gamma} = \gamma + k$. O índice de reprodução basal efetivo no MdSIR assume então a forma:

$$\bar{R}_0(t) = \frac{\beta(p)}{(\gamma + k)} \tag{4}$$

Como o percentual de circulação pode variar com o tempo $\big(p = p(t)\big)$, o índice de reprodução basal efetivo também passa a ser função do tempo $\big(\bar{R}_0(t)\big)$. Para efeito de simplificação de notação, trataremos, por enquanto, o percentual de circulação apenas como p, embora sabendo que este é um parâmetro função do tempo. É mostrado na equação (4) que o parâmetro k contribui para diminuir o índice de reprodução basal efetivo. Além da quarentena de parte dos infectados, este parâmetro pode representar na dinâmica uma série de ações protetivas, tais como o distanciamento de pessoas em lugares públicos e o uso obrigatório de máscaras. Fundamentalmente, a efetuação de testes em massa para identificar infectados assintomáticos e colocá-los em quarentena ampliaria o valor do k, reduzindo assim o índice de reprodução basal efetivo $\big(\bar{R}_0(t)\big)$. É importante lembrar que na ausência total de isolamento social, $p = 1$, ou seja, teríamos 100% de circulação. Se ainda considerarmos que não há quarentena de infectados, teríamos $k = 0$. Nesta situação, a equação (4) seria $\bar{R}_0(t) = \beta(1)/(\gamma + k) = \beta_0/\gamma = \gamma R_0/\gamma = R_0$. Portanto, o índice de reprodução basal efetivo seria igual ao índice de reprodução basal R_0 na ausência de isolamento social. Uma vez que a epidemia se torna realidade em alguma cidade, há sempre medidas de prevenção, sendo a principal delas o isolamento social, de forma que o percentual de circulação p torna-se menor do que 1. Com isso, veremos posteriormente que a taxa de crescimento $\beta(p)$ também diminui, reduzindo assim o índice de reprodução basal efetivo $\bar{R}_0(t)$.

3.2.2 Determinação da Taxa de Crescimento Variável

O estudo apresentado nesta subseção teve como base experimental os dados reais provenientes da região de Hubei (China). O principal motivo reside na importância capital desses dados em razão dos mesmos serem oriundos de um ciclo epidêmico completo, ou seja, de uma epidemia com início, expansão e rápido fim, provocado por uma eficiente estratégia de controle do tipo *lockdown*.

Cada epidemia que assola a humanidade tem suas próprias características e variáveis de controle específicas. No caso do novo coronavírus, o percentual de circulação (p) tem se mostrado como sendo a principal variável de controle em várias localidades do planeta e também aqui no Brasil. Portanto, torna-se fundamental considerar esse efeito no sistema dinâmico. Inicialmente necessitamos realizar a identificação dos parâmetros β e γ originais, ou seja, antes de considerarmos qualquer limitação de circulação. Vamos considerar que o parâmetro γ não sofre alteração em função do percentual de circulação, de forma que o seu valor é constante e equivalente a: $\gamma = 1/T_{inf}$, sendo T_{inf} o tempo de infecção. Neste trabalho foi utilizado o valor indicado em Prem *et al*. (2020), em estudo sobre a epidemia em Hubei (China): $T_{inf} = 5.5 dias \rightarrow \gamma = 0.1818$. Chamaremos de β_0 a taxa de crescimento inicial e seu valor numérico foi também sugerido em Prem *et al*. (2020): $R_0 = 2.25; \beta_0 = \gamma R_0 = 0.4091$. Este é o valor inicial constante, equivalente a $\beta(p)$, para $p = 1$. Nesse estudo inicial adotamos $k = 0$, já que o objetivo é focado na determinação de como varia o $\beta(p)$. Esclarecemos que em Prem *et al*. (2020) os parâmetros são indicados na forma de intervalos numéricos e que adotamos os seus respectivos valores médios.

Inicialmente foram realizadas simulações para validar os parâmetros escolhidos $(\beta_0 \text{ e } \gamma)$. Na Figura 12a é apresentada uma confrontação de uma simulação com dados reais de Hubei, indicando que no início da epidemia, os parâmetros adotados de fato conferem uma boa aproximação do modelo com os dados reais. A partir de certo tempo começaram a existir limitações de circulação de pessoas. Este tempo é muito bem identificado a partir dos dados reais e será chamado de instante de corte (t_c). Por exemplo, observando a Figura 12a, percebemos claramente que, a partir de $t = t_c = 15\ dias$, há um descolamento entre dados reais e simulação, ou seja, a simulação considerando parâmetros constantes só apresenta boa aproximação com os dados reais para $t \leq t_c$. As limitações de circulação impõem que apenas uma parcela da população pode circular livremente. Como a área da região permanece a mesma, na prática, o que há é uma diminuição da densidade populacional, entre as pessoas suscetíveis a entrarem em contato com o vírus. Esta diminuição da densidade acarreta também diminuição da taxa de crescimento β, conforme fica evidente nos estudos sobre Hubei. Nesta cidade, importantes limitações de circulação

foram impostas à população, resultando em um eficiente controle epidêmico. Na Figura 12b é apresentada uma confrontação entre resultados reais e de simulação, para o número total de casos acumulados no tempo.

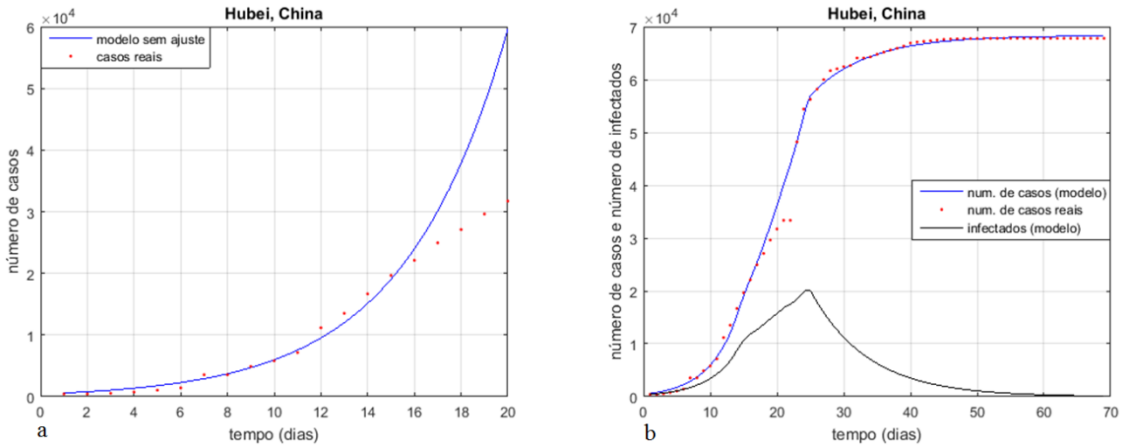


Figura 12. a) Confrontação entre simulação e dados reais, considerando $\beta = \beta_0$;
b) Variações impostas ao β a fim de reproduzir em simulação os dados reais de Hubei:
$t \leq 15$ $(\beta = \beta_0)$; $t > 15$ $(\beta = 0.63\beta_0)$; $t > 20$ $(\beta = 0.6\beta_0)$; $t > 22.2$ $(\beta = 0.56\beta_0)$;
$t > 25$ $(\beta = 0.16\beta_0)$; $\beta_0 = \gamma R_0$.

Ao observarmos a Figura 12, percebemos que o modelo reproduziu bem os dados reais. Porém, para haver essa boa aproximação entre dados reais e modelo, houve a necessidade de adotarmos diferentes valores para a taxa de crescimento β: inicialmente, $\beta = \beta_0 = \gamma R_0$; para $t > 15, \beta = 0.63\gamma R_0$; $t > 20, \beta = 0.6\gamma R_0$; $t > 22.5, \beta = 0.56\gamma R_0$; $t > 25, \beta = 0.16\gamma R_0$. Estas variações para o β foram obtidas por tentativa e erro, até que a resposta em simulação se ajustasse bem aos dados reais (Figura 12b). Este resultado mostra que a limitação de circulação no isolamento social imposto foi se intensificando, promovendo diminuições sucessivas na taxa de crescimento β, até que optaram por uma limitação drástica, após o vigésimo quinto dia, resultando na interrupção da epidemia: o número de infectados (Figura 12b) vai a zero e o número de casos acumulados passa a ser constante. Este é o exemplo ideal a ser seguido por qualquer região que esteja enfrentando uma situação de colapso do seu sistema de saúde. Uma vez constatada que é real essa variação, o problema passa a ser encontrar a função $\beta(p)$, definindo como a taxa de crescimento varia com o percentual de circulação.

Consideramos uma cidade com um número de habitantes qualquer, como por exemplo, $N_h = 211000$. Em seguida, diversas simulações (Figura 13a) foram realizadas com o modelo da equação (1), o SIR original, considerando os parâmetros $\beta = \beta_0$ e γ já definidos, simulações estas realizadas impondo variações no número de habitantes, na

forma: $N_h = \rho 211000, \rho = 1, 0.85, 0.7, \ldots, 0.1$. Os valores máximos das diversas curvas do número acumulado de casos (após o ciclo epidêmico completo) foram armazenados em um arquivo. Posteriormente, assumindo o número de habitantes constante e equivalente ao total da população ($N_h = 211000$), diversas simulações foram realizadas impondo variações ao β, até que fossem encontrados (utilizamos o algoritmo da bissecção) valores que permitiram números de casos acumulados finais iguais aos armazenados anteriormente, que geraram a Figura 13a. As diversas curvas do número de casos dessas últimas simulações podem ser vistas na Figura 13b.

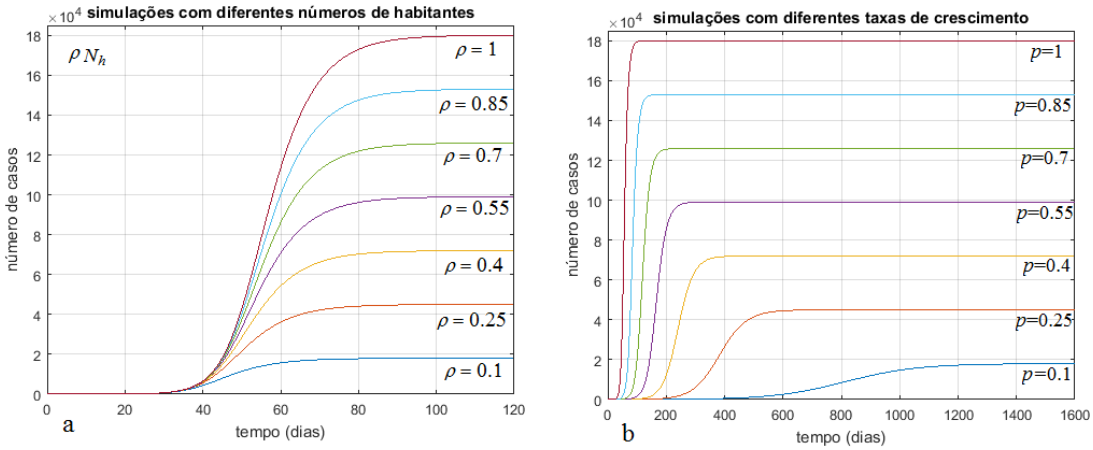


Figura13. a) Simulações com população variável; b) Simulações com o β variável em função do percentual de circulação p.

A Figura 14 destina-se a auxiliar na compreensão da ideia que originou a Figura 13. A hipótese aqui assumida é que, após o ciclo epidêmico completo, o número de casos para densidade constante seja o mesmo número de casos com densidade variável, embora o tempo para se completar o ciclo seja muito maior, para densidades menores. Explicando um pouco mais detalhadamente, vamos considerar uma epidemia começando em uma determinada cidade, com um $R_0 > 1$. A primeira linha de gráficos da Figura 14 considera a cidade diminuindo de área e também de população, de forma que a densidade populacional permanece constante. Neste caso, o R_0 também permanece constante. A segunda linha dessa figura considera a cidade diminuindo de população, mas mantendo a área constante, de forma que a densidade populacional diminui. Esta segunda linha de gráficos seria semelhante a uma cidade que adota 100% de circulação (gráfico da esquerda, $p = 1$), 50% de circulação (gráfico central, $p = 0.5$) e 25% de circulação (gráfico da direita, $p = 0.25$). A hipótese assumida preconiza que, ao final do ciclo epidêmico, os números acumulados de casos dos cenários relativos a uma mesma coluna seriam idênticos. Obviamente, nos

gráficos da segunda linha, a diminuição da velocidade de contaminação em razão da diminuição da densidade populacional (densidade de suscetíveis) tende a deslocar o tempo final do ciclo muito para a direita, conforme visto na Figura 13b. Esta hipótese não deve ser verdadeira para densidades muito pequenas, sendo esta a razão da proposição da zona de aderência, conforme explicado na próxima seção.

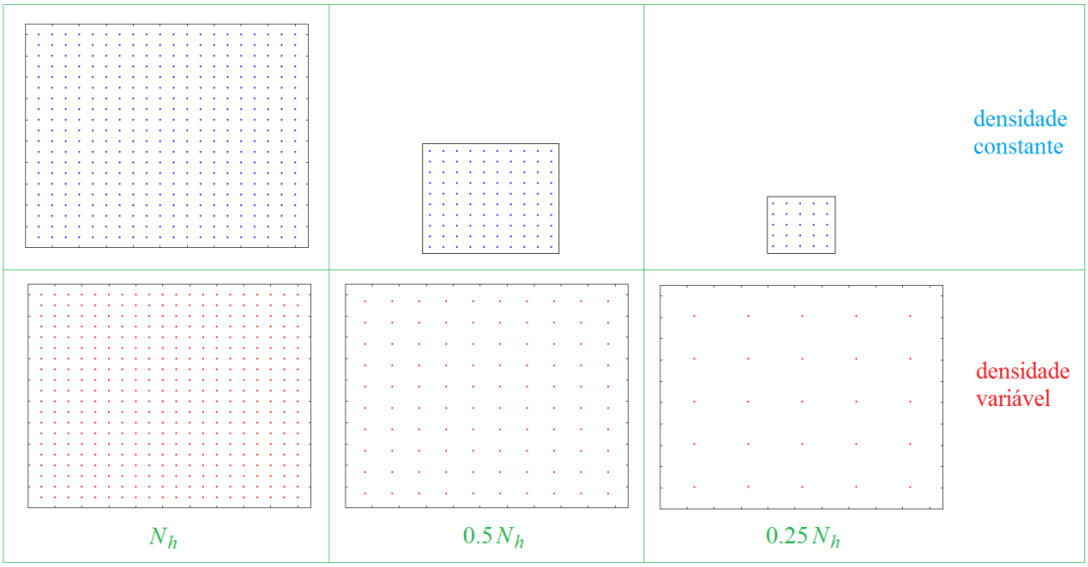


Figura 14. Variação do número de habitantes considerando área variável (densidade constante) e área constante (densidade variável).

Foram considerados mais três pontos que não entraram na Figura 13 por razões de escala: $\rho = p = [0.04; 0.06; 0.08]$. Os valores $\beta(p_i)$ para cada curva da figura 13b foram então identificados a partir do algoritmo numérico da bissecção. Tem-se assim um total de dez pontos $[p_i, \beta(p_i)]$, os quais foram ajustados com um polinômio de sétima ordem:

$$\beta(p) = a_1 p^7 + a_2 p^6 + \cdots + a_7 p + a_8 \quad (5)$$

O polinômio da equação (5) é específico para os parâmetros: $T_{inf} = 5.5 dias \rightarrow \gamma = 0.1818$; $R_0 = 2.25; \beta_0 = \gamma R_0 = 0.4091$, obtidos para Hubei (China). Este polinômio pode ser normalizado na forma: $\beta_n(p) = \beta(p)/(\gamma R_0)$. Dessa forma, uma taxa de crescimento equivalente a:

$$\bar{\beta}(p) = \beta_n(p)\gamma_b R_{0b} \quad (6)$$

se aplicaria para outras cidades ou regiões com diferentes parâmetros, equivalentes a γ_b e R_{0b}. Na Figura 15 são apresentadas curvas de taxa de crescimento em função do percentual de circulação. Na Figura 15a vemos a variação da taxa de crescimento $\beta(p)$

obtida com a equação (6) e considerando o $R_{0b} = R_0$ e $\gamma_b = \gamma$, parâmetros originais de Hubei, enquanto na Figura 15b vemos curvas para diferentes valores de R_{0b}, equivalentes a R_0, $0.85R_0$ e $1.15R_0$. Podemos perceber que mudanças de R_0 para um R_{0b} qualquer provoca deslocamentos na vertical da curva $\beta(p)$, de forma que o acréscimo/decréscimo do índice de reprodução basal provoca também acréscimo/decréscimo da taxa de crescimento, isso para toda a variação entre 0 e 1 do percentual de circulação p.

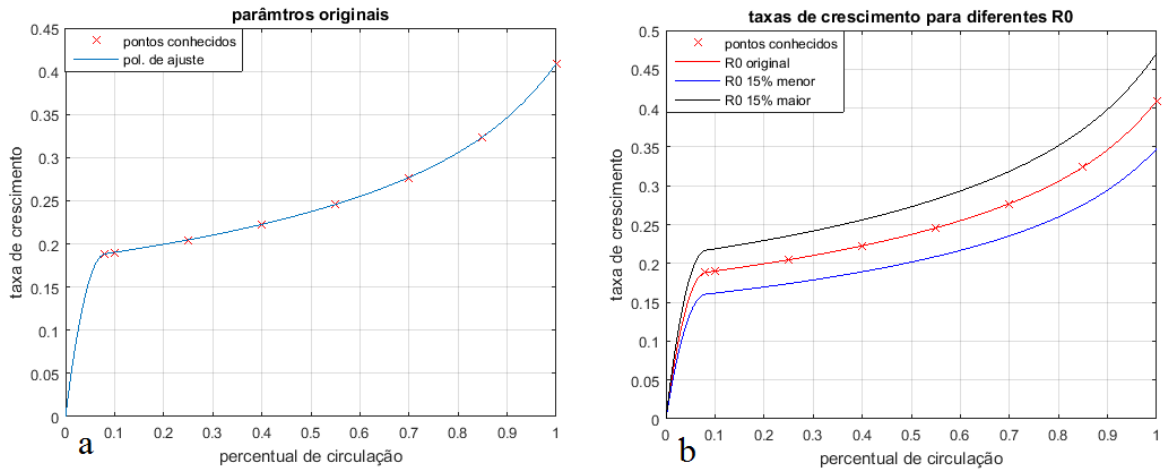


Figura 15. a) $\beta(p)$ com o R_0 original identificado para Hubei;
b) $\beta(p)$ para $R_{0b} = R_0$; $R_{0b} = 1.15R_0$ e $R_{0b} = 0.85R_0$.

Ressaltamos que para valores abaixo de 8% de circulação, as curvas foram ajustadas por uma função quadrática, levando a taxa de crescimento para zero quando o percentual de circulação tende a zero. Essa região de percentuais de circulação muito próximos de zero foi denominada região ou zona de aderência. Sabemos que essa região existe: quando o percentual de circulação é próximo de zero, caracterizando assim um bloqueio quase total, a epidemia é interrompida, o número de infectados busca o valor zero e o número de casos se estabiliza em um valor final constante, fenômeno que aconteceu em Hubei. Entretanto, é difícil conhecer esse limite inferior para o percentual de circulação inicial da zona de aderência. Adotamos 8% para esse limite ($p_{lim} = 0.08$). A taxa de crescimento em função do percentual de circulação pode ser gerada a partir do seguinte algoritmo:

$\beta(p) = \frac{1}{\gamma R_0}(a_1 p^7 + a_2 p^6 + \cdots + a_7 p + a_8)\gamma_b R_{0b}$;

$se\ p \leq p_{lim}$

$$x_f = p_{lim}\ ;\ y_f = \beta(x_f)\ ; a = \beta'(x_f)\ ; \tag{7}$$

$$A = \begin{bmatrix} x_f^2 & x_f \\ 2x_f & 1 \end{bmatrix}; B = \begin{bmatrix} y_f \\ a \end{bmatrix};$$

$$\vec{\alpha} = A^{-1}\vec{B}\ ;$$

$$\beta(p) = \alpha_1 p^2 + \alpha_2 p\ ;$$

$fim\ se$

onde $\beta'(x_f)$ corresponde à derivada do polinômio $\beta(p)$ aplicada ao ponto x_f. Observamos que a taxa de crescimento na zona de aderência é representada por um polinômio de ordem 2 que passa pela origem e tem a mesma derivada do polinômio de ordem 7, no ponto (x_f, y_f). Essa é uma curva idealizada: sabemos que a taxa de crescimento vai buscar o valor zero quando o percentual de circulação tende a zero, mas não sabemos exatamente como é esta curva. Assim como não conhecemos precisamente qual é o valor da constante p_{lim}. Adotamos $p_{lim} = 0.08$ e os resultados são coerentes, conforme será visto a seguir.

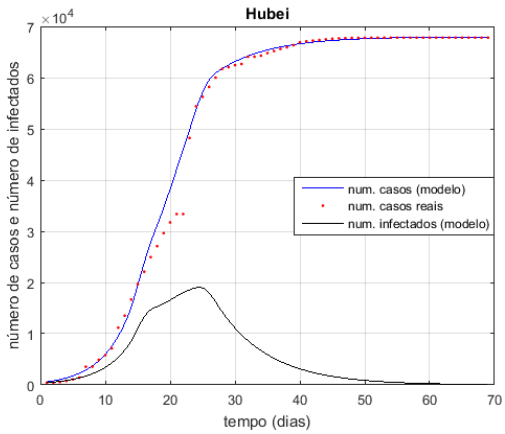
Como teste inicial, simulamos novamente a situação em Hubei, não mais variando diretamente a taxa de crescimento β e sim, o percentual de circulação p, com o $\beta(p)$ obtido a partir do algoritmo (7). Na Figura 16 é apresentada uma confrontação desta simulação com os dados reais. Ao lado direito dessa figura são apresentadas as variações impostas ao p que permitiram a boa aproximação entre simulação e dados reais. As mudanças em p não foram abruptas, descontínuas. Idealizamos uma spline cúbica unindo um degrau ao outro em p, conforme exemplo visto na Figura 16 (lado direito). É natural supor esta curva unindo de forma suave dois degraus em p porque, na prática, uma parcela da população de uma cidade ou região não entraria instantaneamente em isolamento social. Podemos perceber que o bloqueio em Hubei foi muito severo e por isso proporcionou um ótimo desempenho, tendo atingido um percentual de circulação final de 1.3% ($p = 0.013$).

É importante observarmos como variou o índice de reprodução basal efetivo para o caso de Hubei. Como o k foi considerado nulo, da equação (4), temos $\bar{R}_0(t) = \beta(p)/\gamma$. Portanto, considerando as variações impostas ao p ao longo do tempo t, encontramos o

$\beta(p)$ com o algoritmo (7) e assim, o $\bar{R}_0(t)$ estará determinado, já que o γ é um parâmetro constante conhecido. O $\bar{R}_0(t)$ para Hubei pode ser visto na Figura 17.

A análise das Figuras 16 e 17 nos indica que as medidas de restrição de circulação da população começaram no dia 15 e se intensificaram a partir do dia 21, com o *lockdown* completo se estabelecendo a partir do dia 25. Os dados reais de Hubei têm início no dia 22/01/2020 e terminam no dia 30/03/2020. Porém, no dia 22/01/2020 já havia 444 casos confirmados de pacientes com COVID-19, bem como 17 óbitos em toda a província chinesa. Portanto, a epidemia já vinha se desenvolvendo antes deste primeiro dia de dados reais. Outro fator importante a ressaltar é que o modelo tenta reproduzir os dados reais confirmados (pontos em vermelho na Figura 16). Porém, é fato estabelecido e aceito por pesquisadores do assunto, que há subnotificação de casos, cuja principal causa consiste no significativo número de assintomáticos que não são identificados via testes laboratoriais. Esta subnotificação factual é na realidade desconhecida, podendo variar muito em cada cidade ou país, dependendo principalmente se há ou não políticas de testagem em massa da população. Portanto, ao olharmos os dados reais e as simulações que os seguem, devemos ter em mente que o formato das curvas deve estar coerente com a realidade, embora com amplitudes numéricas maiores na realidade de cada cidade ou região pesquisada.

Outra importante observação é que o percentual de circulação p foi o único parâmetro considerado variável, até o presente momento. Este parâmetro de fato tem sido a principal variável de controle epidêmico, mas não é a única. Outras medidas de prevenção também ajudam no controle epidêmico, como são os casos do uso obrigatório de máscara e do distanciamento social em lugares públicos. Porém, como o p foi o único parâmetro deixado livre para variar, este assimilou todos os efeitos dinâmicos oriundos das outras medidas de prevenção. Portanto, é muito provável que o percentual de circulação real em Hubei tenha variado com amplitudes maiores do que as apresentadas na tabela da Figura 16. Este assunto será abordado com maior profundidade no Capítulo V.



$$R_{0b} = R_0$$

Degraus	Intervalo de transição
$p \rightarrow [1\ a\ 0.42]$	$\Delta t \rightarrow [15\ a\ 17]$
$p \rightarrow [0.42\ a\ 0.05]$	$\Delta t \rightarrow [21\ a\ 25]$
$p \rightarrow [0.05\ a\ 0.013]$	$\Delta t \rightarrow [25\ a\ 27]$

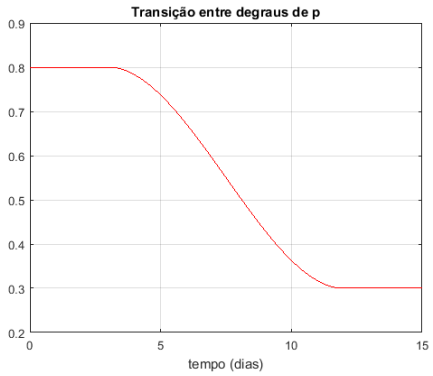


Figura 16 - O controle epidêmico aplicado em Hubei (China).

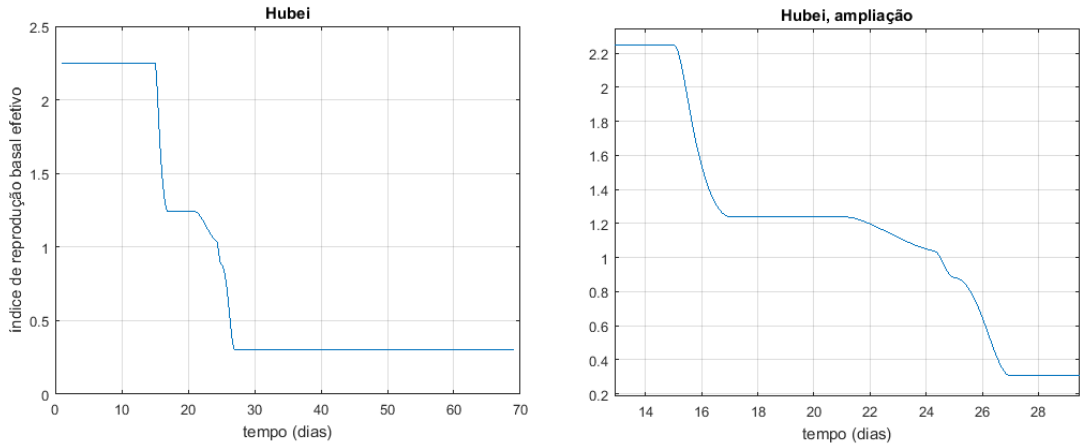


Figura 17. Índice de reprodução basal efetivo $\bar{R}_0(t)$ para Hubei.

3.3 Confrontações entre Dados Reais e Simulações

Uma vez conhecida a função genérica (algoritmo (7)) que expressa a taxa de crescimento em função do percentual de circulação $\big(\beta(p)\big)$, passamos a analisar o que vem acontecendo em diversas cidades ou regiões. Vale ressaltar que o modelo MdSIR permite também identificar qual é o R_{0b} de uma determinada região, agindo da seguinte forma: usamos inicialmente o R_0 identificado para Hubei; realizamos simulações variando o R_0 para um valor R_{0b} até que a resposta simulada se aproxime bem dos dados reais e assim, o R_{0b} passa a ser o valor identificado para a região em estudo. Este procedimento é aplicado apenas para os primeiros dias de dados de cada região estudada, ou seja, para um tempo menor ou igual ao instante de corte $(t \leq t_c)$, devendo-se isto às barreiras de circulação que os diversos governos passaram a adotar e estas alteram o valor desse parâmetro, conforme será visto no Capítulo V. As figuras a seguir mostram confrontações entre dados reais e simulados para algumas das cidades e regiões estudadas na pesquisa que originou este estudo. Foram analisadas quatro comunidades espanholas: Catalunha, Astúrias, Aragon e Andaluzia. Além destas comunidades da Espanha, foram desenvolvidos estudos também para as cidades de Madri e Nova York. Na Espanha, as medidas para a limitação de circulação tiveram início aproximadamente no vigésimo quarto dia a partir do início da epidemia, quando foi detectado o primeiro caso confirmado, em 20/02/2020. Os dados reais vão até o tempo $t = 78\,dias$ (05/05/2020). Nas comunidades de Astúrias, Aragon e Andaluzia o intervalo de dias simulado é menor do que 78 porque foram descartados dos seus respectivos arquivos de dados reais os dias iniciais com número de casos nulos. Os dados reais vistos nos gráficos correspondem aos pontos em vermelho. As simulações consideraram um tempo final maior do que o tempo final dos dados reais, de forma a se fazer uma previsão do que provavelmente aconteceria para tempos futuros. Na Tabela II há informações sobre os valores identificados do R_{0b} de cada caso, bem como sobre as variações identificadas para a variável de controle p (percentual de circulação), com os seus respectivos intervalos de tempo usados nas transferências entre os degraus em p. É importante ressaltar que, no presente capítulo, todas as simulações foram realizadas com o MdSIR (sistema de equações (3)), considerando $k = 0$ e com o $\beta(p)$ obtido a partir do algoritmo (7).

Na Figura 18 mostramos simulações e dados reais para a cidade de Madri. Observamos flutuações no número de casos reais, que podem ser efeitos de atrasos na divulgação dos dados, concentrando mais confirmações em alguns dias do que em outros. É possível encontrar as variações do percentual de circulação que reproduzem fielmente essas flutuações. Porém, daremos preferência por uma curva média de aproximação dos

dados reais, conforme visualizado na Figura 18. A preferência por esta curva média encontra a sua razão na não confiabilidade plena da aquisição temporal dos dados reais.

Na Figura 19 mostramos simulações e dados reais para quatro comunidades espanholas. Observamos que as referidas flutuações no número de casos reais estão presentes também nestes casos, assim como foram observadas na cidade de Madri. Conforme já adiantado anteriormente, foram identificados percentuais de circulação que reproduzem o comportamento médio das curvas de dados reais. Constatamos algumas evidências a partir da análise dos resultados para Madri e demais comunidades espanholas:

i) Os parâmetros R_{0b} de cada caso (Tabela II) são próximos do R_0 original identificado para Hubei, com diferenças que chegaram no máximo a 30% (caso de Madri);

ii) As variações identificadas para o percentual de circulação p em cada caso permitiram uma boa aproximação entre dados reais e simulações, mostrando assim que o modelo proposto é preditivo, desde que os parâmetros sejam corretamente identificados;

iii) As medidas de limitação de circulação adotadas, em todos os casos, conseguiram reverter o crescimento da epidemia;

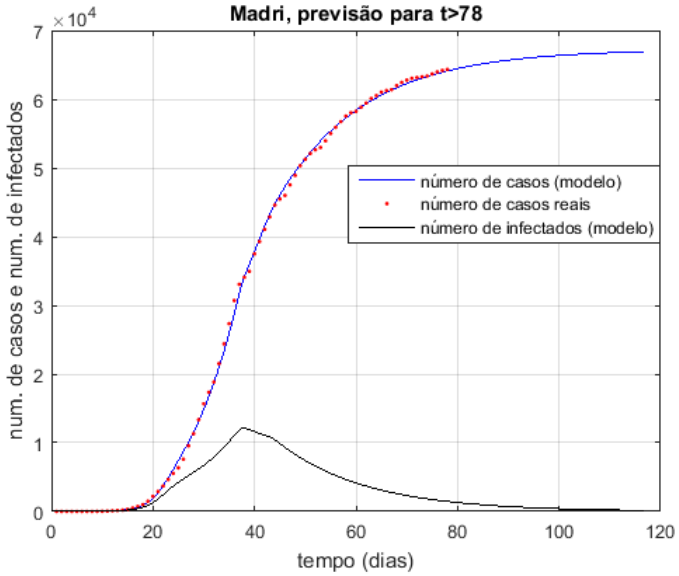


Figura 18. Simulações e dados reais para a cidade de Madri.

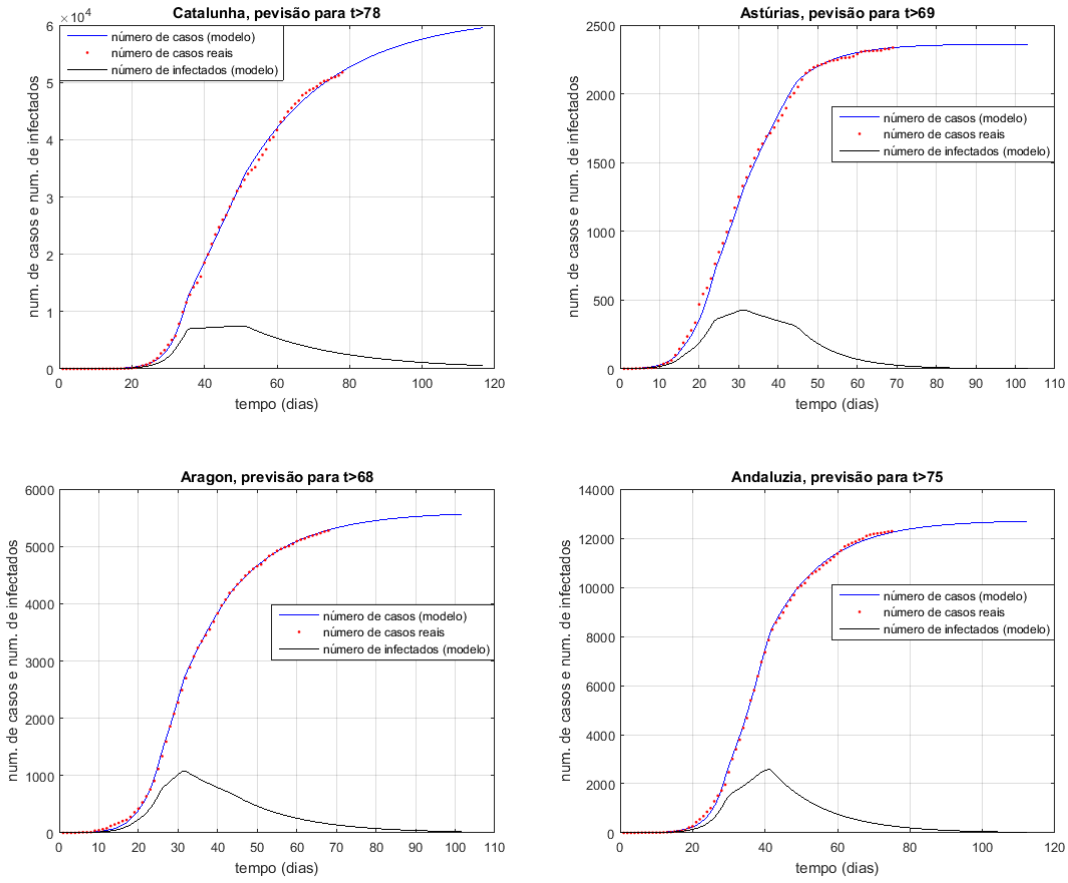


Figura 19. Simulações e dados reais para algumas comunidades espanholas.

Na Figura 20 mostramos simulações e dados reais para a cidade de Nova York. Os dados reais começaram no dia 02/03/2020, quando foi registrado o primeiro caso, encerrando-se no dia 02/05/2020, perfazendo um total de 62 dias de dados. Uma primeira constatação para Nova York foi que o R_{0b} foi 51% maior do que o R_0 original de Hubei (ver Tabela II). Este fato pode estar associado a uma maior densidade populacional, mas há outros fatores que também contribuíram para que a epidemia tenha se mostrado mais agressiva em termos de velocidade de contaminação em Nova York, associados, por exemplo, ao fato da cidade receber visitantes de todo o planeta em janelas de entrada a partir dos seus grandes aeroportos. Como nos casos anteriores, a partir do último dado experimental (último ponto em vermelho nas curvas) temos a continuação da simulação, numa tentativa de prever o comportamento epidêmico no futuro próximo. Podemos perceber que as medidas de limitação de circulação começaram a ser eficientes para reverter o crescimento da contaminação a partir de $t = 47\ dias$. Assim como nos casos da Espanha, o modelo também reproduziu bem os dados reais para a cidade de Nova York.

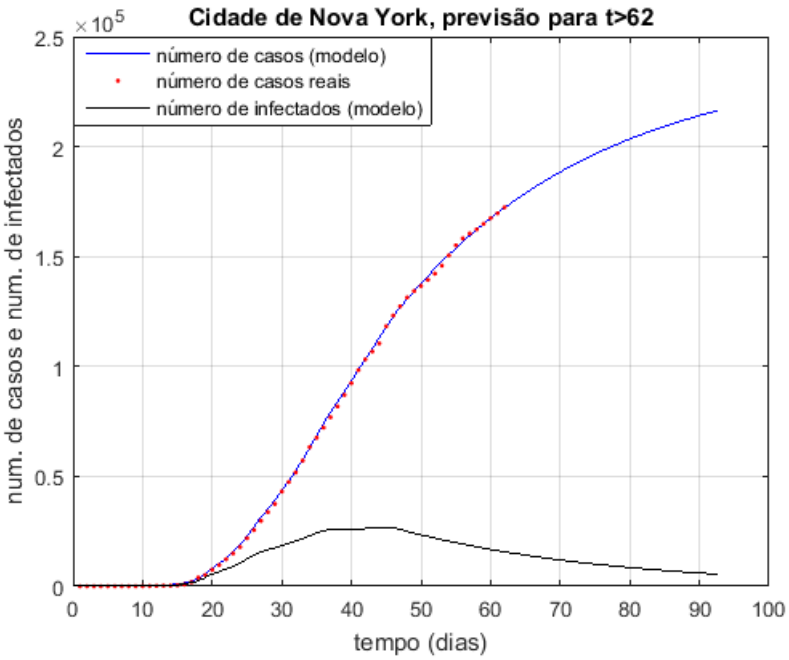


Figura 20. Simulação e dados reais para a cidade de Nova York.

Tabela II. Valores dos R_{0b} identificados, percentuais de circulação (1 corresponde a 100%) e seus respectivos intervalos de transição/efetivação (dias inicial e final da transição).

Cidade / região e R_{0b}	Percentual de circulação (p)	Intervalo de transição (Δt)
Madri R_{0b} =1.32R_0	1 a 0.22	20 a 28
	0.22 a 0.033	35 a 38
	0.033 a 0.024	42 a 45
Catalunha R_{0b} =1.067R_0	1 a 0.9	32 a 34
	0.9 a 0.06	35 a 36
	0.06 a 0.038	50 a 52
Astúrias R_{0b} =1.15R_0	1 a 0.78	15 a 17
	0.78 a 0.065	23 a 25
	0.065 a 0.039	30 a 32
	0.039 a 0.018	43 a 46
Aragon R_{0b} =1.12R_0	1 a 0.9	18 a 20
	0.9 a 0.4	25 a 27
	0.4 a 0.036	29 a 32
	0.036 a 0.028	42 a 45
Andaluzia R_{0b} =1.05R_0	1 a 0.41	28 a 32
	0.41 a 0.029	37 a 42
Cidade de Nova York R_{0b} =1.51R_0	1 a 0.4	19 a 21
	0.4 a 0.048	25 a 27
	0.048 a 0.034	35 a 37
	0.034 a 0.026	45 a 47

Lembramos novamente que nestas simulações, o parâmetro k (taxa de conversão de infectados para a quarentena) foi considerado nulo. Caso considerássemos um $k > 0$, mesmo que um pequeno valor, tipo $k = 0.02$, por exemplo, os percentuais de circulação

vistos na Tabela II teriam seus valores aumentados: este parâmetro contribui para diminuir o contato de infectados com suscetíveis, contribuindo assim com a diminuição do índice de reprodução basal efetivo $\left(\bar{R}_0(t)\right)$. Esse tema será mais investigado a partir do próximo capítulo.

3.4 Simulações com Controle

Conforme já informado anteriormente, a variável adotada como controle epidêmico no Brasil e em diversas localidades do planeta tem sido o percentual de circulação. Portanto, o objetivo principal dessa seção é mostrar um estudo sobre como o controle a partir do percentual de circulação pode interromper a epidemia em uma determinada cidade ou região. As administrações públicas têm o poder de modificar essa variável de controle. Adotaremos uma cidade hipotética, com 211000 habitantes. Consideramos ainda que os parâmetros do modelo são os mesmos identificados para Hubei: $T_{inf} = 5.5\, dias \rightarrow \gamma = 0.1818$; $R_0 = 2.25$; $\beta_0 = \gamma R_0 = 0.4091$. Conforme já explicado anteriormente, esses parâmetros iniciais são utilizados pelo algoritmo (7), o qual se encarrega de determinar as variações da taxa de crescimento em função do percentual de circulação $\left(\beta(p)\right)$.

A Figura 21 mostra curvas com o número de infectados, em simulações considerando diversos percentuais de circulação. Percebemos que as amplitudes nos picos diminuem com o decréscimo do percentual de circulação, assim como os instantes dos picos são deslocados para a direita. Este é o efeito que tem sido divulgado nas mídias como achatamento da curva de contaminação.

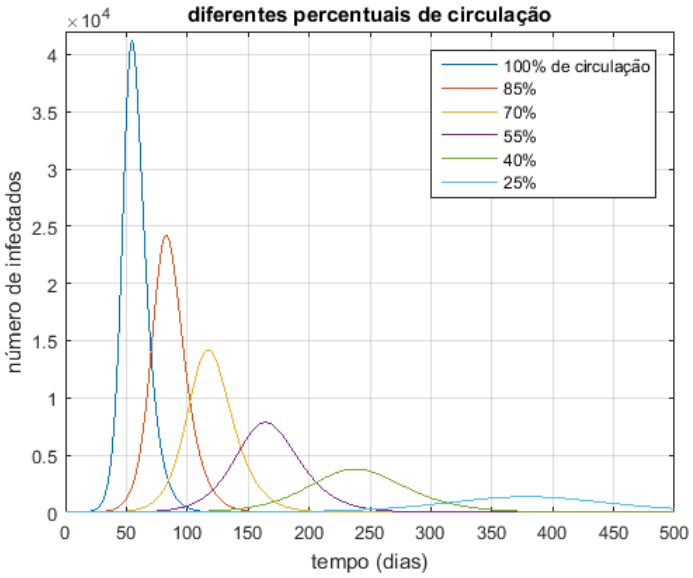


Figura 21. Simulações para diversos percentuais de circulação.

A Figura 22 põe em evidência o fenômeno de aderência, que pode acontecer também no início de uma epidemia localizada em uma determinada cidade ou região. É intuitivo supor que esse fenômeno ocorra quando as três condições a seguir ocorrerem simultaneamente:

i) Se o número de casos $C = (I + R)$ durante o intervalo Δt_{lim} crescer abaixo de um determinado valor C_{lim}, ou seja, se $\frac{C_f - C_i}{\Delta t_{lim}} < C_{lim}$, com C_i e C_f equivalentes aos números de casos inicial e final do intervalo Δt_{lim};

ii) Se $\bar{R}_0(t) = \frac{\beta(p)}{(\gamma+k)} < 1.05$;

iii) Se o número de infectados é menor do que um determinado limite: $I < I_{lim}$.

Para a cidade hipotética de 211000 habitantes utilizada nas simulações desta seção e considerando $k = 0$, este fenômeno de aderência aconteceu para um percentual de circulação menor ou igual a 10%. Os limites adotados foram: $I_{lim} = 5$; $C_{lim} = 1$; $\Delta t_{lim} = 35\ dias$. Quando $p = 0.1$, $\beta(p) = 0.1903$ (obtido com o algoritmo (7)), conferindo $\bar{R}_0(t) = 1.046$, valor que ficou abaixo do limite estipulado de 1.05. Nesta situação, todos os três itens que formam os critérios para o modo de aderência foram satisfeitos em $t = 35\ dias$. A dinâmica é então "capturada", o número de casos se estabiliza em um valor constante, indicando assim o fim da epidemia. Este processo de captura da dinâmica em modo aderência foi programado forçando o p, que neste exemplo era 0.1, a buscar um valor pequeno (0.02) a partir de uma spline cúbica unindo esses dois degraus. O $\beta(p)$ é assim forçado a seguir a trajetória de aderência (curva quadrática vista na Figura 15). Ressaltamos que este modo de aderência aconteceu neste exemplo porque a contaminação localizada na cidade estava em seu início e, neste início, o percentual de circulação já estava baixo, em 10% ($p = 0.1$), quase caracterizando um controle do tipo *lockdown*. Acreditamos que este fenômeno de aderência exista na prática, ocorrendo também no final do ciclo epidêmico, quando o número de infectados já é muito pequeno. Por exemplo, em uma determinada cidade, no final do ciclo epidêmico, em razão de isolamento por quarentena, os últimos cinco infectados não contaminaram mais ninguém e tornaram-se simultaneamente recuperados, ou seja, o final da epidemia não necessariamente se arrastaria por mais tempo, seguindo o lento decaimento exponencial. O fato é que este fenômeno de aderência afeta muito pouco a dinâmica, vindo a produzir algum efeito no início ou no final do ciclo epidêmico, desde que as condições para que ele ocorra sejam satisfeitas.

