

CIÊNCIAS EXATAS E DA TERRA

Identificação de parâmetros na modelagem dinâmica da COVID-19***Parameter identification in the dynamic modeling of COVID-19***Sebastião Cícero Pinheiro Gomes¹, Igor Oliveira Monteiro², Carlos Rodrigues Rocha³**RESUMO**

O presente artigo aborda a identificação paramétrica em um modelo dinâmico da epidemia por COVID-19, localizada em uma determinada cidade ou região. O modelo dinâmico utilizado é uma adaptação do SIR (Suscetíveis, Infectados, Recuperados), que recebeu o nome MdSIR, adaptação esta específica para reproduzir dinamicamente a COVID-19. Foram identificados três parâmetros que variam ao longo do tempo: o índice de reprodução basal; a taxa de quarentena de infectados; o percentual de circulação. Utilizou-se o método de Nelder-Mead Simplex na identificação paramétrica. O procedimento de identificação paramétrica foi utilizado considerando-se os dados reais de diversas cidades brasileiras. Simulações realizadas para essas cidades mostraram que o modelo cujos parâmetros variam ao longo do tempo reproduziu muito bem os dados reais.

Palavras-chave: COVID-19; modelagem dinâmica; identificação paramétrica; MdSIR.

ABSTRACT

This article discusses parametric identification in a dynamic model of the COVID-19 epidemic, located in a specific city or region. The dynamic model used is an adaptation of the SIR (Susceptible, Infected, Recovered), which received the name MdSIR, an adaptation specifically to dynamically reproduce COVID-19. Three parameters were identified that vary over time: the basal reproduction number; the quarantine rate of infected; the circulation percentage. The Nelder-Mead Simplex method was used for parametric identification. The parametric identification procedure was used considering the real data from several Brazilian cities. Simulations carried out for these cities showed that the model whose parameters vary over time well reproduced the real data.

Keywords: COVID-19; dynamic modeling; parametric identification; MdSIR.

¹ Universidade Federal do Rio Grande - FURG, Rio Grande/RS - Brasil. E-mail: scpgomes@gmail.com

² Idem. E-mail: igoromonteiro@gmail.com

³ Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - IFRS, Campus Rio Grande/RS - Brasil. E-mail: cticarlo@gmail.com



1. INTRODUÇÃO

O planeta foi surpreendido no início de 2020 com o anúncio de que viveríamos uma pandemia, originária de um vírus que recebeu o nome de coronavírus ou SARS-CoV-2, cujo primeiro foco surgiu em Hubei (China), no final de 2019. A infecção causada por este vírus ficou conhecida como COVID-19, que surpreendeu principalmente em razão da velocidade com a qual se tornou pandemia, realidade já presente em fevereiro de 2020. O presente trabalho se insere no esforço de tentar compreender melhor do ponto de vista dinâmico a epidemia por COVID-19, localizada em uma determinada cidade ou região. Trata-se de uma continuação do artigo GOMES *et al.* (2020), embora a sua leitura não dependa da leitura prévia deste primeiro artigo.

Apesar de muito recente, já há uma vasta bibliografia sobre modelos e simulações voltados à COVID-19. Em GOMES *et al.* (2020) há uma revisão bibliográfica relativamente detalhada sobre o tema, abordando principalmente aspectos de modelagem e simulação. Os modelos dinâmicos utilizados, em sua maioria, são propostas de variações das estruturas SIR ou SEIR (Suscetíveis, Expostos, Infectados, Recuperados). GIORDANO *et al.* (2020) propõem uma variação do SIR para estudar a epidemia na Itália. Eles subdividem o compartimento R (recuperados) em vários outros compartimentos, sendo esta a principal alteração do modelo proposto com relação ao SIR tradicional. Neste artigo também se encontra uma extensa revisão bibliográfica sobre o tema COVID-19.

O presente artigo usa o modelo MdSIR (modelo SIR modificado), originalmente proposto em GOMES *et al.* (2020). Porém, o MdSIR é utilizado com variações dos seus principais parâmetros. Estas variações ao longo do tempo foram determinadas utilizando-se um algoritmo de identificação paramétrica, baseado no método de Nelder-Mead Simplex. (LAGARIAS *et al.* 1998). Este algoritmo baseia-se 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. O tempo que separa estes dois períodos foi denominado instante de corte (t_c) e é identificado observando-se os dados reais, quando o crescimento do número de casos deixa de ser exponencial. Os parâmetros do modelo são identificados e mantidos constantes, para $t \leq t_c$. Após este 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 artigo 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.

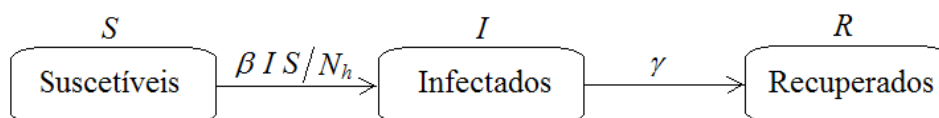


2. MODELO DINÂMICO

Esta seção destina-se a apresentar o modelo dinâmico utilizado no presente trabalho, intitulado MdSIR e que corresponde a uma adaptação do tradicional SIR a fim de melhor reproduzir efeitos dinâmicos característicos da epidemia por COVID-19, localizada em uma determinada cidade ou região.

A Figura 1 mostra a estrutura do SIR, modelo idealizado originalmente por Kermarck e McKendrick, em 1927. Esta é a mais simples estrutura de todos os modelos compartimentais. O sistema de equações (1) contém as três equações diferenciais que descrevem a dinâmica deste modelo. Essa estrutura de modelo muito simples necessita apenas de dois parâmetros físicos: a taxa de crescimento β e a taxa de recuperação γ . N_h é o número de habitantes da cidade ou região em estudo. Sintetizando, o β corresponde à taxa de conversão de suscetíveis (S) em infectados (I), enquanto o γ é a taxa de conversão de infectados (I) para recuperados (R). É importante lembrar que nesta estrutura o R representa tanto os que se recuperaram clinicamente quanto os que vieram a óbito.

Figura 1 - Diagrama de blocos representando o modelo SIR.



Fonte: Elaborada pelo autor.

$$\dot{S} = \frac{-\beta SI}{N_h} \quad (1)$$

$$\dot{I} = \frac{\beta SI}{N_h} - \gamma I$$

$$\dot{R} = \gamma I$$

O parâmetro γ seria clínico, equivalente ao inverso do tempo de infecção, ou seja, $\gamma = 1/T_{inf}$. Assim, pode-se definir o índice de reprodução basal (R_0) como sendo a taxa de crescimento multiplicada pelo tempo de infecção, ou seja, $R_0 = \beta T_{inf} = \beta/\gamma$. Portanto, o R_0 pode significar o número de pessoas para as quais um único infectado pode transmitir a doença, antes de se tornar recuperado. Este é o mais importante dos parâmetros de uma determinada epidemia, que indica o seu nível de agressividade em termos de velocidade de disseminação. 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.

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 intitulado zona ou região de aderência. Conforme explicado detalhadamente em GOMES *et al.* (2020), 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 em GOMES *et al.* (2020) consiste na determinação da variação da taxa de crescimento (β) em função do percentual de circulação (p). Define-se 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. Conforme explicado em GOMES *et al.* (2020), o modelo MdSIR passa então a ter a seguinte forma:

$$\begin{aligned}\dot{S} &= \frac{-\beta(p, R_0, \gamma) SI}{N_h} \\ \dot{I} &= \frac{\beta(p, R_0, \gamma) SI}{N_h} - (\gamma + k)I \\ \dot{R} &= (\gamma + k)I\end{aligned}\tag{2}$$

No MdSIR, a taxa de crescimento é função do percentual de circulação p , do índice de reprodução basal R_0 e da taxa de recuperação γ , de uma determinada cidade ou região. É importante ressaltar que o R_0 que entra na função que calcula o $\beta(p, R_0, \gamma)$ é o valor original, como se a cidade ou região estivesse sem restrição de circulação. O algoritmo que calcula o $\beta(p, R_0, \gamma)$ pode ser visto em detalhes em GOMES *et al.* (2020). 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:

$$\bar{R}_0(p) = \frac{\beta(p, R_0, \gamma)}{(\gamma + k)}\tag{3}$$



Portanto, o parâmetro k contribui para diminuir o índice de reprodução basal efetivo $\bar{R}_0(p)$.

Além da quarentena de parte dos infectados, o parâmetro k pode representar na dinâmica uma série de ações preventivas, 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. Em Gomes *et al.* (2020) há maiores detalhes sobre o modelo dinâmico utilizado no presente trabalho.

3. 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 fazendo-se um pré-tratamento nos dados, antes de submetê-los ao algoritmo de identificação paramétrica, conforme será visto a seguir.

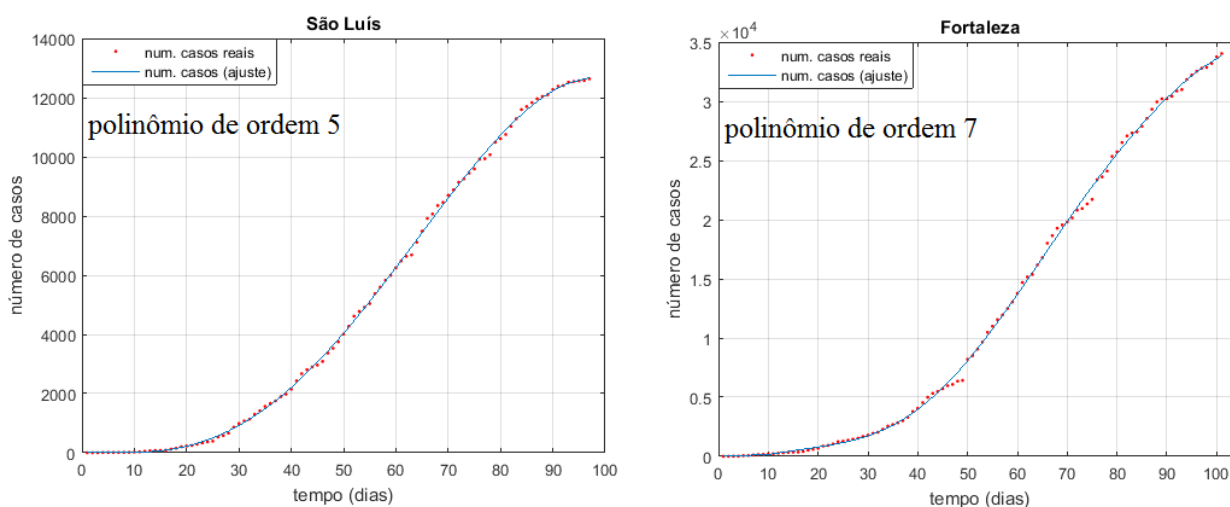
Mostra-se como exemplo os casos de São Luís e Fortaleza, vistos na Figura 2. Foram utilizados polinômios de ajuste, que seguem a tendência dos dados, mas evitando variações bruscas. As ordens dos polinômios de ajuste foram escolhidas observando-se visualmente a qualidade do ajuste, ou seja, escolhia-se a menor ordem que possibilitasse um bom ajuste. Às vezes acontece do polinômio gerar valores negativos para os primeiros pontos. Para contornar este problema, forçaram-se 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 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.



Utilizou-se o Nelder-Mead Simplex (LAGARIAS *et al.* 1998) para a identificação paramétrica, programado em ambiente MATLAB. Minimiza-se 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. Assume-se então o erro na forma: $e_i = |y_{ri} - y_{si}|$, sendo que os valores y_{ri} são formados a partir do polinômio de ajuste (Figura 2). O algoritmo de minimização é aplicado de forma a reduzir o erro e_i a um valor menor do que a tolerância adotada.

Figura 2 – Número acumulado de casos: dados reais e polinômio de ajuste.



Fonte: Elaborada pelo autor.

