Parameter calibration and uncertainty quantification in an SEIR-type COVID-19 model using approximate Bayesian computation

Thiago G. Ritto[1,3], Americo Cunha Jr [2], David A.W. Barton[3]

[1]*Department of Mechanical Engineering, Universidade Federal do Rio de Janeiro*
Cidade Universitária - Ilha do Fundão, 21945-970, Rio de Janeiro, Brazil
tritto@mecanica.ufrj.br
[2]*Institute of Mathematics and Statistics, Universidade do Estado do Rio de Janeiro*
Rua São Francisco Xavier, 524, Maracanã, 20550-900, Rio de Janeiro, Brazil
americo.cunha@uerj.br
[3]*Faculty of Engineering, University of Bristol , Beacon House, Queens Road, BS8 1QU, Bristol, UK*
david.Barton@bristol.ac.uk

Abstract. The present paper applies the approximate Bayesian computation (ABC) for parameter estimation and uncertainty quantification in an SEIR-type model with data of hospitalisation and deaths from the city of Rio de Janeiro. The analysed model considers eight compartments: susceptible, exposed, infectious, asymptomatic, hospitalised, recovered and deceased (SEIAHRD). ABC is employed to update the prior probability density function of the model parameters, where a two objective optimisation problem is formulated (data of healthcare and deaths) and eleven parameters are identified. The transmission rate is allowed to vary over time (to change its baseline). The applied model seems to be consistent with the available data.

Keywords: COVID-2019, nonlinear dynamics, stochastic modelling, ABC, SIR model

1 Introduction

Since the COVID-2019 outbreak in December 2019 [1] many researchers have contributed with a variety of models to deal with COVID data and the epidemic spread understanding and predictions. We are particular interested in susceptible–exposed–infectious–removed (SEIR) compartment models [2]. These models are helpful to analyse different scenarios, to make predictions, and to support decisions. They have a good balance between simplicity (fast to run simulations) and complexity (good to represent the physics of the problem).

Pacheco et al. [3] analysed an SEIR-type model and investigated different scenarios for Brazil. The importance of social isolation and hospital infrastructure was highlighted. Vyasarayani and Chatterjee [4] studied an SEIR model with and an additional compartment for quarantine. They considered time delays for latency and an asymptomatic phase. Yu et al. [5] proposed a fractional SEIR model based on the coupling effect of inter-city networks. None of the above investigations detail the calibration procedure or consider a stochastic model to propagate uncertainties.

Kucharski et al. [6] analysed an SEIR model incorporating uncertainty in case observation using a Poisson observed process (newly symptomatic cases, reported onsets of new cases, reported confirmation of cases) and a binomial observation process (infection prevalence on evacuation flights). The calibration process is not detailed. He et al. [7] analysed an SEIR model with hospitalisation and quarantine. The model parameters are identified using the particle swarm optimisation (PSO) algorithm which is a population-based stochastic optimisation. Stochastic infection is considered introducing a white Gaussian noise. Jha et al. [8] considered a multiple coupled partial differential equations governing the evolution of susceptible, exposed, infectious, recovered and deceased individuals. The Bayesian learning approach is implemented to calibrate the model parameters. They considered additive Gaussian noise to construct the likelihood function and assumed log-Normal priors.

Uncertainties play a major role in epidemiological models. Model parameters, the model itself, and data are uncertain. It is crucial to take into account uncertainties and perform a robust calibration procedure. Probability

theory might be used in this endeavour [9], and the Bayesian learning strategy [10, 11] is convenient because prior knowledge is updated consistently with data and uncertainty quantification occurs automatically.

This paper applies the approximate Bayesian computation (ABC) [12] to calibrate the parameters and quantify uncertainties in an SEIR-type model [13]. The analysed model considers eight compartments: susceptible, exposed, infectious, asymptomatic, hospitalised, recovered and deceased (SEIAHRD) [14].