Podemos vislumbrar o fenômeno de aderência associando-o, por analogia, à dinâmica de um bloco deslizando em uma superfície sob a ação de uma força externa. Ao anularmos esta força externa, o bloco continuará em movimento desacelerado, mas devido às não linearidades do atrito com a superfície, este bloco irá interromper o seu movimento abruptamente, quando atingir uma velocidade muito próxima de zero. Por analogia, no final do ciclo epidêmico, o número de casos vem crescendo, mas com aceleração negativa, ou seja, a velocidade da contaminação vem diminuindo. Esta velocidade da contaminação corresponde à derivada da curva do número de casos: $d(I+R)/dt$. Quando esta velocidade estiver muito próxima de zero, ela se anulará abruptamente, quando o número de infectados também se tornará nulo $(I=0)$ e o número de recuperados se tornará constante $(R=cte)$. Este fenômeno deve acontecer para $\bar{R}_0(t)<1$ e o que nós estamos admitindo é que ele ocorra também para valores muito próximos de 1, no início ou no final do ciclo epidêmico. Adotamos como limite $\bar{R}_0(t)<1.05$ para gerar a Figura 22, mas é muito provável que este limite seja ainda mais próximo de 1. Este fenômeno não pode acontecer no meio do ciclo epidêmico porque, neste caso, o I é sempre muito maior do que o $I_{lim}=5$ (valor adotado). Por isso é que esse fenômeno é pouco significativo do ponto de vista dinâmico, de forma que todos os resultados de confrontação entre simulação e experimento do presente trabalho seriam rigorosamente iguais, caso desconsiderássemos esse fenômeno de aderência na dinâmica do modelo.

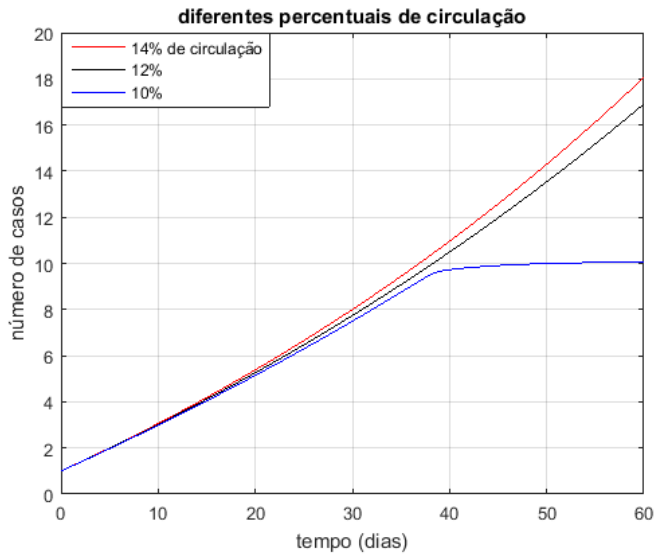


Figura 22. Simulações que ressaltam a ocorrência do fenômeno de aderência.

Na Figura 23 mostramos a aplicação do controle por limitação de circulação, iniciado no instante de controle $t_{cont}=25\,dias$. A cidade já estava, desde o início da epidemia, com um percentual de circulação $p=0.65$, mas esse nível de circulação permitida não impediu o crescimento do número de casos. Ressaltamos que muitas cidades brasileiras atualmente

(junho de 2020) estão com esse tipo leve de restrição de circulação. Em $t = t_{cont} = 25\ dias$ optamos por um controle mais severo da epidemia, forçando o percentual de circulação a assumir o valor do percentual de controle $p_{cont} = 0.015$, ou seja, apenas 1.5% da população podendo circular livremente. Consideramos o intervalo de tempo para a efetivação do controle equivalente a $\Delta t_{cont} = 5\ dias$ (durante esse intervalo, uma spline cúbica une os degraus em p, de 0.65 a 0.015). A análise da Figura 23 mostra que este controle foi eficiente, o número de casos se estabiliza próximo ao valor 600 e a curva de infectados tende a zero. Comparando este resultado com o que ocorreu de fato em Hubei (Figura 16), percebemos que os formatos das curvas são semelhantes, já que ambos correspondem a estratégias de controle do tipo *lockdown*.

Na Figura 24 evidenciamos a chamada segunda onda de contaminação, que pode acontecer caso haja um relaxamento da limitação de circulação antes do tempo ideal. Um controle igual ao caso da figura anterior foi aplicado em $t = t_{cont} = 25\ dias$ (primeiro pico) e estava atuando de forma eficaz na redução da contaminação. Porém, antes do tempo ideal, que neste caso seria em $t = 65\ dias$ (40 dias após a aplicação do primeiro controle) houve o relaxamento da limitação de circulação em $t = 55\ dias$, quando p passou de 0.015 para 0.8 (liberação para que 80% das pessoas voltassem a circular livremente). Esta medida, antes do tempo ideal, provocou a volta do crescimento da contaminação, gerando assim a segunda onda. Foi necessária então a aplicação de um segundo controle em $t = t_{cont} = 85\ dias$, o qual atuou de forma eficaz, estabilizando o número de casos e levando o número de infectados ao valor zero. Em ambos os controles consideramos o intervalo de efetivação igual a $\Delta t_{cont} = 5\ dias$.

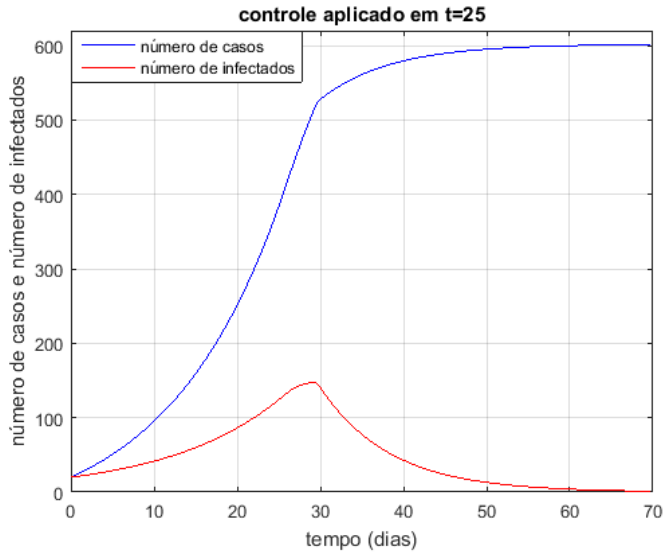


Figura 23. Simulação com a aplicação do controle de circulação em $t = 25$ (p era 0.65 passou a ser 0.015 durante $\Delta t_{cont} = 5\ dias$).

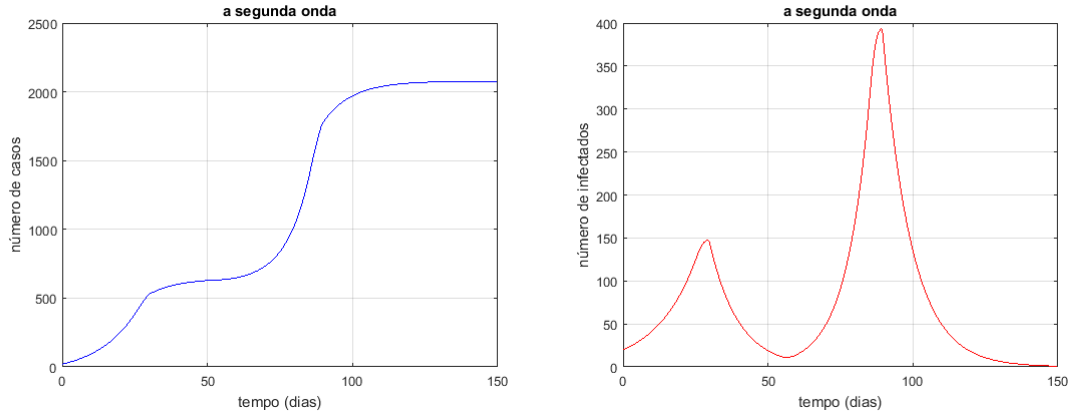


Figura 24. Simulação ressaltando o perigo de relaxarmos a limitação de circulação antes do tempo ideal: p era 0.65; em $t = 25\ dias$ passou a 0.015; em $t = 55\ dias$ passou a 0.8; em $t = 85\ dias$ voltou a 0.015.

3.5 Comentários Finais

O estudo mostrado neste capítulo foi realizado com a utilização de dados reais de diversas cidades ou regiões, algumas não descritas aqui em razão da natural limitação de espaço. São amplamente divulgadas na imprensa brasileira e internacional informações sobre possíveis subnotificações de casos nesta pandemia, de forma que os dados reais para estudos devem ser tratados com cautela. Neste contexto, Hubei mostrou-se como um exemplo ótimo para estudo, pois há um ciclo completo, incluindo um eficiente controle epidêmico. Seguimos uma premissa básica norteadora no presente estudo de modelagem: optar por um modelo com a estrutura mais simples possível, diminuindo a quantidade de parâmetros de difícil estimação, reduzindo assim erros paramétricos. Utilizamos então o modelo SIR, que possui a mais simples estrutura dentre os modelos compartimentais. Porém, foram necessárias algumas alterações que o possibilitaram se tornar mais preditivo e com boa reprodutibilidade com relação aos dados reais observados. Verificamos de imediato, tão logo começaram os primeiros estudos com base em dados reais, que a taxa de crescimento era variável ao longo do tempo. Essa variação se deve, principalmente, a mudanças impostas ao percentual de circulação da população. Por exemplo, uma diminuição do percentual de circulação acarreta uma diminuição da densidade populacional de suscetíveis ao vírus, provocando, consequentemente, uma diminuição no parâmetro taxa de crescimento. Essa foi a motivação para a determinação de uma função na forma de um algoritmo, proposta no presente capítulo e que expressa a taxa de crescimento em função

do percentual de circulação. Outra adaptação imposta ao modelo SIR tradicional foi a consideração da chamada zona ou região de aderência, fenômeno que deve ocorrer na prática caso algumas condições sejam reunidas, sendo a principal delas a verificação de um índice de reprodução basal efetivo menor do que um número muito próximo de 1. A outra adaptação proposta foi a consideração de uma taxa de quarentena de infectados, considerada nula no presente capítulo, mas que será analisada em detalhes a partir do próximo capítulo. Estas adaptações impostas ao modelo SIR geraram o MdSIR, com o qual foram realizadas diversas simulações sobrepostas a dados reais. Os resultados mostraram um bom caráter preditivo do MdSIR, permitindo a análise em simulação de diversos cenários, inclusive com controle epidêmico a partir da variável percentual de circulação da população. Os resultados do presente capítulo permitiram-nos concluir que nem sempre ampliar a complexidade estrutural de um modelo dinâmico pode conduzir a bons desempenhos. Há modelos compartimentais muito mais complexos estruturalmente, plenos de conexões entre os seus diversos compartimentos, cujas relações dinâmicas entre estes são realizadas usando parâmetros de difícil estimação. As inevitáveis incertezas paramétricas podem reduzir significativamente o desempenho do modelo quanto à sua reprodutibilidade e ao seu caráter preditivo. Portanto, para a epidemia com o novo coronavírus, o presente estudo mostra que promover adaptações em um modelo com estrutura mais simples pode produzir bons resultados.

No próximo capítulo analisaremos os efeitos de uma taxa de quarentena de infectados k diferente de zero, bem como expandiremos o estudo a diversas cidades brasileiras.

Referências

ALMEIDA A. C. L. *et al*. Análise do efeito das medidas de contenção à propagação da COVID-19 em Belo Horizonte. Relatório técnico. Março, 2020. https://ufmg.br/storage/2/5/a/7/25a7163c7fb5575ab6d81b5a05bfd844_15863100172762_298779967.pdf.

CHOWELL G. et al. Real-time forecasting epidemic trajectories using computational dynamic ensembles. **Epidemics** 30 (2020) 100379.

DORN, R. C. **Analysis of Dengue Dynamics through the number of reproducibility based on epidemiological data**. Doctoral Thesis in Physics, Universidade Federal da Bahia, 2016. https://blog.ufba.br/pgif/files/2017/04/T27-IF-UFBA.pdf.

EPSTEIN J. M. *et al*. Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations. **PLoS ONE** 3(12): e3955. (2008) doi:10.1371/journal.pone.0003955.

FRANCO, C. M. R. Modelos Matemáticos em Epidemiologia Aplicação: Evolução Epidêmica da COVID-19 no Brasil e no Estado da Paraíba. Relatório técnico, 05/2020, http://www.ces.ufcg.edu.br/portal/phocadownload/userupload/COVID-19_MODELO_SIR.pdf

GAI, C., IRON, C. and KOLOKOLNIKOV, T. Localized outbreaks in an S-I-R model with diffusion. **Journal of Mathematical Biology** (2020) 80:1389–1411.

GDISPEN, UFPel, Modelagem Matemática do COVID-19: Atualização de 15.05.2020. Relatório técnico, grupo https://wp.ufpel.edu.br/fentransporte/2020/05/15/modelagem-matematica-do-COVID-19-atualizacao-de-15-05-2020/.

GIORDANO G. *et al*. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. **Nature Medicine**, 2020, https://doi.org/10.1038/s41591-020-0883-7.

GOMES S. C. P., Monteiro I. O., Rocha C. R. Modelagem dinâmica da COVID-19 com aplicação a algumas cidades brasileiras. Revista Thema, 2020a.

HAMZAH B. F. A. *et al*. Corona Tracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction. [Submitted]. **Bull World Health Organ**. E-pub: 19 March 2020. doi: 10.2471/BLT.20.255695.

HUBBS C. Social Distancing to Slow the Coronavirus: Modeling the flattening of the COVID-19 peaks. Toward data science, Mar, 2020, https://towardsdatascience.com/social-distancing-to-slow-the-coronavirus-768292f04296.

KABIR K. M. A., KUGA K., TANIMOTO J. Analysis of SIR epidemic model with information spreading of awareness. **Chaos, Solitons and Fractals** 119 (2019) 118–125.

KERMACK, W.; MCKENDRICK, A. Contributions to the Mathematical Theory of epidemics. Proceedings of the Royal Society A, Mathematical, Physical and Engineering Science, 1927, https://doi.org/10.1098/rspa.1927.0118.

KRAEMER M. U. G. *et al.* The effect of human mobility and *c*ontrol measures on the COVID-19 epidemic in China. **Science**, Mar, 2020, doi: 10.1126/science.abb4218.

LI Q. *et al*. Early Transmission Dynamics in Wuhan, China,of Novel Coronavirus–Infected Pneumonia. **The New England Journal of Medicine**, March 26, 2020.

LI R. *et al*. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). **Science**, DOI:10.1126/science.abb3221, (2020).

LONDOÑO, S. L. M. **Estimação do Número de Reprodução Basal em Modelos Compartimentais**. Dissertação de Mestrado, Programa de Estatística, UNICAMP, 2014. http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/305840/1/MercadoLondono_SergioLuis_M.pdf.

PREM K. *et al*. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. www.thelancet.com/public-health Published online March 25, 2020 https://doi.org/10.1016/S2468-2667(20)30073-6.

QUINTANS, J. C. O. and SILVA. J. A. Model of dynamic epidemic system with numerical simulation in Python. **Biomatemática**, 28 (2018), 101–114.

RACHAH A. and TORRES D. F. M. Analysis, simulation and optimal control of a SEIR model for Ebola virus with demographic effects. DOI: 10.1501/Commua1_0000000841. Preprint of a paper whose final and definite form is with Commun. **Fac. Sci.Univ. Ank.Ser. A1 Math**. Stat., ISSN: 1303-5991. Article accepted for publication 01 May2017.

REIS R. F. *et al*. Características da pandemia do COVID-19 e impacto das políticas de isolamento no Brasil, Minas Gerais e Juiz de Fora. Nota técnica, http://www.ufjf.br/pgmc/files/2020/03/nt01pgmc-30-03-2020.pdf

SABETI, M. **Discrete epidemic study SIR with age structure and application of pulse and constant vaccination**. Tese de Doutorado, Universidade Federal de Pernambuco, 2011. https://repositorio.ufpe.br/bitstream/123456789/1321/1/arquivo2722_1.pdf

SONG H., LIU S. and JIANG W. Global dynamics of a multistage SIR model with distributed delays and nonlinear incidence rate. **Mathematical Methods in the Applied Sciences**, Volume 40, Issue 6, 2016.

SCHULZ R. A., ARAÚJO C. H. C., COSTICHE S. W. S. COVID - 19: A model for studying the evolution of contamination in Brazil. Preprint, https://arxiv.org/pdf/2003.13932.pdf.

VILLAVERDE J. F. and JONES C. I. Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities. Technical report, May, 2020. https://web.stanford.edu/~chadj/Covid/Dashboard.html.

ZHOU J., YANG, Y. and ZHANG, T. Global stability of a discrete multigroup SIR model with nonlinear incidence rate. **Mathematical Methods in the Applied Sciences**, Volume 40, Issue 14, 2017.

ZHEN, Z. *et al*. Wave propagation in a diffusive SIR epidemic model with spatiotemporal delay. **Mathematical Methods in the Applied Sciences**, Volume 41, Issue 16, 2018.

WHO-China Joint Missionon Coronavirus Disease 2019. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-COVID-19-final-report.pdf.

WU J. T. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. **Nature Medicine**, 2020, https://doi.org/10.1038/s41591-020-0822-7.

Capítulo IV

Segundo Estudo – O Parâmetro k

4.1 Introdução

No presente capítulo investigamos o efeito da taxa de quarentena de infectados (parâmetro k) sobre a dinâmica epidêmica, bem como realizamos estudos de casos com diversas cidades brasileiras. Por enquanto, consideramos que o único parâmetro variável é a taxa de crescimento β, que depende fundamentalmente do percentual de circulação p. Para uma cidade qualquer, que denominamos cidade b, cujos parâmetros sejam R_{0b} e γ_b, o algoritmo (7) calcula a taxa de crescimento como função de p, R_{0b} e γ_b. Portanto, podemos reescrever as equações do MdSIR na forma:

$$\dot{S} = \frac{-\beta(p, R_{0b}, \gamma_b)SI}{N_h}$$

$$\dot{I} = \frac{\beta(p, R_{0b}, \gamma_b)SI}{N_h} - (\gamma_b + k)I \tag{8}$$

$$\dot{R} = (\gamma_b + k)I$$

Em GOMES *et al.* (2020a) há mais detalhes sobre o MdSIR. Como por enquanto consideramos os parâmetros R_{0b} e γ_b constantes, escrevemos, no capítulo anterior, o β apenas como função do p, o que será sempre verdade considerando uma cidade específica. Porém, generalizando para uma infinidade de cidades, cada uma com os seus parâmetros específicos, o β pode ser escrito como função de p, R_{0b} e γ_b, conforme visto na equação (8). Como o p pode variar com o tempo, o índice de reprodução basal efetivo (equação (4)) passa a ser função do tempo e é reescrito na forma:

$$\bar{R}_0(t) = \frac{\beta(p, R_{0b}, \gamma_b)}{(\gamma_b + k)} = \frac{\beta(p, R_{0b}, \gamma_b)}{\bar{\gamma}_b} \tag{9}$$

onde o $\gamma_b = 1/T_{inf}$ corresponde à taxa de recuperação da cidade b e $\bar{\gamma}_b$ é a nova taxa de recuperação efetiva da cidade b, que engloba a taxa de quarentena k. Portanto, podemos escrever:

$$\bar{\gamma}_b = (\gamma_b + k) = \frac{1}{\bar{T}_{inf}} = \left(\frac{1}{T_{inf}} + k\right) = \left(\frac{1 + kT_{inf}}{T_{inf}}\right) \rightarrow \bar{T}_{inf} = \frac{T_{inf}}{1 + kT_{inf}} \quad (10)$$

sendo $\bar{T}_{inf}$ o novo tempo de infecção efetivo da cidade b. A conclusão dessa análise inicial é que, para $k > 0$ temos: $\bar{\gamma}_b > \gamma_b$ e $\bar{T}_{inf} < T_{inf}$. Portanto, a taxa de quarentena de infectados (k) contribui para diminuir o intervalo de tempo de infecção efetivo $(\bar{T}_{inf})$, contribuindo consequentemente com o aumento da taxa de recuperação efetiva $(\bar{\gamma}_b)$ e por isso, age para diminuir o índice de reprodução basal efetivo $\bar{R}_0(t)$. Da equação (9) temos:

$$\bar{R}_0(t) = \frac{\beta(p, R_{0b}, \gamma_b)}{\bar{\gamma}_b} = \beta(p, R_{0b}, \gamma_b)\bar{T}_{inf} \quad (11)$$

O $\bar{T}_{inf}$ seria o que alguns autores chamam de intervalo de geração, que corresponde ao tempo entre os eventos de infecção em um par infectante-infectado (GANYANI *et al*., 2020), ou seja, é o intervalo de tempo decorrido entre o instante no qual um indivíduo é infectado até o instante da transmissão para um outro indivíduo. Quando dividimos o índice de reprodução basal efetivo pelo intervalo de geração, temos a taxa de crescimento (velocidade de propagação da infecção): $\beta(p, R_{0b}, \gamma_b) = \bar{R}_0(t)/\bar{T}_{inf}$. É importante ressaltar que o intervalo de geração é diferente do intervalo serial, definido como sendo o tempo entre o início dos sintomas em um par infectante-infectado (NISHIURA *et al*., 2020). Lembramos que a unidade física de $\beta(p, R_{0b}, \gamma_b)$ é $1/dias$ e que a unidade física de $\bar{T}_{inf}$ é $dias$, de forma que o $\bar{R}_0(t)$ é um número adimensional, representando o número de pessoas para as quais um único infectado transmite a virose. Considerando R_{0b} e γ_b constantes, o $\beta(p, R_{0b}, \gamma_b)$ varia apenas em função do p, conforme preconiza o algoritmo (7). As equações (9) e (11) mostram que, como o $k > 0$ contribui para diminuir o $\bar{T}_{inf}$, ele atua para diminuir o $\bar{R}_0(t)$ e passa a fazer parte dos efeitos dinâmicos que contribuem para mitigar a transmissão do vírus.

Esta análise inicial também corrobora para mostrar a importância da testagem em massa da população de uma cidade que está sofrendo com a epidemia. Quanto mais assintomáticos são identificados e postos em quarentena, mais o valor do k é ampliado, reduzindo assim o índice de reprodução basal efetivo $\bar{R}_0(t)$. Em síntese, na ausência de qualquer medida de prevenção em uma cidade b, teríamos $k = 0$ e $p = 1$ e assim, a taxa de

crescimento seria constante e igual a $\beta(p, R_{0b}, \gamma_b) = \gamma_b R_{0b}$ e ainda, teríamos $\bar{R}_0(t) = R_{0b}$ e $\bar{\gamma}_b = \gamma_b$, ou seja, os parâmetros efetivos seriam iguais aos originais da cidade b. O isolamento social aplicado à cidade b provoca diminuição do percentual de circulação p, que por sua vez diminui o $\beta(p, R_{0b}, \gamma_b)$, diminuindo consequentemente o $\bar{R}_0(t)$. De forma semelhante, a identificação e posterior colocação em quarentena dos infectados detectados (sintomáticos e assintomáticos) amplia o valor do k, que amplia a taxa de recuperação efetiva $\bar{\gamma}_b$ e que, consequentemente, diminui o $\bar{R}_0(t)$. O objetivo final de qualquer controle epidêmico é sempre diminuir o índice de reprodução basal efetivo.

4.2 Simulações com Controle

Nesta seção fazemos algumas simulações semelhantes às efetuadas no capítulo anterior, mas no presente capítulo consideramos uma taxa de quarentena $k > 0$. Adotamos uma cidade com 211000 habitantes e com os parâmetros originais de Hubei: $T_{inf} = 5.5\ dias \rightarrow \gamma = 0.1818$; $R_0 = 2.25$; $\beta_0 = \gamma R_0 = 0.4091$. Lembramos que, neste caso, $\gamma_b = \gamma$ e $R_{0b} = R_0$, ou seja, a cidade b tem os mesmos parâmetros de Hubei.

Inicialmente, vamos observar o efeito que a taxa de quarentena de infectados k provoca nas curvas respostas do modelo, relativas ao número de infectados. Na Figura 25 são mostradas simulações para realçar este efeito. Foram utilizados dois percentuais de circulação, equivalentes a 100% ($p = 1$, Figura 25a) e 85% ($p = 0.85$, Figura 25b). As curvas foram obtidas para k variando de 0 a 0.08. Podemos perceber que as amplitudes nos picos diminuem e os instantes de pico são deslocados para a direita, quando o valor do k é ampliado. Esta figura evidencia a importância da realização dos testes em massa, a fim de conseguir identificar infectados (inclusive assintomáticos) e colocá-los em quarentena. A análise destas duas figuras que compõem a Figura 25 permite ainda concluir que o efeito conjunto da redução do percentual de circulação (p) e do aumento da taxa de quarentena de infectados (k), ambos contribuem para o esforço de controlar a epidemia, diminuindo as amplitudes nos picos e deslocando os instantes de pico para a direita. Este resultado já era esperado, pois as análises feitas na seção anterior já tinham comprovado que ambos os efeitos, diminuição do p e ampliação do k, contribuem para a redução do índice de reprodução basal efetivo $\bar{R}_0(t)$.

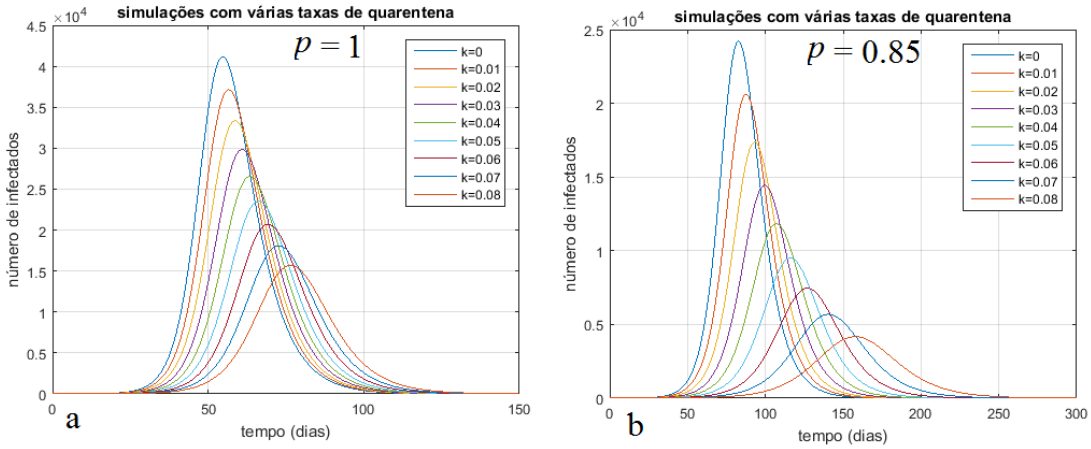


Figura 25. Simulações que realçam o efeito de considerarmos parte dos infectados em quarentena.

A Figura 26 mostra curvas com o número de infectados, em simulações considerando diversos percentuais de circulação. Na Figura 26a consideramos $k = 0$ e na Figura 26b $k = 0.02$. As variações no percentual de circulação foram de 1 a 0.4, com decremento de 0.15. Assim como nos casos das simulações anteriores, percebemos que as amplitudes nos picos diminuem com o decréscimo do percentual de circulação, bem como os instantes de pico são deslocados para a direita. Este é o efeito que tem sido divulgado nas mídias como achatamento da curva de contaminação. Novamente podemos perceber que o efeito conjunto relativo à diminuição do p e a ampliação do k potencializa o controle epidêmico, promovendo diminuição das amplitudes do número de infectados e deslocando os instantes de pico para a direita.

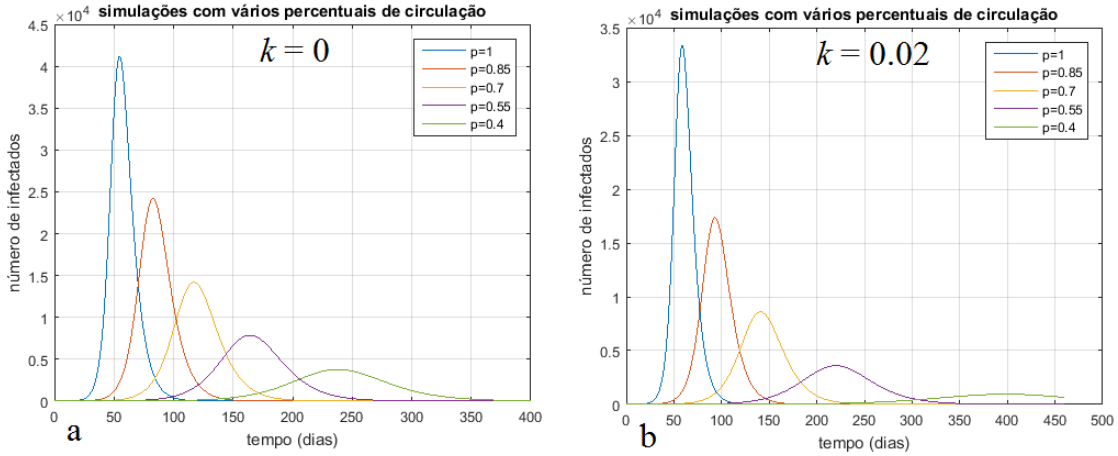


Figura 26. Simulações para diversos percentuais de circulação.

A exemplo do que foi feito no capítulo anterior, no presente capítulo também realizamos um teste ressaltando a captura da dinâmica na zona de aderência.

Reescrevemos as condições para a ocorrência do fenômeno, para torná-las compatível com as definições feitas na seção anterior. Podemos supor que o fenômeno de aderência acontecerá se:

$i)$ O número de casos $C = (I + R)$ durante o intervalo Δt_{lim} crescer abaixo de um determinado valor C_{lim}, ou seja, se $\frac{C_f - C_i}{\Delta t_{lim}} < C_{lim}$, com C_i e C_f equivalentes aos números de casos inicial e final do intervalo Δt_{lim};

$ii)$ $\bar{R}_0(t) = \frac{\beta(p, R_{0b}, \gamma_b)}{\gamma_b + k} < 1.035$;

$iii)$ O número de infectados for menor do que um determinado limite: $I < I_{lim}$.

Para a cidade hipotética de $N_h = 211000$ habitantes utilizada nas simulações desta seção, idealizamos a seguinte simulação: vetor de estado inicial $[S; I; R] = [N_h - 20; 20; 0]$; percentual de circulação inicial $p = 0.85$; percentual de circulação de controle $p_{cont} = 0.15$; instante de aplicação do controle $t_{cont} = 20\, dias$; intervalo de tempo de efetivação do controle $\Delta t_{cont} = 5\, dias$. Portanto, o percentual de circulação inicial era de 85% e passou a 15% quando da aplicação do controle no vigésimo dia, com intervalo de tempo para efetivação igual a 5 dias. Consideramos a taxa de quarentena de infectados $k = 0.06$. Os resultados desta simulação podem ser vistos na Figura 27. Quando o controle é aplicado, há um ponto de inflexão na curva do número de casos, que tende a se estabilizar em um valor constante (Figura 27a). Chamamos a atenção para o modo de aderência, que começou em $t = 98\, dias$. A partir desse instante, a curva do número de infectados vai rapidamente para zero (Figura 27b): a dinâmica entra no modo de aderência com a artificial modificação do percentual de circulação para 2% (Figura 27c). O índice de reprodução basal efetivo era inicialmente $\bar{R}_0(t) = 1.34$ e precisou de 5 dias para se tornar 0.805, quando o controle foi ativado em $t_{cont} = 20\, dias$, passando a 0.33 quando a dinâmica entra no modo de aderência em $t = 98\, dias$ (Figura 27d). Ressaltamos que esse modo de aderência só aconteceu após as três condições citadas anteriormente serem satisfeitas, quando o número de infectados se tornou inferior a cinco, no final do ciclo epidêmico. Fica evidente que este não é um fenômeno significativo do ponto de vista dinâmico e talvez seja esta a razão do mesmo não ter sido explorado, no meio científico, até o presente momento. Ressaltamos ainda que, no item ii das condições de ocorrência da aderência, adotamos um limite mais próximo de 1 do que o adotado no capítulo anterior.

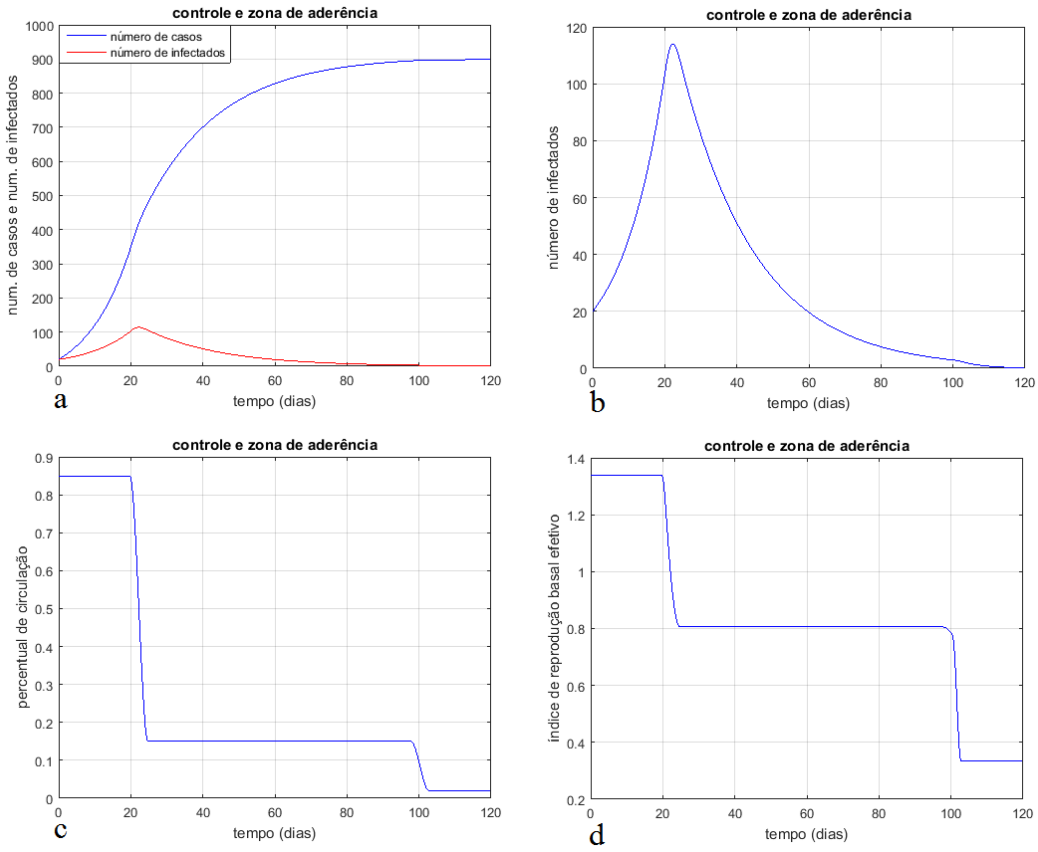


Figura 27. Simulações que ressaltam o controle e a ocorrência do fenômeno de aderência (utilizamos $k = 0.06$).

A quase totalidade das cidades brasileiras está praticando um controle a partir de um percentual de circulação no intervalo: $0.4 \leq p_{cont} \leq 0.6$. Isto configura o que chamamos de controle parcial. Na Figura 28 podemos visualizar os resultados de uma simulação com $p_{cont} = 0.45$, aplicado em $t_{cont} = 20\ dias$ e considerando $\Delta t_{cont} = 5\ dias$. A condição inicial foi idêntica à da simulação anterior: $[S; I; R] = [N_h - 20; 20; 0]$. Consideramos uma taxa de quarentena de infectados $k = 0.06$ e ainda, antes da aplicação do controle havia 100% de circulação livre $(p = 1)$. Observamos que quando o controle é ativado, há um ponto de inflexão na curva do número de casos, que passa a crescer com aceleração negativa. Antes da aplicação do controle, tínhamos $\bar{R}_0(t) = 1.692$ que, após a aplicação do controle, passou a ser $\bar{R}_0(t) = 0.95$. Este valor do $\bar{R}_0(t)$ abaixo mas muito próximo de 1 provoca uma lenta desaceleração na curva do número de casos e, desta forma, são necessários oito meses até que a epidemia chegue ao seu fim. Na realidade, conforme será visto no Capítulo 6, o percentual de circulação de controle (p_{cont}) não é constante, mas oscila entre 0.4 e 0.6 na maioria das cidades, chegando a ficar acima de 0.6 em alguns casos. Esta é a principal razão de vivenciarmos a situação atual (julho/2020), na qual as epidemias localizadas se arrastam há quatro meses com desaceleração lenta na maioria das cidades.

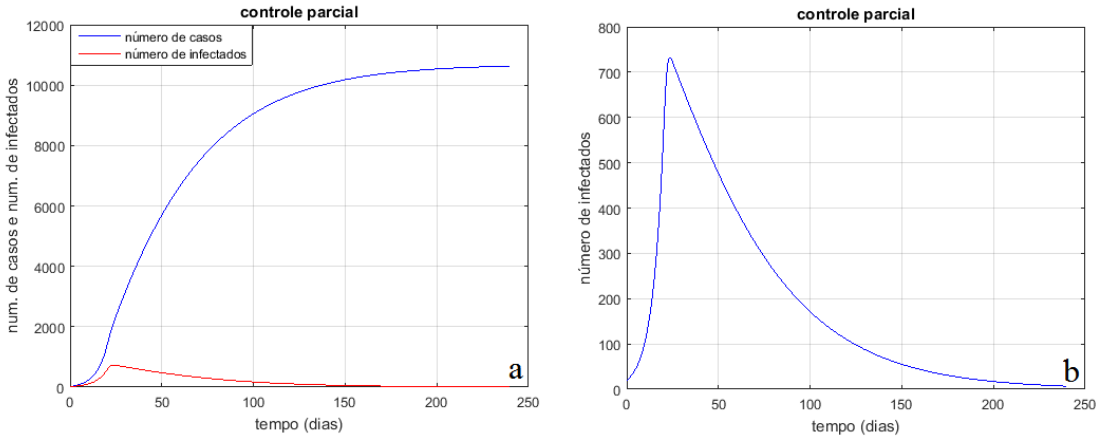


Figura 28. Simulação com a aplicação do controle de circulação em $t_{cont} = 20\ dias$: p era 1 passou a ser 0.45 durante $\Delta t_{cont} = 5\ dias$; utilzamos $k = 0.06$.

4.3 Simulações e Dados Reais de Cidades Brasileiras

O objetivo desta seção é confrontar simulações a dados reais, de forma a mostrar o desempenho do modelo em reproduzir a evolução epidêmica real. Sete capitais brasileiras foram escolhidas para a validação dos resultados de simulações: Fortaleza, Manaus, Porto Alegre, Recife, Rio de Janeiro, São Luís e São Paulo. Os dados reais destas cidades foram obtidos do site Brasil.io (https://brasil.io/dataset/covid19/caso_full/). As figuras foram geradas com o mesmo padrão, mostrando o número de casos reais, o número de casos previsto pelo modelo e a curva de infectados do modelo. Além dessa confrontação entre simulações e dados reais, há ainda a confrontação entre o percentual de circulação previsto pelo modelo e o percentual de circulação realmente praticado, dado obtido do site inloco (https://mapabrasileirodacovid.inloco.com.br/pt/), que usa informações a partir de telefonia móvel. Para facilitar a explicação dos resultados, adotamos as variáveis P_{mod} e P_{real}, significando a média dos dez últimos valores dos percentuais de circulação do modelo e real, respectivamente. O procedimento para a identificação dos parâmetros do modelo foi por tentativa e erro, a partir da confrontação entre dados reais e simulações. Nos primeiros dias da epidemia, quando na maioria das cidades ainda não havia restrições de circulação, identificamos os parâmetros R_{0b} e k para cada cidade, sendo R_{0b} o valor do índice de reprodução basal identificado para a epidemia na cidade b em questão. Optamos por expressar o R_{0b} em função do R_0 identificado para Hubei, conforme explicitado no título de cada figura. Posteriormente, quando as medidas de restrição de circulação começam a entrar em vigor nas cidades, foram determinados os valores de p e consequentemente de $\beta(p, R_{0b}, \gamma_b)$ (usando o algoritmo (7)) que melhor permitiram aproximar as respostas do

modelo aos dados reais. Adotamos $\gamma_b = \gamma$, ou seja, consideramos que todas as cidades têm a mesma taxa de recuperação de Hubei, equivalente a $\gamma_b = \gamma = 1/T_{inf} = 1/5.5 = 0.1818$. Assim como no capítulo anterior, no presente capítulo, a tentativa de aproximar simulações dos dados reais foi realizada identificando o p e consequentemente o $\beta(p, R_{0b}, \gamma_b)$ por tentativa e erro e assumimos que, por enquanto, este é o único parâmetro que varia.

Na Figura 29a são mostrados resultados para Fortaleza, cujos dados disponíveis vão do dia 16/03/2020 ao dia 18/05/2020. A identificação do p e consequentemente do $\beta(p, R_{0b}, \gamma_b)$, por tentativa e erro, considerou dados até o ponto indicado na figura, correspondente ao dia 08/05/2020. Nesta figura, a partir do dia 08/05/2020, a resposta do modelo seria uma previsão para um futuro próximo. Podemos perceber um leve descolamento da curva número de casos do modelo com relação aos dados reais, indicando uma possível pequena redução do número de casos reais, cuja provável razão seja a intensificação da limitação de circulação. Na Figura 29b vemos resultados com uma identificação paramétrica que promove um ajuste fino do modelo com os dados reais. Neste caso, o modelo reproduz até as flutuações de curto período. Porém, na sequência dos testes com dados reais e simulações será feita a opção por um ajuste médio, o qual representa melhor o fenômeno no mais longo período. Este é o caso da Figura 30, também para Fortaleza. É perceptível que o modelo reproduz bem os dados reais. Entretanto, percebemos que o percentual final de circulação (média dos últimos dez dias) previsto pelo modelo foi $P_{mod} = 25\%$, enquanto o percentual final real (identificado via telefonia móvel) foi $P_{real} = 43.5\%$. Porém, os dados de circulação real disponibilizados no site são para todo o estado, de forma que Fortaleza pode estar com uma circulação real menor após a intensificação da limitação de circulação. De qualquer modo, não há notícias de que Fortaleza tenha tido um P_{real} menor do que 40%. Esta diferença entre P_{real} e P_{mod} pode ser um indício de que outras medidas de prevenção, tais como o uso de máscaras e o isolamento social em lugares públicos, devem estar sendo assimiladas na variação do percentual de circulação p, único parâmetro deixado livre para variar, nas simulações do presente capítulo.

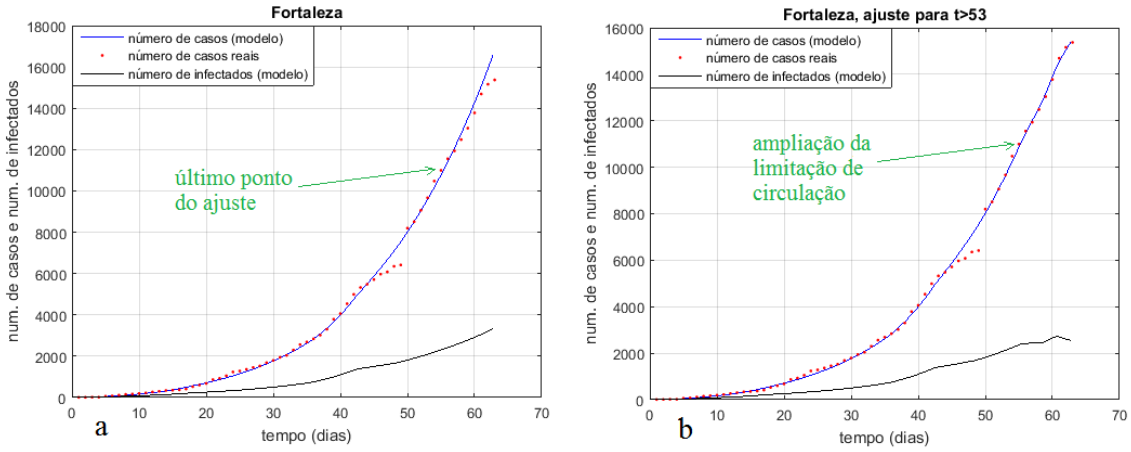


Figura 29. Simulações usando $k = 0.04$; $R_{0b} = 1.17R_0$; a) Simulação e dados reais: ajuste usando os dados até $t = 53\ dias$; $P_{mod} = 48\%$; b) incluindo ajuste fino para $t > 53\ dias$.