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. Este marco inicial das medidas de limitação de circulação será intitulado de instante de corte (t_c). Analisando-se os dados reais é possível identificar o t_c de cada cidade. Os pontos em vermelho na Figura 3 mostram a curva crescente do número de casos reais. As derivadas em cada ponto desta curva crescem até o instante de corte, quando a derivada torna-se praticamente constante, 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 $t \leq t_c$. Faz-se então necessária a identificação dos parâmetros R_0 e k , válidos apenas neste intervalo. Para isso, mantém-se o $p=1$ (não há limitação de circulação para $t \leq t_c$) e identificam-se os parâmetros R_0 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 dos parâmetros obedece à seguinte equação:

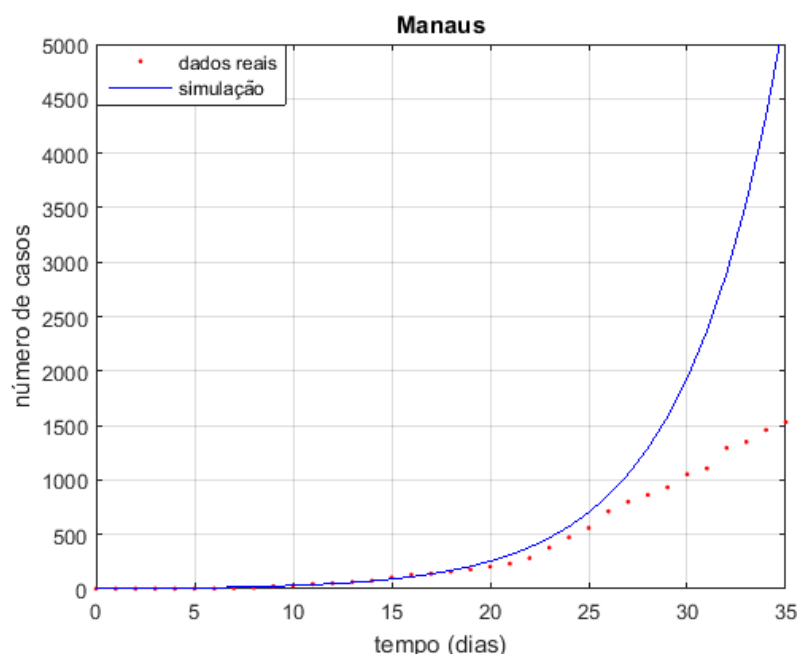


$$e_j = \frac{1}{t_c} \sqrt{\sum_{i=1}^{t_c} (y_{ri} - y_{si})^2} \quad (4)$$

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.

A Figura 3 mostra um exemplo de simulação com os parâmetros iniciais mantidos constantes, para a cidade de Manaus. Os parâmetros identificados foram: $R_0=2.4975$ e $k=0.028$. É importante lembrar que nesse caso, $p=1$. Percebe-se 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 3 - Simulação com os parâmetros constantes.



Fonte: Elaborada pelo autor.

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_0ek e variando-se p ; variando-se simultaneamente os três parâmetros.

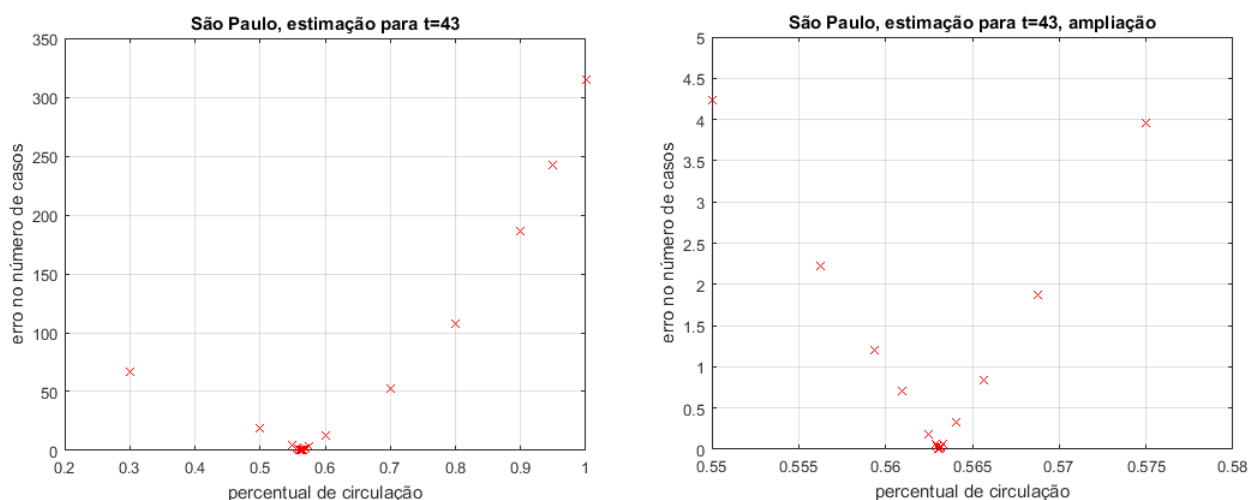
Inicialmente analisou-se o primeiro caso, ou seja, manteve-se R_0ek constantes (identificados para $t \leq t_c$) e, a cada novo dia de dados reais, identificou-se o percentual



de circulação p minimizando-se 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 4 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.

Uma vez que o processo de minimização é aplicado a cada um dos pontos para $t > t_c$, determina-se assim a variação do percentual de circulação p 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 confrontando-se simulação com o número de casos reais, conforme visto na Figura 5, para a cidade de São Paulo. Percebe-se claramente que o modelo reproduziu muito bem os dados reais. Permaneceram constantes o $R_0=2.194$ e $k=0.0212$, valores identificados na etapa anterior, para $t \leq t_c$.

Figura 4 – Diversas iterações no processo de estimação do percentual de circulação (o valor que minimiza o erro foi igual a 0.563).



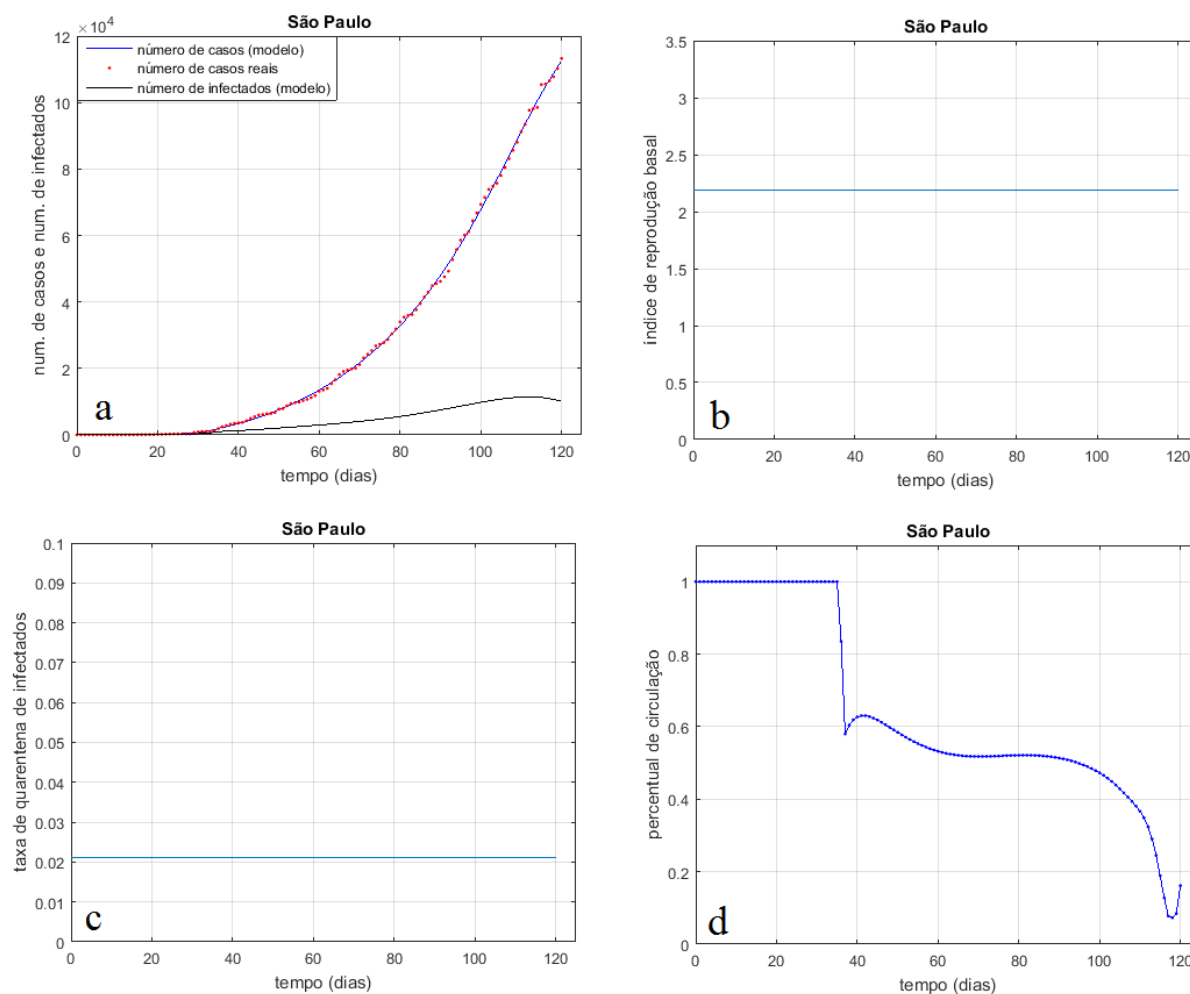
Fonte: Elaborada pelo autor.

Uma simulação semelhante à da Figura 5 foi realizada considerando-se identificação paramétrica com os dados reais da cidade de Fortaleza. Os resultados podem ser vistos na Figura 6. Neste caso, os parâmetros identificados na fase anterior (para $t \leq t_c$) e mantidos constantes foram: $R_0=2.632$ e $k=0.056$. No caso de Fortaleza, o instante de corte foi $t_c=8$. Percebe-se que o modelo também reproduziu bem os dados reais. Entretanto, uma conclusão importante pode ser extraída da análise dos resultados das Figuras 5 e 6, especificamente correlata à variação do percentual de circulação em ambos os casos. Na Figura 5 (São Paulo), o percentual de circulação final foi $p=0.1614$ (16.14%), enquanto na Figura 6 (Fortaleza) foi $p=0.059$ (5.9%). De imediato, conclui-se que estes percentuais de circulação estão abaixo dos praticados efetivamente nas respectivas cidades.



O Portal inloco (disponível em: <<https://mapabrasileirodacovid.inloco.com.br/pt/>>; acesso em: 30 jun. 2020) 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 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 , as contribuições destas outras ações preventivas foram assimiladas e incorporadas nesta variação, de forma que o valor final do p ficou menor do que o praticado realmente nas cidades analisadas.