ABC is a Bayesian likelihood-free strategy where the prior probability density function of the parameters is updated with data. It can be used to propagate uncertainty throughout the model and compute probabilistic envelopes of the response. ABC is employed to update the prior Uniform probability density function of the model parameters. A two-objective optimisation problem is formulated (healthcare and deaths data) and eleven parameters are identified. The transmission rate is allowed to vary over time (to change its baseline). Data from the city of Rio de Janeiro are used to test the new methodology.

The paper is organised as follows. Section 2 depicts the SEIR deterministic and stochastic models. ABC is presented in section 3. The results are shown in section 4: sensitivity analysis, parameter calibration, uncertainty quantification, and other simulations.

2   SEIARHD COVID-19 model

The SEIR model presented here is taken from [14] and was inspired in [15]. Infection spreads via direct contact between a susceptible and infectious individual. Delay is modelled as an exposed group: there is a latent period until an infected person becomes able to transmit. Among the infectious, most individuals are asymptomatic; only a fraction display symptoms after incubation. Disease-related deaths are considered when infectious.

The model considers the dynamics of susceptible (S), exposed (E), infectious (I), asymptomatic (A), hospitalised (H), recovered (R) and deceased (D) individuals. Variable C is the cumulative infected individuals. The model has eight parameters:

- N_0 - initial susceptible population (number of individuals)
- β - transmission rate (days^{-1})
- α - latent rate (days^{-1})
- f_E - symptomatic fraction (non-dimensional)
- γ - recovery rate (days^{-1})
- ρ - hospitalisation rate (days^{-1})
- δ - death rate (days^{-1})
- κ_H - hospitalisation recovery-factor (non-dimensional)

The deterministic dynamic model with constant parameters is written as [14]

$$
\begin{aligned}
\dot{S}(t) &= -\beta S(t)(I(t) + A(t))/N(t)\,, \\
\dot{E}(t) &= \beta S(t)(I(t) + A(t))/N(t) - \alpha E(t)\,, \\
\dot{I}(t) &= f_E \alpha E(t) - (\gamma + \rho + \delta)I(t)\,, \\
\dot{A}(t) &= (1 - f_E)\alpha E(t) - (\gamma + \delta)A(t)\,, \\
\dot{H}(t) &= \rho I(t) - (\gamma + \kappa_H \delta)H(t)\,, \\
\dot{R}(t) &= \gamma(I(t) + A(t) + H(t))\,, \\
\dot{D}(t) &= \delta(I(t) + A(t) + \kappa_H \delta H(t))\,, \\
\dot{C}(t) &= \alpha E(t)\,, \\
N(t) &= N_0 - D(t)\,,
\end{aligned}
\tag{1}
$$

with initial conditions $\{S_0, E_0, I_0, A_0, H_0, R_0, D_0, C_0\}$.

The parameters $\{\beta, \alpha, f_E, \gamma, \rho, \delta, \kappa_H, N_0\}$ are modelled as random variables $\{\hat{\beta}, \hat{\alpha}, \hat{f}_E, \hat{\gamma}, \hat{\rho}, \hat{\delta}, \hat{\kappa}_H, \hat{N}_0\}$ that must be identified. Furthermore, $\hat{\beta}$ is described with the aid of an underling time dependent function as will be

detailed further in this section. The stochastic model with time dependent transmission rate is written as