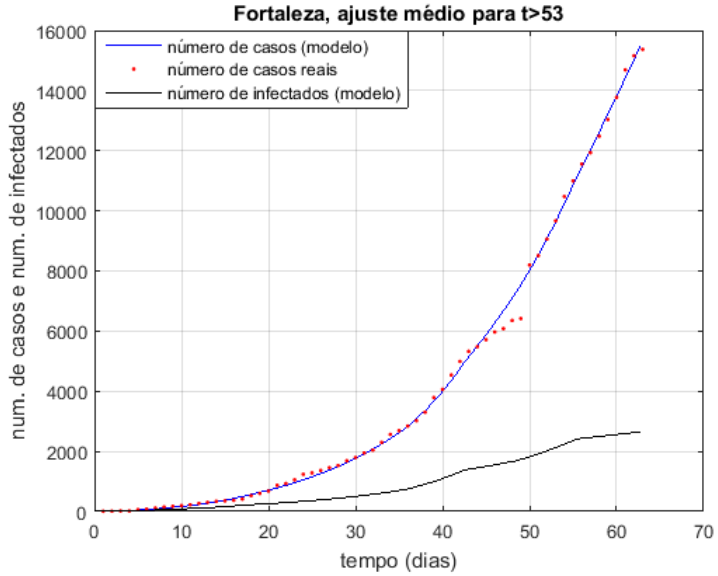


Figura 30. Simulação e dados reais para Fortaleza, com os parâmetros do modelo identificados para ajuste médio: $k = 0.04$; $R_{0b} = 1.17R_0$; $P_{mod} = 25\%$ e $P_{real} = 43.5\%$ (valores médios, últimos 10 dias).

Na Figura 31 mostramos simulação e dados reais para Manaus. Conforme já explicado anteriormente, a opção foi por um ajuste médio de parâmetros, de forma que o modelo não segue flutuações de curto período nos dados reais. Percebemos que o modelo reproduz bem o comportamento médio da curva de casos reais, sendo o $P_{mod} = 45\%$ e o $P_{real} = 55\%$. Como na maioria das grandes cidades brasileiras, a epidemia ainda estava evoluindo de forma crescente em Manaus até o final dos dados reais, em 17/05/2020. O dia 1 em Manaus é 13/03/2020.

Na Figura 32 mostramos simulação e dados reais para Porto Alegre, cidade que chama a atenção pelo baixo número de casos, comparado a outras grandes cidades brasileiras. Em Porto Alegre, os dados vão de 11/03/2020 (dia 1) a 17/05/2020. Observamos que a taxa de quarentena de infectados identificada foi $k = 0.08$, equivalente ao dobro das taxas identificadas para outras cidades. É importante notar que essa taxa engloba também atitudes preventivas que diminuem o contato de infectados com suscetíveis, tais como o uso de máscaras e o distanciamento social em lugares públicos. Conforme explicado anteriormente, a ampliação do k provoca diminuição do $\bar{R}_0(t)$, sendo esta a razão da atenuação do número de casos e a consequente inversão no crescimento do número de infectados. Para Porto Alegre, P_{mod} coincidiu com o P_{real}, ambos iguais a 55%.

Na Figura 33 mostramos simulação e dados reais para Recife (dia 1 em 12/03/2020 e dia final em 17/05/2020). Novamente, o modelo reproduziu bem o comportamento médio da dinâmica real. Neste caso, $P_{mod} = 55\%$ e $P_{real} = 55\%$. Simulação e dados reais para a cidade do Rio de Janeiro (dia 1 em 06/03/2020 e dia final em 17/05/2020) são mostrados na Figura 34, com $P_{mod} = 42\%$ e $P_{real} = 50\%$. Percebemos que o modelo reproduziu bem os dados reais e este comportamento se repetiu para todas as confrontações entre modelo e dados reais: São Luís (dia 1 em 20/03/2020 e dia final em 17/05/2020), Figura 35, com $P_{mod} = 40\%$ e $P_{real} = 55\%$; São Paulo (dia 1 em 25/02/2020 e dia final em 17/05/2020), Figura 36, com $P_{mod} = 54\%$ e $P_{real} = 55\%$.

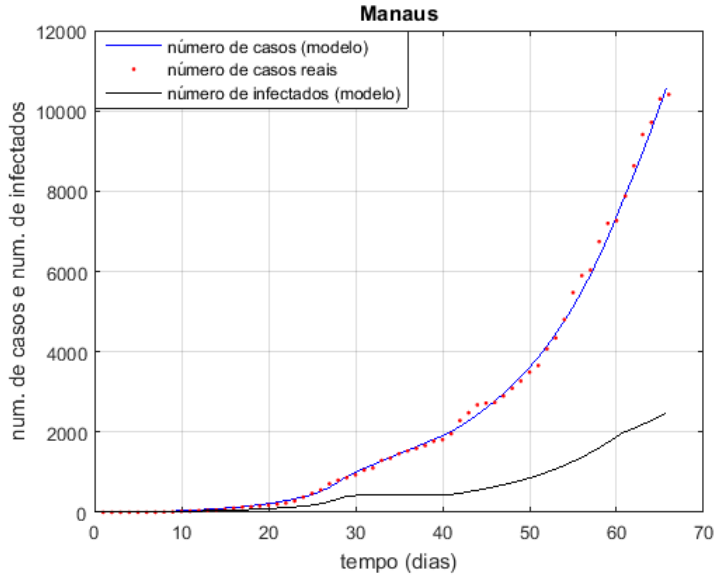


Figura 31 Simulação e dados reais para Manaus, com os parâmetros do modelo identificados para ajuste médio: $k = 0.028$; $R_{0b} = 1.11R_0$; $P_{mod} = 45\%$ e $P_{real} = 55\%$ (valores médios, últimos 10 dias).

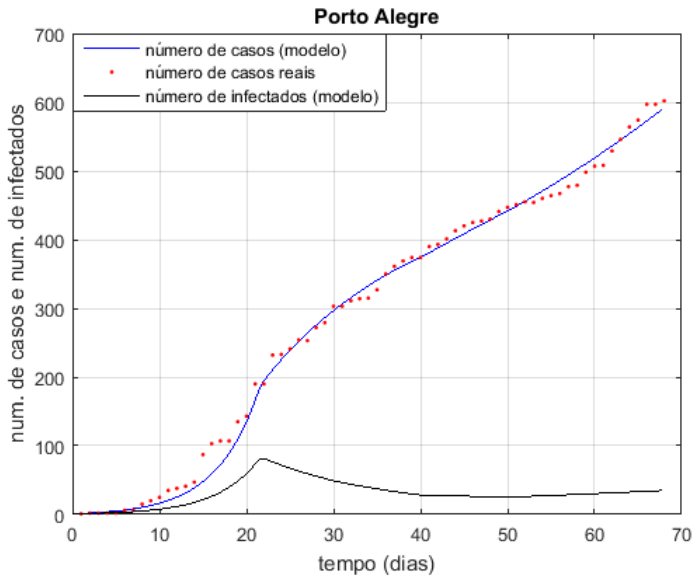


Figura 32. Simulação e dados reais para Porto Alegre, com os parâmetros do modelo identificados para ajuste médio: $k = 0.08$; $R_{0b} = 1.14R_0$; $P_{mod} = 55\%$ e $P_{real} = 55\%$ (valores médios, últimos 10 dias).

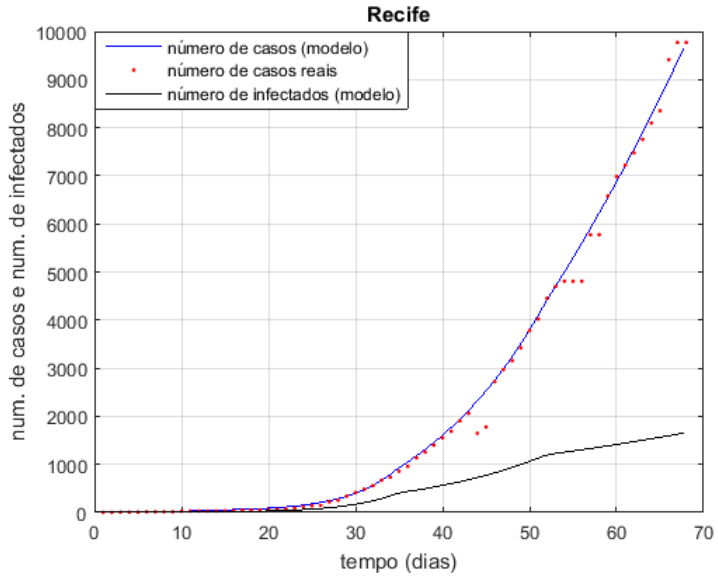


Figura 33 Simulação e dados reais para Recife, com os parâmetros do modelo identificados para ajuste médio: $k = 0.033$; $R_{0b} = 0.96R_0$; $P_{mod} = 55\%$ e $P_{real} = 55\%$ (valores médios, últimos 10 dias).

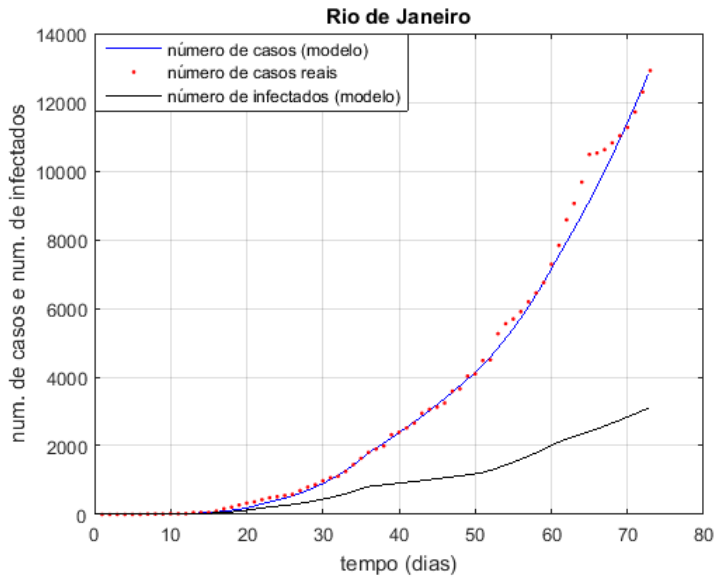


Figura 34. Simulação e dados reais para a cidade do Rio de Janeiro, com os parâmetros do modelo identificados para ajuste médio: $k = 0.04$; $R_{0b} = 1.13R_0$; $P_{mod} = 42\%$ e $P_{real} = 50\%$ (valores médios, últimos 10 dias).

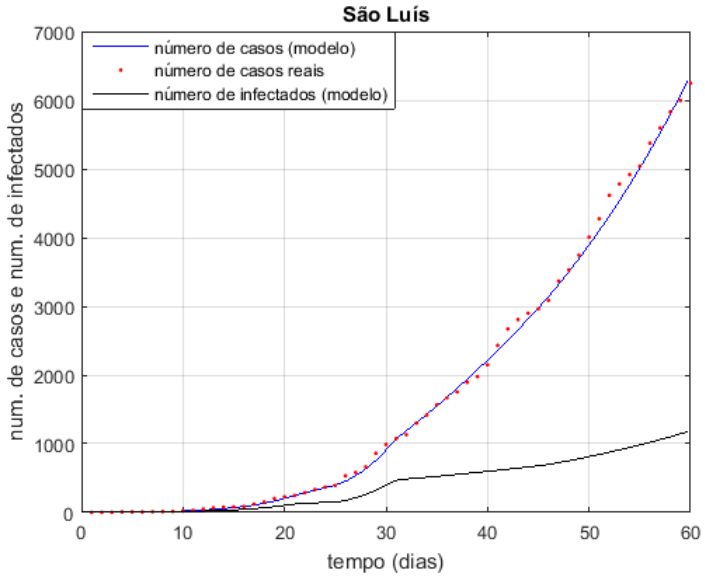


Figura 35. Simulação e dados reais para a cidade do São Luís, com os parâmetros do modelo identificados para ajuste médio: $k = 0.03$; $R_{0b} = 1.12R_0$; $P_{mod} = 40\%$ e $P_{real} = 55\%$ (valores médios, últimos 10 dias).

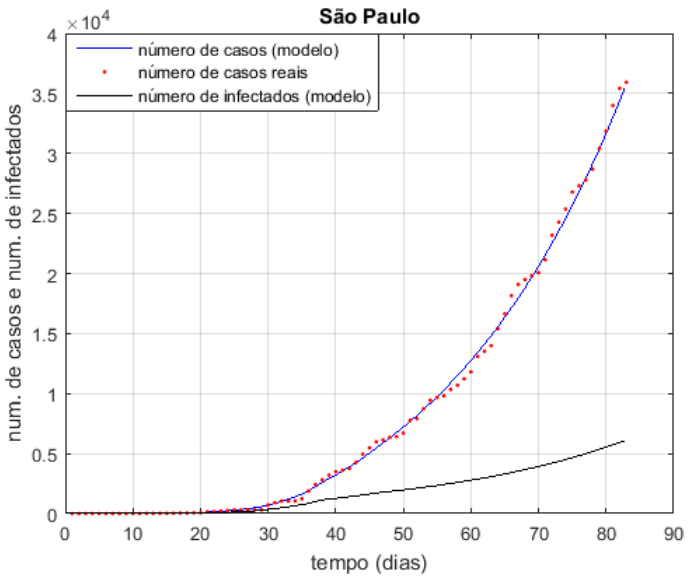


Figura 36. Simulação e dados reais para a cidade de São Paulo, com os parâmetros do modelo identificados para ajuste médio: $k = 0.02$; $R_{0b} = 0.97R_0$; $P_{mod} = 54\%$ e $P_{real} = 55\%$ (valores médios, últimos 10 dias).

A Figura 37 mostra as cidades e os seus respectivos percentuais de circulação médios, relativos aos últimos dez dias de dados reais e dados simulados. Ao lado da tabela vemos estes valores numéricos na forma gráfica. Percebemos que os percentuais identificados com o modelo foram menores em média do que os reais observados por telefonia móvel. A principal explicação para essa diferença reside no fato de apenas um parâmetro estar livre para variar, justamente o percentual de circulação p. São vários os efeitos dinâmicos que contribuem para mitigar uma epidemia, diminuindo o índice de reprodução basal efetivo. Como os outros parâmetros são considerados constantes, o p acaba assimilando esses outros efeitos dinâmicos.

Percentuais médios de circulação

Cidade	P_{mod} (%)	P_{real} (%)
1. Fortaleza	25	43.5
2. Manaus	45	55
3. Porto Alegre	55	55
4. Recife	55	55
5. Rio de Janeiro	42	50
6. São Luís	40	55
7. São Paulo	54	55

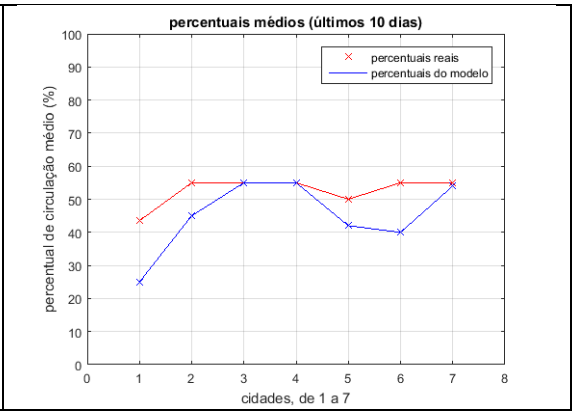


Figura 37. Percentuais médios de circulação reais e previstos pelo modelo (média dos últimos 10 dias de dados).

4.4 Comentários Finais

O estudo apresentado neste capítulo reforça a tese de que um modelo dinâmico de estrutura simples pode reproduzir bem a dinâmica evolutiva da COVID-19 em uma determinada cidade ou região, desde que, evidentemente, os parâmetros sejam bem estimados. Enfocamos principalmente os efeitos dinâmicos decorrentes da taxa de quarentena de infectados, agindo para reduzir o índice de reprodução basal efetivo e assim, contribuindo significativamente para o controle epidêmico. Utilizamos dados reais de algumas cidades brasileiras e verificamos que, em todos os casos, há um período inicial no qual não há efeitos dinâmicos decorrentes de medidas de prevenção, principalmente decorrentes do isolamento social. Neste período inicial, delimitado pelo instante de corte t_c, parâmetros constantes permitem boa aproximação entre simulação e dados reais. Este fato já havia sido detectado no capítulo anterior, quando estudamos os dados de cidades do exterior (China, Espanha e USA). O procedimento adotado foi o seguinte: para $t \leq t_c$, foram identificados os parâmetros R_{0b} e k de cada cidade e estes foram mantidos constantes para o restante do tempo de simulação. A taxa de recuperação também foi mantida constante e igual à identificada para Hubei: utilizamos $\gamma_b = \gamma$, para todas as cidades. Para $t \leq t_c$ utilizamos $p = 1$, ou seja, como não havia medida de prevenção antes do instante de corte, o percentual de circulação estava em 100%. Para $t > t_c$, foram identificadas, por tentativa e erro, variações do p a fim de aproximar simulações e dados reais. De fato, com essa estratégia de atuação, verificamos que o modelo reproduziu bem os dados reais das diversas cidades que foram escolhidas para teste no presente capítulo.

Foi possível detectar que dois problemas ainda permanecem. O primeiro deles é mais ou menos óbvio e já era esperado, tratando-se da dificuldade que surge na identificação paramétrica, utilizando a estratégia de tentativa e erro. Torna-se uma tarefa árdua, modificar o parâmetro, realizar a simulação, verificar o erro e, se necessário, repetir todo o processo, até conseguir baixar o erro a um nível aceitável. Outro problema remanescente é decorrente da metodologia empregada, que permite apenas um único parâmetro variar no tempo a fim de que o modelo se ajuste aos dados reais, precisamente o percentual de circulação p. Sendo o único responsável por adaptar o modelo aos dados reais, este parâmetro acaba por assimilar outros efeitos dinâmicos decorrentes de medidas de prevenção, tais como o uso obrigatório de máscaras e o distanciamento social. Isto tende a fazer com que o percentual de circulação fique menor do que o realmente praticado nas cidades. No próximo capítulo tentaremos atacar estes problemas, permitindo que mais parâmetros também variem no tempo e ainda, introduzindo na estrutura de modelagem uma

técnica de identificação paramétrica própria também para sistemas não lineares, baseada na minimização do erro entre simulações e dados reais.

Referências

GANYANI T., Kremer C., Chen D., Torneri A., Faes C., Wallinga J. and Hens N.. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data. Euro Surveill. 2020 Apr 30; 25(17): 2000257.

GOMES S. C. P., Monteiro I. O., Rocha C. R. Modelagem dinâmica da COVID-19 com aplicação a algumas cidades brasileiras. Revista Thema, http://dx.doi.org/10.15536/thema.V18.Especial.2020.1-25.1793. Especial, 2020a.

NISHIURA H., Lintona N. M., Akhmetzhanov A. R. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases, 2020.

Capítulo V

Terceiro Estudo – Identificação Paramétrica

5.1 Introdução

O presente capítulo aborda a identificação paramétrica para o modelo dinâmico MdSIR, aplicado à epidemia de COVID-19, localizada em uma determinada cidade ou região. Foram identificados três parâmetros que variam ao longo do tempo: o índice de reprodução basal (R_{0b}); a taxa de quarentena de infectados (k); o percentual de circulação (p). Portanto, a partir do presente capítulo, outros dois parâmetros são deixados livres para variar no tempo, diferentemente do capítulo anterior, quando apenas o percentual de circulação podia variar no tempo. A principal consequência disso é o fato desses dois outros parâmetros assimilarem efeitos dinâmicos oriundos de outras ações de prevenção, tais como o uso obrigatório de máscaras e o distanciamento social em lugares públicos, deixando o parâmetro percentual de circulação para assimilar especificamente as variações no isolamento social da cidade em estudo. O MdSIR pode ser reescrito com a indicação dos parâmetros que variam no tempo, na forma:

$$\dot{S} = \frac{-\beta\left(\gamma_b, R_{0b}(t), p(t)\right)SI}{N_h}$$

$$\dot{I} = \frac{\beta\left(\gamma_b, R_{0b}(t), p(t)\right)SI}{N_h} - [\gamma_b + k(t)]I \tag{12}$$

$$\dot{R} = [\gamma_b + k(t)]I$$

É importante lembrar que, ao considerarmos variações paramétricas, o índice de reprodução basal efetivo passa a ser também função do tempo, sendo calculado a partir da seguinte equação:

$$\bar{R}_0(t) = \frac{\beta\left(\gamma_b, R_{0b}(t), p(t)\right)}{[\gamma_b + k(t)]} \tag{13}$$

O parâmetro taxa de crescimento (β) depende agora de três parâmetros: dois que variam no tempo $\left(R_{0b}(t) \text{ e } p(t)\right)$ e um que não varia no tempo (γ_b), mas que pode ser diferente para cidades diferentes. Observamos ainda que a taxa de recuperação efetiva corresponde a $\bar{\gamma}(t) = \gamma_b + k(t)$, de forma que admitir a variação do $k(t)$ significa variar também o $\bar{\gamma}(t)$, porém mantendo um valor mínimo constante equivalente ao γ_b, já que $k(t) \geq 0$.

Permitir que três parâmetros variem no tempo amplia o potencial do modelo para se adaptar à realidade do fenômeno que tenta reproduzir. Utilizamos o método de Nelder-Mead Simplex (LAGARIAS *et al*. 1998) na identificação paramétrica, procedimento este utilizado considerando os dados reais de diversas cidades brasileiras. Este algoritmo se baseia na minimização do erro entre dados reais e simulados: o erro converge para zero enquanto os parâmetros convergem para os seus respectivos valores identificados.

Os dados reais são divididos em dois períodos, sendo o primeiro deles relativo ao início da epidemia, quando ainda não havia limitação de circulação da população na cidade em estudo. No segundo período a cidade já está com medidas preventivas do tipo isolamento social. Conforme já explicado nos capítulos anteriores, o tempo que separa estes dois períodos foi denominado instante de corte (t_c) e é identificado a partir da observação dos dados reais, quando o crescimento do número de casos deixa de ser exponencial. Os parâmetros do modelo são identificados e mantidos constantes, para $t \leq t_c$. Como para este primeiro intervalo ainda não há medidas de prevenção, consideramos um percentual de circulação equivalente a 100%, ou seja, $p = 1$. Após este primeiro período, os parâmetros são identificados a cada novo dia de dados reais. A estrutura de identificação paramétrica e simulação foi testada em diversas cidades brasileiras. No presente capítulo são apresentados os resultados para as seguintes cidades: Belém (PA), Brasília (DF), Fortaleza (CE), Pelotas (RS), Porto Alegre (RS), Rio de Janeiro (RJ), Salvador (BA), São Luís (MA) e São Paulo (SP). Em síntese, pode ser dito que os resultados são considerados bons. O modelo com a adaptação paramétrica específica para cada cidade reproduziu bem os dados reais.

5.2 Identificação Paramétrica e Simulações

A identificação paramétrica é um problema crítico em modelagem dinâmica em razão de alguns fatores correlatos à qualidade dos dados experimentais disponíveis, bem como a frequentes problemas de convergência com os métodos numéricos utilizados. A identificação paramétrica geralmente resume-se a um problema de otimização de alguma função objetivo vinculada à minimização de erros entre simulação e dados reais. Pode ser

dito que uma boa identificação depende fundamentalmente de uma boa observação, ou seja, se a qualidade dos dados observados é boa, aumentam as chances de se conseguir uma boa identificação paramétrica. Especificamente sobre a COVID-19, existem dois principais problemas com os dados reais disponibilizados: subnotificação; retardos na informação. Quem precisa trabalhar com os dados não pode fazer nada para reduzir os efeitos da subnotificação, já que a correção deste problema passaria por uma maior testagem da população, inclusive dos assintomáticos. O retardo na informação pode provocar mudanças bruscas na variação do número acumulado de casos, ou seja, pode haver dias com poucas notificações e outros dias com maior concentração de notificações. Este último problema pode ser minimizado com a realização de um pré-tratamento nos dados, antes de submetê-los ao algoritmo de identificação paramétrica, conforme será visto a seguir.

Mostramos como exemplo os casos de São Luís e Fortaleza, vistos na Figura 38. Foram utilizados polinômios de ajuste, que seguem a tendência dos dados, mas evitando variações bruscas. As ordens dos polinômios de ajuste foram escolhidas observando-se visualmente a qualidade do ajuste, ou seja, escolhemos a menor ordem que possibilitasse um bom ajuste. Às vezes acontece do polinômio gerar valores negativos para os primeiros pontos. Para contornar este problema, forçamos os valores do polinômio a serem iguais aos dados reais nestes primeiros pontos. Isto foi necessário nos primeiros oito pontos para Fortaleza, mas não foi necessário para São Luís. Os dados reais desta e de todas as demais figuras com resultados de cidades brasileiras foram obtidos do site Brasil.io (https://brasil.io/dataset/covid19/caso_full/).

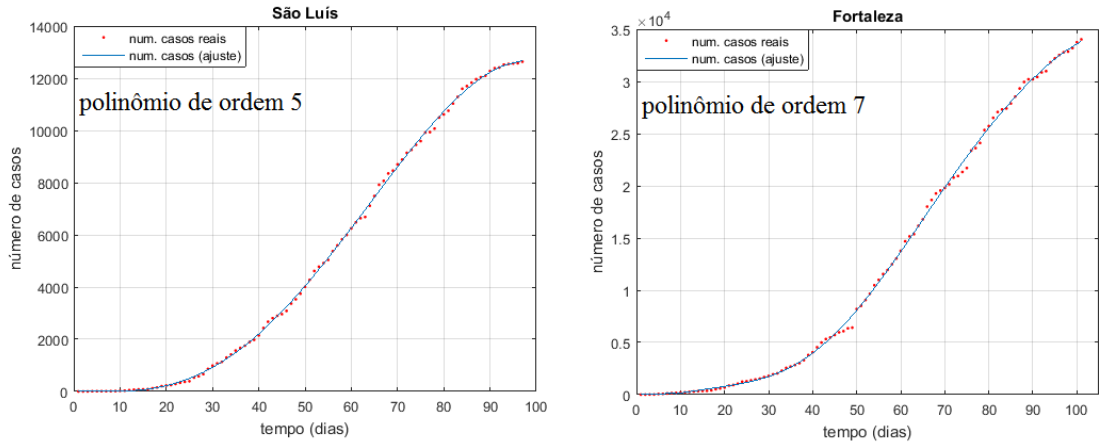


Figura 38. Número acumulado de casos: dados reais e polinômio de ajuste.

Os polinômios de ajuste foram utilizados na identificação paramétrica, em substituição aos dados originais do número de casos. Eles corrigem distorções correlatas a atrasos na informação dos dados e seguem a tendência média, negligenciando flutuações de mais alta frequência, as quais, certamente, não são oriundas de efeitos dinâmicos.

Utilizamos o Nelder-Mead Simplex (LAGARIAS *et al*. 1998) para a identificação paramétrica, programado em ambiente MATLAB. Minimizamos o erro entre o número de casos reais (y_{ri}) e o número de casos simulados (y_{si}), com $i = 1, \dots, n_p$, onde n_p corresponde ao número de pontos. Assumimos então o erro na forma: $e_i = |y_{ri} - y_{si}|$, sendo que os valores y_{ri} são obtidos a partir do polinômio de ajuste (Figura 38). O algoritmo de minimização é aplicado de forma a reduzir o erro e_i a um valor menor do que a tolerância adotada.

No início da epidemia em cada cidade não havia limitação de circulação da população, de forma que $p = 1$, até o instante no qual os governos locais começaram a impor as primeiras medidas para limitar a circulação, tais como o fechamento do comércio e a interdição das aulas nas escolas do município. Conforme adiantado anteriormente, este marco inicial das medidas de limitação de circulação é intitulado de instante de corte (t_c). A análise dos dados reais permite identificar o t_c de cada cidade. Os pontos em vermelho na Figura 39 mostram a curva crescente do número de casos reais. As derivadas em cada ponto desta curva variam até o instante de corte, quando a derivada se torna praticamente constante (aceleração nula), de forma que a curva deixa o crescimento exponencial e passa a crescer como uma reta. Este é um aspecto marcante e presente em todas as cidades analisadas.

O primeiro problema que se impõe é: conhecido o t_c, quais são os parâmetros que melhor permitem uma boa aproximação entre dados reais e simulação, no intervalo $t \leq t_c$. Há então a necessidade de identificação dos parâmetros R_{0b} e k, válidos apenas neste intervalo. Para isso, mantemos o $p = 1$ (não há limitação de circulação para $t \leq t_c$) e identificamos os parâmetros R_{0b} e k que possibilitam uma boa aproximação entre dados reais e simulação. Como os dados reais são diários, t_c é um número inteiro. O erro gerado para a utilização na estimação destes parâmetros obedece à seguinte equação:

$$e_j = \frac{1}{t_c}\sqrt{\sum_{i=1}^{t_c}(y_{ri} - y_{si})^2} \qquad (14)$$

Foram permitidas no máximo 80 iterações no processo de minimização, de forma que j varia de 1 até no máximo 80. Na maioria dos casos, o algoritmo de minimização convergiu para um erro mínimo (menor do que 0.1, precisão preestabelecida neste caso) em no máximo 20 iterações.

Na Figura 39 vemos um exemplo de simulação com os parâmetros iniciais mantidos constantes, para a cidade de Manaus. Os parâmetros identificados foram: $R_{0b} = 2.4975$ e $k = 0.028$. É importante lembrar que nesse caso, $p = 1$. Podemos perceber que, a partir do tempo $t = 16\,dias$, a simulação começa a apresentar um número de casos maior do que o real. Este descolamento entre dados reais e simulados se verifica exatamente no instante de corte t_c, que neste caso foi igual a 16.

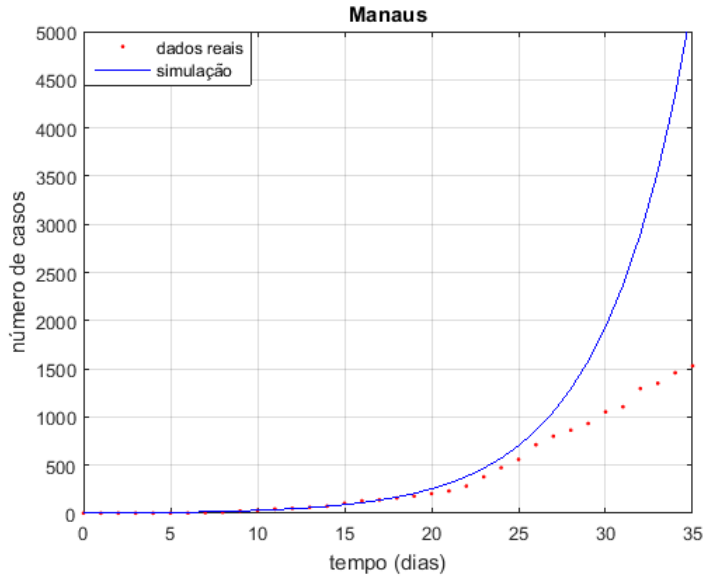


Figura 39. Simulação com os parâmetros constantes.

Conforme visto anteriormente, os parâmetros se mantêm constantes até o instante de corte t_c. Após esse tempo, são necessárias estratégias para identificá-los ponto a ponto. Foram analisadas duas situações distintas: mantendo-se constantes R_{0b} e k e variando-se $p(t)$; variando-se simultaneamente os três parâmetros $R_{0b}(t), k(t)$ e $p(t)$.

Inicialmente analisamos o primeiro caso, ou seja, mantivemos R_{0b} e k constantes (identificados para $t \leq t_c$) e, a cada novo dia de dados reais, identificamos o percentual de circulação $p(t)$ minimizando o erro $e_i = |y_{ri} - y_{si}|$, para $i > t_c$. O valor inicial atribuído a p foi 1, em todos os casos. A Figura 40 mostra diversas iterações no processo de minimização do

erro, para o caso $i = t = 43$. Esta figura evidencia a rápida convergência do processo iterativo, uma característica do método de Nelder-Mead Simplex.

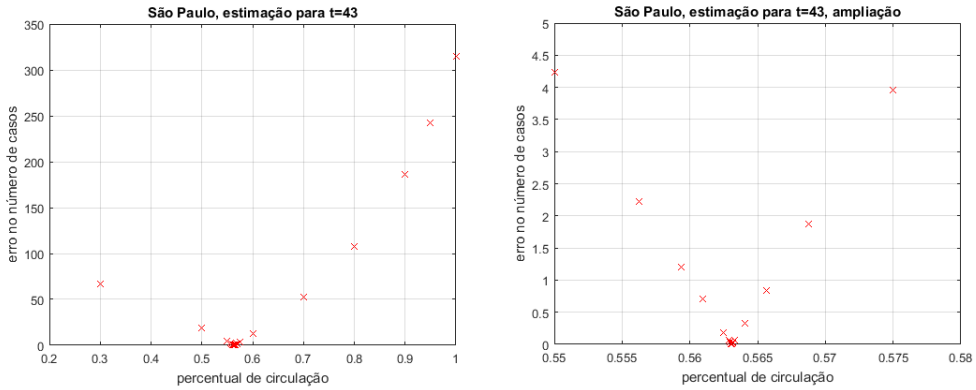


Figura 40. Diversas iterações no processo de estimação do percentual de circulação $p(t)$ (o valor que minimiza o erro foi $p(43) = 0.563$).

Uma vez que o processo de minimização é aplicado a cada um dos pontos para $t > t_c$, determinamos assim a variação do percentual de circulação $p(t)$ que força o modelo a reproduzir os dados reais, ou seja, a curva número de casos do modelo coincide com a curva polinomial que ajustou o número de casos reais. Porém, os resultados gráficos são mostrados a partir da confrontação entre simulação e número de casos reais, conforme visto na Figura 41, para a cidade de São Paulo. Esta figura mostra uma confrontação entre dados reais e simulação (Figura 41a), o índice de reprodução basal inicial (R_{0b}, Figura 41b), a taxa de quarentena de infectados (k, Figura 41c) e o percentual de circulação ($p(t)$, Figura 41d). Podemos perceber claramente que o modelo reproduziu muito bem os dados reais. Permaneceram constantes o $R_{0b} = 2.194$ e $k = 0.0212$, valores identificados na etapa anterior, para $t \leq t_c$.

Uma simulação semelhante à da Figura 41 foi realizada considerando a mesma estratégia de identificação paramétrica, agora aplicada aos dados reais da cidade de Fortaleza. Os resultados podem ser vistos na Figura 42. Neste caso, os parâmetros identificados na fase anterior (para $t \leq t_c$) e mantidos constantes foram: $R_{0b} = 2.632$ e $k = 0.056$. No caso de Fortaleza, o instante de corte foi $t_c = 8\,dias$. Podemos perceber que o modelo também reproduziu bem os dados reais. Entretanto, uma conclusão importante pode ser extraída da análise dos resultados das Figuras 41 e 42, especificamente correlata à variação do percentual de circulação $p(t)$ em ambos os casos. Na Figura 41 (São Paulo), o percentual de circulação final foi $p(120) = P_{mod} = 16.14\%$, enquanto na Figura 42 (Fortaleza) foi $p(100) = P_{mod} = 5.9\%$. De imediato, concluímos que estes percentuais de circulação estão abaixo dos praticados efetivamente nas respectivas cidades. O portal inloco

(https://mapabrasileirodacovid.inloco.com.br/pt/) divulga 62.3% e 61% para os estados de São Paulo e Ceará, respectivamente, dados do dia 24/06/2020, último dia de simulação. Mesmo admitindo a possibilidade das capitais estarem com maior isolamento social do que nos interiores dos estados, ainda assim a diferença entre o previsto em simulação e o real é muito significativa. A explicação para essa diferença está associada ao fato de que outras ações preventivas, tais como o uso obrigatório de máscara e o distanciamento social em lugares públicos, também contribuem para diminuir a velocidade de propagação da epidemia. Como foi permitida apenas a variação do parâmetro $p(t)$, as contribuições destas outras ações preventivas foram assimiladas e incorporadas nesta variação, de forma que o valor final do $p(t)$ se mostra menor do que o praticado realmente nas cidades analisadas.

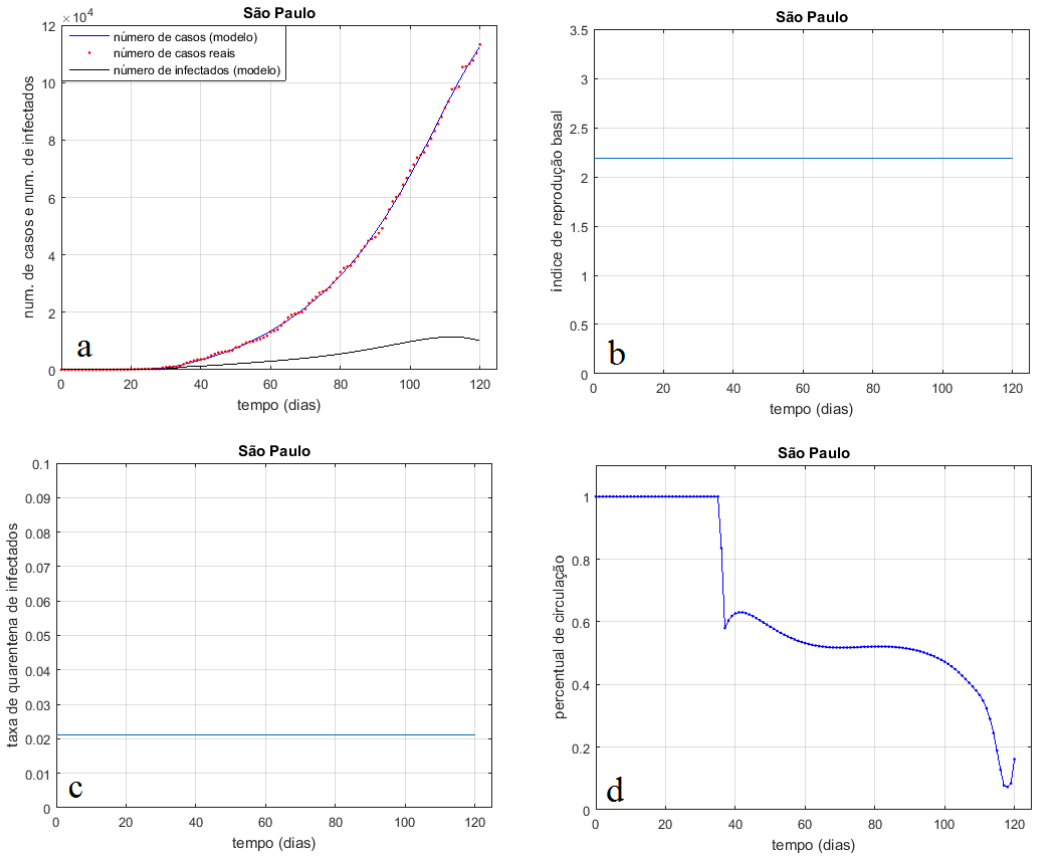


Figura 41. Simulação e dados reais para a cidade de São Paulo, de 25/02 a 24/06, mantendo-se constantes R_{0b} e k.

É importante lembrar que os resultados mostrados no capítulo anterior têm dados reais cuja data final do arquivo foi 17/05/2020. Por exemplo, nos resultados para a cidade de São Paulo, vistos na Figura 36, o arquivo conta com 83 dias de dados reais. Outra observação importante é que, no capítulo anterior, o parâmetro $p(t)$ foi identificado por tentativa e erro, de forma que o mesmo era naturalmente constante por intervalos. Já no atual capítulo, os dados reais das cidades têm data final em 24/06/2020, ou seja, com um

maior número de dias de dados reais e ainda, o parâmetro $p(t)$ varia ponto a ponto, continuamente no tempo. Assim acontece também com a variação dos demais parâmetros na continuação do presente capítulo. Evidentemente, a variação contínua do parâmetro $p(t)$ efetuada no atual capítulo deve expressar melhor a realidade contínua deste parâmetro, praticada nas cidades cujos dados serviram de testes.

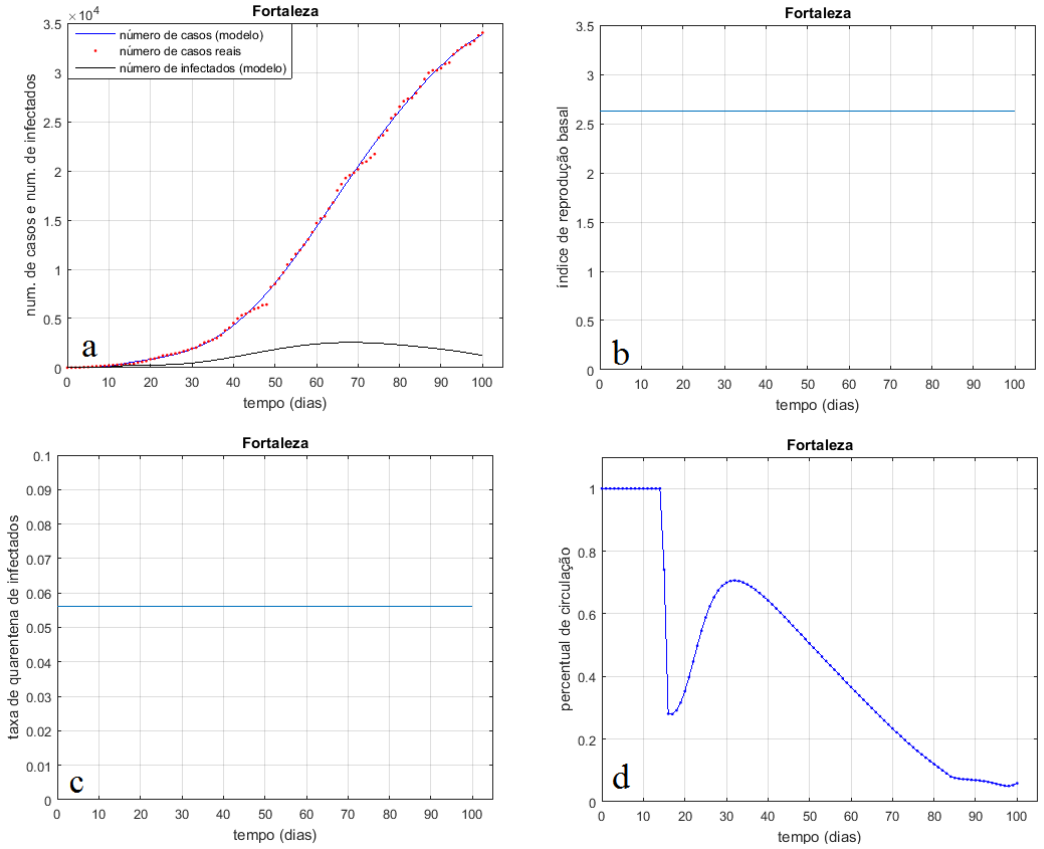


Figura 42. Simulação e dados reais para a cidade de Fortaleza, de 25/02 a 24/06, mantendo-se constantes R_{0b} e k.

O próximo passo consiste no estudo do segundo caso, para o qual consideramos, após o instante de corte $(t > t_c)$, variações simultâneas nos três parâmetros, $R_{0b}(t)$, $k(t)$ e $p(t)$. Assim como no caso anterior, o processo de minimização é aplicado individualmente a cada novo ponto com o número real de casos. Na Figura 43 mostramos o processo iterativo convergente da estimação paramétrica com minimização do erro, processo este aplicado em um único ponto, no caso, para $t = 50\ dias$. Podemos perceber que a convergência acontece com oscilações, tanto no erro quanto nos parâmetros, mas com amplitudes decrescentes. O erro tende a zero e isso significa que a resposta do modelo é muito aproximada (a precisão estabelecida foi 10^{-3}) ao padrão de comparação (valor do polinômio de ajuste).