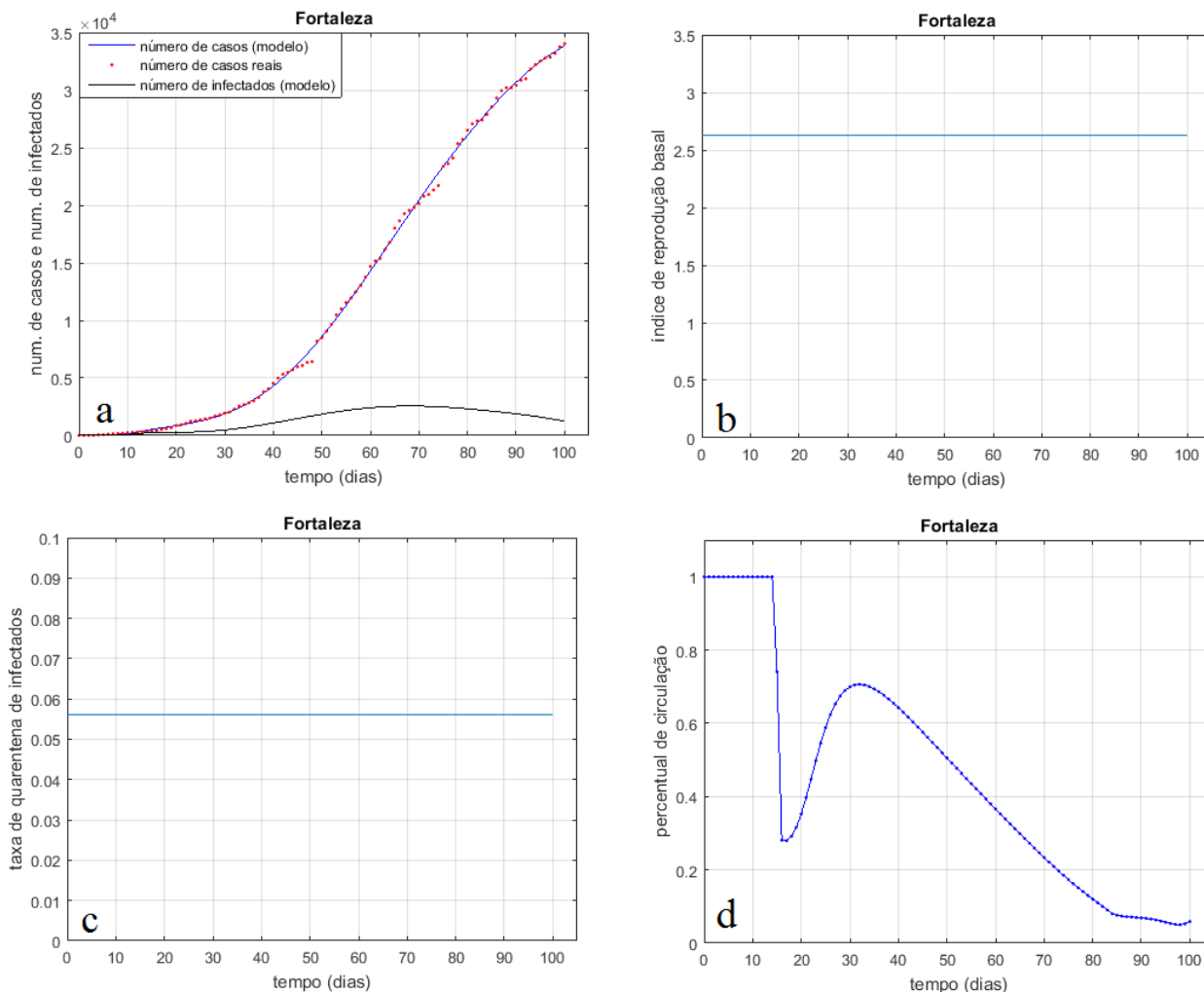
Figura 5 – Simulação e dados reais para a cidade de São Paulo, de 25/02 a 15/06, mantendo-se constantes R_0 e k .



Fonte: Elaborada pelo autor.



Figura 6 – Simulação e dados reais para a cidade de Fortaleza, de 25/02 a 24/06, mantendo-se constantes R_0 e k .

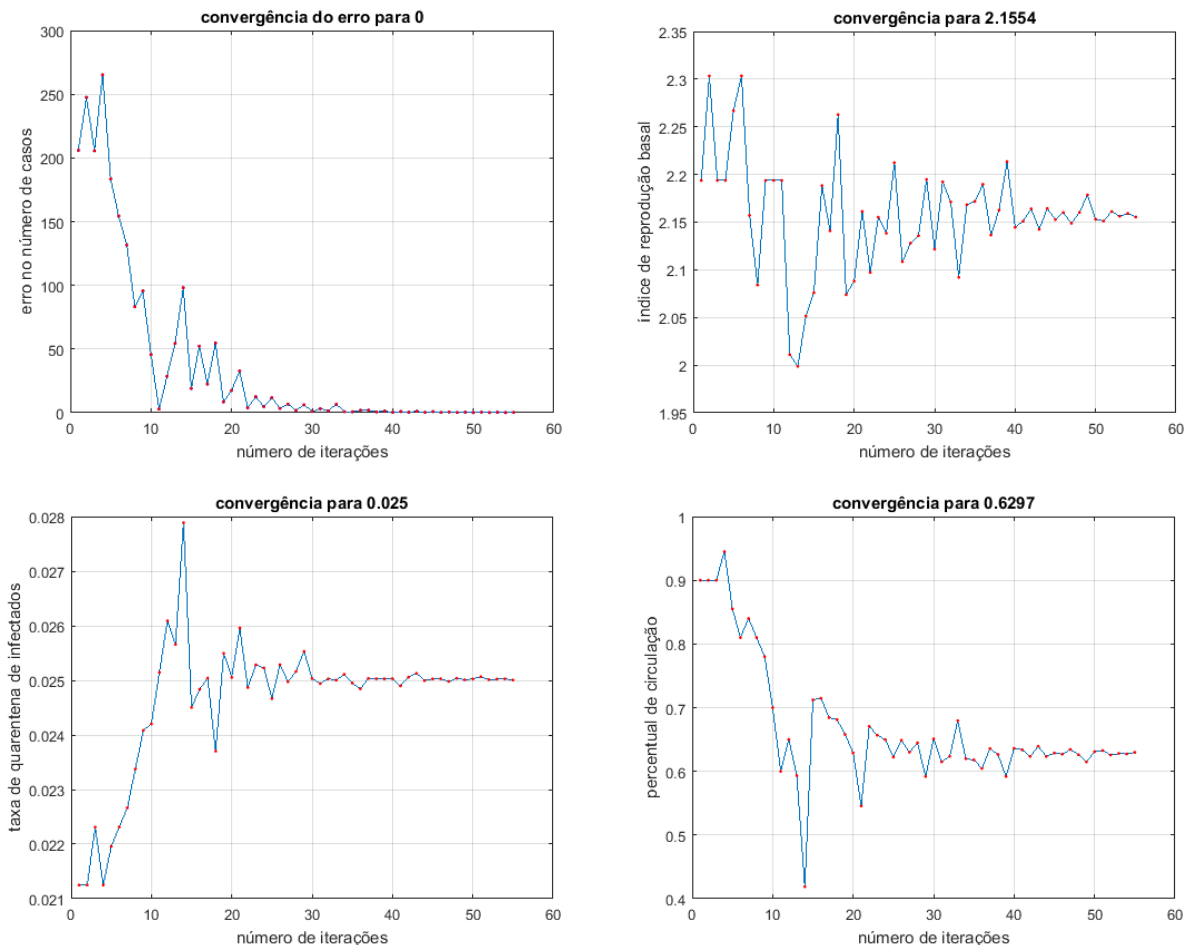


Fonte: Elaborada pelo autor.

O próximo passo consiste no estudo do segundo caso, para o qual se considera, após o instante de corte, variações simultâneas nos três parâmetros, R_0 , k e p . Assim como no caso anterior, o processo de minimização é aplicado individualmente a cada novo ponto com o número real de casos. A Figura 7 mostra o processo iterativo convergente da estimação paramétrica com minimização do erro, processo este aplicado em um único ponto, no caso, para $t=50$. Percebe-se que a convergência acontece com oscilações, tanto no erro quanto nos parâmetros, mas com amplitudes decrescentes. O erro vai pra 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 7 - Cidade de São Paulo, algoritmo de minimização aplicado no tempo $t=50$.

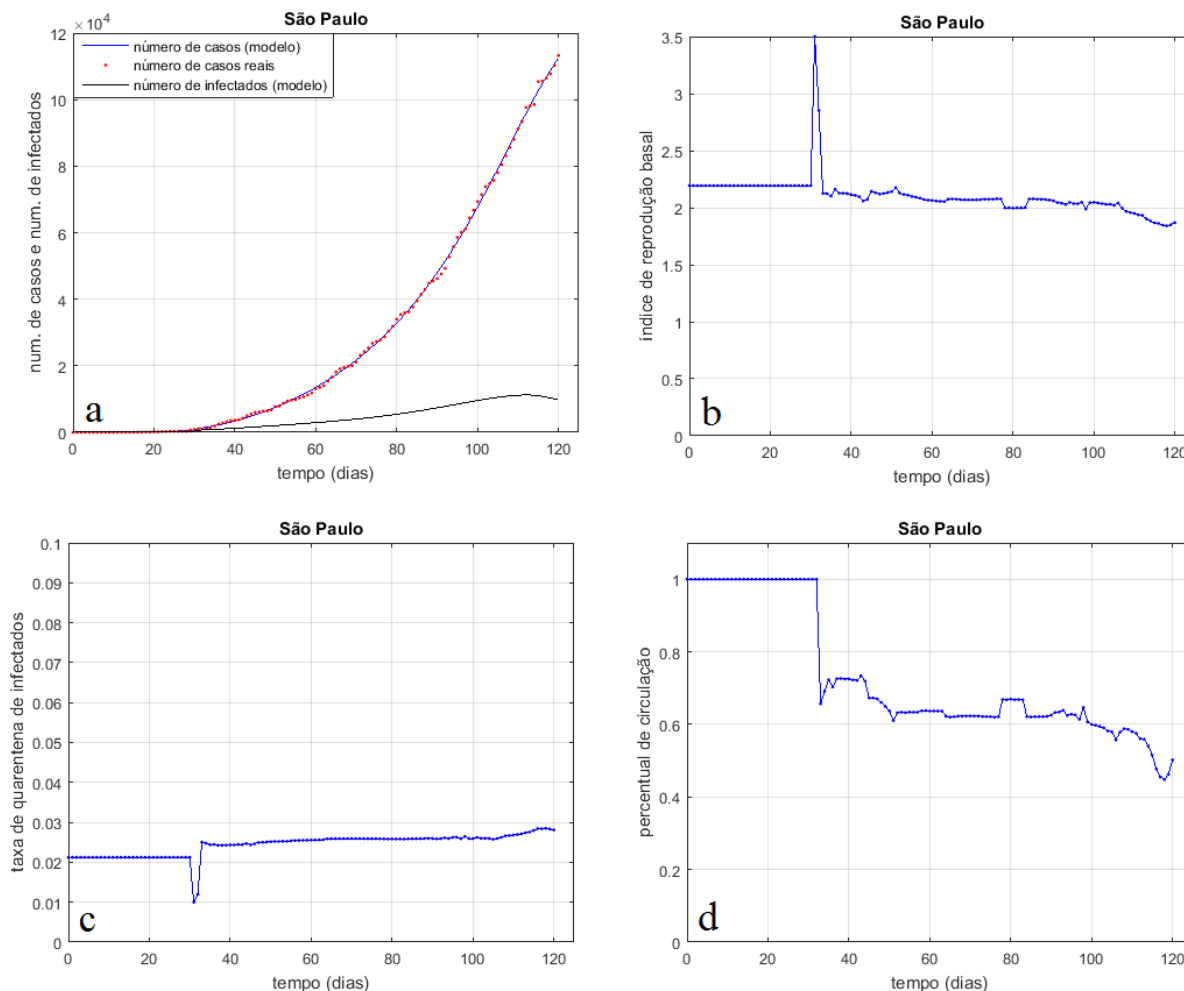


Fonte: Elaborada pelo autor.

O processo de minimização e estimação paramétrica da Figura 7 foi repetido em todos os pontos, para os quais $t > t_c$. Portanto, os três parâmetros agora variam com o tempo: $R_0(t), k(t)ep(t)$. A Figura 8 mostra resultados para a cidade de São Paulo. Nela veem-se uma confrontação entre simulação e experimento, bem como as variações paramétricas ao longo do tempo. É importante comparar a Figura 8 com a Figura 5. As simulações mostram resultados praticamente idênticos (Figuras 5a e 8a), 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 8 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 Figura 8d).



Figura 8 – Simulação e dados reais para a cidade de São Paulo, de 25/02 a 24/06, variando-se os três parâmetros.