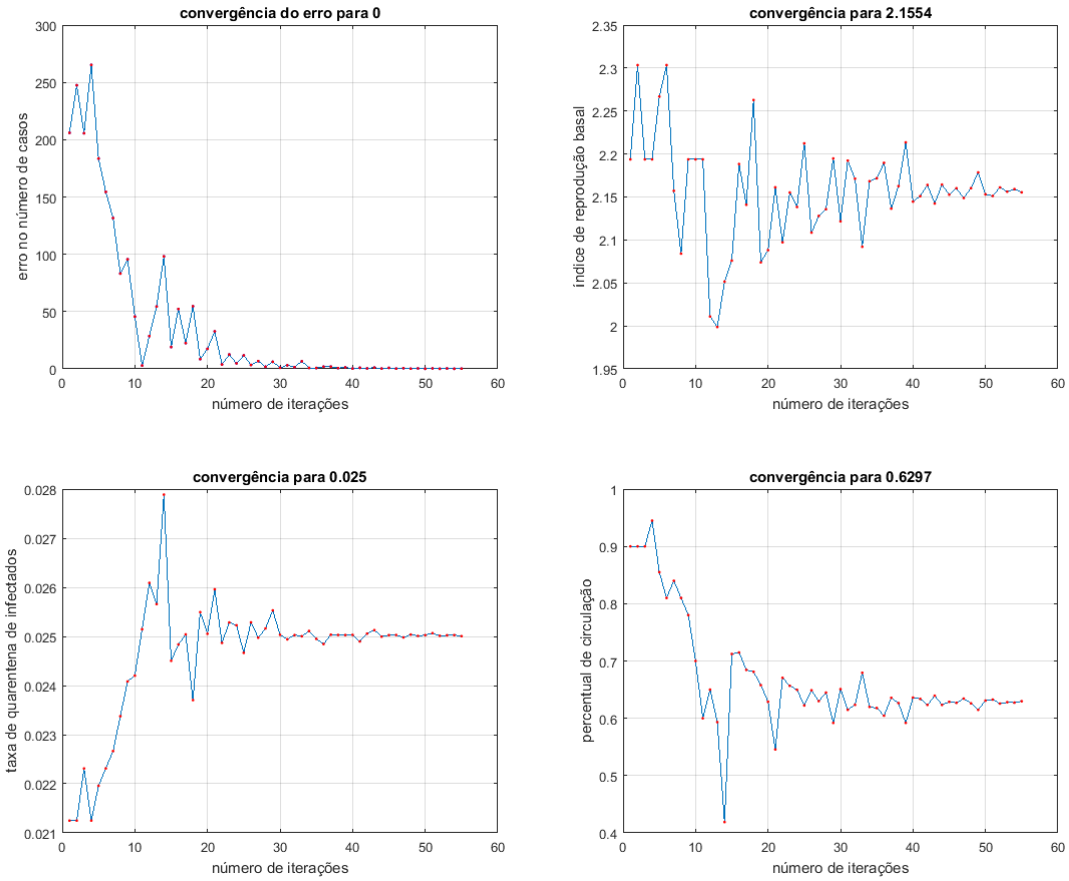


Figura 43. Cidade de São Paulo, algoritmo de minimização aplicado no tempo $t = 50\ dias$.

O processo de minimização e estimação paramétrica da Figura 43 foi repetido em todos os pontos, para os quais $t > t_c$. Portanto, os três parâmetros agora variam com o tempo: $R_0(t), k(t)$ e $p(t)$. A Figura 44 mostra resultados para a cidade de São Paulo. Nela vemos uma confrontação entre simulação e experimento, bem como as variações paramétricas ao longo do tempo. É importante comparar a Figura 44 com a Figura 41. As simulações mostram resultados praticamente idênticos (Figuras 41a e 44a), de forma que o número de casos simulado reproduz muito bem o número de casos real. Quanto aos parâmetros, na parte constante, para $t \leq t_c$, os resultados das duas figuras evidentemente são iguais. Porém, para $t > t_c$, os parâmetros na Figura 44 variam ao longo do tempo: o $R_0(t)$ diminui e o $k(t)$ aumenta, com pequenas variações, de forma que possibilitam valores mais próximos da realidade para o percentual de circulação $p(t)$. Para o estado de São Paulo, o valor do percentual de circulação no dia 24/06/2020 foi $P_{real} = 62.3\%$, enquanto que o previsto pelo modelo foi igual a $P_{mod} = 50.2\%$ (último ponto da Fig. 44d).

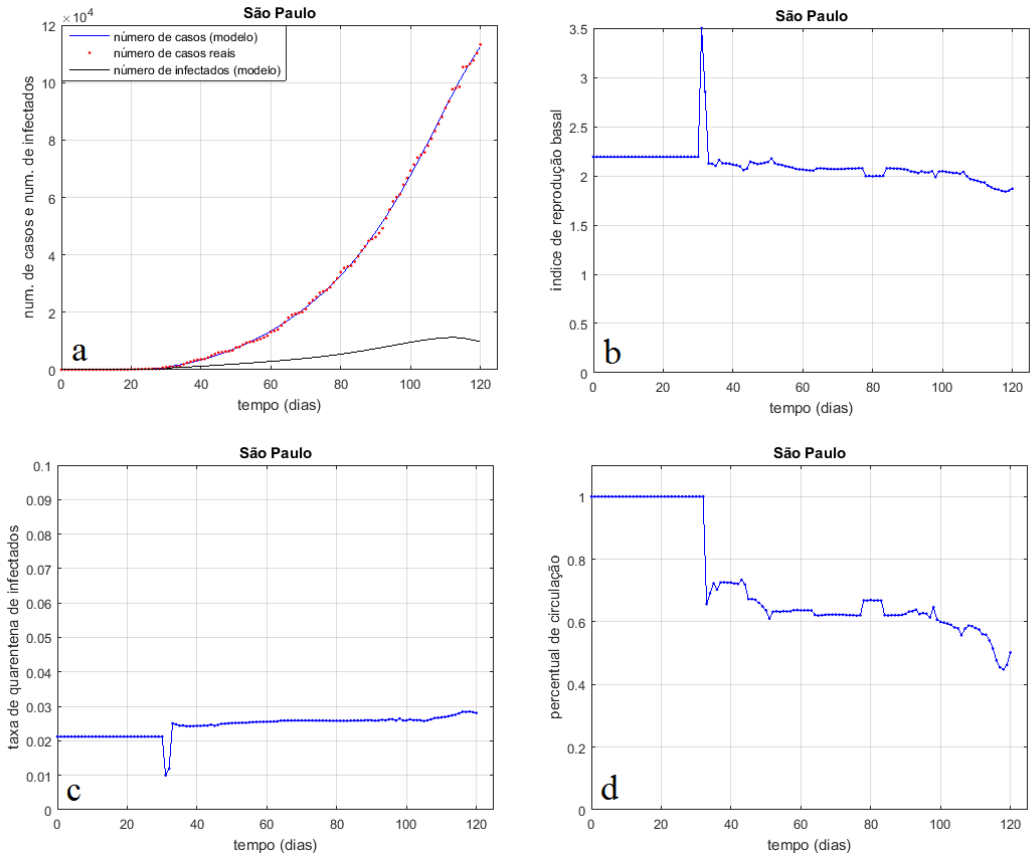


Figura 44. Simulação e dados reais para a cidade de São Paulo, de 25/02 a 24/06, com os três parâmetros variando ao longo do tempo.

Resultados semelhantes ao da Figura 44 foram obtidos para a cidade de Fortaleza, os quais podem ser vistos na Figura 45. Estes resultados devem ser comparados aos da Figura 42. Percebemos claramente que as Figuras 45a e 42a são praticamente idênticas, ou seja, o modelo com os três parâmetros variáveis também reproduz muito bem os dados reais. Isto sugere uma conclusão óbvia: diferentes valores dos parâmetros podem conduzir a um mesmo resultado. Giordano *et al.* (2020) também chegaram à mesma conclusão, em trabalho realizado utilizando dados reais da Itália. Eles usaram uma variação do modelo SIR, mas com estrutura muito mais complexa, na qual o compartimento dos recuperados é subdividido em vários outros compartimentos. O último ponto da Figura 45d, equivalente ao percentual de circulação previsto pelo modelo, foi $P_{mod} = 45\%$, enquanto que o percentual de circulação real deste último dia identificado para o estado do Ceará foi a $P_{real} = 56.7\%$. É interessante observar que o último ponto da Figura 42d foi $P_{mod} = 5.9\%$, valor bem abaixo do P_{real}. A questão que se impõe naturalmente é: por que o P_{mod} da Figura 45d se aproxima bem mais do P_{real} do que o P_{mod} da Figura 42d? Observamos na Figura 45a que há um ponto de inflexão na curva de crescimento do número de casos, indicando que um conjunto de ações preventivas atuou para diminuir a velocidade de crescimento da contaminação,

diminuição esta acentuada principalmente a partir de $t > 70\ dias$. No caso da Figura 42, só o parâmetro $p(t)$ varia, de forma que todas as ações preventivas são assimiladas nesta variação. Já na Figura 45 variam também o $R_{0b}(t)$ e o $k(t)$ e estas variações assimilam outras ações preventivas, deixando a principal ação preventiva correlata ao percentual de circulação para a variável $p(t)$. Esta questão é analisada com maiores detalhes na próxima seção, quando são introduzidos estudos com mais cidades brasileiras.

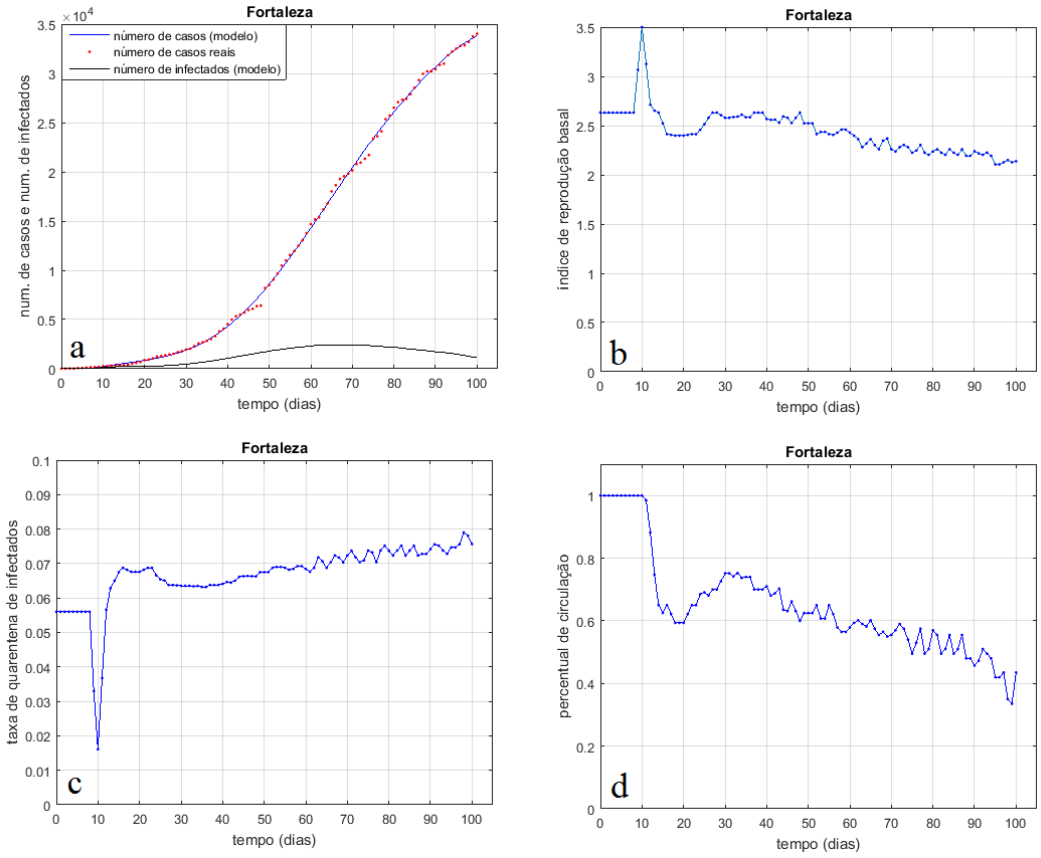


Figura 45. Simulação e dados reais para a cidade de Fortaleza, de 25/02 a 24/06, com os três parâmetros variando ao longo do tempo.

5.3 Estudo de Casos

Nesta seção, novos casos são analisados, relativos a diversas cidades brasileiras. O objetivo é testar exaustivamente a estrutura de identificação paramétrica e simulação, bem como identificar características específicas sobre a evolução epidêmica nestas cidades. Todas as simulações são realizadas com a estratégia de variação paramétrica nos três parâmetros, conforme explicado na seção anterior.

Ressaltamos mais uma vez que o $R_{0b}(t)$ que entra no cálculo da função $\beta(\gamma_b, R_{0b}(t), p(t))$ equivale ao valor para o caso em que $p = 1$, ou seja, trata-se do índice de

reprodução basal inicial, no caso da cidade não contar com restrições de circulação de pessoas. Este $R_{0b}(t)$ é identificado com o algoritmo de minimização, conforme explicado na seção anterior. O índice de reprodução basal efetivo, considerando todas as medidas preventivas, tem cálculo a partir da equação (13), deixando claro assim que este índice também varia no tempo.

Na Figura 46a vemos o $\bar{R}_0(t)$ obtido com a equação (13), para a cidade de São Paulo. Ao lado, a Figura 46b revela o $R_{0b}(t)$ identificado ao longo do tempo. É importante salientar o porquê do $\bar{R}_0(t)$ ser menor do que o $R_{0b}(t)$. Primeiro, se não houvesse nenhuma medida preventiva, inclusive com o percentual de circulação $p = 1$ (ausência de isolamento social), $k(t)$ seria nulo e da equação (13) teríamos: $\bar{R}_0(t) = \beta(\gamma_b, R_{0b}(t), 1)/\gamma_b = \gamma_b R_{0b}(t)/\gamma_b = R_{0b}(t)$ e, no caso de São Paulo, teríamos $\bar{R}_0(t) = R_{0b}(t) = 2.194$, valor constante visto nos primeiros pontos da Figura 46b. Com as medidas preventivas, o γ_b permanece constante, mas o $\beta\big(\gamma_b, R_{0b}(t), p(t)\big)$ diminui quando diminuem o $p(t)$ e/ou o $R_{0b}(t)$, conforme explicado no Capítulo III (ver também GOMES *et al*. (2020b)). Portanto, conforme mostra a equação (13), o $\bar{R}_0(t)$ diminui quando diminuem $p(t)$ e/ou o $R_{0b}(t)$, diminuindo também com o aumento do $k(t)$. A análise da Figura 46a indica que o $\bar{R}_0(t)$ final (dia 24/06/2020) para São Paulo está próximo de 1, mas ligeiramente inferior. É importante frisar que para haver um ponto de inflexão na curva de crescimento do número de casos é necessário que o $\bar{R}_0(t)$ se torne menor do que 1.

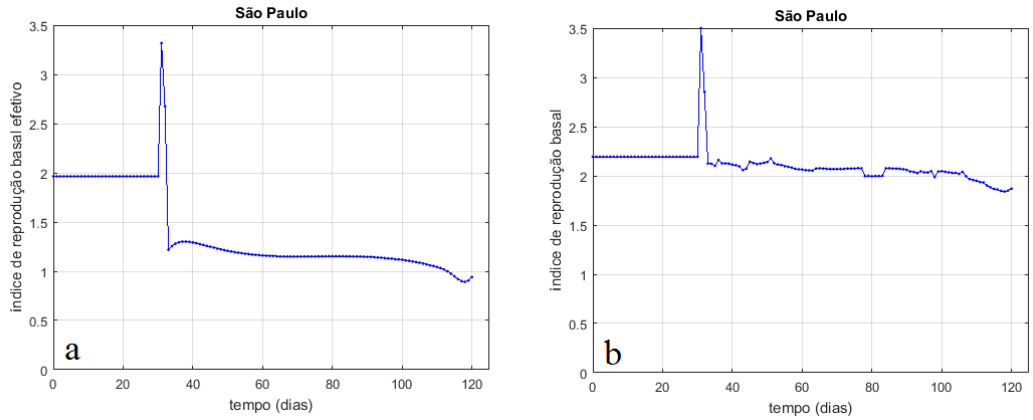


Figura 46. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de São Paulo.

Na Figura 47a vemos o $\bar{R}_0(t)$, enquanto na Figura 47b vemos o $R_{0b}(t)$, para a cidade de Fortaleza. Neste caso, podemos perceber que o $\bar{R}_0(t)$ torna-se menor do que 1, próximo ao tempo $t = 70\ dias$, sendo esta a razão da inflexão na curva de crescimento do número de casos vista na Figura 45a. Porém, é muito tímida esta inflexão porque o $\bar{R}_0(t)$ continua próximo de 1. Para $t = t_c + 1$, há uma descontinuidade na passagem da fase constante

(intervalo $t \leq t_c$) para a fase de variação paramétrica (intervalo $t > t_c$). Na fase anterior, ou seja, para $t \leq t_c$, os parâmetros são constantes. A forte oscilação que percebemos tanto na Figura 46 quanto na Figura 47 não é fruto de efeito dinâmico, tendo origem numérica, devida à descontinuidade na passagem à fase de variação paramétrica. Porém, conforme visto nas figuras, são necessários apenas poucos passos no tempo para o algoritmo de identificação retomar de forma estável a estimação dos parâmetros.

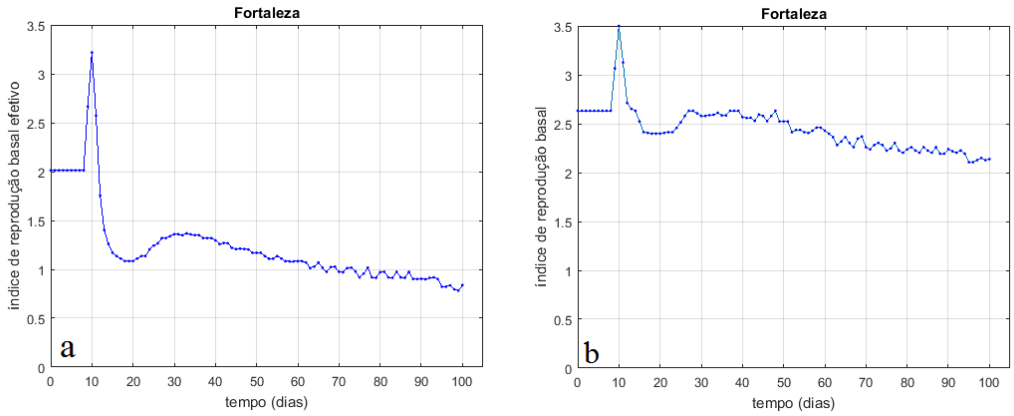


Figura 47. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Fortaleza.

5.4 Extensão do Estudo a Diversas Cidades Brasileiras

Esta seção se destina a apresentar os resultados de análises com estimação paramétrica e simulações, para diversas cidades brasileiras. A metodologia é idêntica à que foi usada na seção anterior, para as cidades de São Paulo e Fortaleza. Em todos os casos, a variação paramétrica foi identificada de forma simultânea nos três parâmetros: $R_0(t)$, $k(t)$ e $p(t)$. As principais explicações sobre a geração dos resultados já foram apresentadas na seção anterior, de forma que os comentários sobre as figuras a seguir serão menos extensos.

Nas Figuras 48 a 61 são mostrados resultados para as cidades de Belém (PA), Brasília (DF), Pelotas (RS), Porto Alegre (RS), Rio de Janeiro (RJ), Salvador (BA) e São Luís (MA). Um fato comum a todos os casos é que o modelo reproduz muito bem os dados reais, conforme pode ser visto nas Figuras 48a, 50a, ..., 60a. Como nos casos anteriores de São Paulo e Fortaleza, o algoritmo de minimização leva o erro em cada ponto para zero, de forma que o modelo segue perfeitamente os pontos do polinômio que se ajusta aos dados reais. A escolha da ordem desse polinômio é muito importante, pois define se o modelo reproduzirá ou não flutuações de mais alta frequência geralmente observadas nos dados reais. A análise dos dados permite verificar que essas flutuações, na maioria dos casos, não refletem efeitos dinâmicos reais e têm origens em problemas vinculados a retardos na

detecção/divulgação do número de novos casos. Portanto, a opção sempre foi por usar um polinômio de ordem adequada (escolha visual), que filtra essas flutuações de mais alta frequência. Os polinômios que melhor serviram a essa finalidade tiveram ordem igual ou superior a cinco. Estes foram comentários gerais, que se aplicam a todos os casos. A seguir, alguns breves comentários específicos de cada caso são adicionados.

Na Fig. 48a (cidade de Belém) percebemos um ponto de inflexão na curva de crescimento do número de casos, que ocorre em aproximadamente $t = 68\,dias$, quando a velocidade de crescimento começa a diminuir e, consequentemente, a aceleração do crescimento torna-se negativa. Este efeito dinâmico é também perceptível na Figura 49a, quando no mesmo tempo $t = 68\,dias$ o $\bar{R}_0(t)$ torna-se menor do que 1. Favoreceram a existência desse efeito dinâmico os seguintes fatos: diminuição no $R_{0b}(t)$ (Figuras 48b e 49b) como consequência de ações preventivas, tais como o uso de máscara e o distanciamento social em lugares públicos; aumento do $k(t)$ (Figura 48c), consequência de testagem e identificação de infectados postos em quarentena; diminuição do $p(t)$ (Figura 48d), consequência de decisões sobre isolamento social.

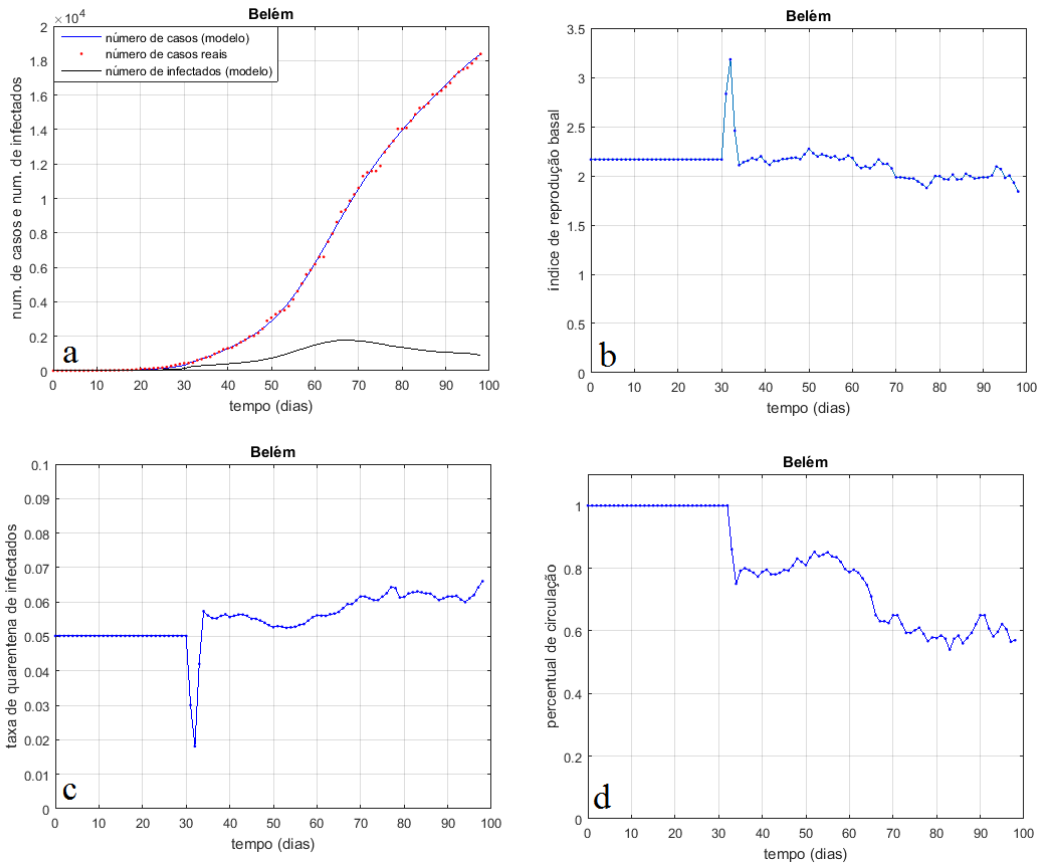


Figura 48. Identificação paramétrica e simulação para a cidade de Belém, de 18/03 a 24/06.

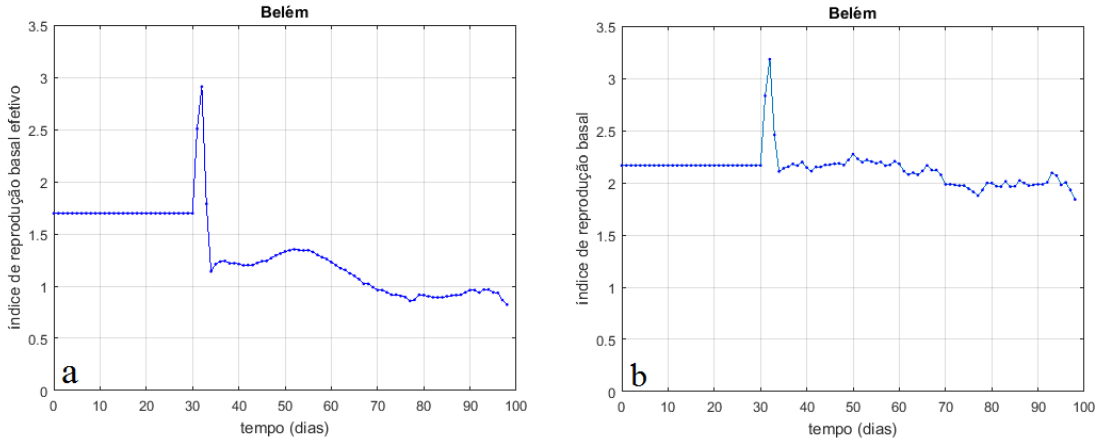


Figura 49. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Belém.

Brasília já esteve em uma situação mais perigosa, com um $\bar{R}_0(t)$ próximo a 1.4, mas nos últimos dias de dados reais passou a valores ligeiramente maiores do que 1 (Figura 51a). Por esta razão não há ponto de inflexão no crescimento do número de casos (Figura 50a) até o dia 24/06/2020.

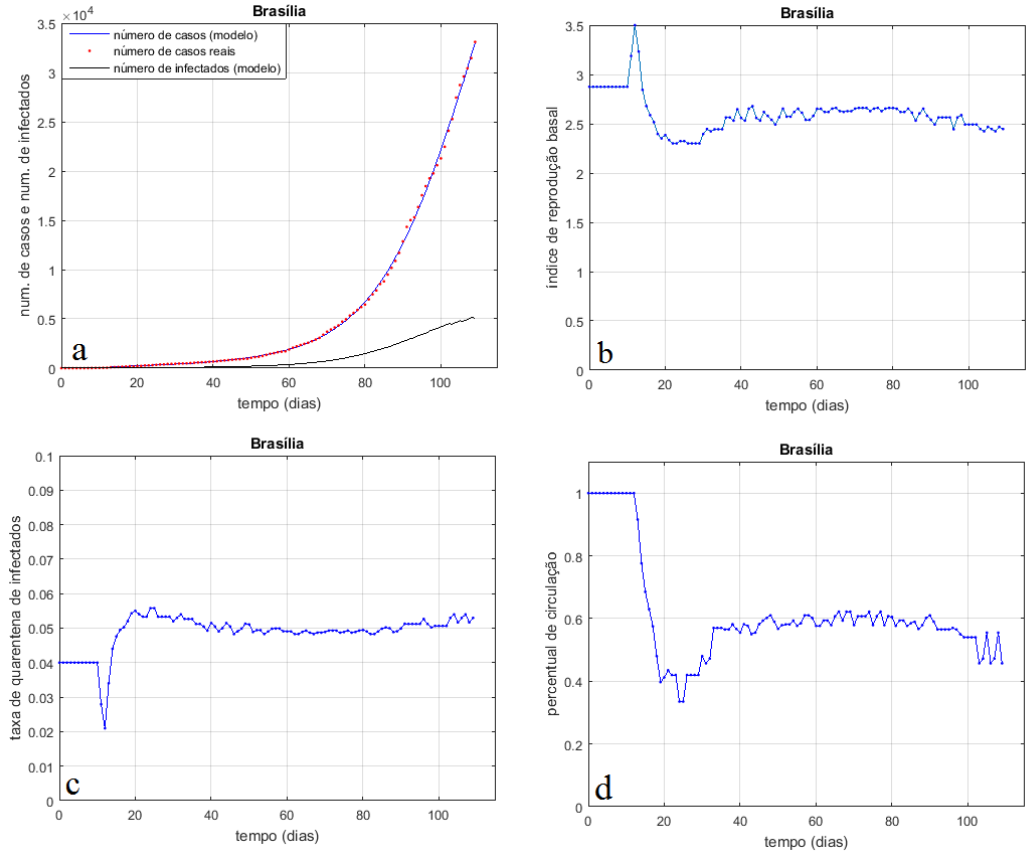


Figura 50. Identificação paramétrica e simulação para a cidade de Brasília, de 07/03 a 24/06.

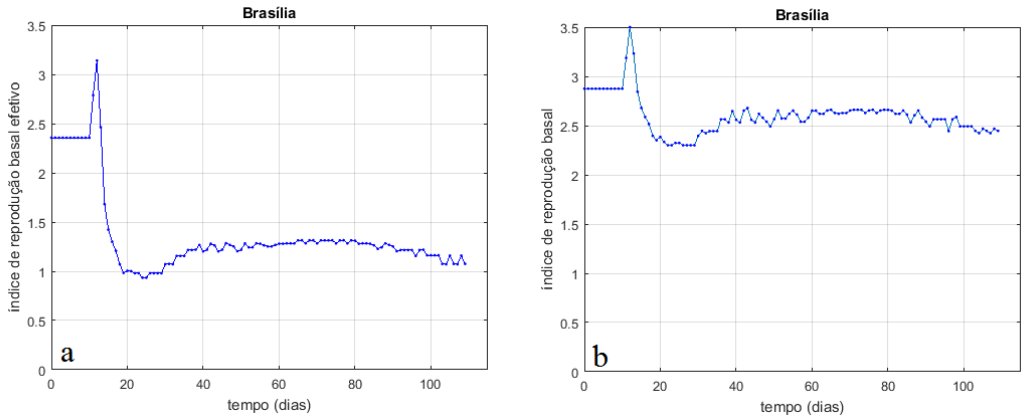


Figura 51. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Brasília.

As situações em algumas cidades de menor porte também foram analisadas, como é o caso de Pelotas (RS), com 342405 habitantes. Nesta cidade as limitações de circulação começaram praticamente coincidentes com o início da epidemia, ou seja, o seu instante de corte é nulo $(t_c = 0)$.

Em Pelotas podemos perceber que há uma inflexão no número de casos em $t = 23\ dias$ (Figura 52a), quando $\bar{R}_0(t)$ torna-se menor do que 1 (Figura 53a). Porém, o $\bar{R}_0(t)$ volta a ficar maior do que 1 para $t \geq 34\ dias$ e assim permanece por um longo tempo, até oscilar um pouco abaixo de 1 em $t = 80\ dias$, voltando em seguida a ficar superior a 1. Uma das causas do $\bar{R}_0(t)$ permanecer maior do que 1 é o elevado percentual de circulação (Figura 52d), que se mantém superior a 60%.

Em Porto Alegre aconteceu um fenômeno diferente, ainda não visto nas cidades em análises neste trabalho. Logo após $t = t_c$, o $\bar{R}_0(t)$ torna-se menor do que 1 (Figura 55a), quando há uma inflexão na curva do número de casos com aceleração passando a negativa (Figura 54a). Esta primeira inflexão acontece em $t = 25\ dias$. Tudo estava indo bem após esse tempo: a curva estava com aceleração negativa e o número de casos tendia a se estabilizar em um valor constante. Porém, o $\bar{R}_0(t)$ volta a ficar maior do que 1 em $t = 56\ dias$, de forma que neste tempo há uma nova inflexão na curva do número de casos, desta feita com aceleração positiva e esta situação se mantém até o último dia dos dados. Provavelmente, esta perda de desempenho no controle epidêmico para a cidade de Porto Alegre esteja associada à flexibilização adotada com relação às medidas de restrição de circulação. Apesar deste problema, como o desempenho foi bom até o tempo $t = 56\ dias$, Porto Alegre permanece, em 24/06/2020, com os melhores índices em termos do número de casos e de óbitos, comparados às outras capitais brasileiras.

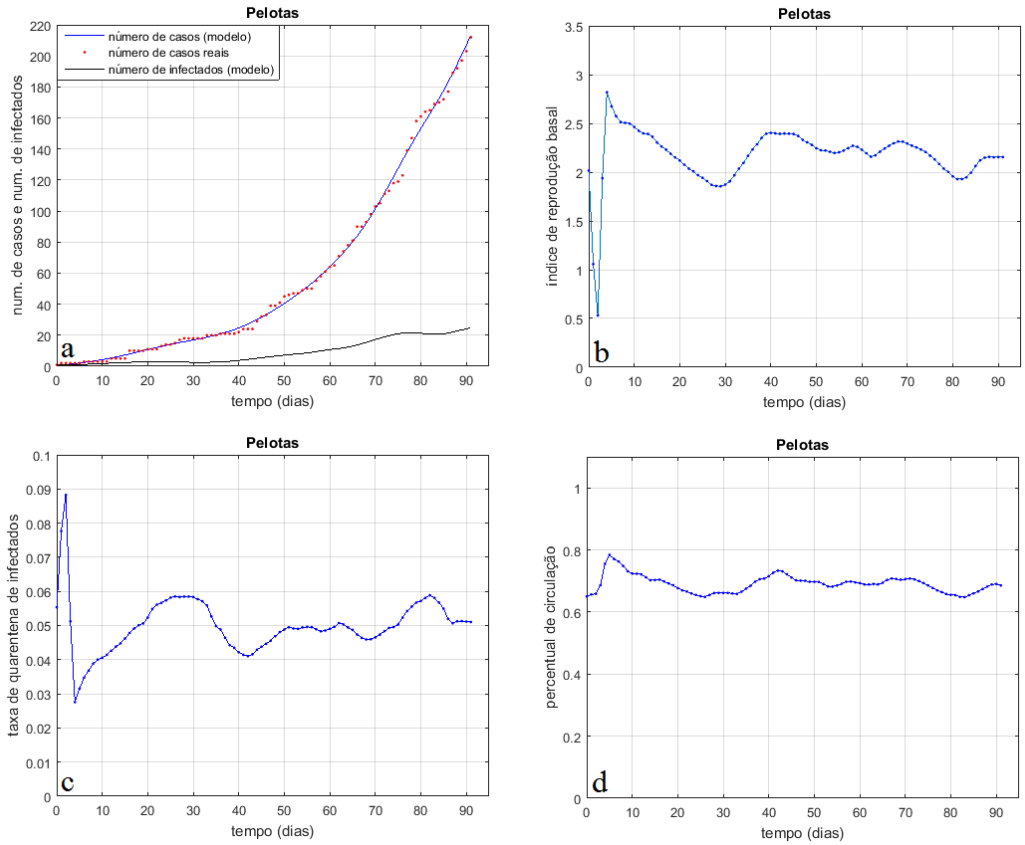


Figura 52. Identificação paramétrica e simulação para a cidade de Pelotas, de 25/03 a 24/06.

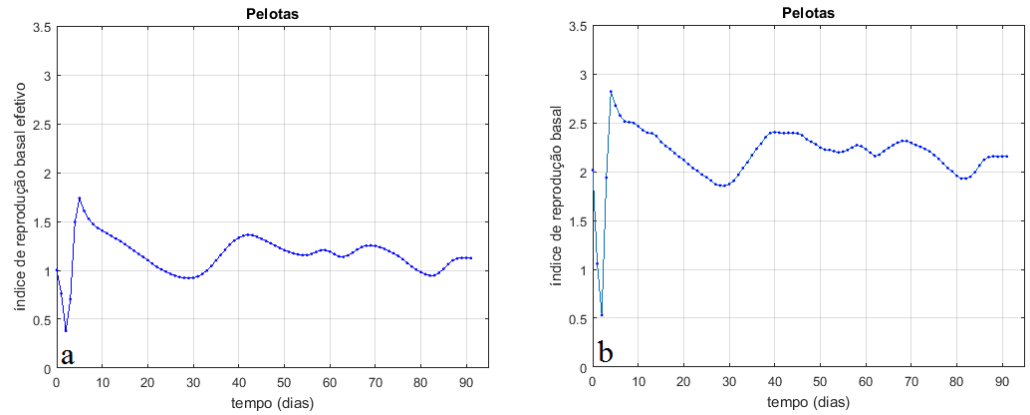


Figura 53. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Pelotas.

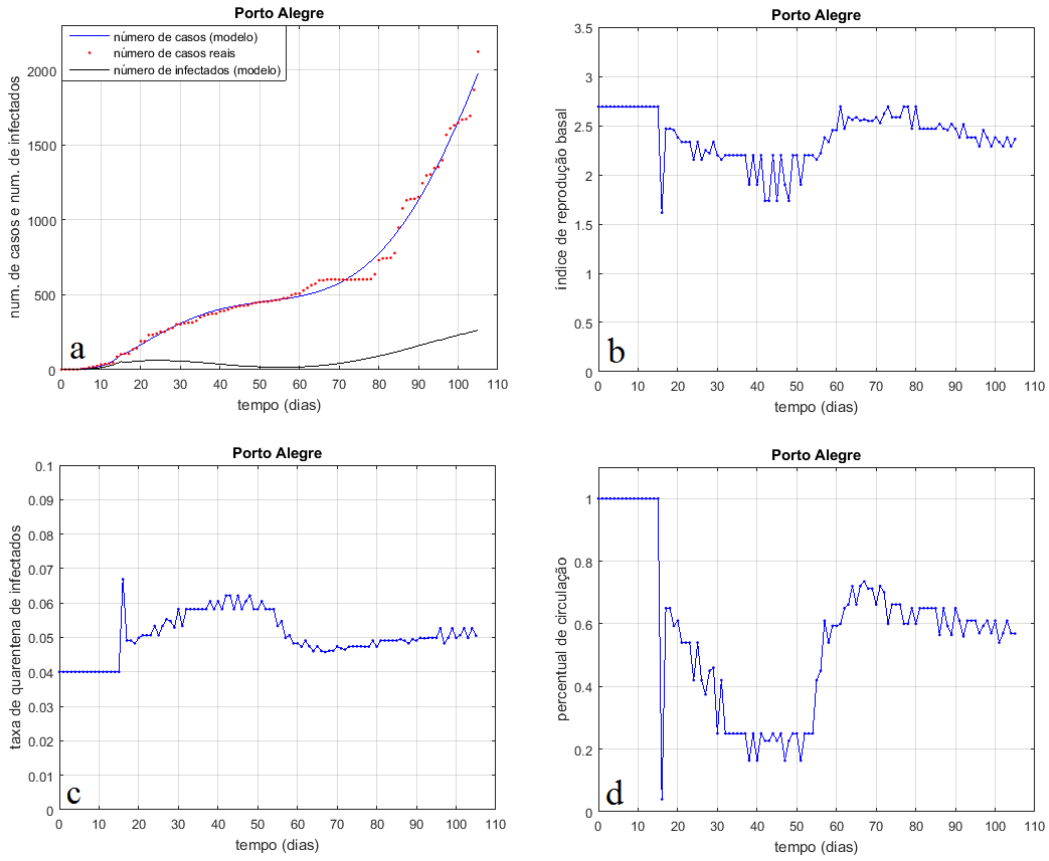


Figura 54. Identificação paramétrica e simulação para a cidade de Porto Alegre, de 11/03 a 24/06.

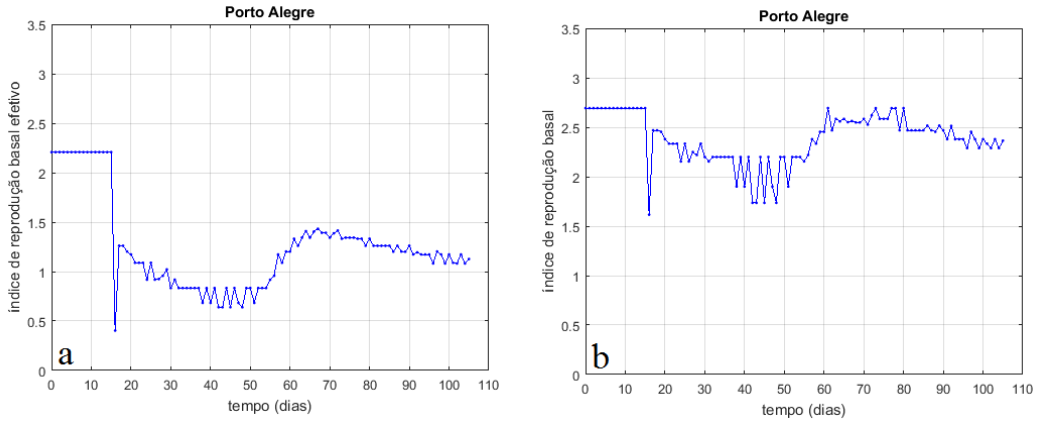


Figura 55. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Porto Alegre.

O Rio de Janeiro esteve com o $\bar{R}_0(t)$ acima de 1 até o tempo $t = 85\ dias$ (Figura 57a) e, antes desse tempo, o crescimento do número de casos foi acelerado (Figura 56a). Após esse tempo, verificamos que houve uma leve inflexão na curva do número de casos, que passa a crescer com aceleração negativa (Figura 56a) durante um curto intervalo de tempo, pois o $\bar{R}_0(t)$ tende a ficar oscilando em torno de 1 para $t \geq 97\ dias$. Isto faz com que

o crescimento do número de casos se aproxime de uma reta, nos últimos dias de dados reais vistos na Figura 56a.

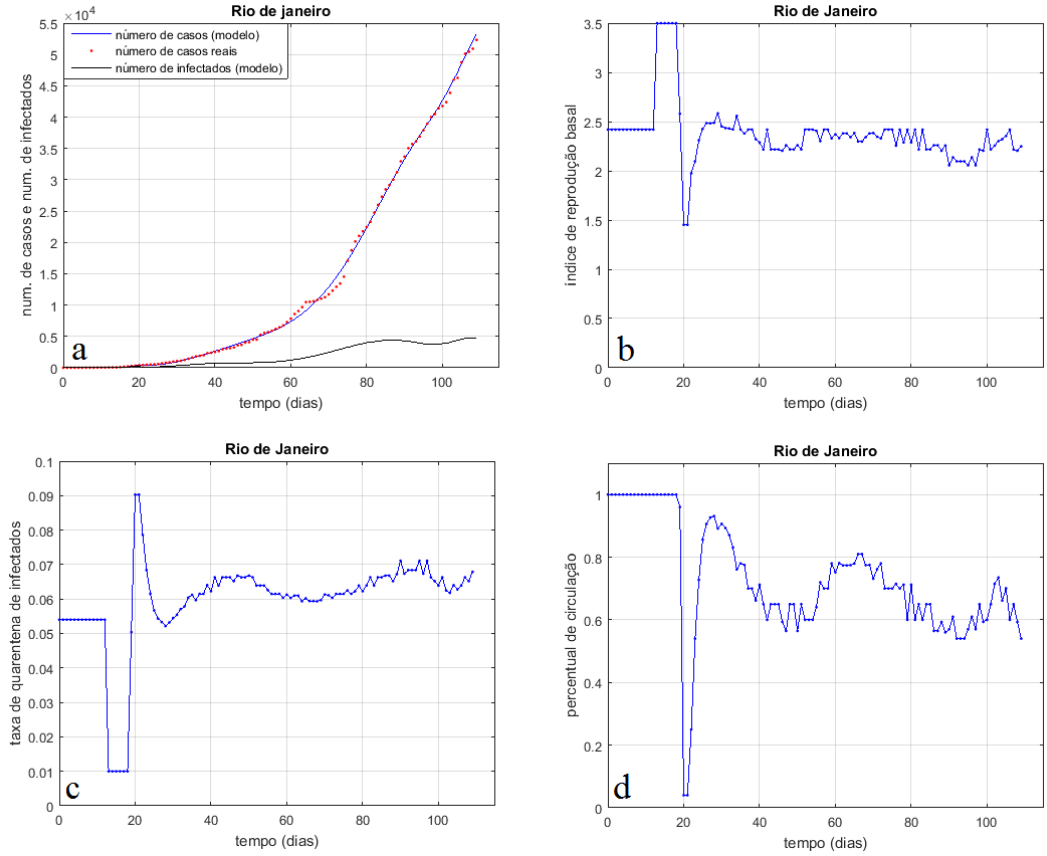


Figura 56. Identificação paramétrica e simulação para a cidade Rio de Janeiro, de 06/03 a 24/06.

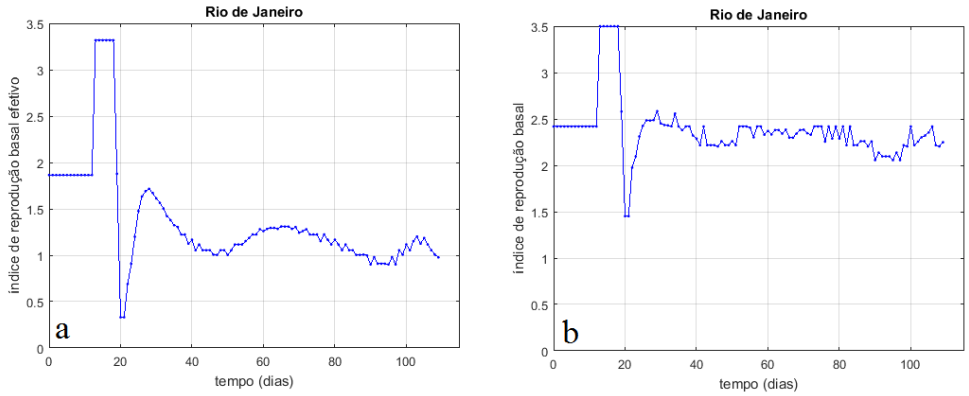


Figura 57. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade do Rio de Janeiro.

Em Salvador há uma diminuição progressiva do $\bar{R}_0(t)$ (Figura 59a), parâmetro que ficou menor do que 1 durante 10 dias $(90 \leq t \leq 100\, dias)$ e depois se tornou levemente

maior do que 1 novamente. O percentual de circulação ficou sempre superior a 60% (Figura 58d), fato que impede uma desaceleração na curva do número de casos (Figura 58a).

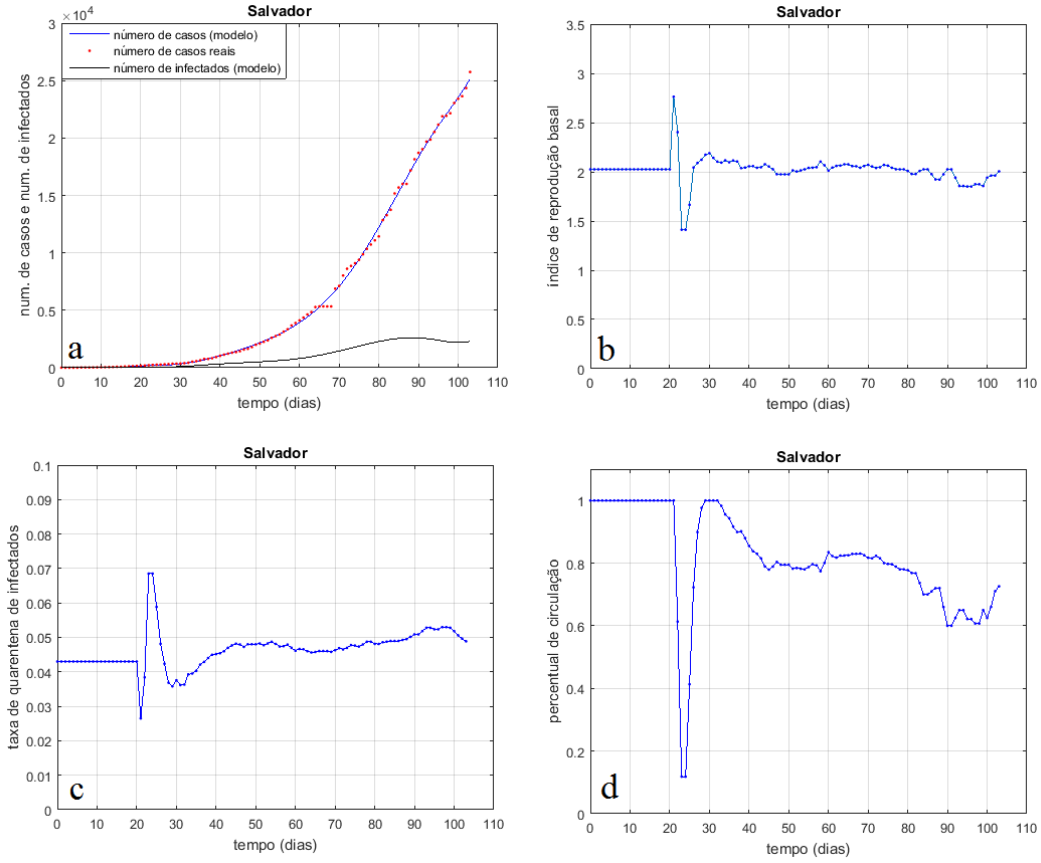


Figura 58. Identificação paramétrica e simulação para a cidade de Salvador, de 16/03 a 24/06.

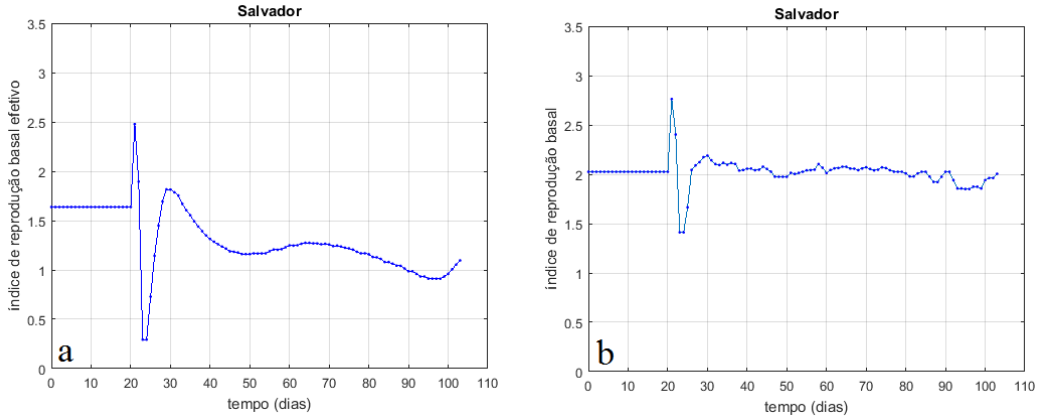


Figura 59. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Salvador.

Em São Luís o $\bar{R}_0(t)$ diminuiu progressivamente até ficar menor do que 1 a partir de $t > 70\ dias$ (Figura 61a). A partir desse tempo há uma inflexão na curva do número de

casos, que passa a crescer com aceleração negativa (Figura 60a). Essa situação melhorou porque o $\bar{R}_0(t)$ continuou diminindo gradativamente, chegando em 24/06/2020 a ficar próximo de 0.5%. Este bom desempenho foi possível porque o $R_{0b}(t)$ diminui ao longo do tempo (Figura 60b), o $k(t)$ aumenta (Figura 60c) e, principalmente, o $p(t)$ diminui (Figura 60d), chegando a 22.67% em 24/06/2020. Informações de jornais afirmam que em São Luís houve decreto de *lockdown* em maio, 2020.

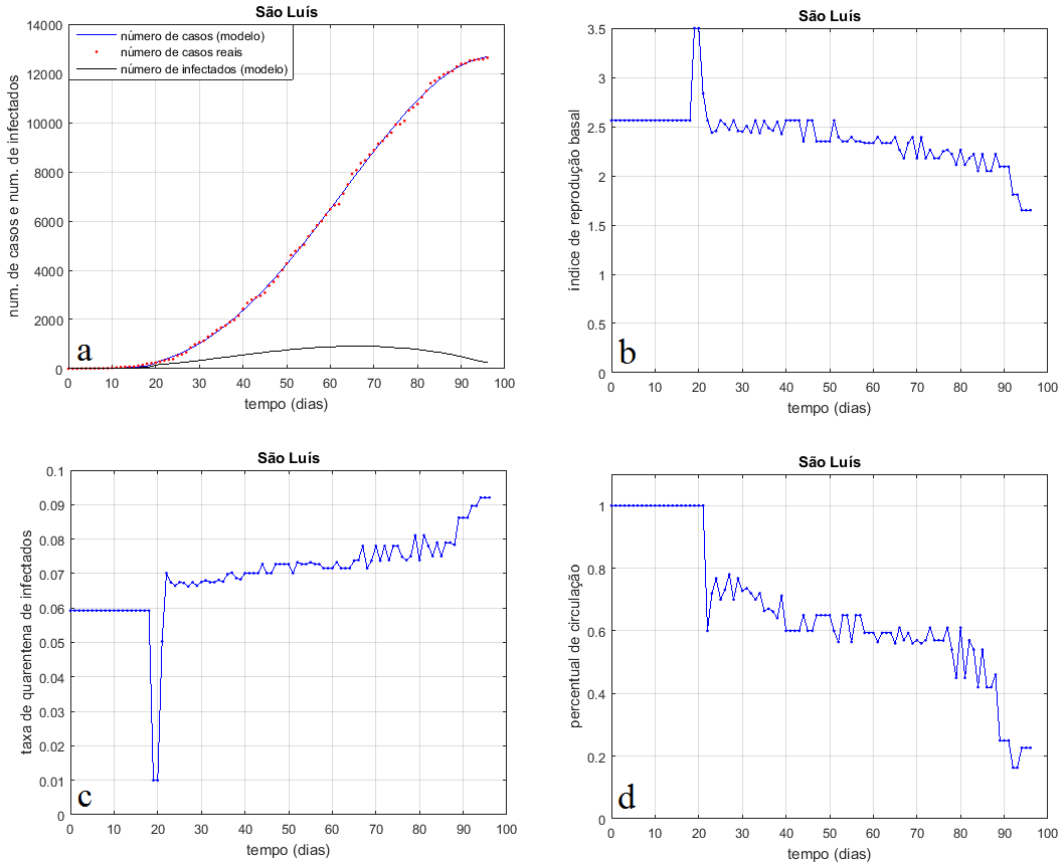


Figura 60. Identificação paramétrica e simulação para a cidade São Luís, de 25/02 a 24/06.

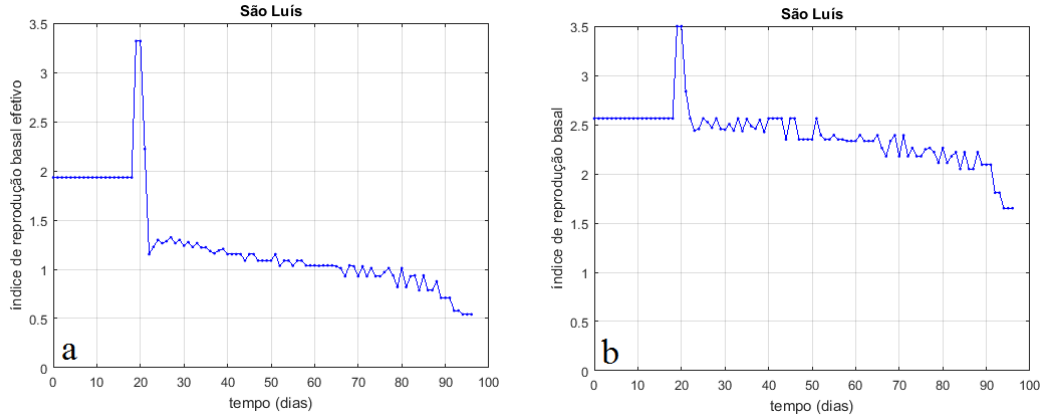


Figura 61. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade do São Luís.

A Tabela III contém os valores numéricos relativos ao último dia de dados reais e simulados, equivalente a 24/06/2020. Caso a intenção fosse fazer previsões para qualquer uma das cidades estudadas no presente trabalho, os parâmetros da Tabela III poderiam ser utilizados. Uma alternativa seria utilizar a média dos últimos cinco dias de dados. Nesta tabela, P_{mod} e P_{real} representam os percentuais de circulação previstos pelo modelo e reais, respectivamente. O P_{real} foi extraído do site inloco (https://mapabrasileirodacovid.inloco.com.br/pt/). Porém, este site disponibiliza gratuitamente apenas os dados de circulação para os estados, não havendo informações específicas para as capitais. Em parte, este fato pode explicar as diferenças mais significativas entre P_{mod} e P_{real} vistas na Tabela III. Na Figura 62 são mostrados os valores de P_{mod} e P_{real} na forma gráfica. No eixo das abcissas, a numeração das cidades obedece à ordem vista na Tabela III. A principal conclusão que podemos extrair dessa comparação é que o modelo prevê percentuais de circulação coerentes com os valores realmente praticados nas cidades. Com a exceção de São Luís, cujo decreto de *lockdown* pode ter acentuado a diferença entre os percentuais de circulação do estado do Maranhão e da sua capital, a máxima diferença entre P_{real} e P_{mod} em módulo foi inferior a 15%.

Tabela III. Valores numéricos de parâmetros relativos ao último dia do conjunto de dados reais e simulados.

Cidade	P_{mod} (%)	P_{real} (%)	$k(t)$	$R_{0b}(t)$	$\bar{R}_0(t)$
1. Belém	57	61.0	0.066	1.841	0.823
2. Brasília	45.72	60.6	0.0529	2.447	1.075
3. Fortaleza	45	56.7	0.0756	2.137	0.84
4. Pelotas	68.56	61.7	0.0510	2.155	1.125
5. Porto Alegre	56.89	61.7	0.0505	2.364	1.126
6. Rio de Janeiro	54	60.2	0.0679	2.248	0.976
7. Salvador	72.6	62.1	0.0489	2.005	1.096
8. São Paulo	50.2	62.3	0.0281	1.869	0.941
9. São Luís	22.67	63	0.092	1.651	0.542

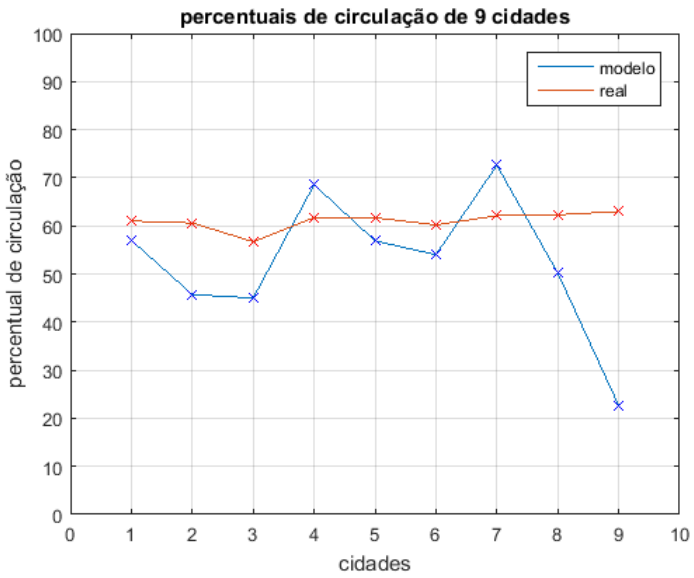


Figura 62. Percentuais de circulação reais e previstos pelo modelo.

5.5 Comentários Finais

As epidemias de COVID-19 nas cidades que foram analisadas no presente capítulo mostraram características comuns, sendo a principal delas a variação paramétrica. Os parâmetros variam porque as medidas de prevenção impostas à sociedade também variam ao longo do tempo. Estas variações paramétricas foram identificadas utilizando-se um algoritmo de minimização do erro entre dados reais e dados simulados, mais especificamente o método de Nelder-Mead Simplex. Simulações realizadas mostraram que o modelo com parâmetros variáveis reproduziu muito bem os dados reais em testes realizados com diversas cidades brasileiras. O principal parâmetro que serve como variável de controle epidêmico é o percentual de circulação, definido como sendo o percentual da população de uma cidade que está circulando livremente, ou seja, não está em isolamento social. Uma das vantagens do modelo e da estrutura de identificação paramétrica propostos é que se torna automaticamente identificado qual é o percentual de circulação que está sendo praticado na cidade em estudo e, de fato, enquanto não houver vacina contra o novo coronavírus, esta é a principal variável de controle epidêmico. Outra vantagem da modelagem proposta é a utilização de uma estrutura simples de modelo dinâmico, baseada no SIR, reduzindo assim o número de parâmetros utilizados. Esta estratégia foi fundamental para o bom desempenho do modelo com identificação de variações paramétricas: menos parâmetros significam menos fontes potenciais de erros na estimação. O estudo com dados reais foi realizado do início da epidemia em cada cidade até o dia 24/06/2020. As dinâmicas

mudarão com o tempo, mas o que fica de importante como contribuição específica do presente capítulo é a estrutura de identificação paramétrica, que pode ser utilizada em qualquer cidade ou região.

Referências

GIORDANO G. *et al*. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 2020, https://doi.org/10.1038/s41591-020-0883-7.

GOMES S. C. P., Monteiro I. O., Rocha C. R. Identificação de parâmetros na modelagem dinâmica da COVID-19. Revista THEMA, 2020b. **DOI**: http://dx.doi.org/10.15536/thema.V18.Especial.2020.26-53.1854. Especial, 2020b.

KERMACK W., Mckendrick A. Contributions to the Mathematical Theory of epidemics. Proceedings of the Royal Society A, Mathematical, Physical and Engineering Science, 1927, https://doi.org/10.1098/rspa.1927.0118.

LAGARIAS J.C., Reeds J. A., Wright M. H., and Wright P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions, SIAM Journal of Optimization, Vol. 9 Number 1, pp. 112-147.

Capítulo VI

Quarto Estudo – Atualização dos Dados e Previsão com o MdSIR

6.1 Introdução

O objetivo do presente capítulo é promover uma atualização dos dados, ou seja, repetir a estratégia de identificação paramétrica e simulação utilizando dados reais que vão até o dia 13/07/2020. Como os dados reais do capítulo anterior vão até o dia 24/06/2020, esta atualização corresponde a 19 novos dias nos arquivos das cidades analisadas. Após esta atualização, uma segunda parte do capítulo será destinada a uma análise de previsão com o MdSIR, utilizando a seguinte estratégia:

a) Com os resultados obtidos no capítulo anterior, temos os valores numéricos dos parâmetros identificados ao longo do tempo;

b) Adotamos os parâmetros relativos ao último dia de dados do capítulo anterior, correspondente a 24/06/2020;

c) Com os parâmetros do item b) considerados constantes, fazemos simulações para os próximos 19 dias, contados a partir do dia 24/06/2020;

d) Comparamos os resultados das simulações com os dados reais do presente capítulo, que vão até o dia 13/07/2020;

e) O intervalo de tempo do dia 24/06/2020 ao dia 13/07/2020 é puramente previsão, ou seja, os resultados de simulações com os parâmetros identificados no dia 24/06/2020 são confrontados aos dados reais que vão até o dia 13/07/2020.

A análise dos resultados indica que o modelo apresenta boa previsibilidade, em média, para os primeiros 15 dias de dados. Para algumas cidades a previsibilidade foi considerada boa para os 19 dias de dados. Porém, ficou evidente que um intervalo de confiança em termos de previsibilidade seria correspondente aos próximos 15 dias. Este intervalo de tempo relativamente pequeno de confiabilidade preditiva deve-se a ações externas que mudam com rapidez em algumas cidades, principalmente correlatas a variações nos índices de isolamento social. Ampliação ou flexibilização do isolamento social impactam diretamente em mudanças no percentual de circulação $p(t)$. Como este parâmetro permanece constante durante o intervalo de previsão, as referidas variações no

isolamento social provocam diminuição do intervalo considerado ainda preditivo para o modelo.

6.2 Atualização para Algumas Cidades Brasileiras

As figuras apresentadas nesta seção seguem o mesmo padrão das apresentadas no capítulo anterior, dispensando assim maiores comentários. Como sempre, os dados reais para o estudo foram obtidos do site Brasil.io (https://brasil.io/dataset/covid19/caso_full/). No presente capítulo foram incorporadas mais duas cidades, com relação às relacionadas no capítulo anterior, de forma que analisamos os resultados das seguintes cidades: Belém (PA); Brasília (DF); Cuiabá (MT); Fortaleza (CE) Pelotas (RS); Porto Alegre (RS); Rio Grande (RS); Rio de Janeiro (RJ); São Luís (MA); São Paulo (SP). A data inicial de cada cidade é diferente, pois coincide com o aparecimento do primeiro caso na respectiva cidade. A data final neste capítulo é 13/07/2020, mas algumas cidades não tinham ainda os seus dados reais atualizados até o horário no qual baixamos os seus arquivos de dados, de forma que, para estas cidades, a data final é 12/07/2020. Com relação aos resultados do capítulo anterior, a presente atualização significa 18 e às vezes 19 dias a mais.

Os resultados da presente atualização estão resumidos nas Figuras 64 a 85. Um fato é comum a todos os casos: com a estimação dos parâmetros que variam ao longo do tempo, o modelo reproduz muito bem o comportamento dinâmico real, conforme pode ser verificado nas confrontações entre simulação e dados reais presentes nas Figuras 64a, 66a, 68a, ...,84a. Nas figuras podemos ver que os parâmetros permanecem constantes no início da simulação. Conforme explicado no capítulo anterior, trata-se do intervalo de tempo $0 \leq t \leq t_c$, no qual a cidade ainda não contava com medidas de prevenção (principalmente sem isolamento social), onde t_c é o instante de corte de cada cidade. Para $t > t_c$ os parâmetros variam, conforme pode ser visto nas referidas figuras. É muito importante observar a variação do índice de reprodução basal efetivo $\bar{R}_0(t)$: haverá inflexão (com aceleração negativa) na curva de crescimento do número de casos se $\bar{R}_0(t) < 1$.

A ampliação no número de dias de dados reais, de 18 ou 19 dias, com relação aos resultados obtidos no capitulo anterior, em alguns casos, provocou a necessidade de alteração na ordem do polinômio de ajuste, utilizado nos dados reais do número de casos. Conforme explicado no capítulo anterior, na seção 5.2 (ver Figura 38), é muito importante escolher adequadamente a ordem desse polinômio porque os parâmetros do modelo serão identificados a partir da minimização do erro entre os valores do polinômio de ajuste e os valores da curva do número de casos da simulação. Sintetizando, a identificação

paramétrica força o modelo a tentar reproduzir a forma polinomial que ajusta os dados reais. Em razão dessa mudança, os resultados do presente capítulo, em alguns casos, não são simplesmente a continuação dos resultados obtidos no capítulo anterior, embora continuem muito semelhantes. Em alguns casos foram necessários utilizar dois polinômios de ajuste, com o cuidado de adotarmos derivadas aproximadamente iguais no ponto de junção dos dois polinômios. Isto se deve ao fato de que, em algumas cidades, o crescimento do número de casos é muito lento, com valores praticamente constantes por 40 dias, por exemplo, vindo a acelerar posteriormente. Na Figura 63a vemos os dois polinômios utilizados para Salvador, enquanto na 63b plotamos os polinômios sobrepostos aos dados reais. Quando polinômios de mais alta ordem foram usados, utilizamos ainda um filtro passa baixa do tipo FIR (*Finite Impulse Response*), a fim de evitar que o polinômio siga as flutuações de mais alta frequência. Em alguns casos, usamos um único polinômio identificado a partir de um determinado tempo, próximo ao início dos dados reais e, de 0 até este tempo, forçamos a que a curva tivesse os seus pontos iguais aos próprios dados reais filtrados com um passa baixa. Em resumo, cada caso é tratado separadamente, com a sua própria estratégia de ajuste. Esta fase de pré tratamento dos dados reais é fundamental, já que o modelo seguirá o ajuste, ou seja, a qualidade do ajuste está diretamente relacionada com a qualidade do modelo em termos preditivos, na seguinte sequência lógica: se o ajuste é bom, ele representa bem o comportamento médio dos dados reais, bem como a tendência deste comportamento para o futuro próximo; os parâmetros do modelo são identificados a fim de que este reproduza o comportamento médio do ajuste; se o modelo reproduz bem o comportamento identificado no ajuste, a sua resposta para tempos futuros deve prever a tendência do fenômeno real. Todo esse esforço deve-se ao fato dos dados reais apresentarem flutuações de alta frequência que não são oriundas de efeitos dinâmicos e sim, de problemas na identificação/detecção dos casos, que podem se acumular em um único dia, deixando outros com pequena variação.

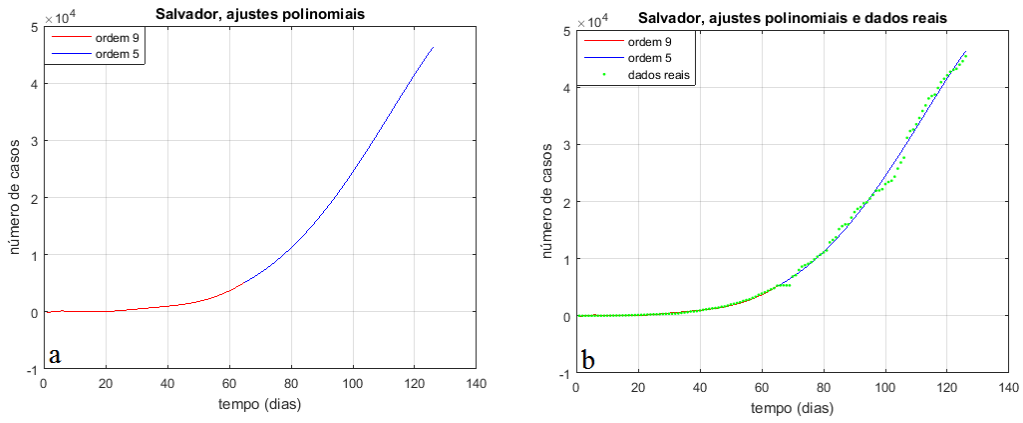


Figura 63. Polinômios de ajuste utilizados em Salvador.

Nas Figuras 64 e 65 são apresentados os resultados para Belém. Na Figura 64a vemos que Belém continua com um crescimento desacelerado do número de casos. O $\bar{R}_0(t)$ se mantém abaixo de 1 (Figura 65a) e isso atesta este crescimento desacelerado.

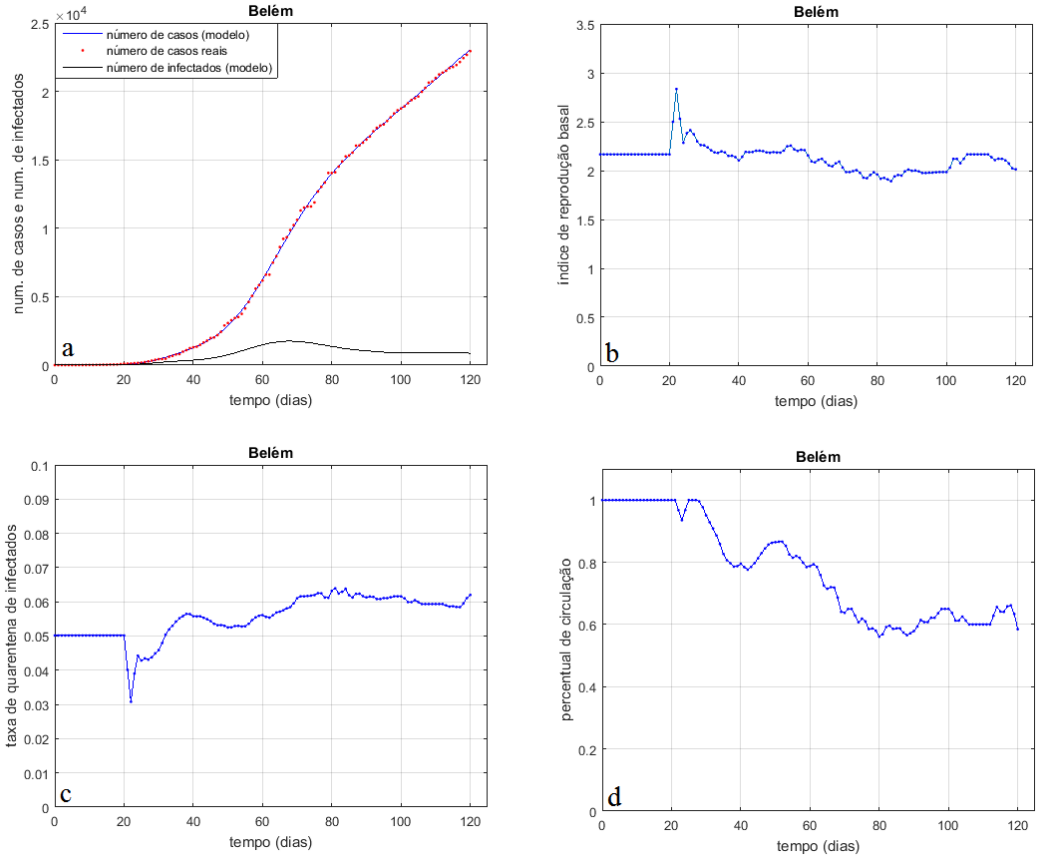


Figura 64. Identificação paramétrica e simulação para a cidade de Belém, de 18/03 a 12/07.

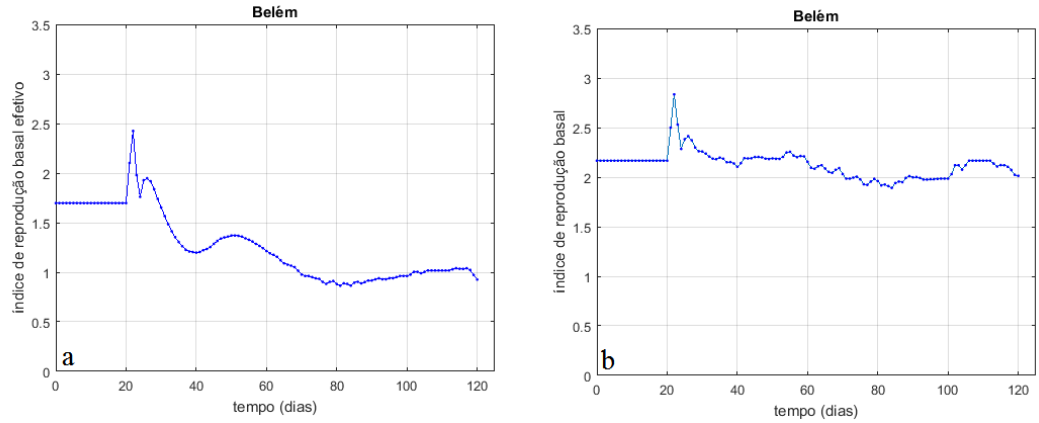


Figura 65. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Belém.

Brasília continua com crescimento acelerado (Figura 66a), embora a aceleração esteja diminuindo ao longo do tempo, tendência observada no comportamento do $\bar{R}_0(t)$ que diminui com o tempo (Figura 67a), se aproximando de 1 no tempo final.

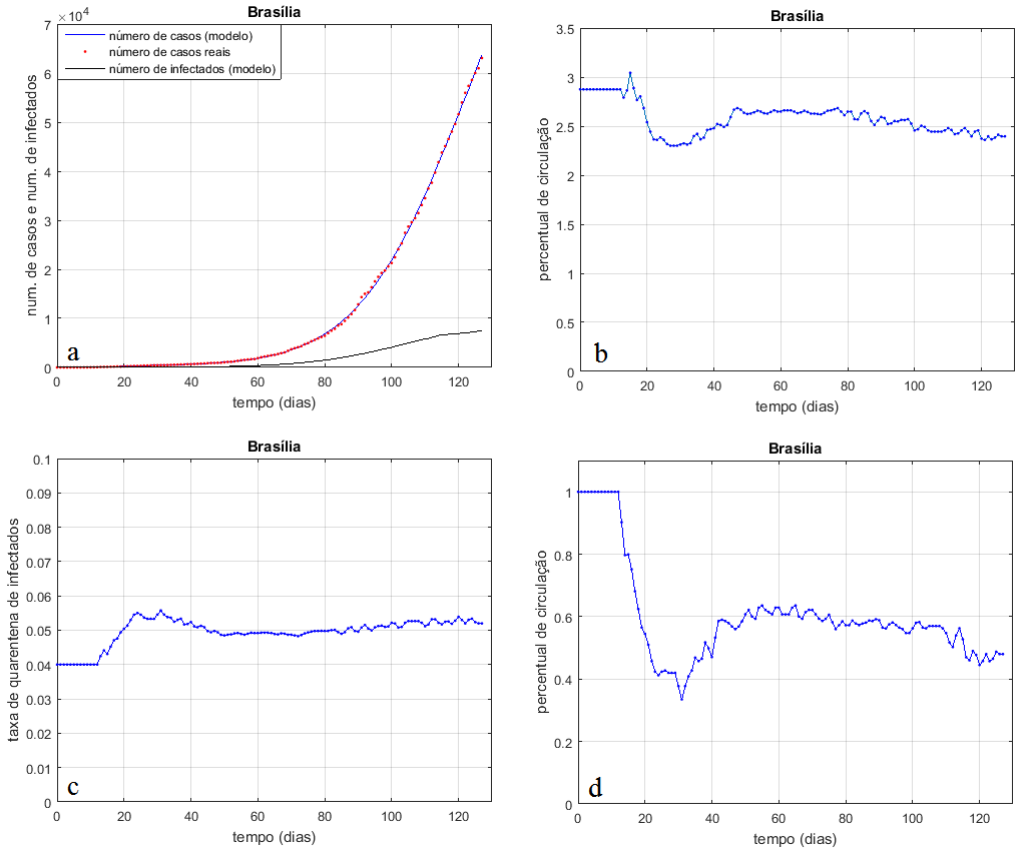


Figura 66. Identificação paramétrica e simulação para a cidade de Brasília, de 07/03 a 24/06.

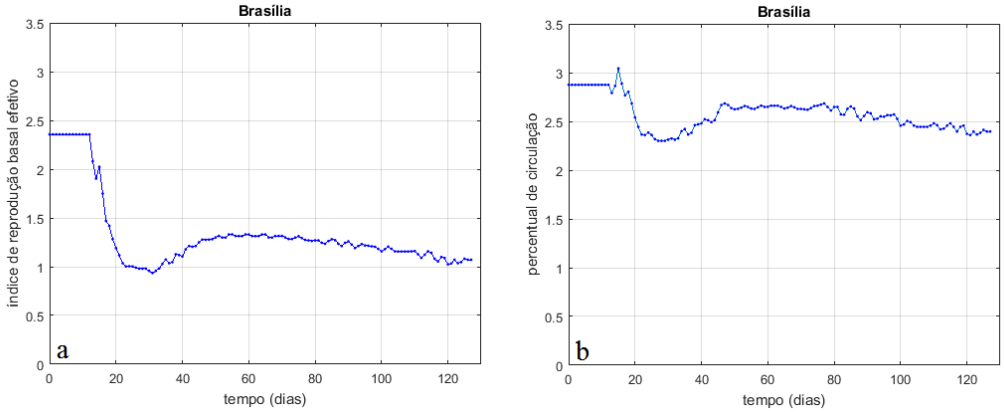


Figura 67. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Brasília.

O $\bar{R}_0(t)$ em Cuiabá se aproxima de 1 próximo ao tempo final, passando a ficar abaixo de 1 nos quatro últimos dias da simulação (Figura 69a), tempo no qual acontece uma inflexão na curva de crescimento do número de casos (Figura 68a).

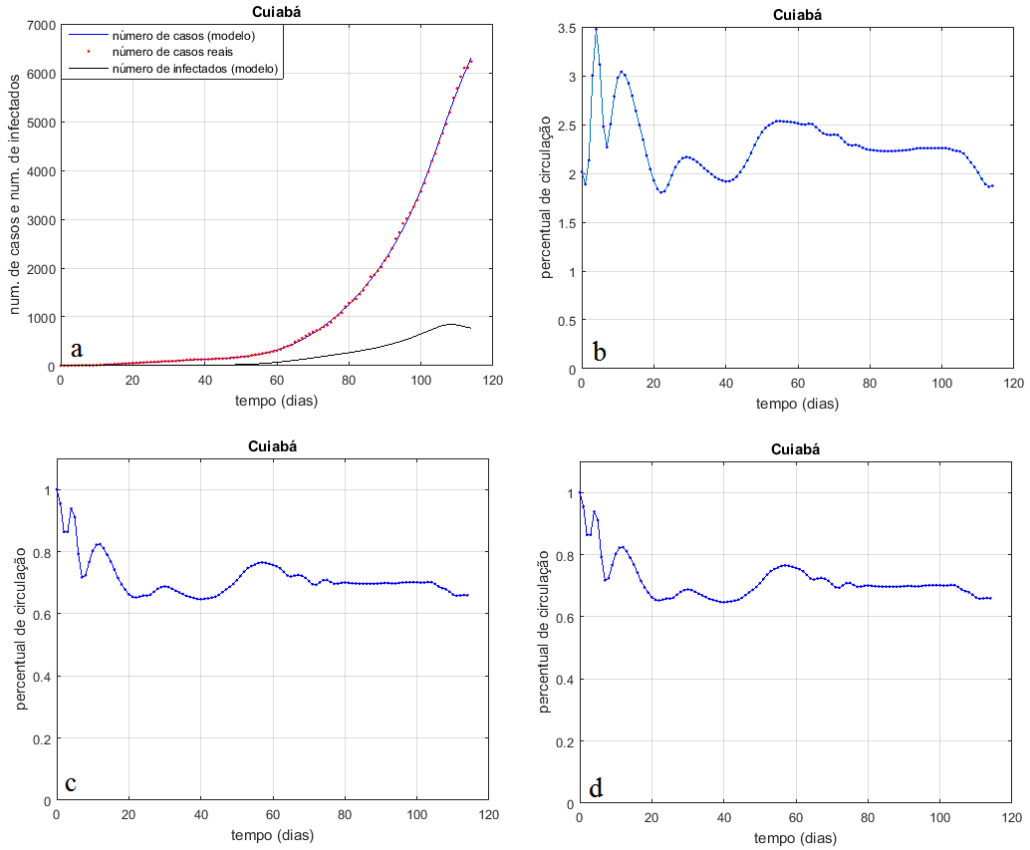


Figura 68. Identificação paramétrica e simulação para a cidade de Cuiabá, de 20/03/2020 a 12/07/2020.

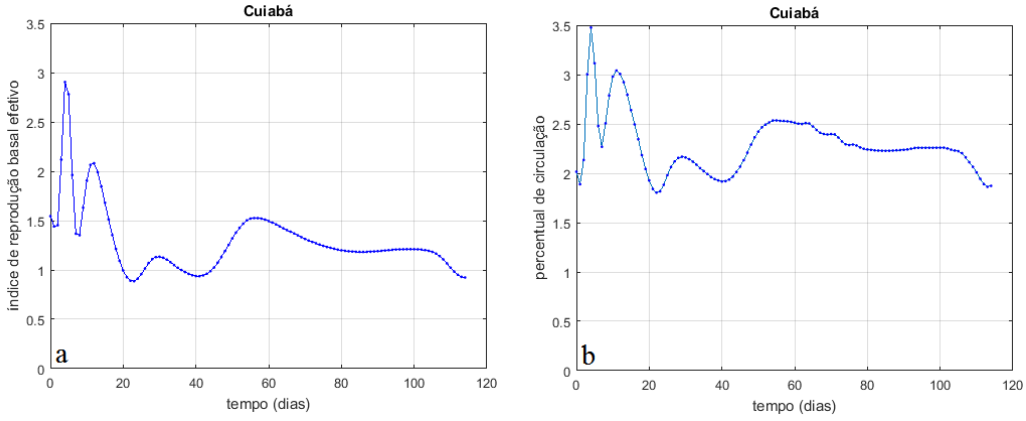


Figura 69. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Cuiabá.

Fortaleza mantém uma desaceleração bem consistente da curva do número de casos (Figura 70a). O ponto de inflexão, quando o crescimento passou a ser desacelerado, aconteceu em $t = 70\ dias$, conforme pode ser visto na Figura 71a, quando o $\bar{R}_0(t)$ passa a ser menor do que 1 e desde então se manteve assim.

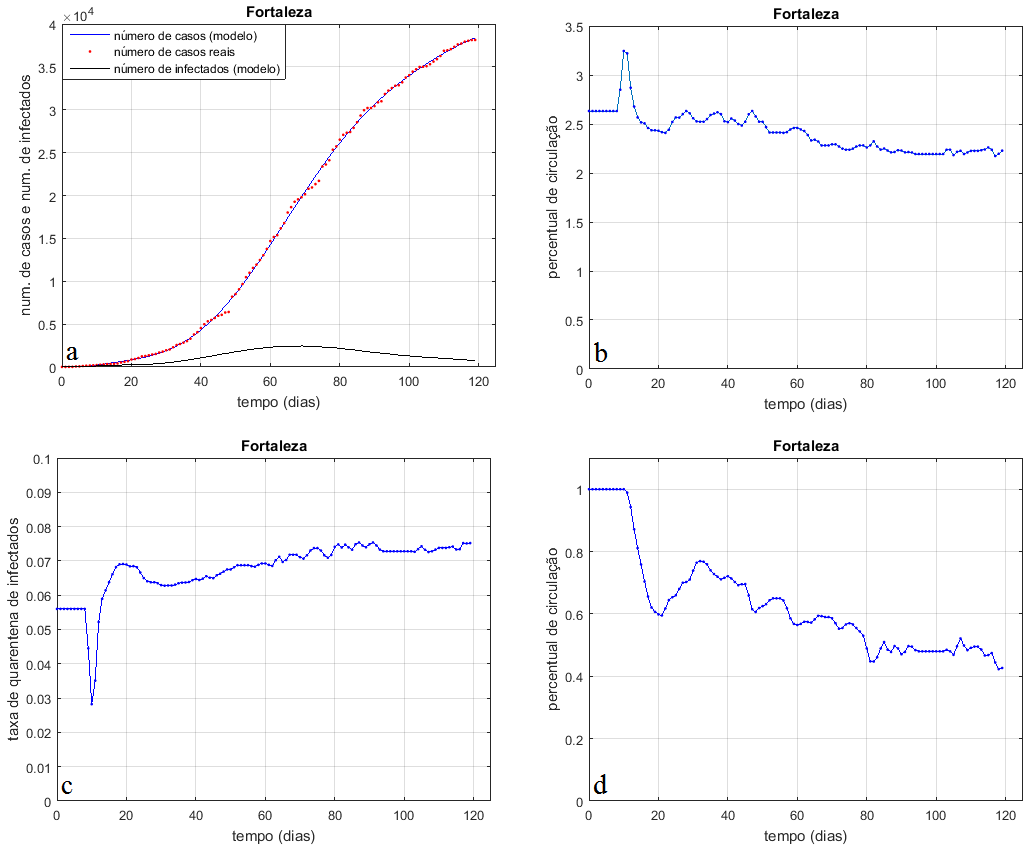


Figura 70. Identificação paramétrica e simulação para a cidade de Fortaleza, de 16/03 a 12/07.

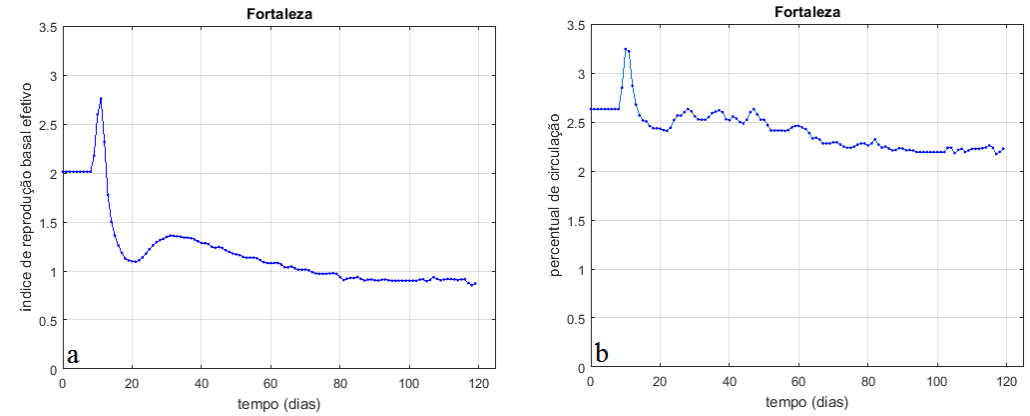


Figura 71. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Fortaleza.

O interior do Brasil passou a apresentar crescimento acelerado da contaminação e este comportamento atualmente (13/07/2020) se verifica também nas cidades do interior do RS. A cidade de Pelotas é um desses casos, conforme pode ser visualizado na Figura 72a. O $\bar{R}_0(t)$ acima de 1 reflete esse comportamento (Figura 73a).

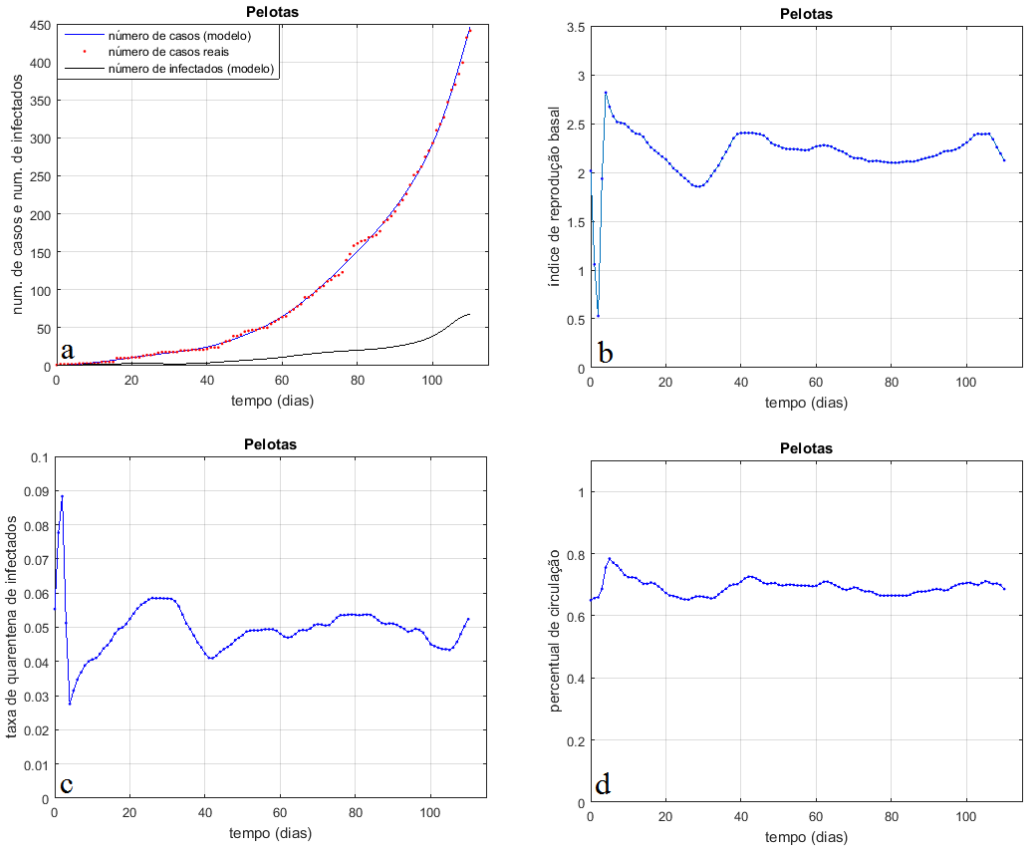


Figura 72. Identificação paramétrica e simulação para a cidade de Pelotas, de 25/03/2020 a 13/07/2020.