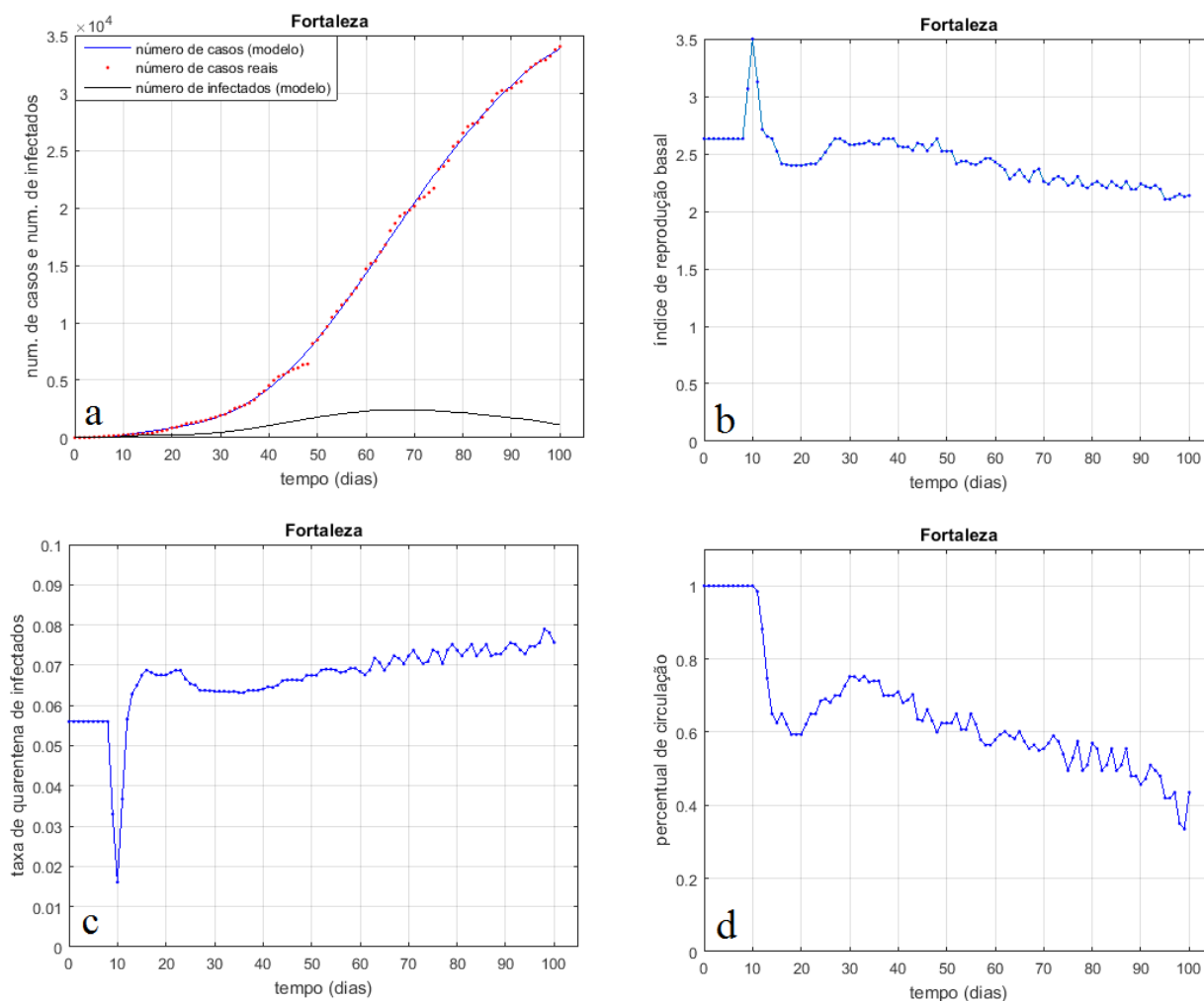
Fonte: Elaborada pelo autor.

Resultados semelhantes ao da Figura 8 foram obtidos para a cidade de Fortaleza, os quais podem ser vistos na Figura 9. Estes resultados devem ser comparados aos da Figura 6. Percebe-se claramente que as Figuras 6a e 9a 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 9d, 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 6d 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 9d se aproxima bem mais do P_{real} do



que o P_{mod} da Figura 6d? Observa-se na Figura 9a 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 $t > 70$. No caso da Figura 6, só o parâmetro p varia, de forma que todas as ações preventivas são assimiladas nesta variação. Já na Figura 9 variam também o $R_0(t)$ 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 intriduzidos estudos com mais cidades brasileiras.

Figura 9 - Simulação e dados reais para a cidade de Fortaleza, de 25/02 a 24/06, variando-se os três parâmetros.



Fonte: Elaborada pelo autor.



4. 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 adotando-se a estratégia de variação paramétrica dos três parâmetros, conforme explicado no seção anterior. Desta forma, o modelo passa a ser escrito a partir das seguintes equações diferenciais:

$$\begin{aligned} \dot{S} &= \frac{-\beta(p(t), R_0(t), \gamma)SI}{N_h} \\ \dot{I} &= \frac{\beta(p(t), R_0(t), \gamma)SI}{N_h} - [\gamma + k(t)]I \\ \dot{R} &= [\gamma + k(t)]I \end{aligned} \quad (5)$$

Ressalta-se mais uma vez que o $R_0(t)$ que entra no cálculo da função β equivale ao valor para o caso em que $p=1$, ou seja, trata-se do índice de reprodução basal no caso da cidade não contar com restrições de circulação de pessoas. Este $R_0(t)$ 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 seguinte equação:

$$\bar{R}_0(t) = \frac{\beta(p(t), R_0(t), \gamma)}{[\gamma + k(t)]} \quad (6)$$

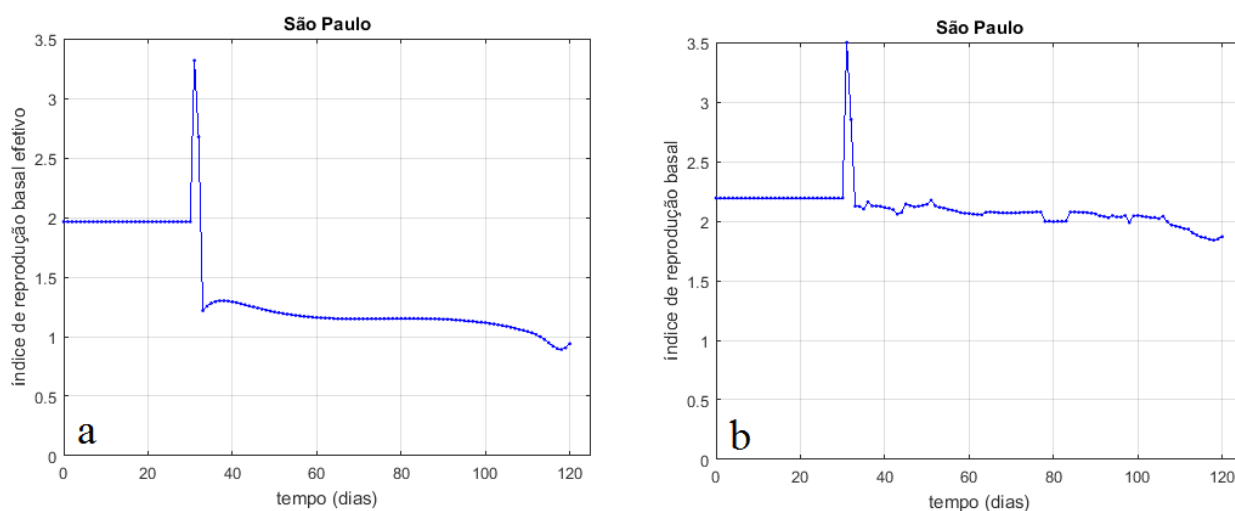
Portanto, o índice de reprodução basal efetivo ($\bar{R}_0(t)$) realmente praticado na cidade (ou região) em análise também varia ao longo do tempo.

A Figura 10a mostra o $\bar{R}_0(t)$ obtido com a equação (6), para a cidade de São Paulo. Ao lado, a Figura 10b mostra o $R_0(t)$ identificado ao longo do tempo. É importante salientar o porquê do $\bar{R}_0(t)$ ser menor do que o $R_0(t)$. Primeiro, se não houvesse nenhuma medida preventiva, inclusive com o percentual de circulação $p=1$ (ausência de isolamento social), $k(t)$ seria nulo e teria-se: $\bar{R}_0(t) = \beta(1, R_0(t), \gamma) / \gamma = \gamma R_0(t) / \gamma = R_0(t)$ e, no caso de São Paulo, teria-se $\bar{R}_0(t) = R_0(t) = 2.194$, valor constante visto nos primeiros pontos da Figura 10b. Com as medidas preventivas, o γ permanece constante, mas o



$\beta(p(t), R_0(t), \gamma)$ diminui quando diminuem o $p(t)$ e/ou o $R_0(t)$ (ver GOMES *et al.* (2020)). Portanto, conforme mostra a equação (6), o $\bar{R}_0(t)$ diminui quando diminuem $p(t)$ e/ou o $R_0(t)$, diminuindo também com o aumento do $k(t)$. A análise da Figura 10a indica que o $\bar{R}_0(t)$ final 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 10 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de São Paulo.

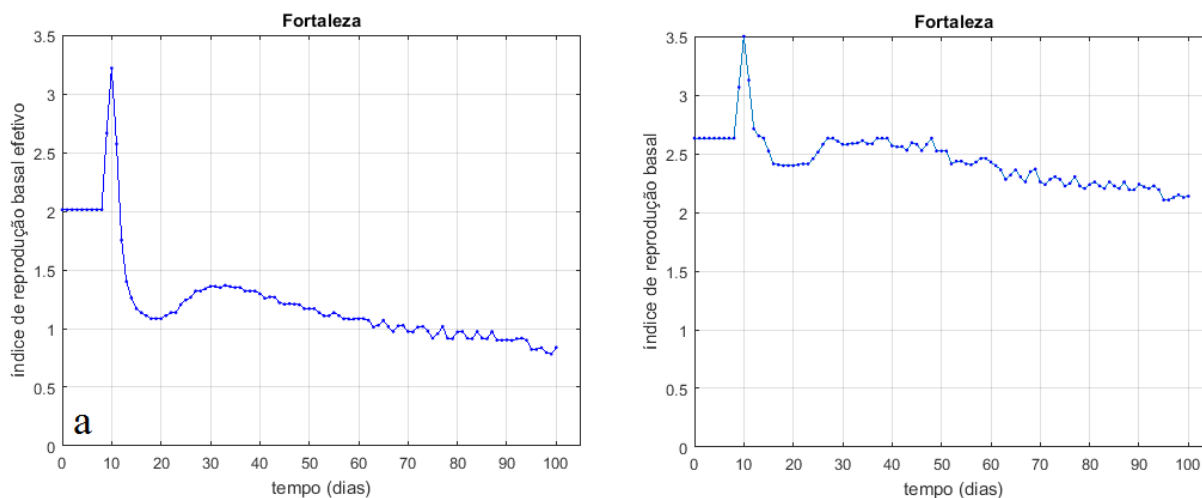


Fonte: Elaborada pelo autor.

Na Figura 11a mostra-se o $\bar{R}_0(t)$, enquanto na Figura 11b mostra-se o $R_0(t)$, para a cidade de Fortaleza. Neste caso, percebe-se que o $\bar{R}_0(t)$ torna-se menor do que 1, próximo ao tempo $t=70$, sendo esta a razão da inflexão na curva de crescimento do número de casos vista na Figura 9a. 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 para a fase de identificação paramétrica. Na fase anterior, ou seja, para $t \leq t_c$, os parâmetros são constantes. A forte oscilação que se percebe tanto na Figura 10 quanto na Figura 11 não é fruto de efeito dinâmico, tendo origem numérica, devida à descontinuidade na passagem à fase de identificação 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 11 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Fortaleza.



Fonte: Elaborada pelo autor.