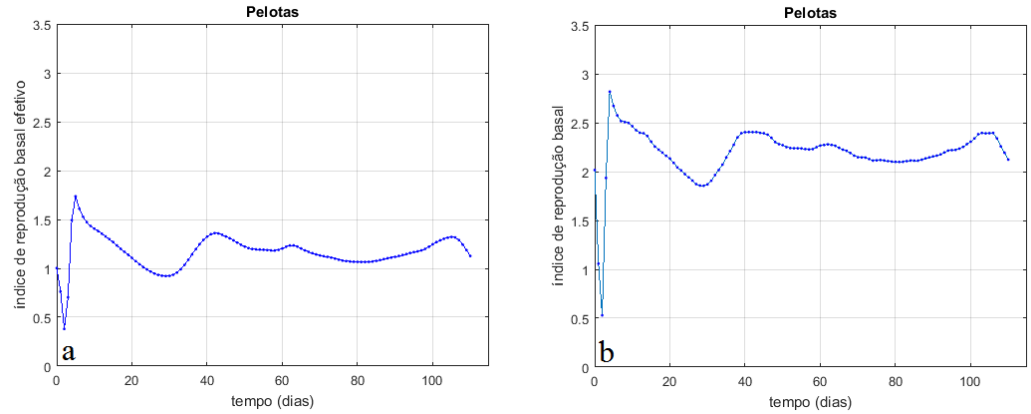


Figura 73. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Pelotas.

Em Porto Alegre a situação identificada no capítulo anterior continua, ou seja, o crescimento do número de casos continua acelerado (Figura 74a), comportamento que pode ser também identificado na manutenção do $\bar{R}_0(t)$ maior do que 1 (Figura 75a). A cidade vivenciou bons resultados na contenção da epidemia, logo após o início das primeiras medidas preventivas de isolamento social: o $\bar{R}_0(t)$ se tornou menor do que 1 e permaneceu assim por 35 dias (Figura 75a). Depois desse intervalo bom, houve provavelmente relaxamento nas condições de isolamento social, de forma que o percentual de circulação ficou próximo a 60% (Figura 74d). Essa mudança fez com que o $\bar{R}_0(t)$ voltasse a ficar maior do que 1 e assim permanecesse até o tempo final.

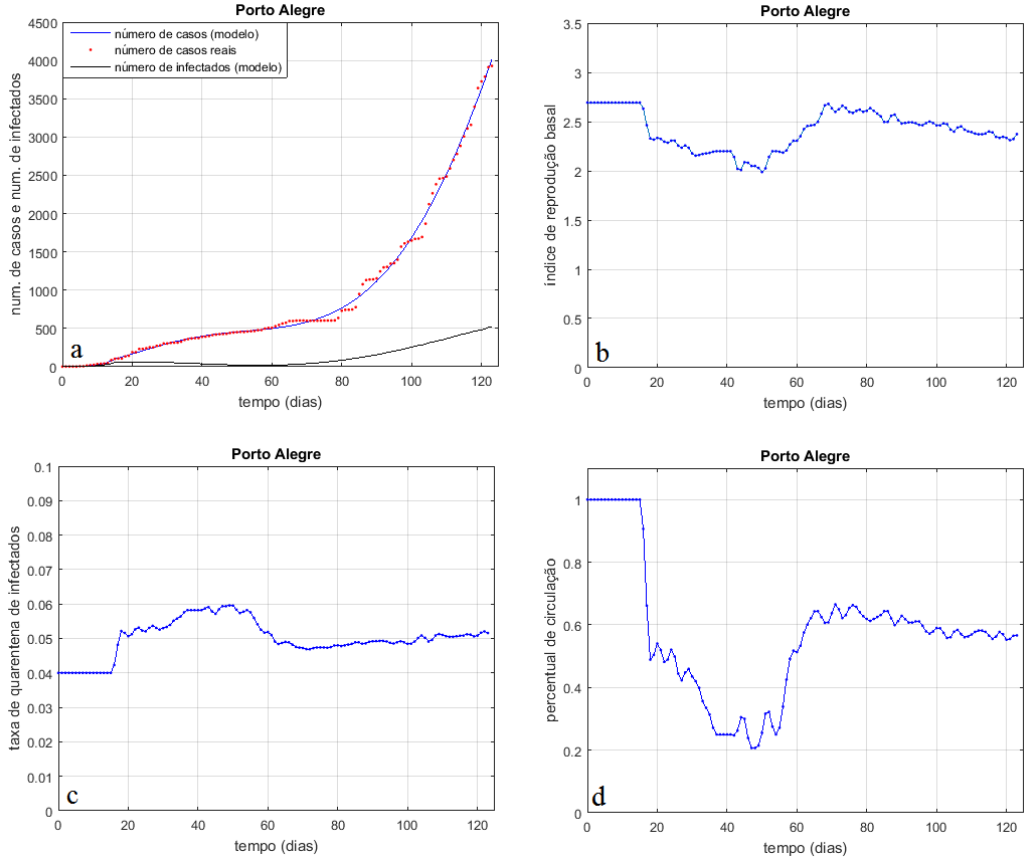


Figura 74. Identificação paramétrica e simulação para a cidade de Porto Alegre, de 11/03/2020 a 12/07/2020.

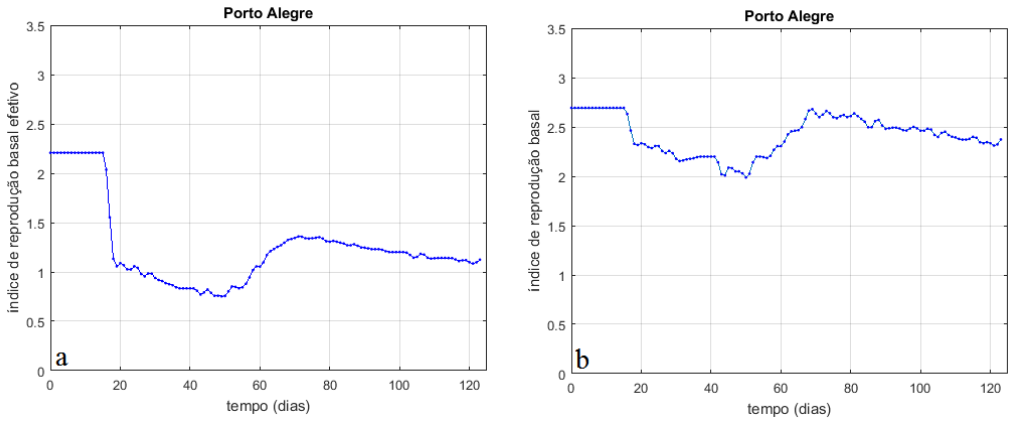


Figura 75. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Porto Alegre.

No Rio de Janeiro o ponto de inflexão aconteceu em $t = 100$, quando o $\bar{R}_0(t)$ passa a ser menor do que 1 (Figura 77a) e a curva do número de casos começa a ter crescimento desacelerado (Figura 76a). Esta situação tem se mantido, embora o $\bar{R}_0(t)$ esteja oscilando próximo a 1.

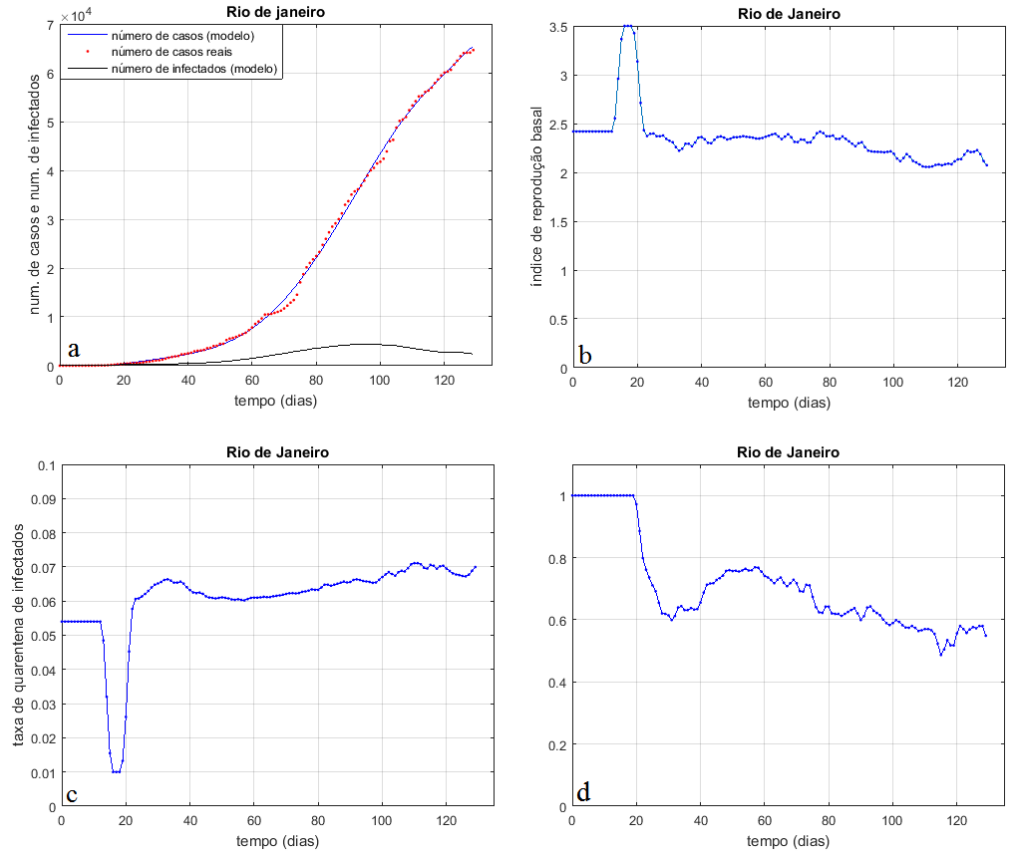


Figura 76. Identificação paramétrica e simulação para a cidade do Rio de Janeiro, de 06/03/2020 a 13/07/2020.

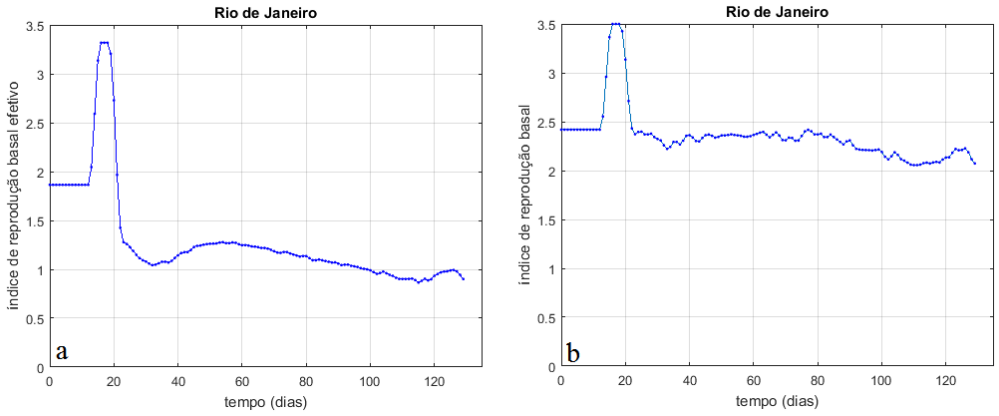


Figura 77. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade do Rio de Janeiro.

Rio Grande (RS) é mais uma cidade do interior que está apresentando crescimento acelerado, conforme pode ser visto na curva do número de casos (Figura 78a). O $\bar{R}_0(t)$ oscila (Figura 79a), mas se mantém superior a 1, denotando que, de fato, o crescimento continua acelerado. Certamente, quem mais contribui para essa situação é o alto percentual de circulação que se mantém próximo a 70% (Figura 78d).

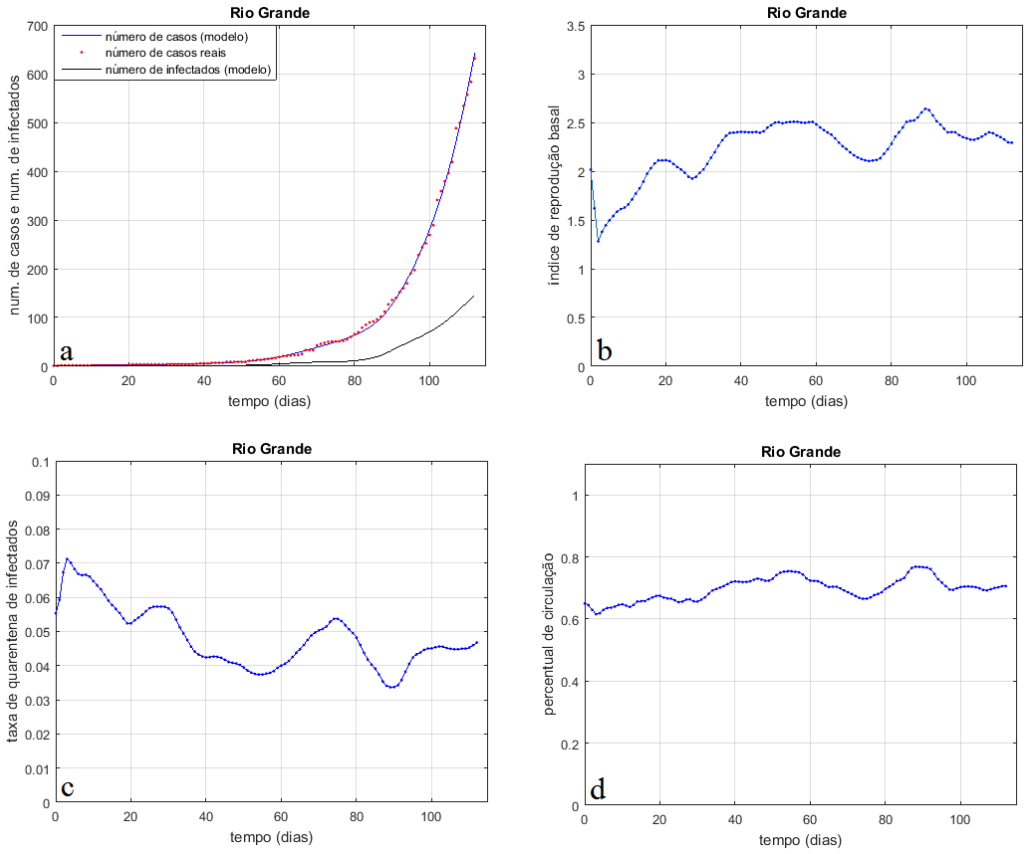


Figura 78. Identificação paramétrica e simulação para a cidade de Rio Grande, de 06/03/2020 a 13/07/2020.

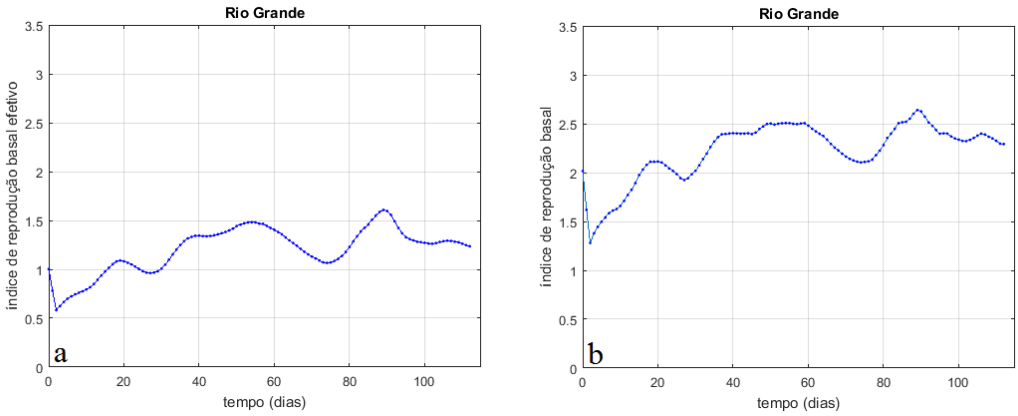


Figura 79. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Rio Grande.

O crescimento acelerado de Salvador parece ter se encerrado nos últimos quatro dias de simulação, quando há uma inflexão na curva do número de casos (Figura 80a), coincidente com o fato do $R_0(t)$ ter se tornado menor do que 1 (Figura 81a). Entretanto, é necessário esperar mais cinco ou seis dias para ver se de fato essa desaceleração irá permanecer.

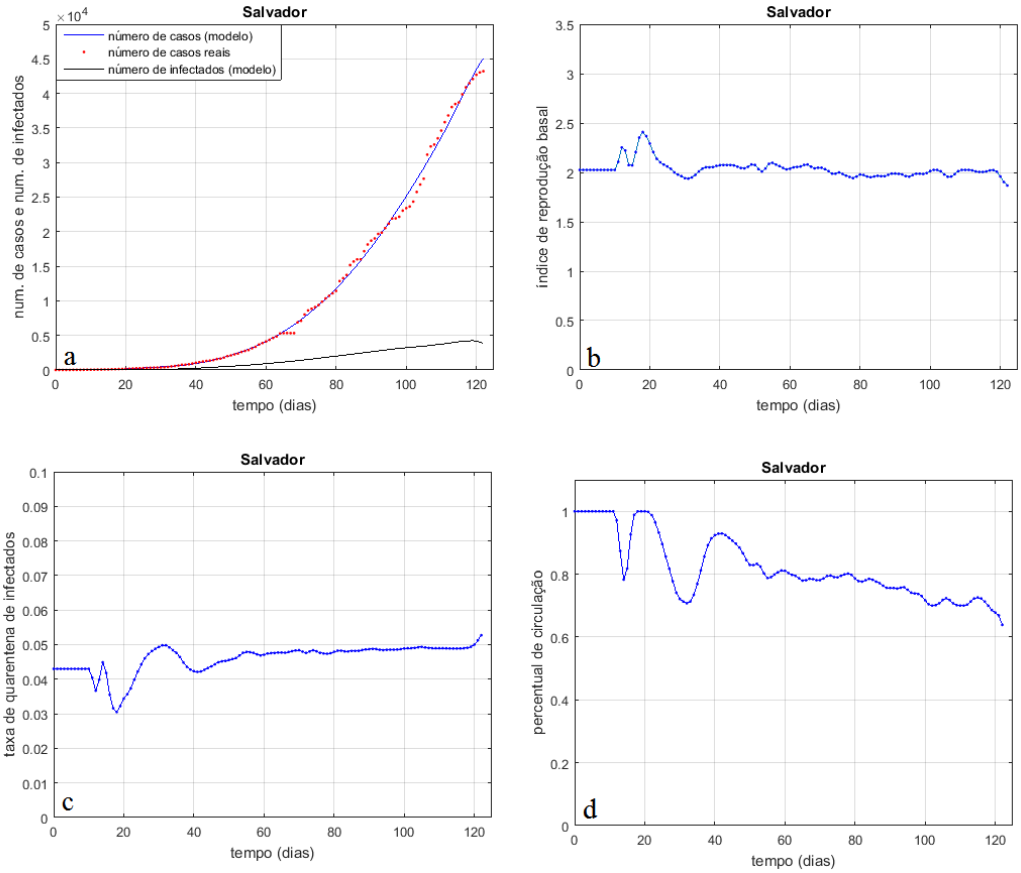


Figura 80. Identificação paramétrica e simulação para a cidade de Salvador, de 13/03/2020 a 13/07/2020.

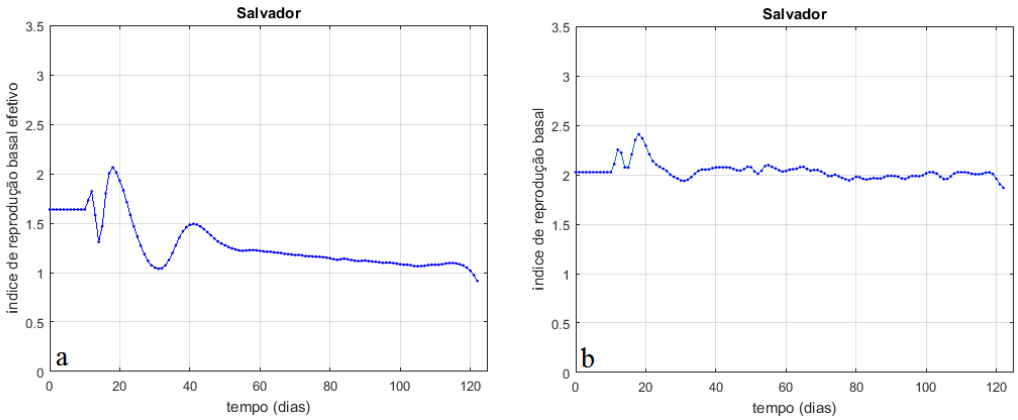


Figura 81. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Salvador.

Depois de um longo crescimento acelerado, São Paulo finalmente apresentou um ponto de inflexão com desaceleração na curva do número de casos (Figura 82a). O $\bar{R}_0(t)$ passou a ser menor do que 1 (Figura 83a), aproximadamente 15 dias antes do tempo final de simulação e, de fato, esta desaceleração parece consistente. A contribuição decisiva para essa desaceleração foi a diminuição do percentual de circulação (Figura 82d), que nos últimos dias da simulação passou a ficar abaixo de 50%.

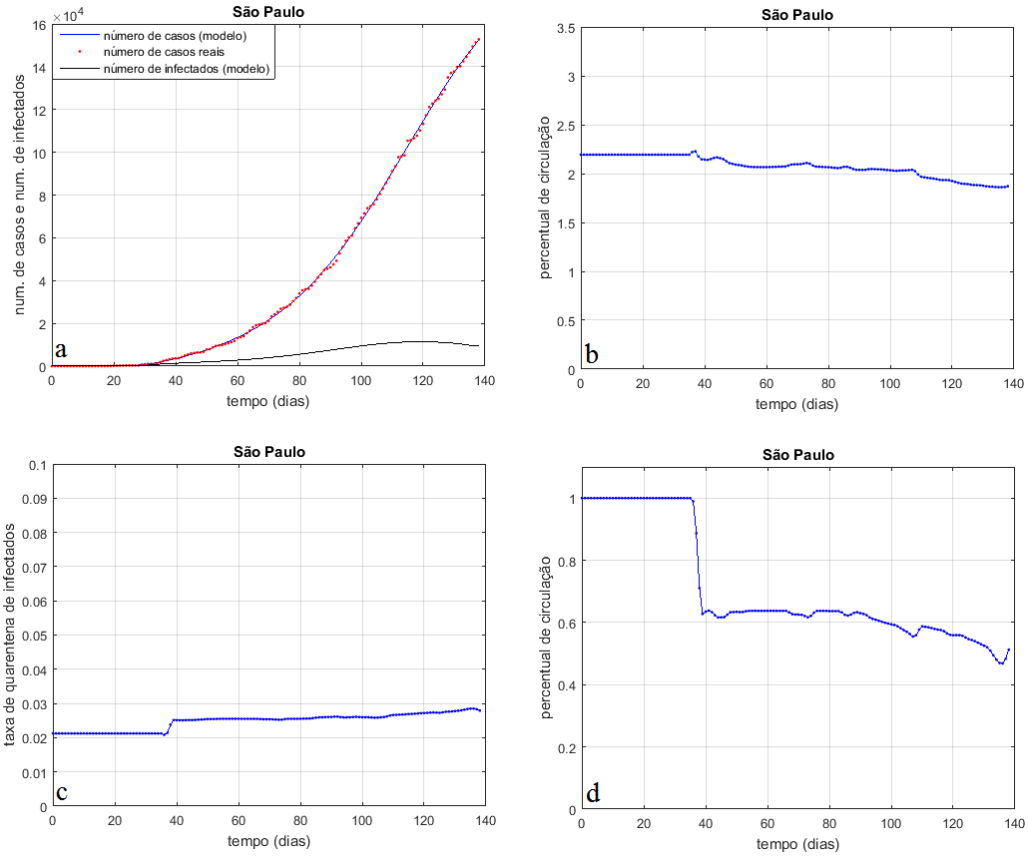


Figura 82. Identificação paramétrica e simulação para a cidade de São Paulo, de 25/02/2020 a 13/07/2020.

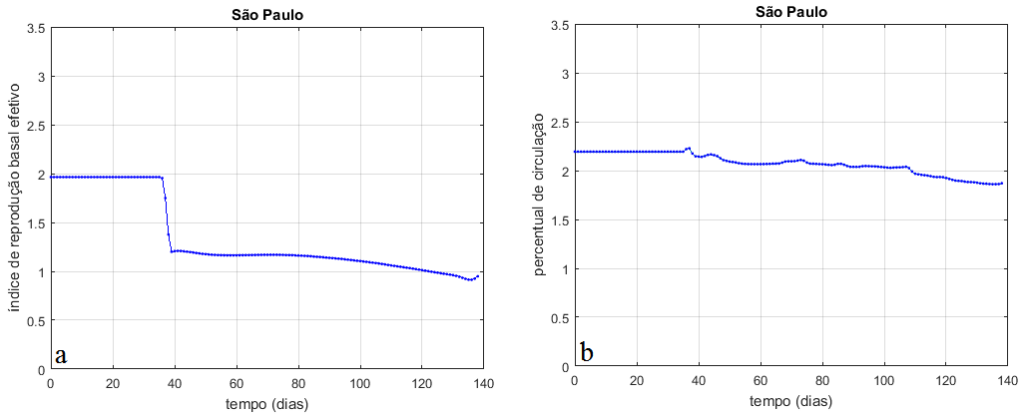


Figura 83. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de São Paulo.

São Luís vinha em uma boa trajetória desacelerada do número de casos (Figura 84a). Porém, provavelmente em razão de flexibilização do isolamento social, o percentual de circulação cresce significativamente a partir de $t = 98\,dias$ (Figura 84d) e há uma nova aceleração no crescimento do número de casos, denotada também pelo fato do $\bar{R}_0(t)$ voltar a ser maior do que 1 em $t = 102\,dias$ (Figura 85a).

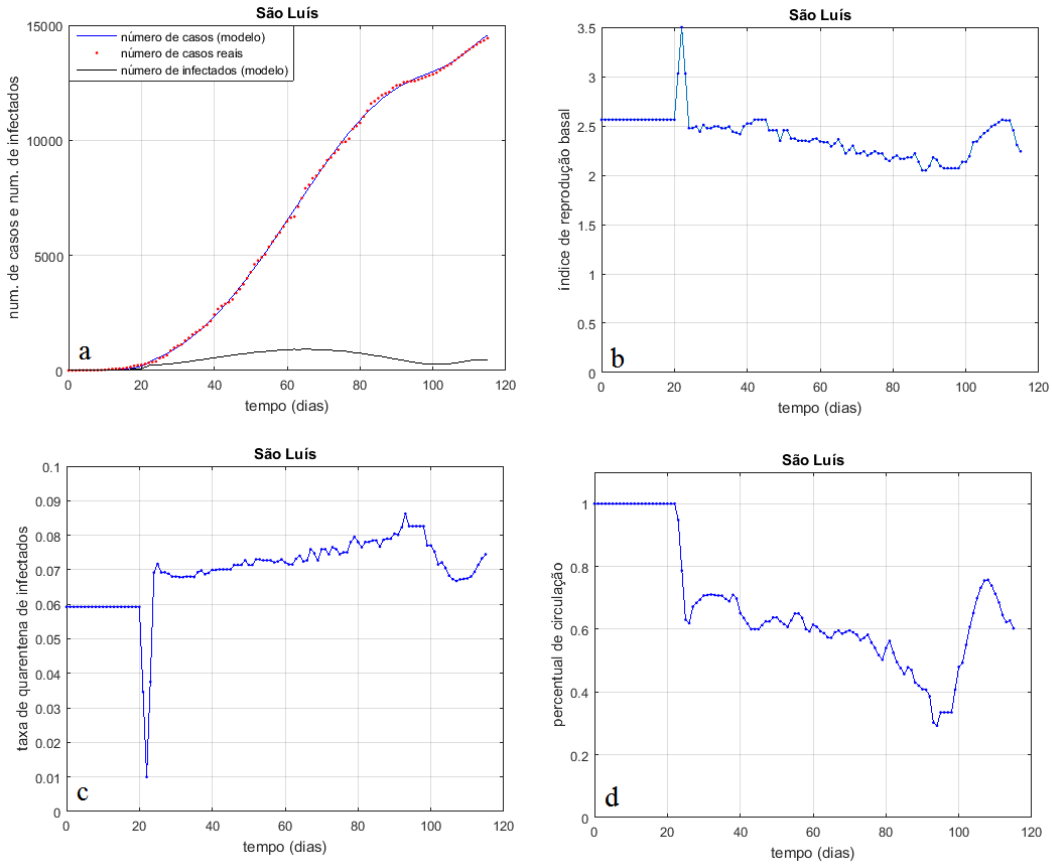


Figura 84. Identificação paramétrica e simulação para a cidade de São Luís, de 20/03/2020 a 13/07/2020.

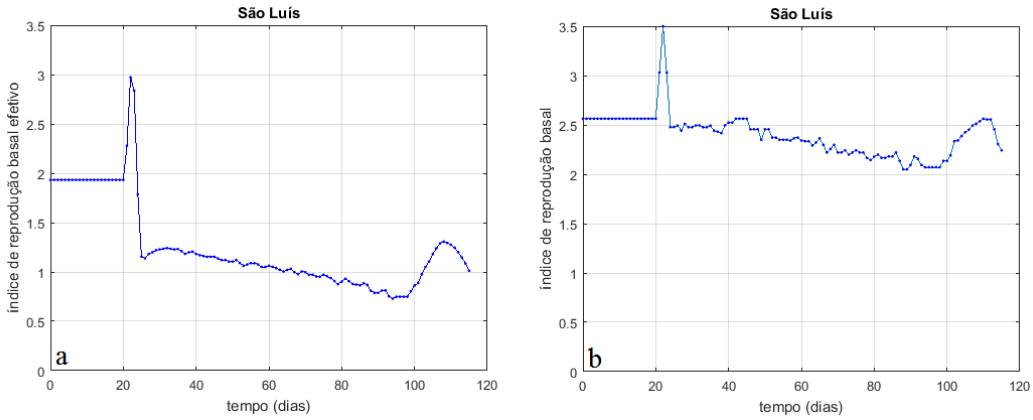


Figura 85. $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de São Luís.

A Tabela IV resume os resultados das simulações confrontadas aos dados reais, mais especificamente comparando os percentuais de circulação real e previsto pelo modelo. Além desta comparação, há também os valores da taxa de quarentena de infectados $k(t)$, do índice de reprodução basal $R_{0b}(t)$ e do índice de reprodução basal efetivo $\bar{R}_0(t)$. Obviamente, conforme visto nas figuras anteriores, todos os parâmetros da Tabela IV variam no tempo, de forma que os valores numéricos que compõem esta tabela correspondem ao último dia de simulação e dados reais de cada cidade. Lembramos que o $R_{0b}(t)$ é o índice de reprodução basal caso considerássemos a cidade sem nenhuma medida de prevenção, equivalente a considerar $k(t) = 0$ e $p(t) = 1$. Já o $\bar{R}_0(t)$ corresponde ao índice de reprodução basal efetivamente praticado na cidade, incluindo todas as medidas de prevenção.

Assim como no caso da Tabela III, na Tabela IV P_{real} corresponde ao percentual de circulação identificado via telefonia celular para todo o estado ao qual pertence a cidade em questão, dados obtidos do site inloco (https://mapabrasileirodacovid.inloco.com.br/pt/), enquanto que o P_{mod} corresponde ao percentual de circulação previsto pelo modelo para esta mesma cidade. Considerando a média de todas as cidades, essa média para o P_{real} foi 59.18, enquanto a média do P_{mod} foi 58.33. Na Figura 86 vemos uma confrontação entre P_{real} e P_{mod} para todas as cidades, representadas na abcissa com os números de 1 a 11, numeração correspondente à mesma ordem disposta na Tabela IV. Tanto a Tabela IV quanto a Figura 86 permite-nos concluir que os percentuais de circulação previstos pelo modelo são coerentes com os praticados realmente nas cidades. A maior diferença em termos absolutos foi igual a 11.39%. Como o site inloco só disponibiliza gratuitamente dados de isolamento social para os estados, assumimos que estes dados, por aproximação, são válidos para as capitais, muito embora possa haver alguma capital (como é o caso de Fortaleza) para a qual está em vigor um isolamento mais rigoroso, medida tomada apenas

para a capital, de forma que é possível que haja mais isolamento social em Fortaleza do que no restante do estado do Ceará.

Tabela IV. Valores numéricos de parâmetros relativos ao último dia do conjunto de dados reais e simulados (atualização com relação à Tabela III).

Cidade	P_{mod} (%)	P_{real} (%)	$k(t)$	$R_{0b}(t)$	$\bar{R}_0(t)$
1. Belém	58.472	61.3	0.0620	2.013	0.926
2. Brasília	48	58.7	0.052	2.396	1.071
3. Cuiabá	65.955	61.8	0.0587	1.872	0.924
4. Fortaleza	42.667	52.3	0.0752	2.228	0.0873
5. Pelotas	68.605	59.2	0.0524	2.123	1.126
6. Porto Alegre	56.625	59.2	0.0516	2.373	1.124
7. Rio Grande	70.59	59.2	0.0468	2.292	1.236
8. Rio de Janeiro	55.5	58.9	0.0688	2.086	0.916
9. Salvador	63.658	59	0.053	1.860	0.923
10. São Luís	60.25	61	0.0745	2.242	1.012
11. São Paulo	51.263	60.4	0.0279	1.871	0.949

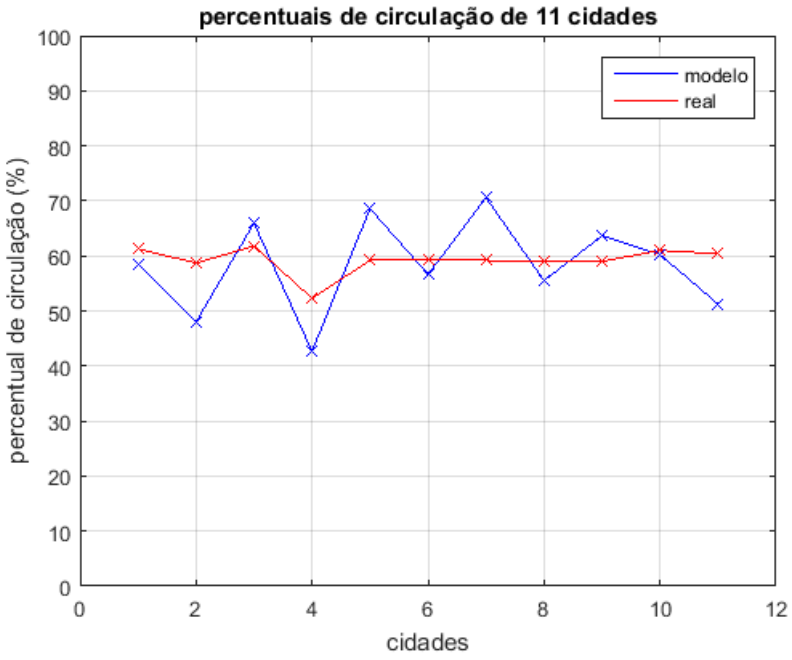


Figura 86. Percentuais de circulação reais e previstos pelo modelo, dados relativos a 11 cidades.

6.4 Previsão com o MdSIR

Esta seção é dedicada a um exercício de previsão com o MdSIR. Os parâmetros do modelo são considerados constantes e iguais aos identificados no último dia de simulação dos resultados obtidos no capítulo anterior, ou seja, os parâmetros são os apresentados na Tabela III. Tomando como exemplo a cidade de Belém, os parâmetros são:

$$p(t) = p = 0.57$$

$$k(t) = k = 0.066 \quad (15)$$

$$R_{0b}(t) = R_{0b} = 1.841$$

Com estes parâmetros da Tabela III realizamos simulações com os tempos finais equivalentes a 18 ou 19 dias, correspondentes aos dias de atualização do presente capítulo. Por exemplo, o último dia de dados do capítulo anterior para Belém foi 24/06/2020 (os parâmetros (15) são relativos a esse dia). O último dia de dados para Belém no capítulo atual foi 12/07/2020. Portanto, para Belém, o tempo final de simulação será igual a 18 dias. Evidentemente, nós já conhecemos os dados reais de Belém até 12/07/2020. Porém, o modelo não "aprendeu" estes dados: como os parâmetros são os identificados em 24/06/2020 e estes são mantidos constantes, o modelo não "conhece" os dados reais da atualização feita no presente capítulo, de forma que se trata, de fato, de um exercício de previsão.

As figuras apresentadas a seguir foram todas geradas sob o mesmo padrão, de forma a facilitar as análises dos resultados. Nelas, os pontos que aparecem em verde são relativos ao número de casos reais não computados, ou seja, desconhecidos do modelo. Portanto, o exercício de previsão começa a partir do tempo correspondente ao início dos pontos em verde, de forma que o erro de previsão é formado a partir da diferença entre pontos verdes e curva em azul (curva número de casos do modelo). Considerando n_{pp} igual ao número de pontos do intervalo de tempo correspondente à previsão (assumirá valores iguais a 18 ou 19), o erro relativo da previsão (e_p) pode ser calculado a partir da seguinte equação:

$$e_p = \frac{1}{n_{pp} y_{rm}} \left\{ \sum_{i=1}^{n_{pp}} [(y_{ri} - y_{si})^2] \right\}^{\frac{1}{2}} \quad (16)$$

onde $y_{ri}, i = 1, \dots, n_{pp}$ corresponde aos pontos número de casos reais, y_{rm}é o valor médio destes pontos e y_{si} são os pontos número de casos da simulação.

Nas Figuras 87 a 97 podem ser visualizados os resultados das simulações confrontadas aos dados reais, correspondentes às 11 cidades da Tabela IV. Em todos os casos, enquanto a comparação é entre pontos vermelhos e curva azul, os resultados são excelentes, denotando que o modelo reproduz bem os dados reais que serviram de base para a identificação paramétrica, conforme já era conhecido de resultados anteriores. A novidade é a comparação entre os pontos em verde e a curva azul, ou seja, comparação entre dados reais e simulados, no período de previsão, sendo este o período de interesse para a aferição do caráter preditivo do modelo. Passaremos então a analisar o comportamento do modelo em cada cidade, com ênfase nos respectivos períodos de previsão.

Em Belém (Figura 87), a previsão foi boa durante os primeiros 7 dias. A partir do oitavo dia houve um descolamento entre dados reais e simulados: o número de casos reais passou a ser maior, provavelmente em razão de mudanças no isolamento social praticado na cidade.

Em Brasília (Figura 88) a previsão foi boa nos primeiros 14 dias, de forma que apenas nos últimos 5 dias o número de casos reais apresentou uma diminuição com relação ao número de casos simulados.

A previsão para Cuiabá (Figura 89) foi boa para quase todos os dias. As exceções foram os últimos três dias, quando deve ter havido algum problema de retardo na detecção/informação dos dados reais: o penúltimo dia teve o mesmo número de casos que o dia anterior e isto provocou o distanciamento entre modelo e dados reais.

O modelo se aproximou bem dos dados reais nos primeiros 9 dias de previsão para Fortaleza (Figura 90). Porém, aconteceu um fenômeno interessante: no décimo dia houve um acréscimo no número de casos reais um pouco maior do que o padrão de crescimento até então e isso fez com que a curva de dados reais seguisse um pouco acima da simulação, mas com a mesma curvatura, apenas deslocada na vertical.

Com exceção dos dois últimos pontos, em Pelotas (Figura 91) a previsão pode ser considerada boa no restante dos pontos. Isto significa que a cidade não apresentou mudanças significativas em termos de isolamento social, ou seja, os parâmetros identificados no dia 24/06/2020 continuaram aproximadamente válidos até o dia 11/07/2020 (nos dias 12 e 13/07 houve um aumento nos dados reais com relação aos simulados).

Porto Alegre (Figura 92) a previsão apresentou boa precisão para todos os pontos. Isto significa que os parâmetros identificados no dia 24/06/2020 continuaram válidos até o dia 12/07/2020. É interessante observar a Figura 74d para entender o porquê desta boa precisão em todo o intervalo de previsão: o percentual de circulação apresentou oscilações de pequena amplitude sobre uma média constante, para os últimos 20 dias. Como este parâmetro não sofreu alteração significativa, a previsão continuou válida em todo o referido intervalo.

Em Rio Grande (Figura 93) a previsão foi boa em praticamente todos os pontos. A única exceção foi o último ponto, para o qual o número de casos reais apresentou aumento superior ao verificado na simulação.

No Rio de Janeiro houve boa previsão para todos os pontos (Figura 94). O crescimento do número de casos continuou a mesma tendência desacelerada, de forma que o modelo captou e previu bem essa tendência.

Em Salvador (Figura 95) houve um crescimento abrupto no terceiro dia do período de previsão. Após este salto fruto dessa descontinuidade na curva do número de casos reais, esta curva continuou crescendo de forma aproximadamente paralela ao modelo.

Em São Luís (Figura 96), mudanças na política de isolamento social provocaram um aumento do número de casos a partir do sexto dia do período de previsão, provocando um descolamento entre dados reais e simulação.

A simulação conseguiu prever bem o comportamento real em São Paulo (Figura 97) até o décimo dia. A partir do décimo primeiro dia houve um descolamento entre dados reais e simulação: a situação em termos de ações preventivas na capital paulista provocou uma desaceleração um pouco mais acentuada nos dados reais, comparada à desaceleração que continuou acontecendo na simulação.

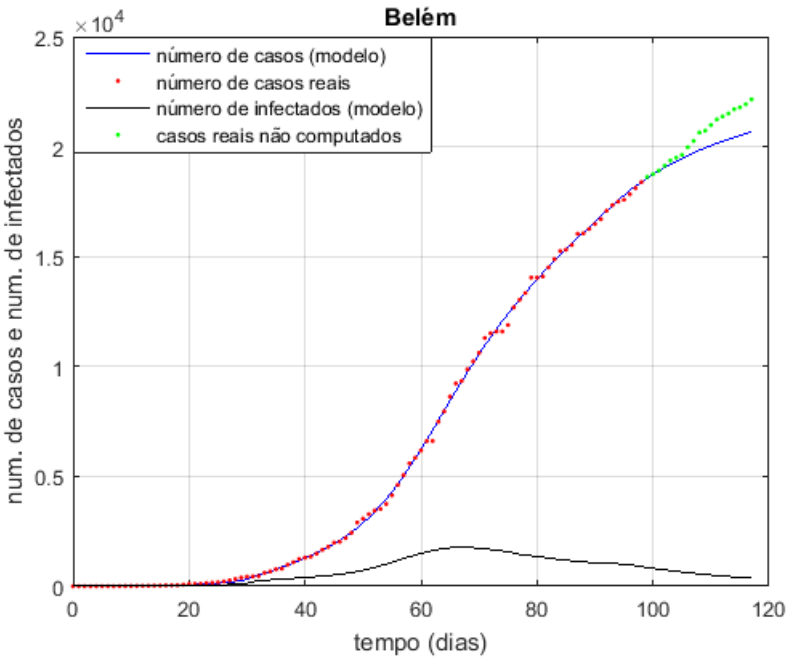


Figura 87. Simulação e dados reais para Belém, dados reais de 18/03 a 12/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 208, erro relativo $e_p = 0.0212$.

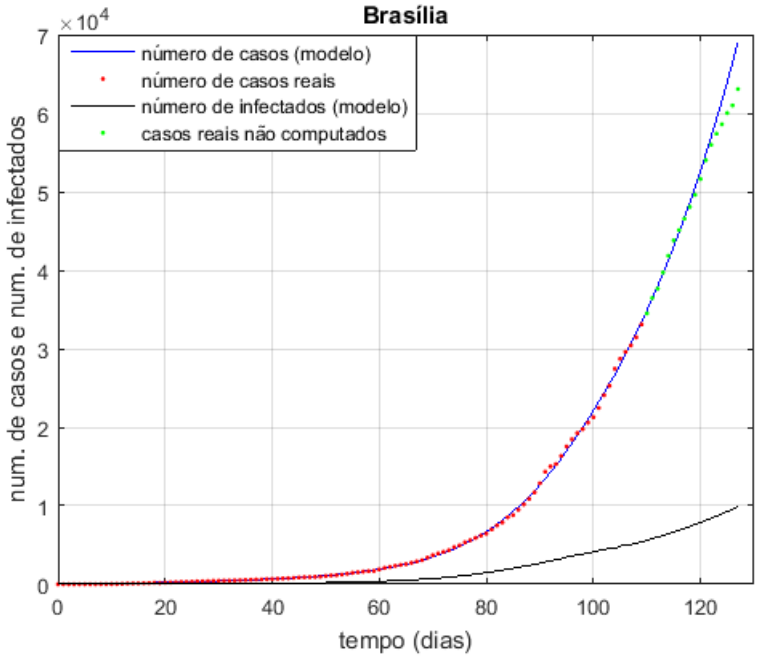


Figura 88. Simulação e dados reais para Brasília, dados reais de 07/03 a 12/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 572, erro relativo $e_p = 0.0413$.