5. EXTENSÃO DO ESTUDO A DIVERSAS CIDADES BRASILEIRAS

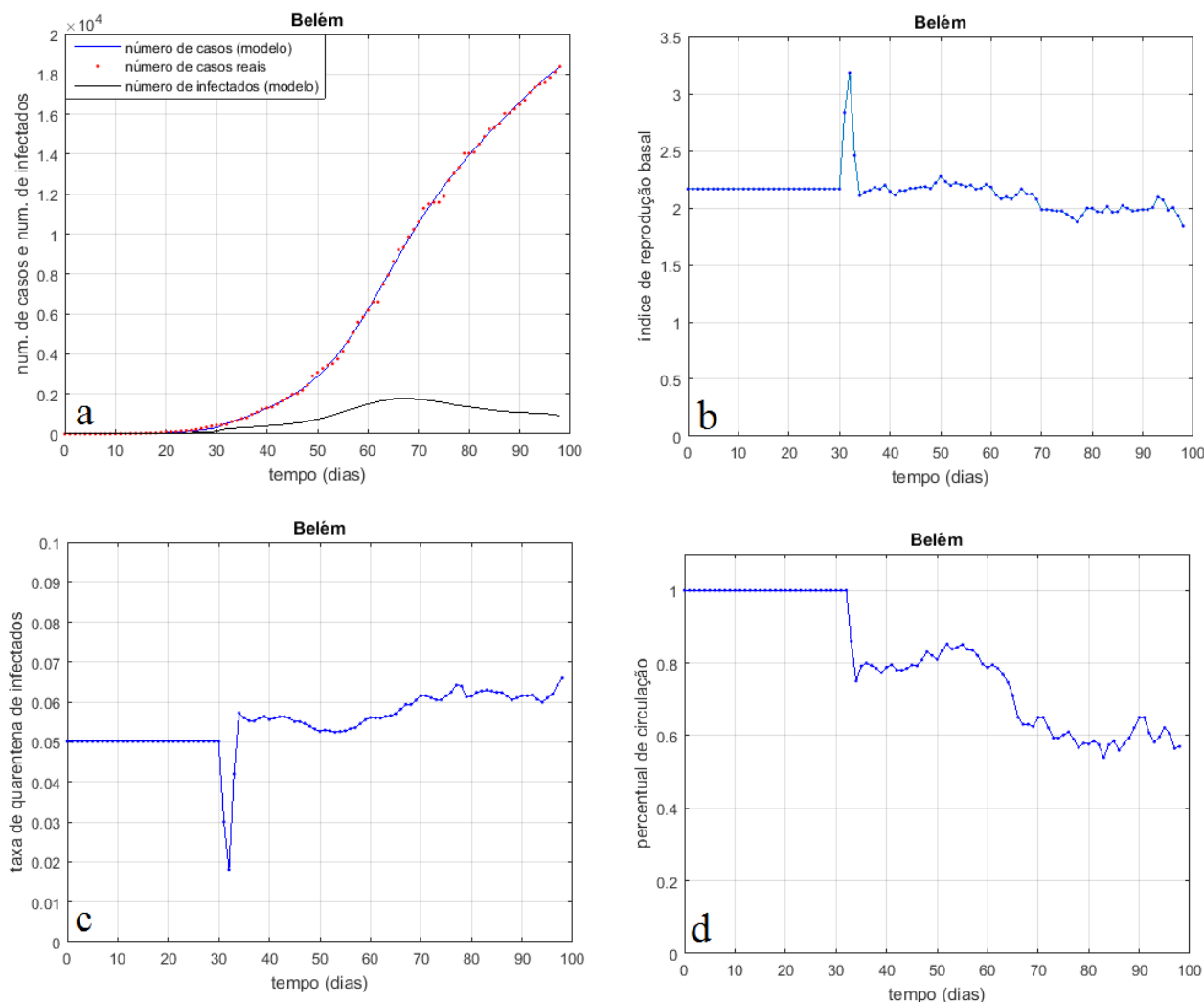
Esta seção destina-se a apresentar os resultados de análises com estimação paramétrica e simulações, para diversas cidades brasileiras. A metodologia é idêntica à que foi usada na seção anterior, para as cidades de São Paulo e Fortaleza. Em todos os casos, assume-se a variação paramétrica simultânea dos 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 12 a 25 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 12a, 14a, ..., 24a. Como nos casos anteriores de São Paulo e Fortaleza, em todos os casos, 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 irá reproduzir 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 Figura 12a (cidade de Belém) percebe-se um ponto de inflexão na curva de crescimento do número de casos, que ocorre em aproximadamente $t=68$, quando a velocidade de crescimento começa a diminuir e, conseqüentemente, a aceleração do crescimento torna-se negativa. Este efeito dinâmico é também perceptível na Figura 13a, quando no mesmo tempo $t=68$ o $\bar{R}_0(t)$ torna-se menor do que 1. Favoreceram à existência desse efeito dinâmico os seguintes fatos: diminuição no $R_0(t)$ (Figuras 12b e 13b) como consequência de ações preventivas, tais como o uso de máscara e o distanciamento social em lugares públicos; aumento do $k(t)$ (Figura 12c), consequência de testagem e identificação de infectados; diminuição do $p(t)$ (Figura 12d), consequência de decisões governamentais sobre limitações de circulação.

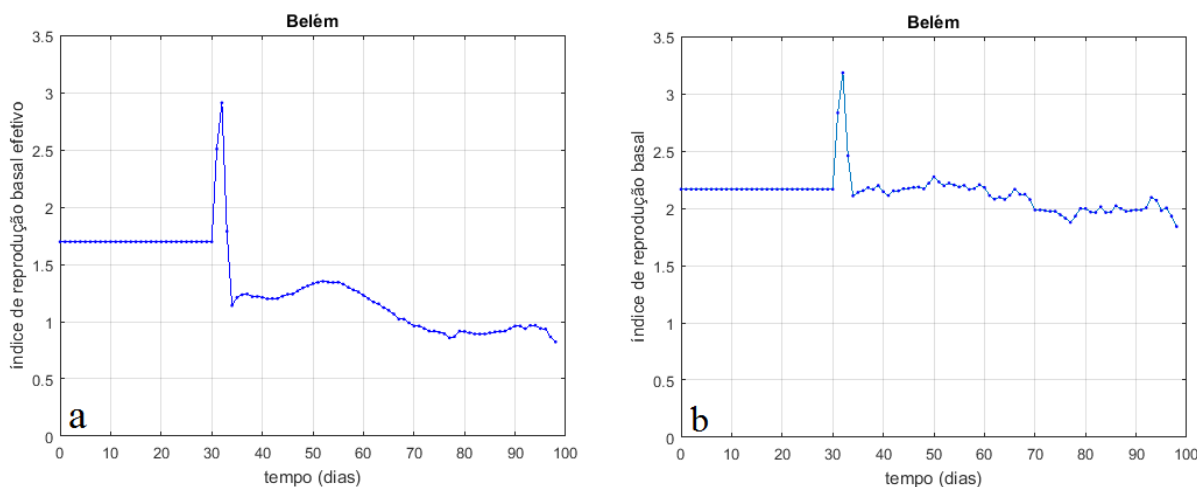
Figura 12 - Identificação paramétrica e simulação para a cidade de Belém, de 18/03 a 24/06.



Fonte: Elaborada pelo autor.

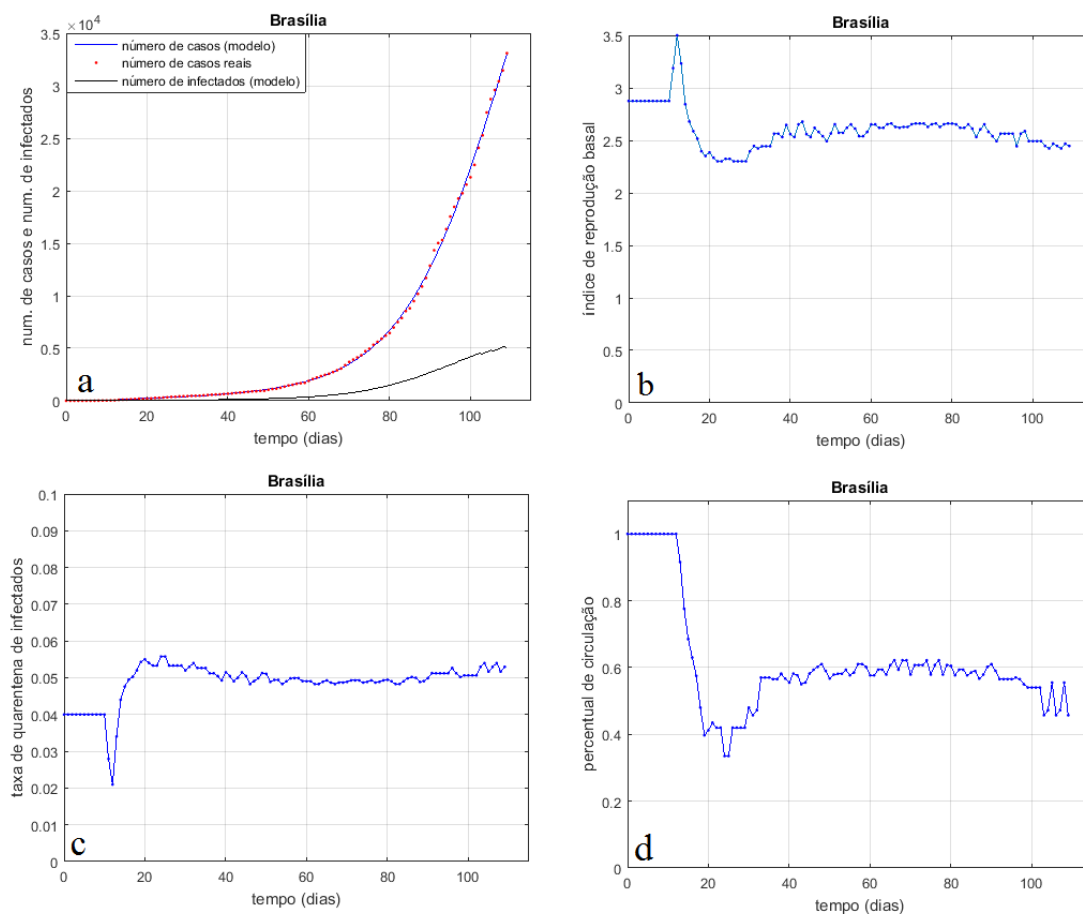


Figura 13 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Belém.



Fonte: Elaborada pelo autor.

Figura 14 - Identificação paramétrica e simulação para a cidade de Brasília, de 07/03 a 15/06.

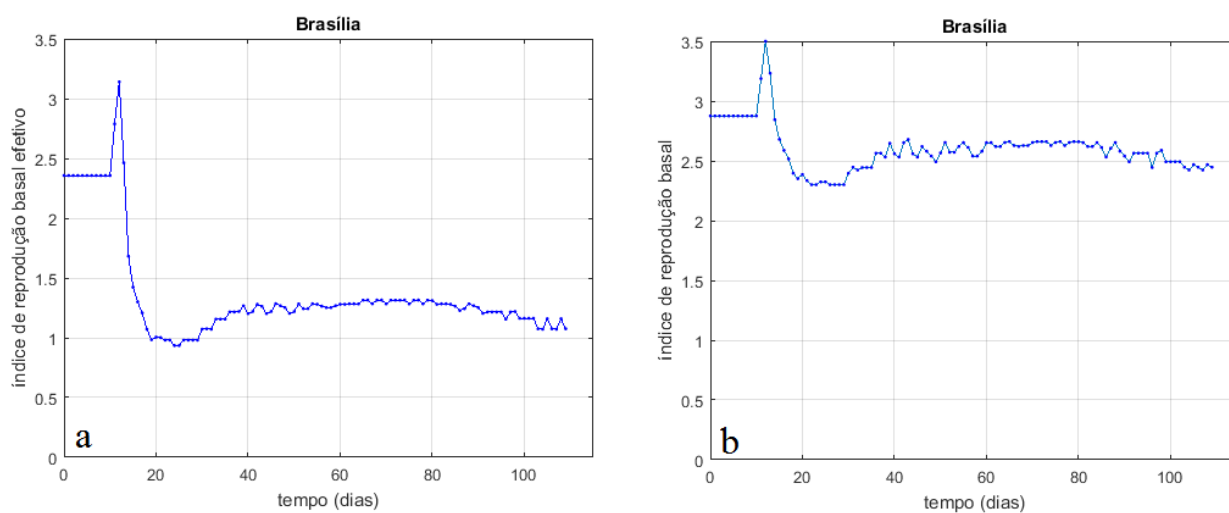


Fonte: Elaborada pelo autor.



Brasília já esteve em uma situação mais perigosa, com um $\bar{R}_0(t)$ próximo a 1.4, mas que nos últimos dias de dados reais passou a valores ligeiramente maiores do que 1 (Figura 15a). Por esta razão não há ponto de inflexão no crescimento do número de casos (Figura 14a) até o dia 24/06/2020.

Figura 15 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Brasília.



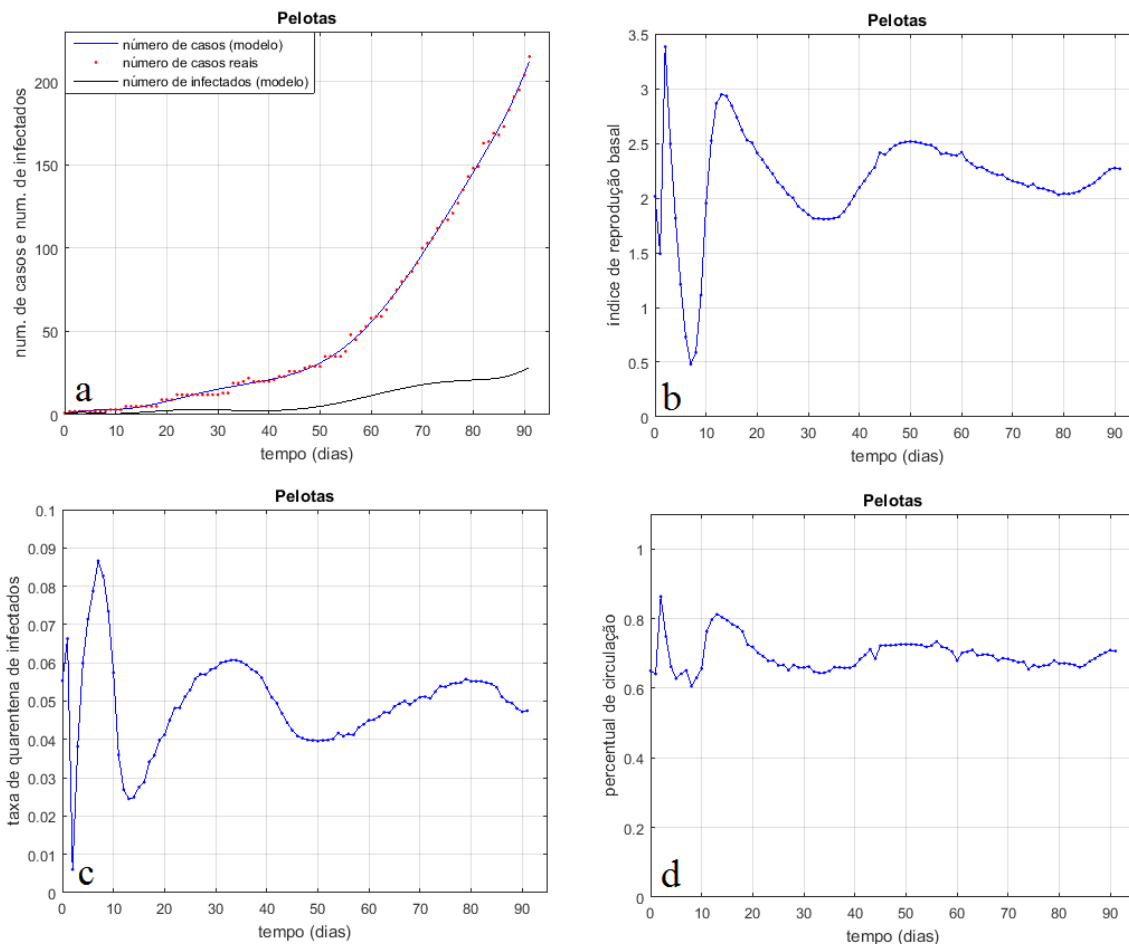
Fonte: Elaborada pelo autor.

As situações em algumas cidades de menor porte também foram analisadas, como é o caso de Pelotas (RS), com 342.405 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 percebe-se que há uma inflexão no número de casos em $t=25$ (Figura 16a), quando $\bar{R}_0(t)$ torna-se menor do que 1 (Figura 17a). Porém, o $\bar{R}_0(t)$ volta a ficar maior do que 1 para $t \geq 40$ e assim permanece até o último dia de dados reais, equivalente a 24/06/2020, de forma que o crescimento do número de casos continua acelerado. Uma das causas do $\bar{R}_0(t)$ permanecer maior do que 1 é o elevado percentual de circulação (Figura 16d), que se mantém superior a 60%.

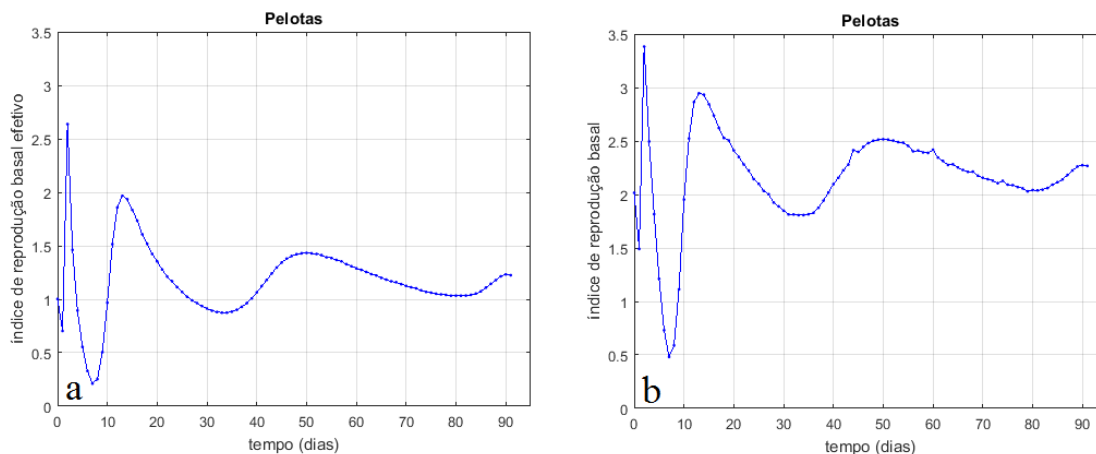


Figura 16 – Identificação paramétrica e simulação para a cidade de Pelotas, de 25/03 a 15/06.



Fonte: Elaborada pelo autor.

Figura 17 – $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Pelotas.

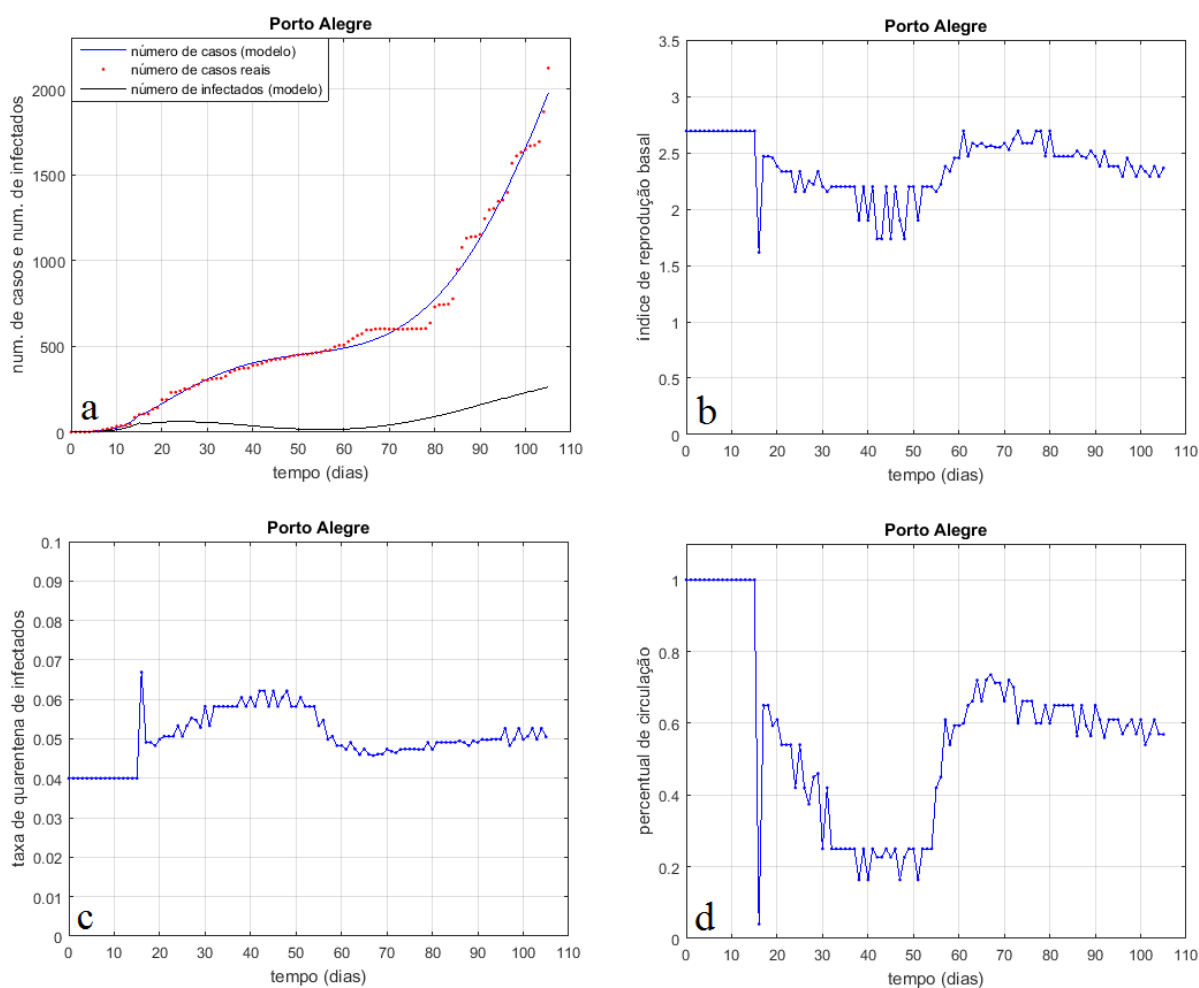


Fonte: Elaborada pelo autor.



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 19a), quando há uma inflexão na curva do número de casos com aceleração passando a negativa (Figura 18a). Esta primeira inflexão acontece em $t=25$. 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 $t=56$, 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 de 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 $t=56$, Porto Alegre permanece, em 24/06/2020, com os melhores índices em termos do número de casos e de óbitos, comparado às outras capitais brasileiras.

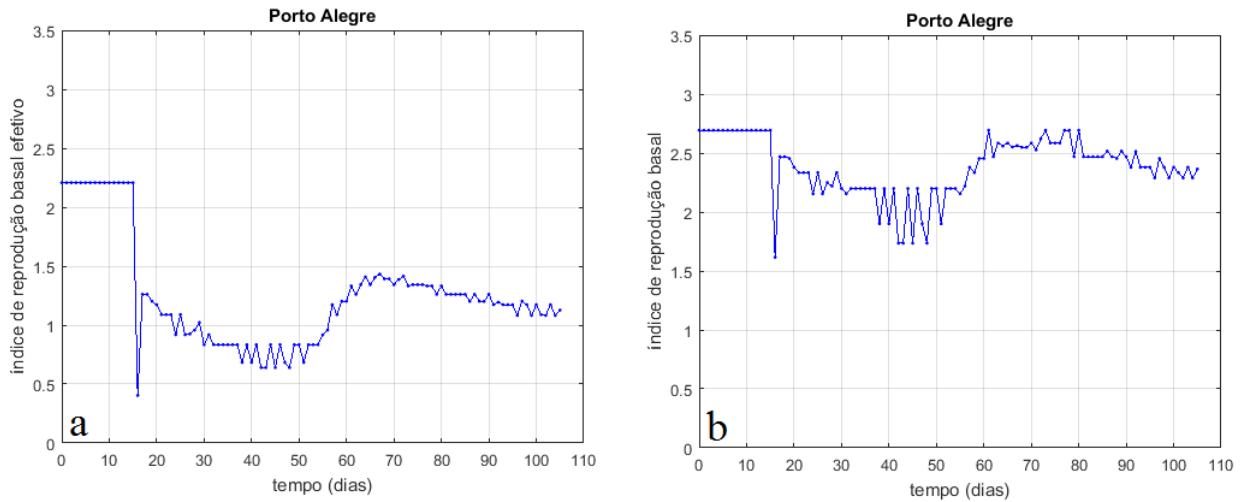
Figura 18 - Identificação paramétrica e simulação para a cidade de Porto Alegre, de 11/03 a 16/06.