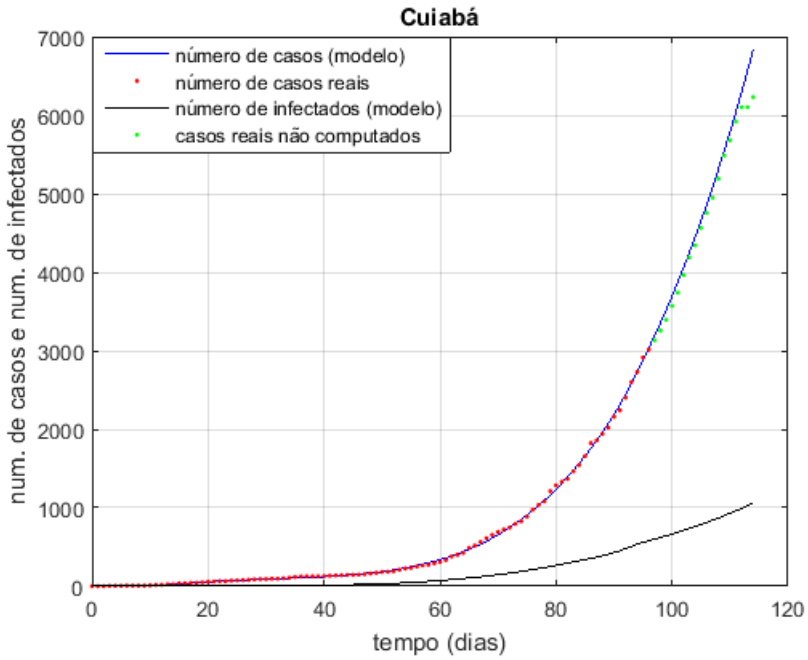


Figura 89. Simulação e dados reais para Cuiabá, dados reais de 20/03 a 12/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 50, erro relativo $e_p = 0.0352$.

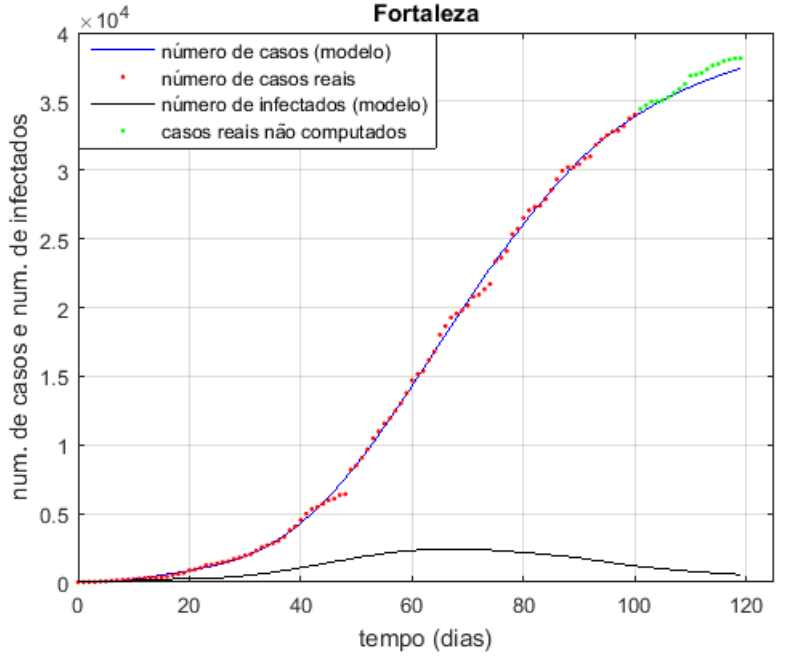


Figura 90. Simulação e dados reais para Fortaleza, dados reais de 16/03 a 13/07, previsão de 24/06/2020 a 13/07/2020, erro médio = 216, erro relativo $e_p = 0.0112$.

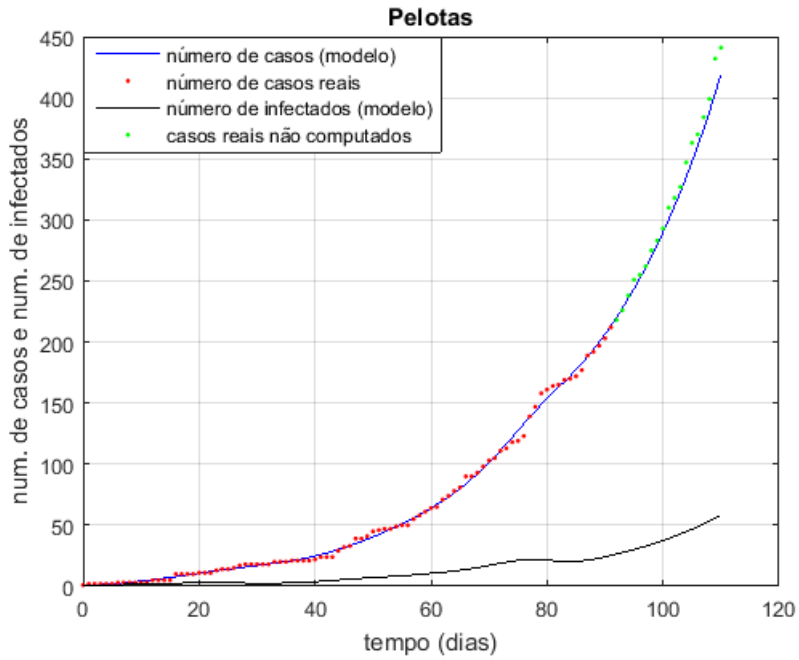


Figura 91. Simulação e dados reais para Pelotas, dados reais de 25/03 a 13/07, previsão de 24/06/2020 a 13/07/2020, erro médio = 2.85, erro relativo $e_p = 0.0231$.

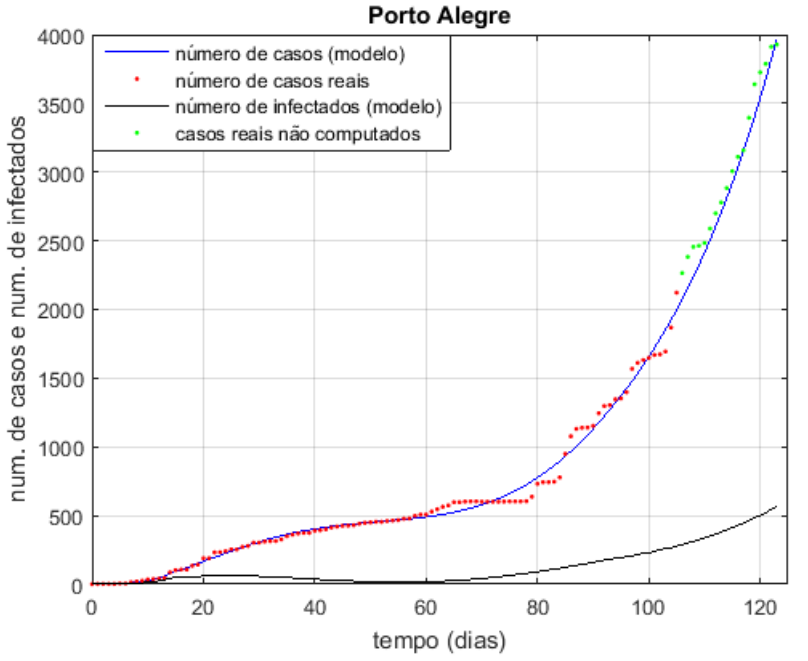


Figura 92. Simulação e dados reais para Porto Alegre, dados reais de 11/03 a 12/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 42, erro relativo $e_p = 0.0390$.

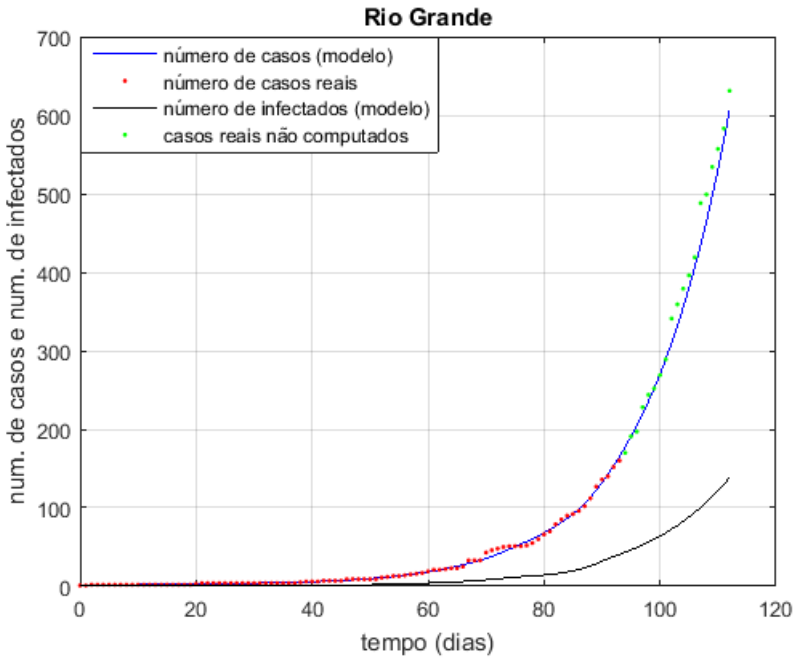


Figura 93. Simulação e dados reais para Rio Grande, dados reais de 23/03 a 13/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 5.36, erro relativo $e_p = 0.0526$.

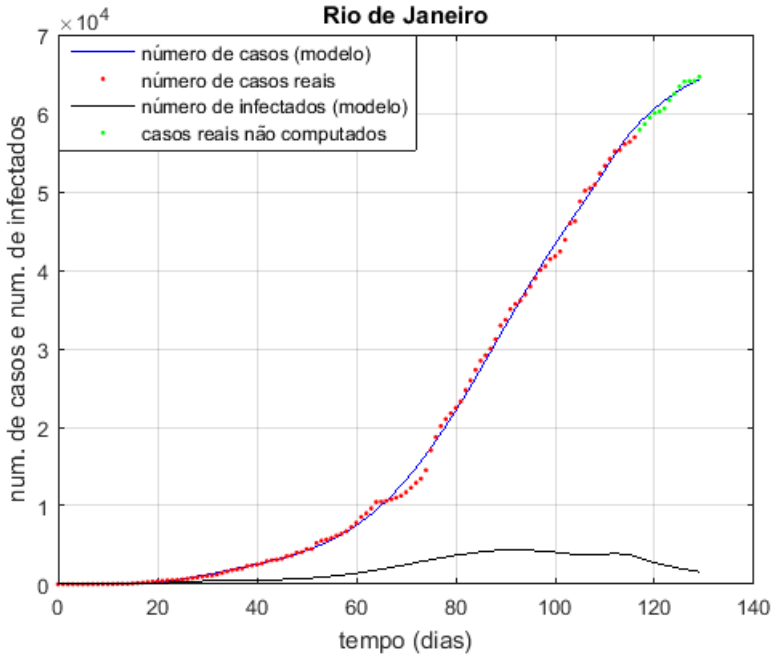


Figura 94. Simulação e dados reais para Rio de Janeiro, dados reais de 06/03 a 13/07, previsão de 23/06/2020 a 13/07/2020, erro médio = 583, erro relativo $e_p = 0.0253$.

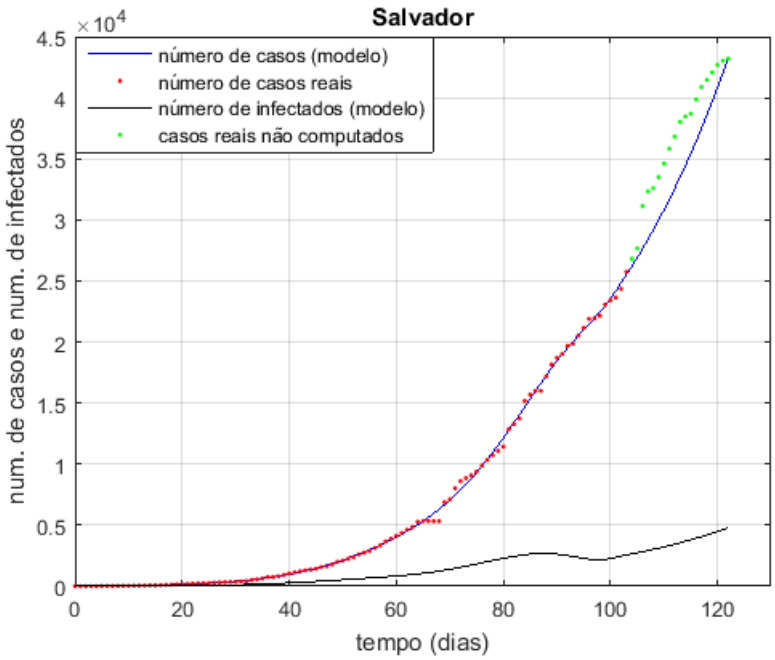


Figura 95. Simulação e dados reais para Salvador, dados reais de 03/03 a 13/07, previsão de 24/06/2020 a 13/07/2020, erro médio = 740, erro relativo $e_p = 0.057$.

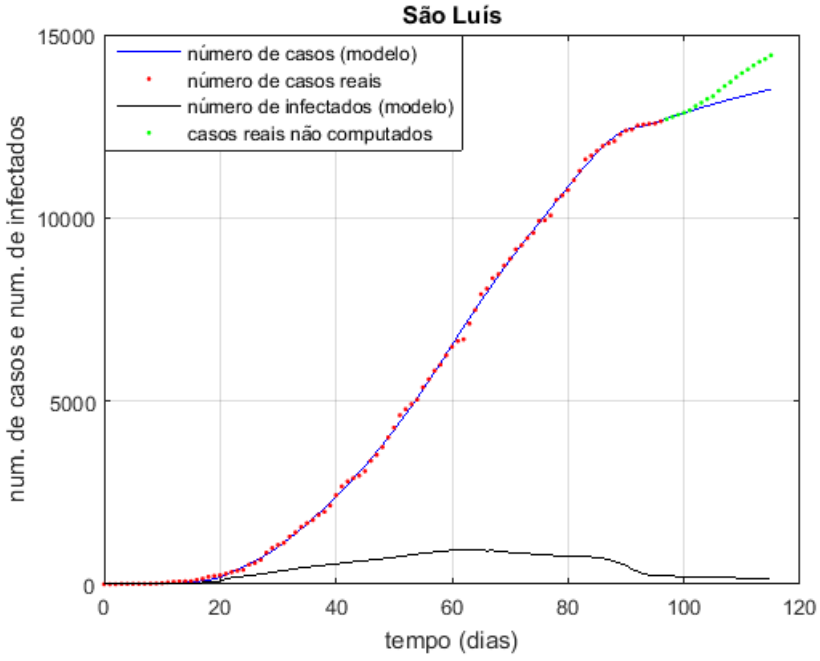


Figura 96. Simulação e dados reais para São Luís, dados reais de 20/03 a 13/07, previsão de 24/06/2020 a 13/07/2020, erro médio = 120, erro relativo $e_p = 0.0158$.

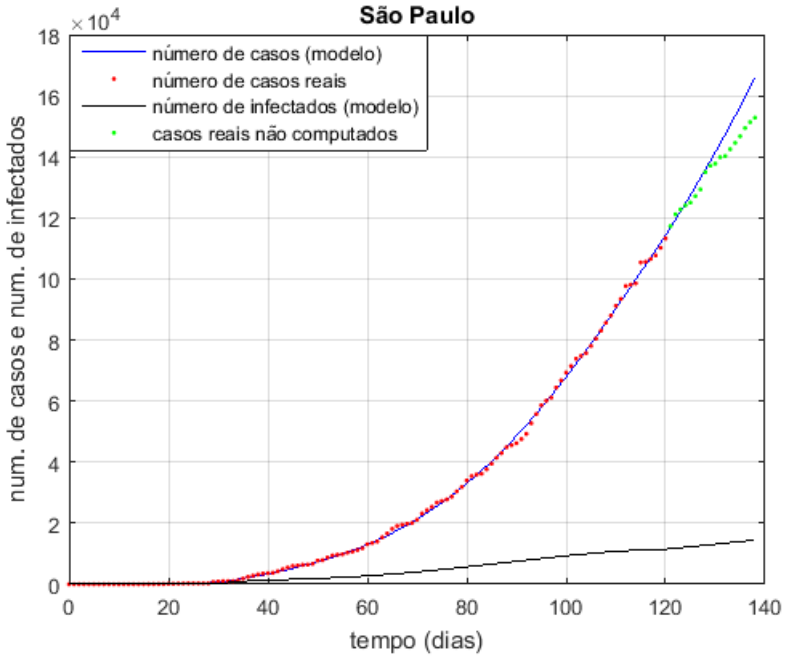


Figura 97. Simulação e dados reais para São Paulo, dados reais de 25/02 a 12/07, previsão de 24/06/2020 a 12/07/2020, erro médio = 1614, erro relativo $e_p = 0.0330$.

Na Tabela V são mostrados os erros relativos das previsões para as onze cidades, erros estes calculados a partir da equação (16). Consideramos duas situações para efeito dos cálculos dos erros: todo o tempo do intervalo de previsão, equivalente a 19 dias (em cinco cidades são 18 dias); apenas os primeiros 10 dias do intervalo de previsão. Na Figura 98 os erros relativos são mostrados em um gráfico de barras. A numeração das cidades na abcissa obedece à mesma ordem vista na Tabela V. A previsão pode ser considerada boa, de uma forma geral. Os erros relativos mostraram-se inferiores a 6%, nos 19 dias de previsão. Fica evidente que os erros diminuem quando consideramos um tempo menor de previsão: a média do e_p para 10 dias foi 1.84%, enquanto que para 19 dias foi 3.2%.

Tabela V. Erros relativos considerando 10 e 19 dias de previsão (média do erro relativo (e_p): em 19 dias = 3.2%; em 10 dias = 1.84%).

Cidade	Erro relativo em 19 dias (%)	Erro relativo em 10 dias (%)
1. Belém	2.22	0.75
2. Brasília	4.13	1.21
3. Cuiabá	3.52	1.47
4. Fortaleza	1.12	0.88
5. Pelotas	2.31	1.29
6. Porto Alegre	3.90	3.43
7. Rio Grande	5.26	2.57
8. Rio de Janeiro	2.11	2.09
9. Salvador	5.70	4.60
10. São Luís	1.58	0.62
11. São Paulo	3.30	1.32

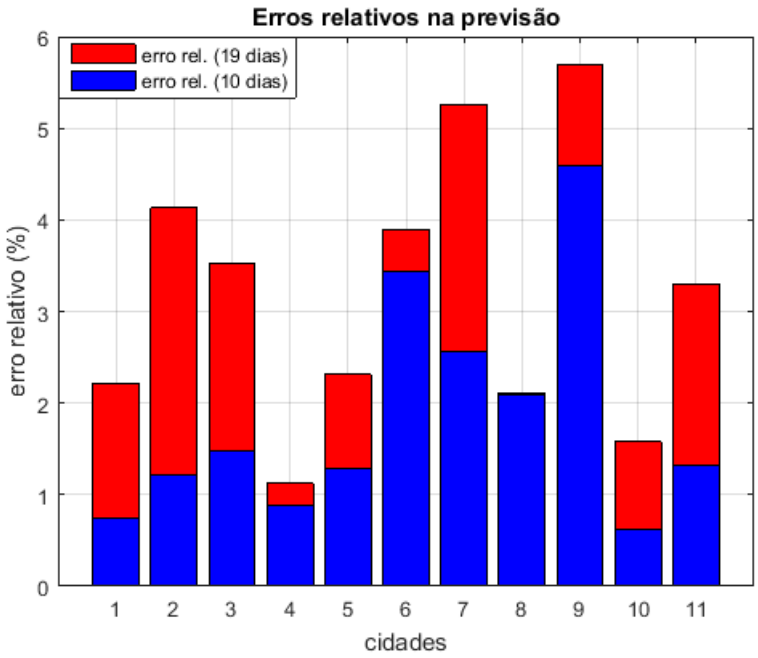


Figura 98. Erros relativos na previsão para as 11 cidades da Tabela V.

6.5 Comentários Finais

Acreditamos que uma das principais utilidades de um modelo dinâmico seja a sua capacidade de testar diferentes cenários, possíveis de acontecer em razão das políticas de prevenção adotadas em países, regiões ou cidades. Dentre as políticas de prevenção, o isolamento social tem sido a principal, usada como variável de controle epidêmico tanto no Brasil quanto no resto do planeta. Neste sentido, o MdSIR serve muito bem a essa finalidade, conforme será demonstrado no próximo capítulo, quando abordaremos e explicaremos em detalhes o simulador desenvolvido e implementado na forma de um aplicativo, o Simcovid 2.0. Entretanto, quando se pretende utilizar um modelo para a realização de previsões com boa precisão, há sem dúvida uma limitação, correlata ao tempo de previsão. O estudo desenvolvido neste capítulo mostrou que uma previsão com boa precisão, que possibilitaria erro relativo inferior a 5%, seria confiavelmente possível para um tempo de previsão menor ou igual a 15 dias. Esta limitação temporal para uma previsão com boa precisão se deve a uma característica específica dessa pandemia de COVID-19: ações externas à dinâmica, oriundas de decisões governamentais principalmente correlatas ao isolamento social, podem alterar as características dinâmicas e, certamente, não há como prever essas ações externas, que podem acelerar ou desacelerar a propagação epidêmica.

Capítulo VII

Simcovid – Simulador para a Dinâmica da COVID-19

7.1 Introdução

Neste capítulo será apresentado o Simcovid, software simulador desenvolvido a partir dos resultados dos estudos dos modelos dinâmicos apresentados nos capítulos anteriores, e que baseia sua simulação no modelo MdSIR. Os aspectos do desenvolvimento do software e da operação do aplicativo são descritos, bem como as motivações e requisitos do projeto.

Logo após terem sido feitos os primeiros estudos de modelos epidemiológicos, e deles ter resultado o modelo MdSIR, surgiu a ideia de criação de um software que pudesse ser facilmente distribuído e instalado por usuários interessados em analisar o comportamento da Covid-19 em sua região. Os programas desenvolvidos em sistemas de álgebra computacional (ou CAS, do inglês *Computer Algebra System*), embora efetivos na resolução de problemas, necessitam que esses softwares sejam instalados nos computadores que executarão a simulação, o que acrescenta dificuldade de utilização, e eventualmente custos de aquisição elevados de licenças de uso, o que poderia afastar o interesse daqueles que não trabalham com esse tipo de software. Além de disponibilizar uma ferramenta para análise da Covid-19, uma motivação adicional para a criação do Simcovid foi a conscientização das pessoas sobre a questão do distanciamento social, até então a forma mais eficaz de redução da contaminação.

Os requisitos do projeto foram identificados a partir dessas motivações e das características do modelo dinâmico e de funcionamento. Seriam eles:

- Ser de fácil instalação (ou ao menos similar a outros softwares de uso comum);
- Ter uma interface gráfica com o usuário simples e intuitiva, porém completa;
- Ter uma descrição, na interface, da natureza dos parâmetros e da simulação em uso;
- Ser multiplataforma, podendo rodar em diferentes sistemas operacionais;
- Usar o modelo MdSIR, o mais próximo possível do modelo implementado em CAS;

- Apresentar resultados de forma visual, preferencialmente, embora os valores numéricos obtidos na simulação pudessem ser consultados, se desejado;
- Ser gratuito aos usuários.

A primeira versão do Simcovid foi desenvolvida no sistema operacional Kubuntu Linux, sendo apenas copiado para um computador com sistema operacional Microsoft Windows e executado sem nenhuma modificação. Além desses testes de funcionamento, o software foi executado com sucesso em um *smartphone* com sistema operacional Android 9.1. A Figura 99 apresenta o protótipo funcional no *smartphone*.

Apesar de funcional, estes primeiros testes foram feitos com todo o ambiente de desenvolvimento da aplicação instalado em cada equipamento utilizado. Para possibilitar a instalação, foram estudadas algumas possíveis ferramentas, obtendo-se sucesso para o caso do Microsoft Windows. Então, foi definido que o foco seria primeiro disponibilizar publicamente o Simcovid para este sistema operacional. Nesta versão, o usuário deveria prover todos os parâmetros da simulação, o que pode ser difícil de obter, principalmente na questão do índice de reprodução basal (R_0).

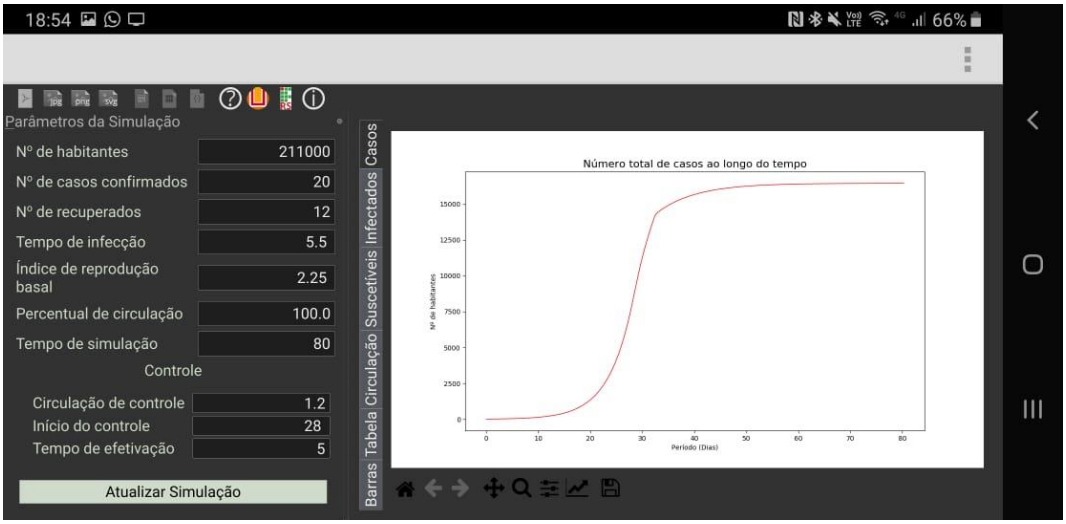


Figura 99. Tela principal do Simcovid 1.0, executando em um *smartphone* Android.

A dificuldade de obtenção dos parâmetros relativos a cidades/regiões brasileiras por parte da maioria dos usuários, bem como os estudos subsequentes de identificação paramétrica, motivou a definição de novas funcionalidades para o Simcovid, que seriam:

- Prover um banco de dados de parâmetros de cidades, para que os usuários possam ter os parâmetros mais condizentes com o cenário a ser analisado, aproximando assim as simulações da realidade da cidade em questão;

- Apresentar dados relativos às cidades relacionadas no banco de parâmetros, de forma a contextualizar a realidade dessas cidades e descrever como os parâmetros foram identificados a partir dos dados coletados;
- Implementar um sistema de atualização automática e transparente do banco de parâmetros das cidades, de forma a manter o banco de dados atualizado com a realidade das cidades mantidas.

Assim, foi implementada a versão 2.0 do Simcovid, que será apresentada nas próximas seções.

7.2 Aspectos do Desenvolvimento do Software

7.2.1 Escolhas de projeto

As ferramentas para o desenvolvimento do aplicativo foram definidas em função dos requisitos do projeto, funcionais e não funcionais. A escolha da linguagem de programação e plataforma de desenvolvimento levou em consideração a disponibilidade de bibliotecas matemáticas necessárias para o porte do código Matlab utilizado para o desenvolvimento das simulações, a disponibilidade de bibliotecas para a criação de interfaces gráficas para usuários, e a possibilidade de uso do aplicativo em diferentes sistemas operacionais com um mínimo de adaptações de código. Assim, foi adotada a linguagem Python como ferramenta primária de desenvolvimento (PSF, 2019).

Python é uma linguagem dinâmica, interpretada e multiplataforma, com muitas aplicações em aplicações científicas graças ao grande número de bibliotecas disponíveis para diversas áreas do conhecimento (LANGTANGEN, 2016; LINGE & LANGTANGEN, 2020). A popularidade da linguagem provê uma grande base de recursos de aprendizagem, documentação e uma ativa comunidade de usuários, que auxilia enormemente no aprendizado e solução de problemas de programação. Além disso, há variedade de APIs (*Application Program Interface*) para diferentes bibliotecas de interfaces gráficas para usuário (BORGES, 2014). Por esses motivos, Python é uma das linguagens de programação mais utilizadas nas áreas de desenvolvimento de software científico, sejam aplicações console/desktop, web, e cada vez mais para sistemas embarcados (CASS, 2020). Por fim, Python é livre, e sua instalação básica é gratuita, assim como muitas distribuições, que além do interpretador, reúnem bibliotecas, APIs e ambientes de desenvolvimento para facilitar o uso de Python.

A parte de modelagem e simulação utilizou essencialmente as bibliotecas NumPy e SciPy, voltadas para a modelagem e processamento numérico (OLIPHANT, 2007; PAULI *et al*., 2020), e Matplotlib/Seaborn para a criação dos gráficos (HUNTER, 2007; WASKOM et al., 2020). Além de adequadas para o processamento numérico, as bibliotecas em questão apresentam similaridades com recursos disponíveis em sistemas de álgebra computacional, como o Matlab e o Octave, o que facilitou o porte dos códigos originais da simulação.

Para a criação da interface gráfica com o usuário foi escolhido o Qt, que é multiplataforma, porém se integra ao layout do sistema operacional hospedeiro da aplicação (QT, 2019). Apesar de ter sido escrito em C++, existem APIs para Python, sendo a PyQt (WILLMAN, 2020) utilizada no Simcovid. Além disso, outras bibliotecas específicas de Python conseguem se integrar ao Qt, fator importante para poder apresentar os gráficos gerados em Matplotlib nos painéis com abas da janela da aplicação.

A exportação dos resultados de simulação foi parcialmente resolvida pelo uso da Matplotlib, uma vez que ela incluiu outras bibliotecas das quais dependia para a parte gráfica. Assim, exportação para os formatos pdf, svg e png foram providos pelas próprias bibliotecas. A exportação dos dados numéricos foi resolvida em parte pela biblioteca numpy, que permitia importar e exportar arquivos no formato textual csv (campos separados por vírgula). Para exportar dados nos formatos json e xlsx (este do Microsoft Excel), foi utilizada a biblioteca Pandas, especializada em processamento de dados estruturados (MCKINNEY, 2010).

Por fim, o formato adotado para os arquivos dos parâmetros das cidades e o arquivo de configuração do aplicativo foi o json, que além do Pandas, tem uma biblioteca nativa do interpretador Python. Outra biblioteca de manipulação de arquivos empregada foi a zipfile, para manipulação dos arquivos no formato zip que são utilizados no processo de atualização dos parâmetros e informações sobre os diferentes estudos de caso (BERNARD, 2016).

7.2.2 Componentes e funcionalidades

O Simcovid utilizou o paradigma da orientação a objetos na sua arquitetura, aproveitando as características nativas da linguagem e da maioria de suas bibliotecas. As classes de componentes foram projetadas de forma a considerar o maior desacoplamento possível entre a modelagem/simulação e a apresentação/visualização, a fim de facilitar futuras manutenções do software, e mesmo mudanças na visualização por adoção de outra interface gráfica sem comprometer o cerne do aplicativo, que é a modelagem e simulação.

Os diferentes componentes do software e seus relacionamentos são apresentados na Figura 100.

O modelo e simulação do software são definidos pelas classes Model e Simulation (abstratas), que são especializadas nas classes ModelMdSIR e SimulationMdSIR. Uma instância de ModelMdSIR representa o modelo dinâmico MdSIR, empregado no Simcovid. Instâncias de SimulationMdSIR utilizam objetos ModelMdSIR para executar a simulação, empregando o método LSODA de resolução de equações diferenciais ordinárias (PETZOLD, 1983), parte da biblioteca ODEPACK (HINDMARSH, 1983), utilizada por Numpy.

A interface gráfica com o usuário é implementada nas classes WndPrincipal e WParametros. A primeira é a responsável pela apresentação e ações da janela principal da aplicação, enquanto a segunda define o layout e as ações do painel de parâmetros de simulação, sendo usada como parte da instância de WndPrincipal que provê a interface com o usuário. Entre essas ações estão a obtenção dos parâmetros do modelo para apresentação no painel correspondente, a atualização do modelo/simulação com os parâmetros definidos no painel e o disparo de execução de simulação.

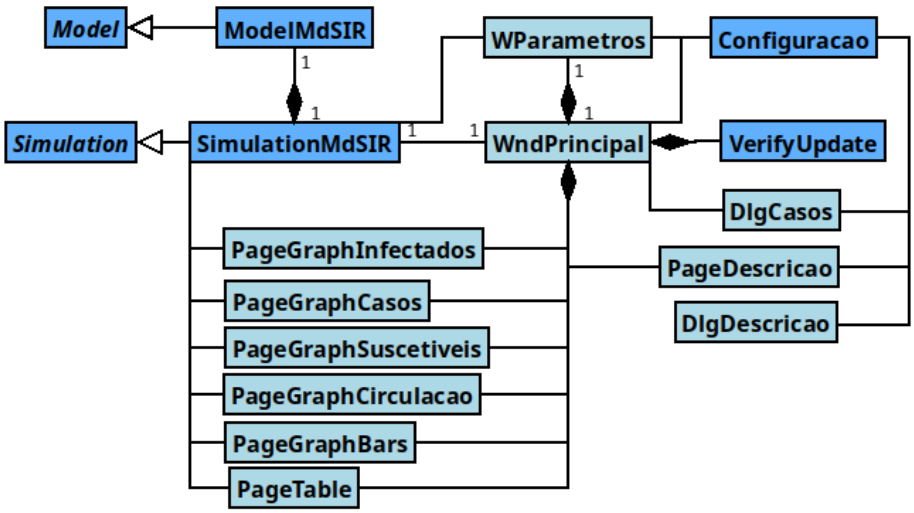


Figura 100. Diagrama de classes do Simcovid 2.0.

O resultado da simulação é retornado em um conjunto de *arrays* Numpy. Estes são acessados por instâncias das classes PageGraphCasos, PageGraphInfectados, PageGraphSuscetiveis, PageGraphCirculacao e PageGraphBars para disponibilizar um painel com abas de visualização gráfica dos resultados da simulação. Uma instância de PageTable cria uma aba no painel de resultados contendo uma tabela onde os valores numéricos do resultado da simulação podem ser lidos.

Uma instância da classe Configuracao é responsável por manter o estado de execução do software, bem como atualizar um arquivo para que o estado possa se manter entre execuções do Simcovid. Além de definições dos locais onde os arquivos da aplicação estão disponíveis e do estilo visual do aplicativo, a instância de configuração também mantém os parâmetros das cidades disponíveis para o modo Estudo de Casos, implementado na versão 2.0 do Simcovid, e seu estado de uso (se os dados já foram utilizados pelo usuário, ou se está atualizado).

A adição do modo Estudo de Casos no Simcovid 2.0 trouxe a necessidade de inclusão de *widgets* complementares à janela principal, como uma janela de listagem das cidades e o estado dos parâmetros destas (DlgCasos) e de descrição do cenário de um caso e da identificação paramétrica deste (DlgDescricao). As descrições de casos são formadas de um arquivo html e png, que são renderizadas na janela DlgDescricao para apresentar de forma estruturada o cenário e os gráficos correspondentes. Além desses componentes, foi criado uma classe para replicar a descrição de casos junto aos gráficos resultantes da simulação, chamada PageDescricao.

Um dos requisitos acrescentados na versão 2.0 do Simcovid foi a necessidade de manter atualizados os arquivos de parâmetros e descrições das cidades listadas no modo Estudo de Casos. Assim, foi implementado um módulo de atualização por uma instância da classe VerifyUpdate. O objeto em questão é executado em um *thread* (linha de execução) paralelo à interface gráfica do Simcovid, sempre que este é iniciado. Ele procura se comunicar com o servidor que hospeda as atualizações do Simcovid e obtém os dados de um arquivo manifesto, que descreve em formato json a versão mais recente de um arquivo no formato zip contendo todos os arquivos de atualização. Caso a data de atualização do manifesto seja mais recente do que a registrada na configuração do aplicativo, a instância de VerifyUpdate baixa o arquivo zip correspondente em um diretório temporário, extraindo os arquivos de parâmetros e de descrição de casos (incluindo as figuras), e substitui aqueles cuja data final de obtenção dos dados usados para identificação paramétrica sejam mais recentes que os registrados na configuração do Simcovid. A configuração do software, então, é atualizada com os novos dados. Isso possibilita também incluir novas cidades. O diagrama apresentado na Figura 101 ilustra esse processo.

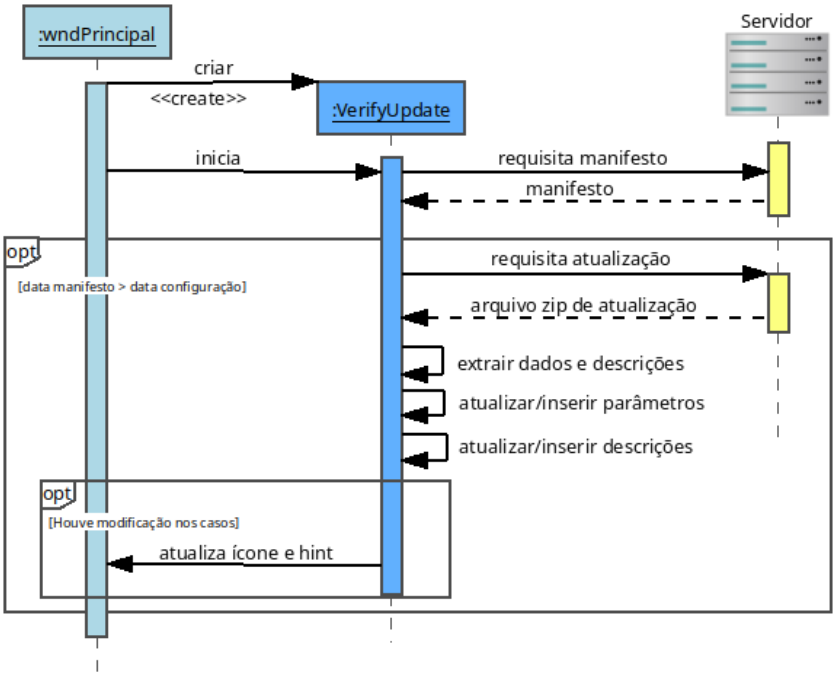


Figura 101. Diagrama de sequência do processo de atualização de cidades.

O protocolo de comunicação com o servidor é o HTTPS. Para implementar o componente de atualização de dados foi utilizado a biblioteca requests, com a json e a zipfile para a leitura e processamento dos arquivos. Todas essas são bibliotecas básicas do Python (BERNARD, 2016).

7.2.3 Empacotamento e instalação no Windows

As aplicações desenvolvidas em Python usualmente dependem do interpretador e das bibliotecas necessárias estarem instaladas no computador em que serão executadas, o que é uma característica dos softwares interpretados. Embora este procedimento seja usual para desenvolvedores e pesquisadores que utilizam ferramentas para escrever os seus softwares, os demais usuários não consideram este um procedimento trivial para instalar e executar aplicações. Assim, foi necessário buscar uma forma de empacotar a aplicação junto a uma ferramenta de instalação, no padrão usualmente empregado em sistemas operacionais como o Windows.

A ferramenta utilizada para o empacotamento da aplicação foi o PyInstaller. Segundo Cortesi (2020), o PyInstaller é um pacote Python, misto de aplicativo e biblioteca, capaz de criar um ambiente isolado a partir do qual todas as dependências para uma aplicação são inclusas, de maneira que esta não depende do que está instalado no sistema operacional hospedeiro para a sua execução. Ao se executar o PyInstaller, este analisa o código fonte da aplicação, detecta as bibliotecas Python empregadas, e monta uma distribuição onde o

código fonte, as bibliotecas e os binários necessários para a execução do programa (incluindo um interpretador Python) são todos reunidos. Além disso, um arquivo executável que dispara a aplicação e configura o ambiente para execução é criado, e ainda é possível definir um ícone gráfico para a aplicação.

Antes de empacotar o Simcovid, porém, foi necessário criar um ambiente virtual de desenvolvimento através do pacote virtualenv, que cria um ambiente isolado Python com um conjunto de bibliotecas independente das instaladas nativamente no sistema operacional, de forma que mesmo que o sistema sofra modificações, o ambiente virtual mantenha as bibliotecas e versões destas necessárias para o correto funcionamento da aplicação (BERNARD, 2016). A criação do ambiente virtual provê um interpretador Python e as bibliotecas internas ao interpretador (*built-in*). As demais bibliotecas necessárias para a aplicação e o PyInstaller devem ser instaladas nesse ambiente virtual. Para tanto, este ambiente provê um comando de gerenciamento de pacotes que pode instalar, atualizar e remover bibliotecas com versões diferentes das instaladas no sistema operacional. Com isso, apenas as bibliotecas realmente necessárias ao Simcovid foram inclusas no ambiente virtual e, com isso, a criação da distribuição do software teve seu tamanho reduzido em 77% do tamanho originalmente produzido sem o ambiente virtual, com as bibliotecas instaladas junto ao sistema operacional.

Após a criação da distribuição na forma de um diretório com um conjunto de arquivos, foi utilizado o software NSIS (*Nullsoft Scriptable Install System*) para criar um instalador para o Microsoft Windows (SZEKELY, 2019). O NSIS utiliza um arquivo de descrição das características do instalador/desinstalador de softwares desejado, tais como ícone do aplicativo, criação de entradas de menu e atalhos da área de trabalho, arquivos necessários para inclusão no aplicativo e ações de registro no sistema operacional. Com isso, a distribuição foi empacotada e comprimida em um arquivo executável, facilmente distribuível e de instalação intuitiva para usuários Windows. Durante a instalação, os arquivos do Simcovid são instalados no sistema do usuário, em um diretório padrão, e são criadas entradas no menu e um atalho na área de trabalho. Assim, após a instalação, o Simcovid se comporta como qualquer outro aplicativo do sistema.

7.3 Utilização do Simcovid

O Simcovid é distribuído como um aplicativo análogo ao do sistema operacional em que é executado. No Windows, por exemplo, ele é um executável, que durante a instalação é vinculado a um item de menu e um atalho com ícone. Para iniciá-lo, então, basta dar um

duplo clique neste ícone. De forma similar, o Simcovid será iniciado conforme o padrão adotado para disparar qualquer aplicativo no sistema operacional hospedeiro.

Uma vez iniciado, o Simcovid logo apresentará a tela inicial, similar à apresentada na Figura 102. Nela, é possível observar dois painéis flutuantes, o de parâmetros da simulação (à esquerda), e o de ajuda para os parâmetros (à direita, inicialmente oculto). A área central é o local onde os resultados da simulação são apresentados, bem como informações sobre as cidades analisadas no modo Estudo de Casos. Acima dessas áreas, estão visíveis as barras de ferramentas e o menu da aplicação. A forma de desenho da janela pode variar um pouco, de acordo com o sistema operacional em que o Simcovid é executado. Porém, a área interna das janelas permanece a mesma.

A utilização do simulador se dá principalmente pelo painel de parâmetros. As ações disponíveis nas barras de ferramentas são complementares a este, ou definem funcionalidades adicionais para o mesmo.

7.3.1 Parâmetros da Simulação e Ajuda

O painel de parâmetros pode ser melhor visualizado na Figura 103a. Os parâmetros da simulação são organizados por grupos. Os dois grupos iniciais definem as características da região em análise. O primeiro grupo corresponde aos dados populacionais: número de habitantes (N_h), número de casos confirmados, número de pacientes recuperados (R) e percentual de circulação (p). O número de infectados (I) será a diferença entre o número de casos confirmados e o número de recuperados. O segundo grupo define as características de contágio: tempo de infecção (T_{inf}) e índice de reprodução basal (R_0). Lembramos que este parâmetro corresponde ao R_{0b} definido nos capítulos anteriores. Além destes parâmetros, o segundo grupo inclui a taxa de quarentena (k), que corresponde ao percentual de infectados que são detectados e são isolados em hospitais ou nas próprias residências, de modo a não contribuírem mais com a contaminação.

Logo a seguir temos a definição de dias de simulação, ou seja, o número de dias para os quais a simulação irá ser realizada, considerando o dia inicial aquele para o qual os valores dos parâmetros dos dois grupos iniciais foram definidos.