Fonte: Elaborada pelo autor.

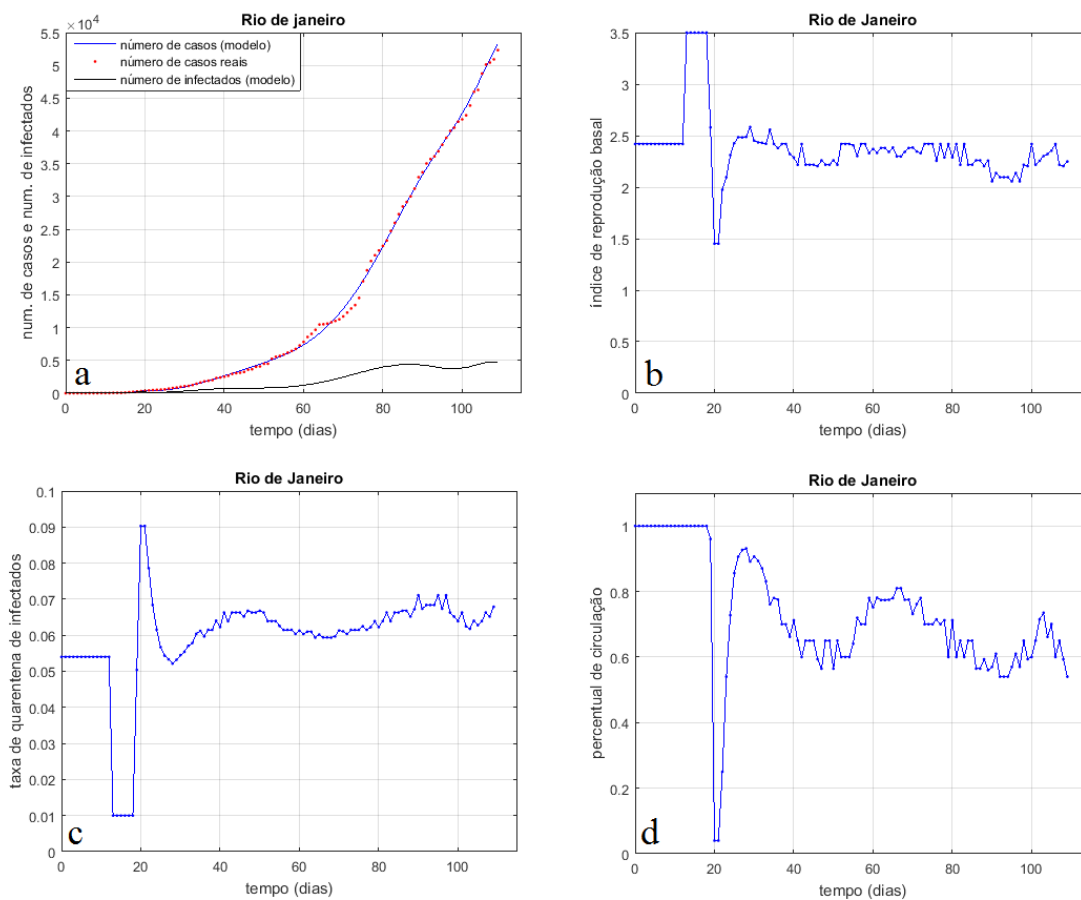


Figura 19 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de são Porto Alegre.



Fonte: Elaborada pelo autor.

Figura 20 - Identificação paramétrica e simulação para a cidade Rio de Janeiro, de 06/03 a 16/06.

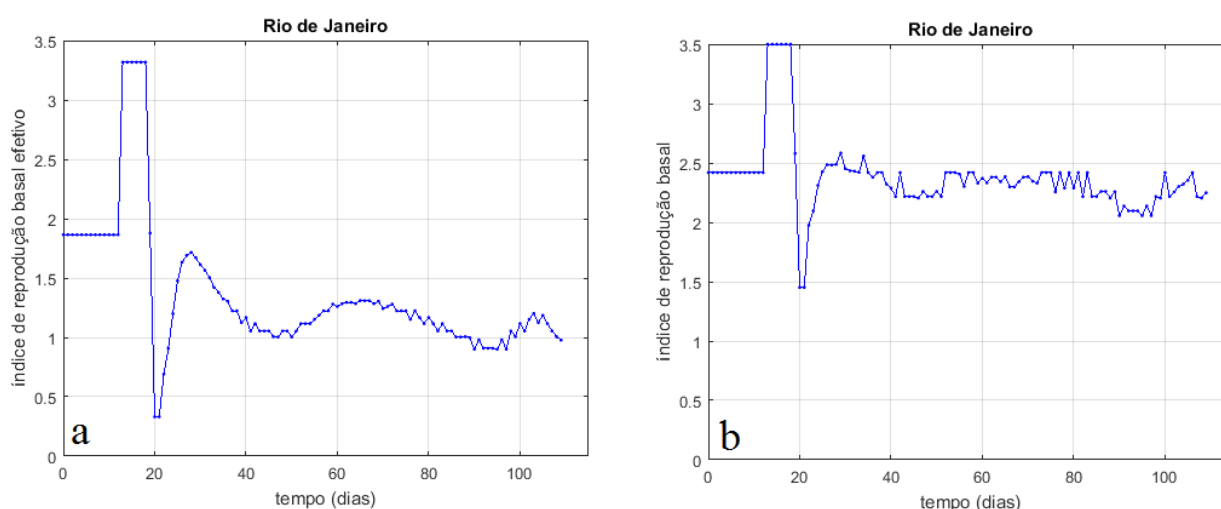


Fonte: Elaborada pelo autor.



O Rio de Janeiro esteve com o $\bar{R}_0(t)$ acima de 1 até o tempo $t=85$ (Figura 21a) e, antes desse tempo, o crescimento do número de casos foi acelerado (Figura 20a). Após esse tempo, percebe-se que houve uma leve inflexão na curva do número de casos, que passa a crescer com aceleração negativa (Figura 24a) durante um curto intervalo de tempo, pois o $\bar{R}_0(t)$ tende a ficar oscilando em torno de 1 para $t \geq 97$. Isto faz com que o crescimento do número de casos se aproxime de uma reta, nos últimos dias de dados reais vistos na Figura 20a.

Figura 21 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade do Rio de Janeiro.



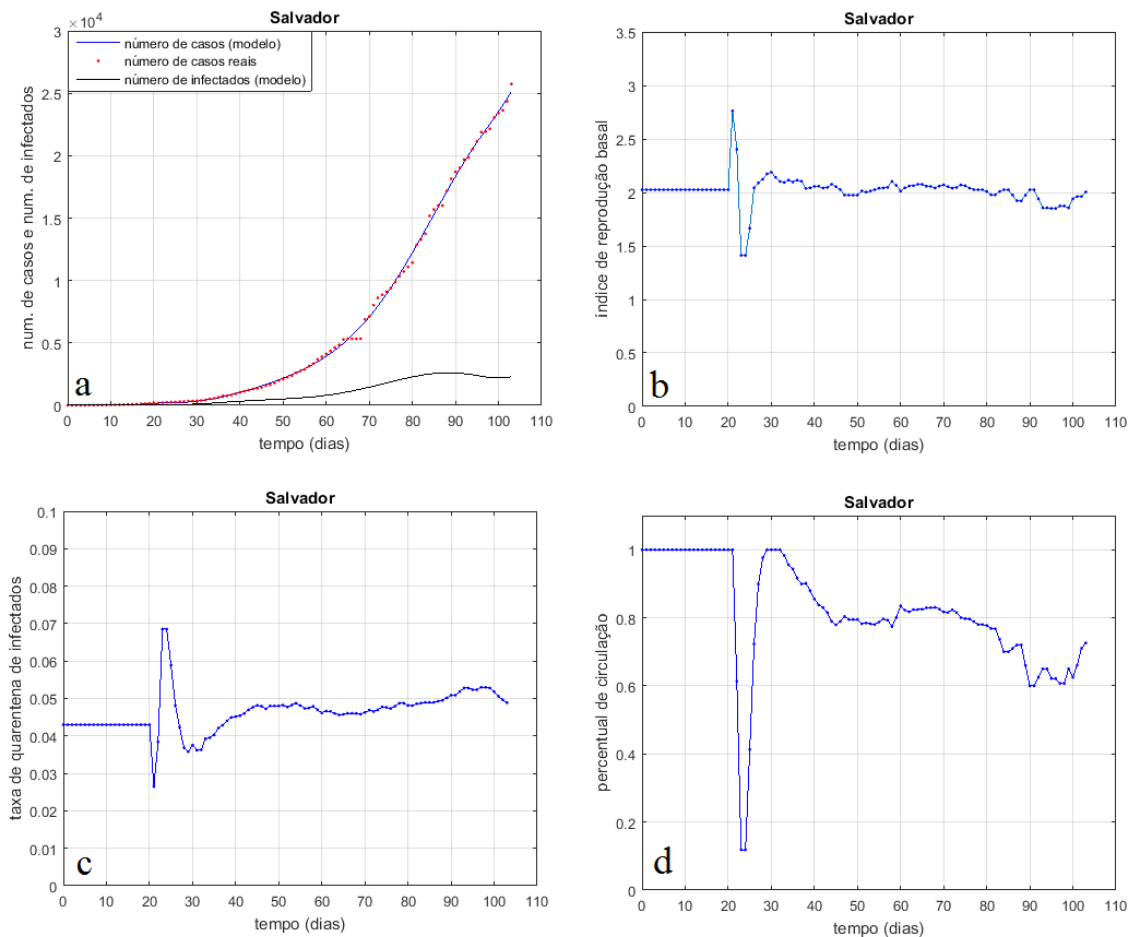
Fonte: Elaborada pelo autor.

Em Salvador há uma diminuição progressiva do $\bar{R}_0(t)$ (Figura 23a), parâmetro que ficou menor do que 1 durante 10 dias ($90 \leq t \leq 100$) e depois se tornou levemente maior do que 1 novamente. O percentual de circulação ficou sempre superior a 60% (Figura 22d), fato que impede uma desaceleração na curva do número de casos (Figura 22a).

Em São Luís o $\bar{R}_0(t)$ diminuiu progressivamente até ficar menor do que 1 a partir de $t > 70$ (Figura 25a). A partir desse tempo há uma inflexão na curva do número de casos, que passa a crescer com aceleração negativa (Figura 24a). Essa situação melhorou porque o $\bar{R}_0(t)$ continuou diminuindo gradativamente, chegando em 24/06/2020 a ficar próximo de 0.5%. Este bom desempenho foi possível porque o $R_0(t)$ diminuiu ao longo do tempo (Figura 24b), o k aumenta (Figura 24c) e, principalmente, o $p(t)$ diminuiu (Figura 24d), chegando a 22.67% em 24/06/2020. Informações de jornais afirmam que em São Luís houve decreto de *lockdown* em maio, 2020.

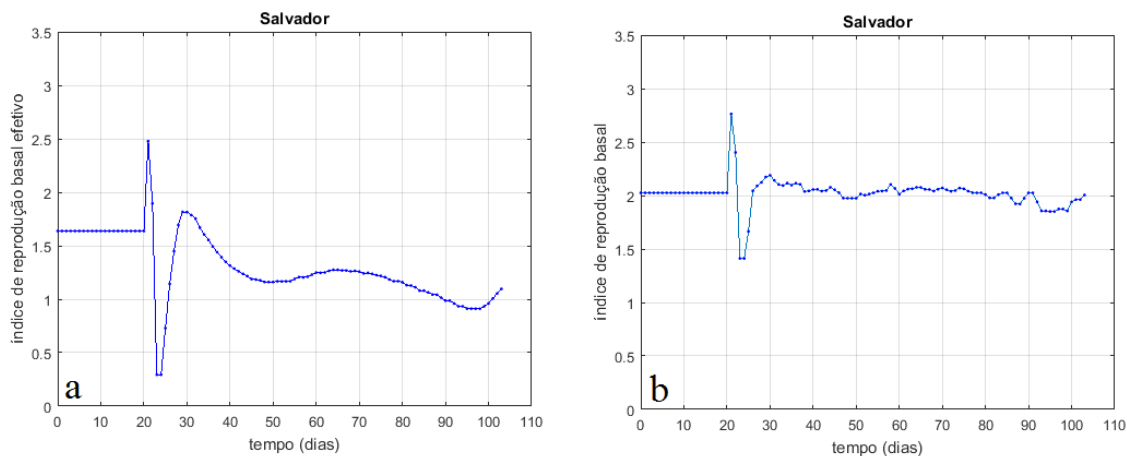


Figura 22 - Identificação paramétrica e simulação para a cidade de Salvador, de 16/03 a 24/06.



Fonte: Elaborada pelo autor.

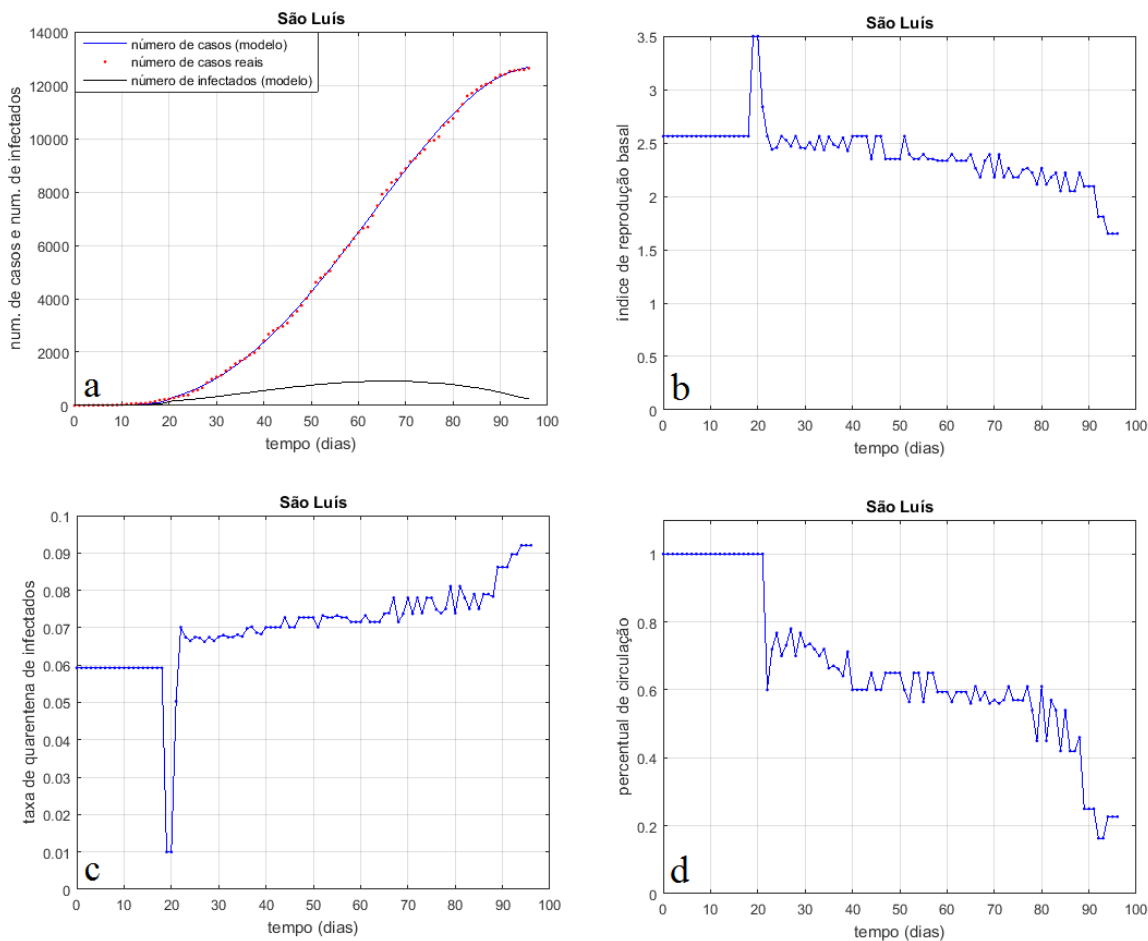
Figura 23 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade de Salvador.



Fonte: Elaborada pelo autor.

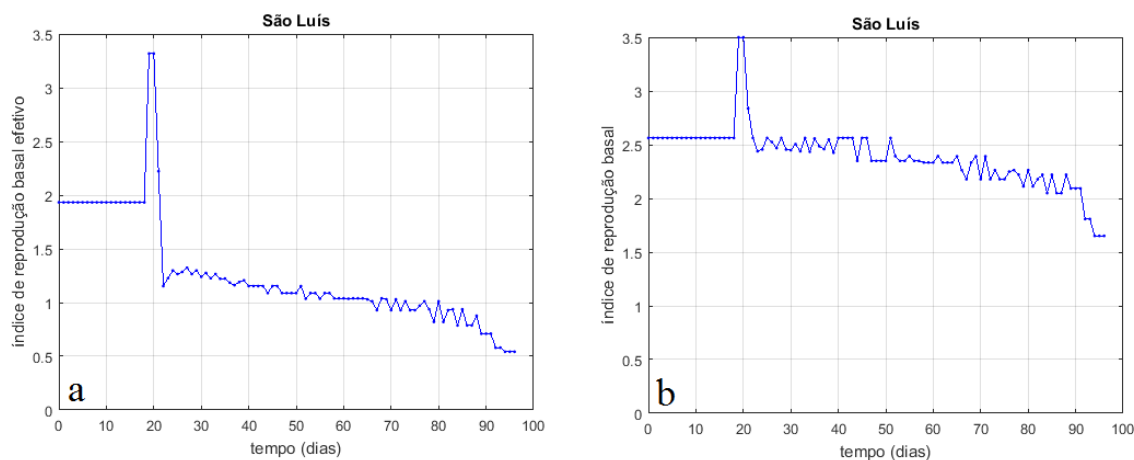


Figura 24 - Identificação paramétrica e simulação para a cidade São Luís, de 25/02 a 15/06.



Fonte: Elaborada pelo autor.

Figura 25 - $\bar{R}_0(t)$ e $R_0(t)$ para a cidade do São Luís.



Fonte: Elaborada pelo autor.

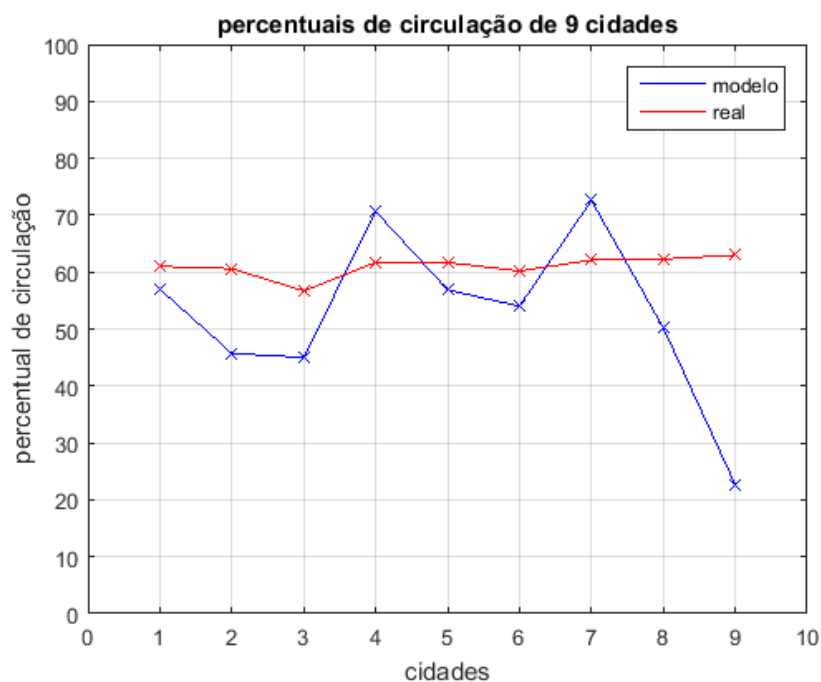


O Quadro I contém os valores numéricos relativos ao último dia de dados reais e simulados, equivalente a 24/06/2020. Caso a intenção fosse fazer previsões para qualquer uma das cidades estudadas no presente trabalho, os parâmetros do Quadro I poderiam ser utilizados. Uma alternativa seria utilizar a 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 in loco. 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 no Quadro I. A Figura 26 mostra os valores de P_{mod} e P_{real} na forma gráfica. No eixo das abcissas, a numeração das cidades obedece à ordem alfabética vista no Quadro I. A principal conclusão que se pode 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%.

Quadro I - Valores numéricos de parâmetros relativos ao último dia do conjunto de dados reais e simulados.

Cidade	P_{mod} (%)	P_{real} (%)	k	R_0	\bar{R}_0
Belém	57	61.0	0.066	1.841	0.823
Brasília	45.72	60.6	0.0529	2.447	1.075
Fortaleza	45	56.7	0.0756	2.137	0.84
Pelotas	70.68	61.7	0.0475	2.267	1.225
Porto Alegre	56.89	61.7	0.0505	2.364	1.126
Rio de Janeiro	54	60.2	0.0679	2.248	0.976
Salvador	72.6	62.1	0.0489	2.005	1.096
São Paulo	50.2	62.3	0.0281	1.869	0.941
São Luís	22.67	63	0.092	1.651	0.542

Fonte: Elaborado pelo autor.

**Figura 26** – Percentuais de circulação reais e previstos pelo modelo.

Fonte: Elaborada pelo autor.

6. CONCLUSÕES

As epidemias por COVID-19 nas cidades que foram analisadas no presente trabalho 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 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 do presente artigo é a estrutura de identificação paramétrica, que pode ser utilizada em qualquer cidade ou região. Em trabalhos futuros, pretende-se investigar a utilização de controle ótimo aplicado à contenção da epidemia em uma determinada localidade.

7. REFERÊNCIAS

GIORDANO, G. *et al.* Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. **Nature Medicine**, v.26, p.855–860, abr. 2020.

GOMES, S. C. P.; MONTEIRO, I. O.; ROCHA, C. R. **Modelagem dinâmica da Covid-19 com aplicação a algumas cidades brasileiras**. [Preprint] Disponível em: <<https://imef.furg.br/images/stories/documentos/2020/artigo2.pdf>>. Acesso em: 30 jun. 2020.

KERMACK, W.; MCKENDRICK, A. A Contribution to the Mathematical Theory of Epidemics. **Proceedings of the Royal Society. Series A, Mathematical, Physical and Engineering Science, Containing Papers of a Mathematical and Physical Character**, v.115, n.772, p.700-721, ago. 1927.

LAGARIAS, J. C.; REEDS, J. A.; WRIGHT, M. H.; WRIGHT, P. E. Convergence properties of the nelder-mead simplex method in low dimensions. **Siam Journal of Optimization**, v.9 n.1, p.112-147, 1998.

Submetido em: **30/06/2020**

Aceito em: **24/07/2020**