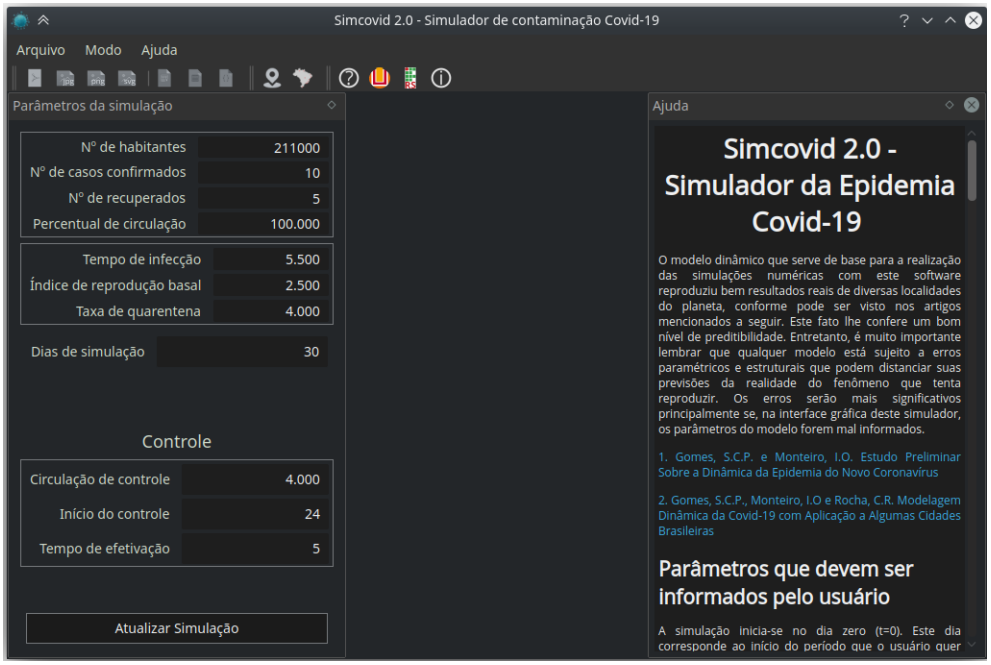


Figura 102. Tela principal do Simcovid 2.0, executando em Linux (KDE Plasma 5.18).

O terceiro grupo de parâmetros de simulação corresponde às ações de controle que podem ser tomadas para a redução da taxa de contaminação, notadamente o isolamento social, pela redução do percentual de circulação livre de pessoas. Assim, temos como definir a circulação de controle (p_c), que seria o percentual a ser assumido para reduzir o contágio; o início do controle (t), ou seja, o dia no qual começará a implantação desse controle, a partir do início da simulação; e o tempo de efetivação do controle (Δt), ou seja, em quantos dias se passará do percentual de circulação inicial para o percentual de circulação de controle.

Ao manter o cursor sobre qualquer parâmetro do painel, aparecerá um texto explicando brevemente o significado daquele valor (*hint*). O mesmo acontece se o usuário solicitar ajuda através do ícone do ponto de interrogação da janela (?). Caso o usuário precise de explicações mais detalhadas sobre os parâmetros e dicas de valores a serem empregados, ele pode solicitar a apresentação do painel de ajuda, mostrado na Figura 103b. Neste painel é apresentado um texto explicativo sobre os parâmetros da simulação, bem como *links* para os artigos que embasam o modelo dinâmico do software. Ao clicar sobre um *link*, o navegador padrão do sistema abrirá este, apresentando o artigo selecionado.

O painel de ajuda não é visível no início da execução do *software*, sendo invocado pelo menu ou pela barra de ferramentas. O motivo principal é prover espaço para a apresentação dos resultados da simulação. A qualquer momento, o usuário pode optar por

fechar o painel, para disponibilizar mais espaço para os gráficos, ou mesmo destacar o painel de ajuda, arrastando o título dele com o cursor. Apesar de não poder ser fechado, o painel de parâmetros pode ser destacado, disponibilizando assim toda a janela para a apresentação dos resultados da simulação, se desejado.

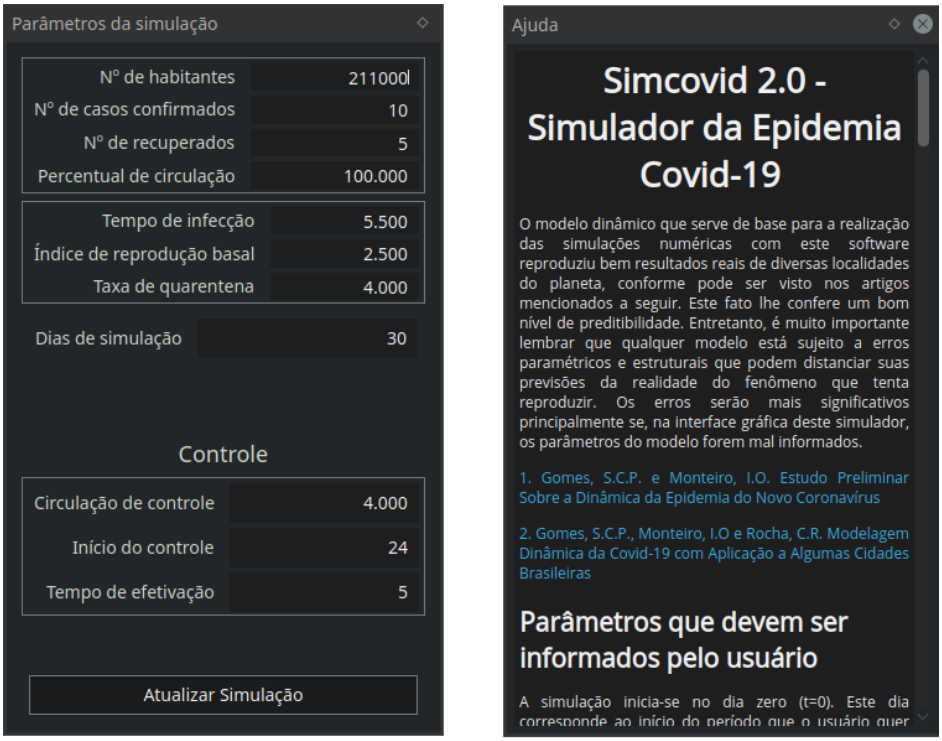


Figura 103. Painel de parâmetros de simulação (a) e de ajuda ao usuário (b).

Uma vez definidos os parâmetros, o botão Atualizar Simulação deve ser pressionado para que a simulação seja executada. Na primeira vez, serão criados painéis com abas, cada qual contendo um gráfico dos resultados, bem como um painel específico com os valores da simulação apresentados em uma tabela. Execuções de simulação subsequentes atualizarão os resultados das abas. A apresentação de resultados é exemplificada na Figura 104.

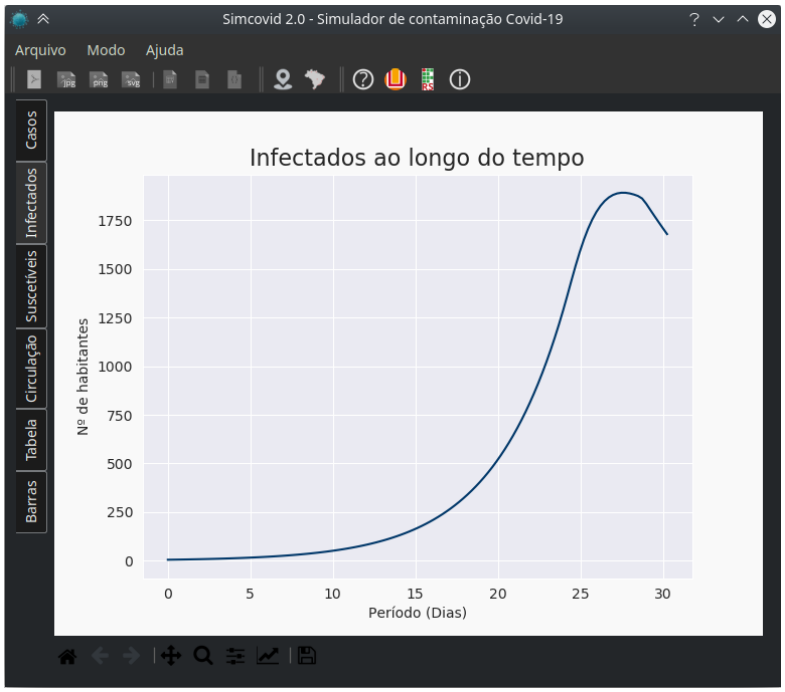


Figura 104. Área de resultados da simulação, em forma de páginas com abas

7.3.2 Resultados de Simulação

Os resultados da simulação são apresentados na forma de gráficos, como os apresentados na Figura 105. O número total de casos ao longo do tempo corresponde à soma dos infectados mais os recuperados $(I + R)$ ao longo dos dias da simulação (Figura 105a). Quando usado controle de circulação, o gráfico da Figura 105b mostra a variação do percentual de circulação a partir do dia inicial de implantação do controle. Eventualmente, é possível observar outro degrau, correspondente a situação de aderência (definida no Capítulo II). O gráfico da Figura 105c corresponde ao número de infectados ao longo do tempo, os quais ainda podem contaminar outros indivíduos da categoria dos suscetíveis. Por fim, o gráfico dos suscetíveis da Figura 105d representa a parcela da população que ainda pode ser contaminada. Podemos observar que esta parcela diminui à medida que o número de casos aumenta. As escalas dos gráficos são automaticamente definidas. Eventualmente, os valores do eixo das abcissas (y) podem ser apresentados em função de um multiplicador mostrado no canto superior esquerdo do gráfico. Nesta situação, os valores em questão são de magnitude elevada, e optamos por esta notação mais compacta para uma melhor apresentação do gráfico.

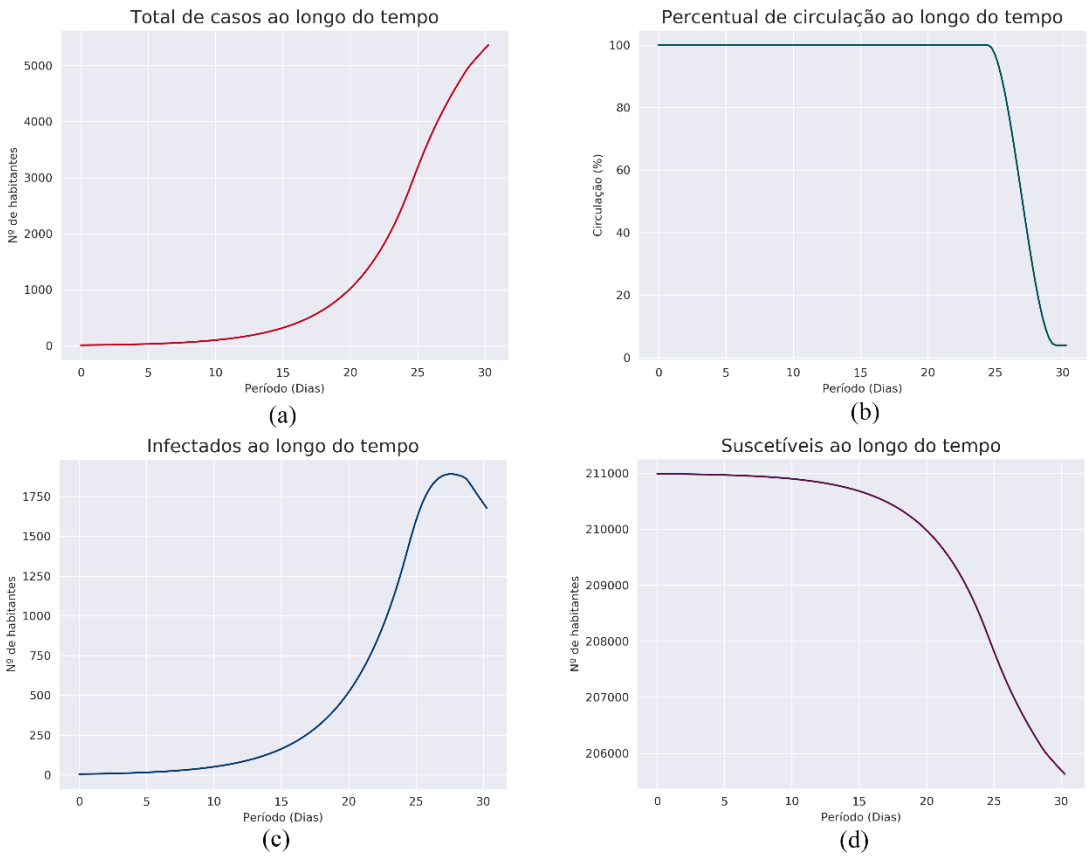


Figura 105. Gráficos da simulação: total de casos (a), percentual de circulação (b), infectados (c) e suscetíveis (d).

Ao passar com o cursor nos gráficos é mostrada a posição deste no canto inferior direito. Assim, colocando o cursor sobre a linha, é possível saber os valores relativos a cada dia da simulação. Além disso, abaixo de cada gráfico há uma barra de navegação (ver Figura 104), que permite fazer zoom e observar em maior detalhe uma região dele, bem como ajustar alguns parâmetros de apresentação, como legendas e título. Há também uma opção para salvar o gráfico em diferentes formatos, da mesma forma que as ações de exportação disponíveis na barra de ferramentas.

Além dos gráficos da Figura 105, há um gráfico de barras, normalizado em relação à população da região de estudos (ou seja, todos os valores são divididos pelo número de habitantes), que apresenta os suscetíveis (S), os infectados (I) e os removidos (R) variando ao longo do tempo. Essa variação é apresentada tanto na forma de uma animação (controlada por botões no canto inferior esquerdo do gráfico) quanto pode ser controlada manualmente por uma barra deslizante abaixo do gráfico. No caso de ser definido um controle de circulação, as barras mudam de cor quando o controle tem início e quando ele

atinge o percentual desejado. Um instante dessa variação no tempo é mostrado na Figura 106.

Figura 106. Gráfico de variação do número de suscetíveis (S), infectados (I) e removidos (R) ao longo do intervalo de simulação, adquirido para o 22º dia.

Uma aba adicional ainda apresenta os dados numéricos dessa simulação, para o caso de ter de se realizar uma análise mais detalhada dos resultados da simulação. Os valores da primeira coluna, em tom diferente das demais, corresponde ao instante da simulação em que os valores foram calculados, em dias. Em função do passo da simulação, os valores dos dias não são inteiros, mas sim correspondem aos instantes para os quais os valores foram calculados. O painel é apresentado na Figura 107.

	Circulação(%)	Nº de casos	Suscetíveis	Infectados	Removidos
3.53	100.00	22	210978	11	11
3.78	100.00	24	210976	12	12
4.03	100.00	25	210975	13	12
4.29	100.00	27	210973	14	13
4.54	100.00	28	210972	14	14
4.79	100.00	30	210970	15	15
5.04	100.00	32	210968	16	16
5.29	100.00	34	210966	17	17
5.55	100.00	36	210964	18	18
5.80	100.00	38	210962	19	19
6.05	100.00	40	210960	20	20
6.30	100.00	43	210957	22	21

Figura 107. Resultados numéricos da simulação, apresentados em uma tabela.

Os resultados da simulação podem ser salvos através das ações de exportação, disponíveis no menu Arquivo|Exportar… ou na barra de ferramentas apresentada na Figura 108a.

Figura 108. Barra de ferramentas de exportação (a) e barra de ferramentas de seleção de modo (b).

Todos os gráficos podem ser salvos como arquivos gráficos nos formatos PNG (bitmap comprimido), PDF e SVG (vetorizados). Os resultados numéricos da simulação podem ser salvos em formato de texto estruturado, como o JSON e o CSV, ou no formato de planilha XSLX. Estes dois últimos formatos podem ser abertos em vários outros softwares de planilhas eletrônicas, como o Microsoft Excel.

7.3.3 Modo Estudo de Casos

O principal recurso adicionado na versão 2.0 do Simcovid foi o modo Estudo de Casos. Nele, o usuário tem acesso a uma relação de cidades brasileiras e os parâmetros correspondentes a estas, podendo então realizar simulações para a previsão do comportamento futuro daquela cidade. A seleção de modo pode ser definida pela barra de ferramentas apresentada na Figura 108b ou pelo menu Modo. O software inicia no modo simulação de regiões, que depende do usuário para o fornecimento dos parâmetros. Ao selecionar o modo Estudo de Casos, uma janela contendo uma relação de cidades é disponibilizada, e para cada cidade é apresentada a data da atualização mais recente de parâmetros e um ícone de descrição. Os itens ainda não utilizados pelo usuário desde a última atualização são marcados com um sinal no canto superior esquerdo. Na Figura 109a, por exemplo, Rio Grande (RS) já havia sido utilizada em simulação, enquanto as demais cidades não. Caso o usuário decida por primeiro ler a descrição de cenário do caso antes de selecioná-lo, é possível clicar na terceira coluna da relação de cidades, que abrirá uma nova janela, como a da Figura 109b, onde são apresentados gráficos com a evolução da pandemia no tempo, bem como a circulação no período de aquisição de dados e a variação do índice de reprodução basal. O texto descreve também o período de aquisição dos dados e as características do processo de identificação paramétrica utilizado para a determinação dos parâmetros do caso em questão.

(a) (b)

Figura 109. Relação de cidades do modo Estudo de Casos (a) e descrição do caso da cidade de Porto Alegre (RS) no dia 09 de julho de 2020 (b).

Após selecionar uma cidade, o painel de parâmetros é atualizado com os valores identificados e os campos correspondentes passam a não ser editáveis. Além disso, um novo quadro é mostrado, incluindo o nome da cidade e o período para o qual os parâmetros foram identificados, bem como um campo indicativo da data de início da simulação. Os campos de controle e de definição do número de dias da simulação são passíveis de edição, para que o usuário possa avaliar cenários futuros com diferentes restrições de circulação em comparação com o cenário sem nenhum controle (padrão quando a cidade é inicialmente selecionada).

A execução de simulação ocorre da mesma forma que no modo simulação de regiões. Os resultados são os mesmos. Porém, adiciona-se uma aba de descrição do caso, contendo o mesmo texto explicativo mostrado na Figura 109b. A Figura 110 ilustra a tela principal no modo estudo de casos, após uma simulação para a cidade de Porto Alegre (RS). Observe que é uma simulação de 30 dias, considerando, neste exemplo, o início desse período como o dia 09 de julho de 2020, que é o último dia do período para o qual os dados foram coletados e a identificação dos parâmetros da simulação foi realizada.

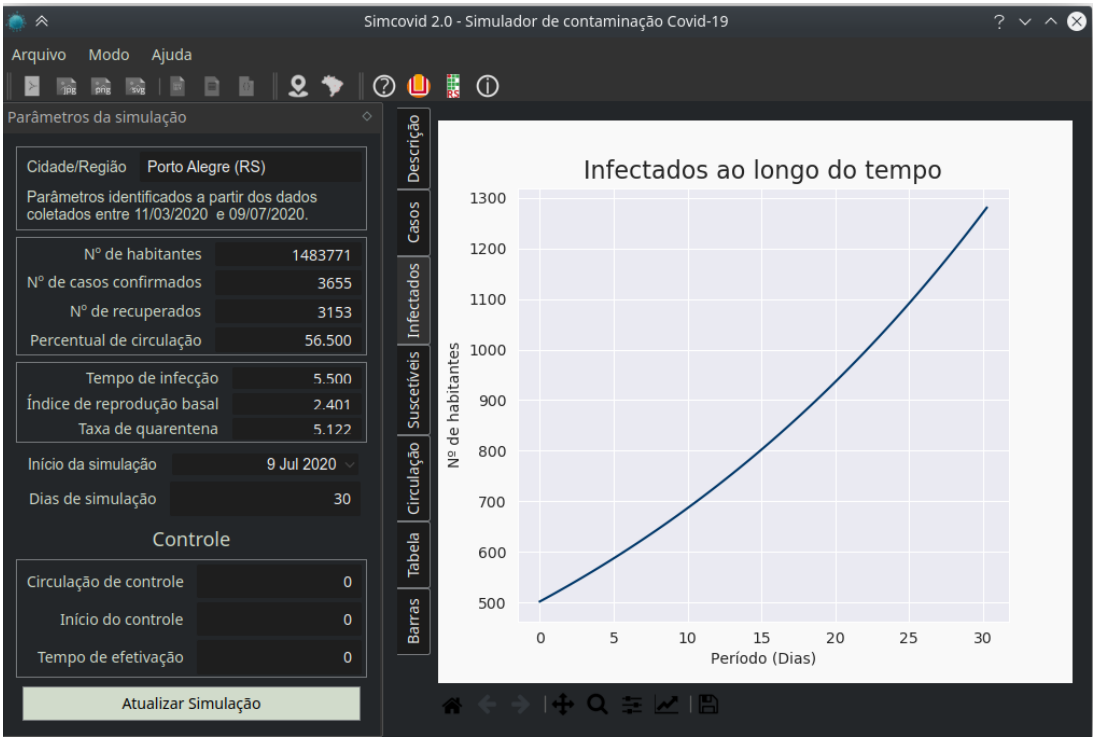


Figura 110. Resultado de simulação no modo estudo de casos.

Os parâmetros das cidades são atualizados a partir de um servidor que os armazena. Na inicialização do software, este verifica a data do último conjunto de parâmetros e, se esta data é mais recente do que a data da última atualização feita, baixa o arquivo correspondente e o processa, atualizando as cidades disponíveis. O processo ocorre de forma transparente para o usuário, que é notificado sobre os novos dados através da mudança de cor do ícone do modo Estudo de Casos. A identificação paramétrica dos parâmetros é feita regularmente pela equipe responsável pelo Simcovid, de forma a manter o software o mais atualizado possível em relação aos cenários das cidades definidas.

Cada atualização pode modificar apenas o conjunto de parâmetros, e apenas para algumas cidades. Da mesma forma, uma atualização pode incluir novas cidades. Por isso, é importante verificar, no momento da seleção da cidade, a data de atualização e o estado dos parâmetros (se novos ou já empregados pelo usuário).

7.4 Testando Cenários com o Simcovid

Uma das facilidades que o Simcovid permite consiste no teste de diferentes cenários para a evolução epidêmica em uma determinada cidade ou região de interesse. Tomemos como exemplo o caso de Porto Alegre, cujos parâmetros e o gráfico com o índice de reprodução basal efetivo podem ser vistos na Figura 111, relativos à situação no dia

27/07/2020. Percebemos que o índice de reprodução basal efetivo está acima de 1, de forma que a simulação com a previsão para tempos futuros de Porto Alegre apresentará um crescimento acelerado da contaminação.

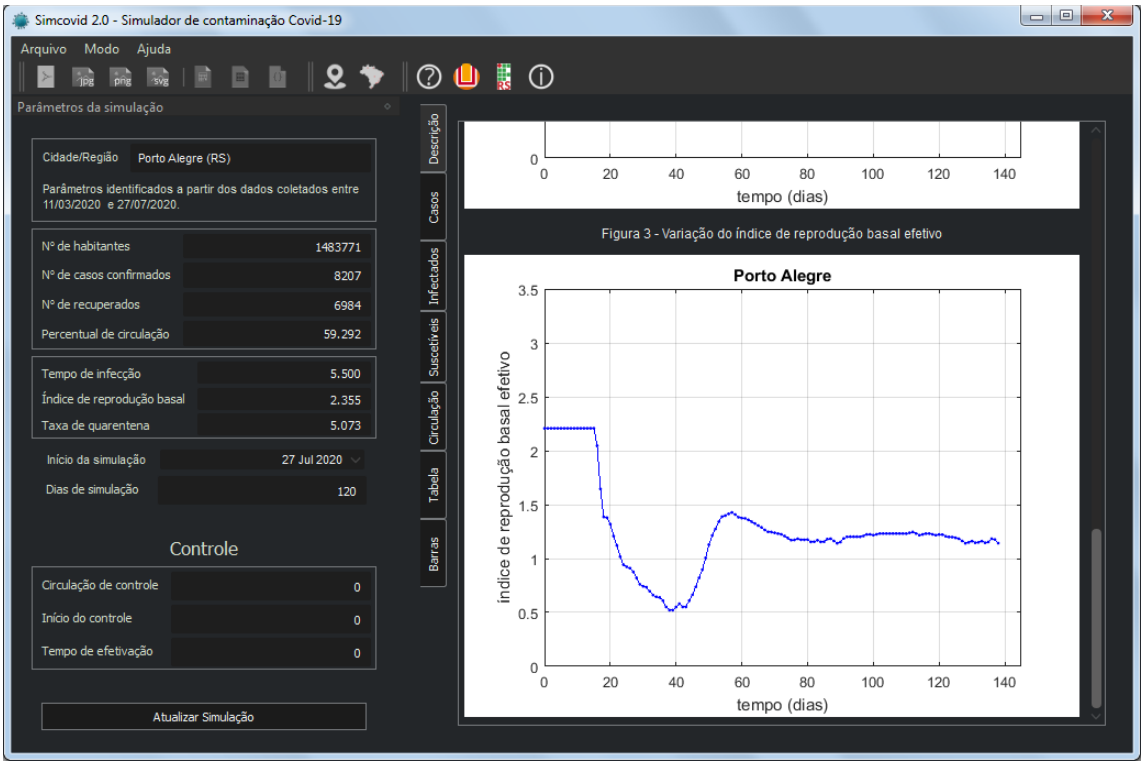


Figura 111. Tela do Simcovid mostrando os parâmetros relativos ao dia 27/07/2020, bem como um gráfico com a evolução do índice de reprodução basal efetivo até esta data.

Analisaremos quatro cenários possíveis para Porto Alegre. O primeiro deles seria deixar a situação exatamente como está em 27/07/2020, ou seja, nenhuma medida de controle adicional seria tomada, de forma que o percentual de circulação do dia 27/07/2020 ($P_{mod} = 59.292\%$) se manteria inalterado durante 120 dias de simulação. Os resultados da simulação para este primeiro cenário podem ser vistos na Figura 112, na qual podemos perceber que o número de casos chegaria a 210700 (Figura 112a), com 12200 infectados no pico, que ocorreria no tempo $t = 110\ dias$ (Figura 112b).

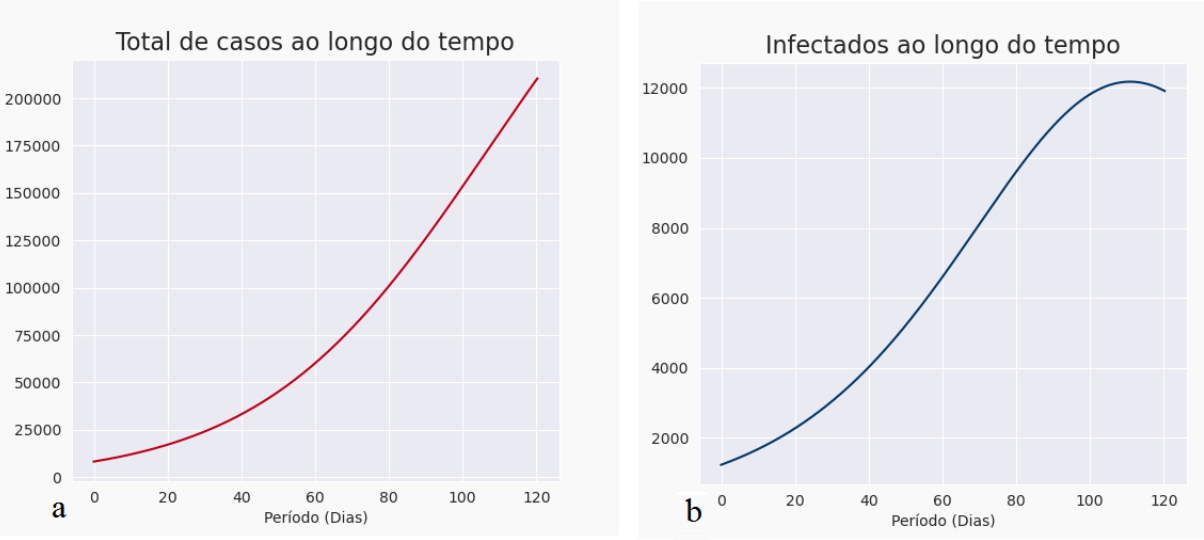


Figura 112. Simulação para Porto Alegre, considerando um tempo final de 120 dias após a data da última atualização dos dados (27/07/2020).

O segundo cenário seria a liberação total da circulação de pessoas. Vamos considerar que, mesmo após a liberação total oficial, 15% da população continuaria em isolamento. Desta forma, o percentual de circulação passaria a ser igual a 85%. Os parâmetros do controle (digitados na interface) para testar este cenário seriam: $p_{cont} = 85\ (\%)$ (percentual de controle); $t_{cont} = 20\ dias$ (instante da aplicação do controle); $\Delta t_{cont} = 5\ dias$ (intervalo de efetivação do controle). Na Figura 113 podemos ver que o resultado dessa flexibilização seria catastrófico: no instante de pico, correspondente a $t = 70\ dias$, Porto Alegre teria 79700 infectados simultâneos e, certamente, não haveria sistema de saúde que suportasse a grande demanda por tratamento decorrente desse elevado número de infectados.

O terceiro cenário seria implementar um controle epidêmico que reduzisse o percentual de circulação dos atuais (em 27/07/2020) 59.292%, para 49%. Vamos considerar esse controle aplicado 20 dias após o 27/07/2020 e ainda, que o controle seja efetivado em 5 dias, ou seja, a cidade de Porto Alegre levaria 5 dias para que o percentual de circulação passasse de 59.292% para 49%. Os parâmetros digitados na interface seriam (conforme visto na Figura 114): $p_{cont} = 49\%$; $t_{cont} = 20\ dias$; $\Delta t_{cont} = 5\ dias$. Ao atualizarmos a simulação vemos que o número de infectados, no pico agora em $t = 105\ dias$, seria igual a 4000. Trata-se de uma redução significativa com relação ao primeiro cenário, mas talvez não suficiente para que o sistema de saúde trabalhe longe do seu limite.

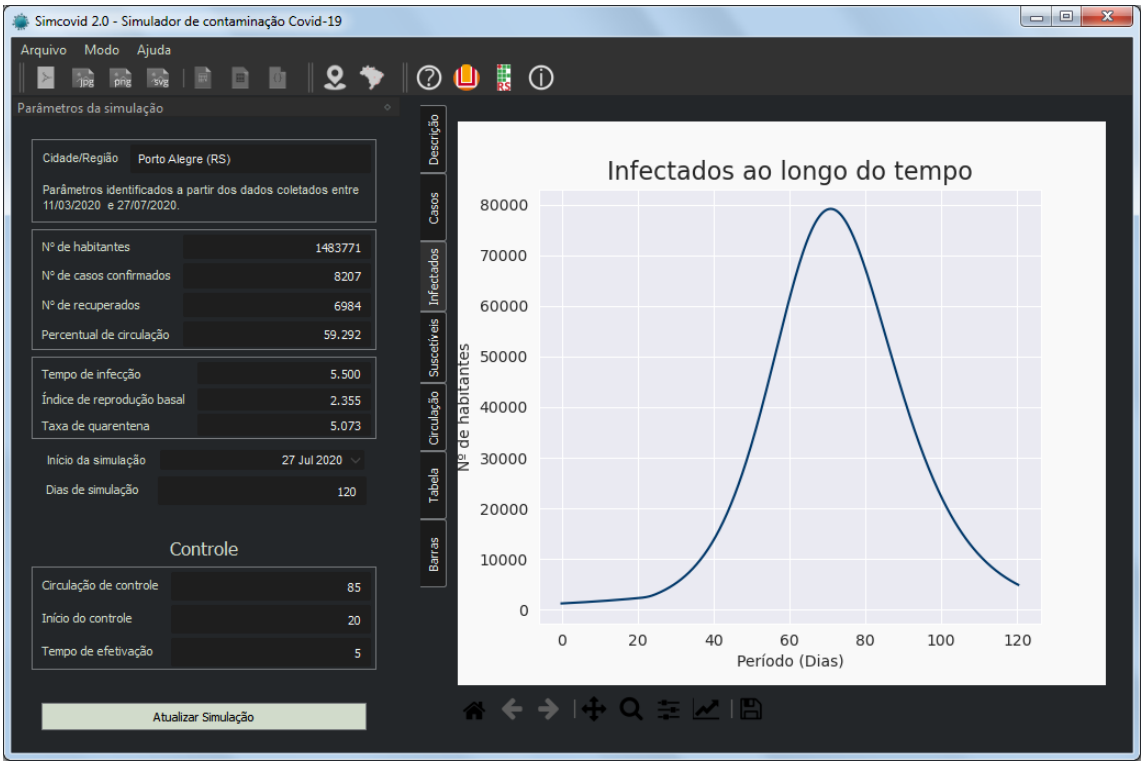


Figura 113. Flexibilização do isolamento social, permitindo um percentual de circulação equivalente a 85%.

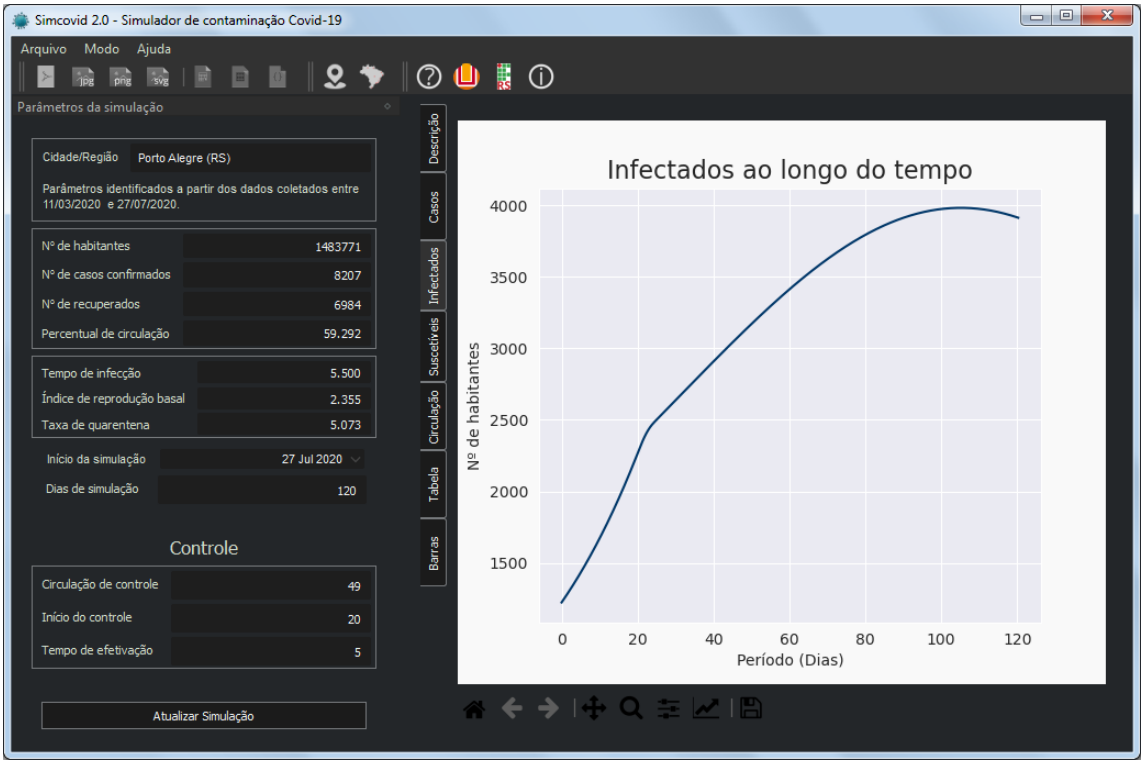


Figura 114. Simulação para Porto Alegre, considerando o terceiro cenário (controle com percentual de circulação a 49%).

O quarto e último cenário avaliado seria a imposição de um controle do tipo *lockdown* total, a exemplo do que foi aplicado em Hubei (China), com as seguintes características: $p_{cont} = 2\%$ (percentual de circulação de controle); $t_{cont} = 20\ dias$ (tempo do início do controle); $\Delta t_{cont} = 5\ dias$ (intervalo de tempo de efetivação do controle). Na Figura 115 vemos os parâmetros do controle digitados na interface, bem como a curva do número de infectados prevista na simulação. Podemos perceber que 30 dias após a implementação do controle do tipo *lockdown* a epidemia em Porto Alegre seria debelada.

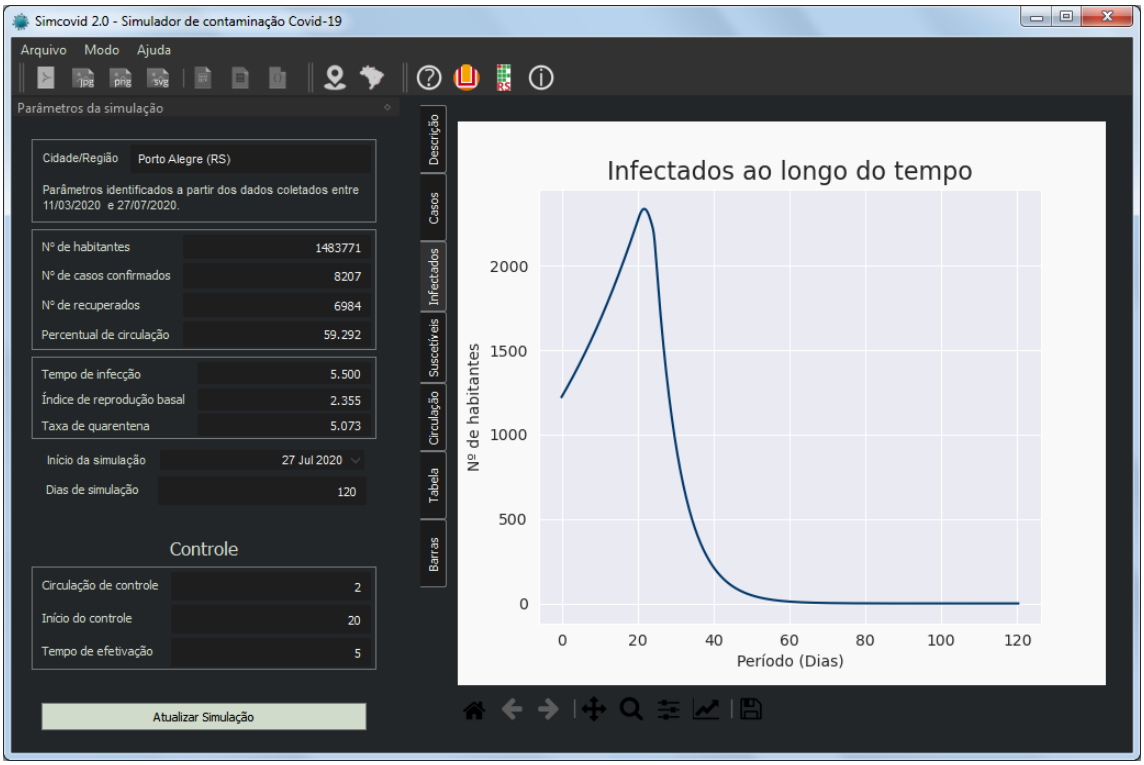


Figura 115. Simulação para Porto Alegre, considerando o quarto cenário (controle do tipo *lockdown* total, com percentual de circulação a 2%).

Não somos especialistas em economia, de forma que não nos arriscaremos a fazer projeções sobre os impactos econômicos específicos de cada cenário apresentado. Entretanto, uma comparação é imediata: no terceiro cenário, cujo pico se verifica em 110 dias, seriam necessários 280 dias ao todo, contados a partir de 27/07/2020, até chegarmos próximos ao final do ciclo epidêmico. Neste caso, a descida do pico é lenta porque o índice de reprodução basal efetivo é próximo de 1 $(\bar{R}_0(t) = 1.061)$. Entretanto, caso a opção fosse adotar o *lockdown*, a epidemia seria finalizada em 30 dias. O que seria mais custoso em termos econômicos, se 280 dias com 49% de circulação de pessoas, ou 30 dias com 2% de circulação, não sabemos dizer. Porém, em termos de vidas humanas, evidentemente, haveria muito menos óbitos se o quarto cenário (*lockdown*) fosse o adotado para Porto

Alegre. Certamente, a análise deve envolver outras hipóteses tais como a possibilidade de uma vacina ser disponibilizada à população antes dos 280 dias necessários no terceiro cenário. A variável de controle passaria a ser a vacinação em massa e a epidemia seria rapidamente interrompida. Outra importante realidade que deve fazer parte do contexto geral de análise é que mais de 99% da população, após o *lockdown*, continuaria suscetível a contrair a infecção, de forma que um segundo surto epidêmico é altamente provável se as fronteiras do município continuarem abertas e se, principalmente, as outras cidades e regiões do país não fizerem também um *lockdown* simultâneo, com coordenação nacional.

Finalizamos esta seção indicando que muitos outros cenários poderiam ser testados com o Simcovid, permitindo assim múltiplas escolhas sobre a estratégia de controle epidêmico a ser adotada para uma determinada cidade ou região em estudo. É importante salientar que a análise sobre diferentes cenários acontece em um contexto de simulações de longo prazo, para as quais não existe mais a boa precisão do modelo. Vimos no capítulo anterior que a previsão com maior precisão seria para um tempo de simulação máximo equivalente a 15 dias. Entretanto, embora sem grande precisão, a análise de longo prazo nos diferentes cenários indica uma tendência, que pode servir como importante alerta. Por exemplo, no primeiro cenário, mesmo admitindo que o erro possa chegar a 15%, lidar com 12200 infectados simultâneos (previsto pelo modelo), ou 14030 (15% a mais), ou ainda 10370 (15% a menos) pode representar, de qualquer forma, problemas com níveis semelhantes de gravidade.

7.5 Comentários Finais

O instalador do Simcovid está disponível para download gratuito no endereço eletrônico http://sites.riogrande.ifrs.edu.br/simcovid/simcovid. Até o momento, o banco de dados de parâmetros e descrição de cenários contém as cidades de Belém (PA), Brasília (DF), Cuiabá (MT), Fortaleza (CE), Pelotas (RS), Porto Alegre (RS), Rio de Janeiro (RJ), Rio Grande (RS), Salvador (BA), São Luís (MA) e São Paulo (SP). Os parâmetros para essas cidades são regulamente atualizados no servidor da aplicação e, consequentemente, atualizados no software, quando o computador do usuário tem acesso à Internet.

Acreditamos que a finalidade inicial do Simcovid tenha sido atingida: prover uma ferramenta relativamente simples para a análise da evolução da Covid-19 em regiões e cidades, que pudesse ser utilizada por qualquer usuário, não ficando restrita a especialistas das áreas de epidemiologia ou de modelagem computacional. Com isso, o usuário pode

avaliar diferentes cenários, fato que pode contribuir para ampliar a conscientização sobre a importância do isolamento social como meio de combate à contaminação.

Entendemos também que o Simcovid ainda é um aplicativo em evolução, podendo trazer novos recursos e formas de apresentação de resultados em futuras atualizações. Além disso, ainda está em estudo a elaboração de instaladores para outros sistemas operacionais como Linux e MacOS, bem como a adequação da interface para *smartphones* Android, visando uma melhor experiência de uso e um instalador compatível com esse sistema operacional.

Referências

BERNARD, J. Utilities. In: Python Recipes Handbook. Apress, 2016.

BORGES, L. E. Python para desenvolvedores. Novatec, 2014.

CASS, S. Top Programming Languages 2020. IEEE Spectrum, 2020. Disponível em <https://spectrum.ieee.org/at-work/tech-careers/top-programming-language-2020>. Acessado em julho/2020.

CORTESI, D. PyInstaller Documentation, Release 3.6. 2020. URL: https://www. pyinstaller. org/documentation. html (visited on 12/06/2020).

HINDMARSH, A. C. ODEPACK, A Systematized Collection of ODE Solvers. IMACS Transactions on Scientific Computation, v. 1, p. 55 – 64, 1983.

HUNTER, J. D. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, v.9, n.3, p. 90-95, 2007, DOI:10.1109/MCSE.2007.55.

LANGTANGEN, H. P. A Primer on Scientific Programming with Python. Texts in Computational Science and Engineering. 5ed. Springer, 2016.

LINGE, S., LANGTANGEN, H. P. Programming for Computations – Python. Texts in Computational Science and Engineering. Springer, 2020.

MCKINNEY, W. Data Structures for Statistical Computing in Python**.** Proceedings of the 9th Python in Science Conference, p. 51-56, 2010.

OLIPHANT, T. E., Python for Scientific Computing. Computing in Science & Engineering, v. 9, n. 3 p.10-20, 2007. DOI:10.1109/MCSE.2007.58.

PETZOLD, L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. SIAM Journal on Scientific and Statistical Computing, v. 4, n. 1, p. 136-148, 1983.

PSF: Python Software Foundation. What is Python? Executive Summary. 2019. Disponível em <https://www.python.org/doc/essays/blurb/>. Acessado em julho/2020.

QT Company. About Qt. 2019. Disponível em <https://wiki.qt.io/About_Qt>. Acessado em julho/2020.

SZEKELY, Amir et at. NSIS (Nullsoft Scriptable Install System). 2019. disponível em <https://github.com/kichik/nsis>. Acessado em 5/7/2020.

VIRTANEN, P. *et al*. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, v. 17, n. 3, p. 261-272, 2020.

WASKOM, M. *et al*. seaborn: statistical data visualization. v.0.10.1. 2020. Disponível em <https://seaborn.pydata.org/>. Acessado em junho/2020.

WILLMAN, J. M. Beginning PyQt. Apress, 2020. DOI 10.1007/978-1-4842-5857-6.