$$\dot{\mathbf{S}}(t) = -\hat{\beta}(t)\mathbf{S}(t)(\mathbf{I}(t) + \mathbf{A}(t))/\mathbf{N}(t)\,,$$
$$\dot{\mathbf{E}}(t) = \hat{\beta}(t)\mathbf{S}(t)(\mathbf{I}(t) + \mathbf{A}(t))/\mathbf{N}(t) - \hat{\alpha}\mathbf{E}(t)\,,$$
$$\dot{\mathbf{I}}(t) = \hat{f}_E\hat{\alpha}\mathbf{E}(t) - (\hat{\gamma} + \hat{\rho} + \hat{\delta})\mathbf{I}(t)\,,$$
$$\dot{\mathbf{A}}(t) = (1 - \hat{f}_E)\hat{\alpha}\mathbf{E}(t) - (\hat{\gamma} + \hat{\delta})\mathbf{A}(t)\,,$$
$$\dot{\mathbf{H}}(t) = \hat{\rho}\mathbf{I}(t) - (\hat{\gamma} + \hat{\kappa}_H\hat{\delta})\mathbf{H}(t)\,, \qquad (2)$$
$$\dot{\mathbf{R}}(t) = \hat{\gamma}(\mathbf{I}(t) + \mathbf{A}(t) + \mathbf{H}(t))\,,$$
$$\dot{\mathbf{D}}(t) = \hat{\delta}(\mathbf{I}(t) + \mathbf{A}(t) + \hat{\kappa}_H\hat{\delta}\mathbf{H}(t))\,,$$
$$\dot{\mathbf{C}}(t) = \hat{\alpha}\mathbf{E}(t)\,,$$
$$\mathbf{N}(t) = \hat{N}_0 - \mathbf{D}(t)\,,$$

where the boldface represents random responses. As the disease spreads, the parameter β might change, and it will be taken in account by the following expression (taken from [16]):

$$\beta(t) = \beta_1 + \frac{\beta_2 - \beta_1}{2}\left(1 + \tanh\left(\rho_\beta\frac{t - t_\beta}{2}\right)\right)\,, \qquad (3)$$

where β_1 is the initial value of β, β_2 the final value, ρ_β defines how fast β reaches β_2, and t_β is the transition time (when $t = t_\beta$ then $\beta = (\beta_1 + \beta_2)/2$). There is a total of eleven random variables to be calibrated: $\boldsymbol{\varphi} = \{\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}_\beta, \hat{t}_\beta, \hat{\alpha}, \hat{f}_E, \hat{\gamma}, \hat{\rho}, \hat{\delta}, \hat{\kappa}_H, \hat{N}_0\}$.

3 Approximate Bayesian computation for parameter estimation and uncertainty quantification

3.1 Approximate Bayesian computation

The approximate Bayesian computation (ABC) [12] does not assume a likelihood function and so the usual assumption of additive independent Gaussian noise is not necessary. The model prediction (with parameter $\boldsymbol{\varphi}^*$) and the experiment are directly compared; for instance

$$\text{error}(\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*)) = \alpha_{\text{er}}\frac{||\mathbf{y}_{\text{hosp}} - \mathbf{y}_{m,\text{hosp}}(\boldsymbol{\varphi}^*)||^2}{||\mathbf{y}_{\text{hosp}}||^2} + (1 - \alpha_{\text{er}})\frac{||\mathbf{y}_{\text{deaths}} - \mathbf{y}_{m,\text{deaths}}(\boldsymbol{\varphi}^*)||^2}{||\mathbf{y}_{\text{deaths}}||^2}\,, \qquad (4)$$

where $\boldsymbol{\varphi}^*$ is the set of parameters of the model, $\mathbf{y}_m$ is the model prediction, $\mathbf{y}$ is the available data. Data from hospitalised individuals $\mathbf{y}_{\text{hosp}}$ and deaths $\mathbf{y}_{\text{deaths}}$ are considered, and $\alpha_{\text{er}} \in [0, 1]$ is the weight put to the hospitalisation data. If $\alpha_{\text{er}} = 0.5$ data from healthcare and deaths have the same weight.

We must combine the prior information about the parameters in a prior distribution $\pi(\boldsymbol{\varphi})$, and define a tolerance ϵ. A simple rejection sample can be implemented, that is, a sample $\boldsymbol{\varphi}^*$ drawn from the prior distribution is accepted only if $\text{error}(\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*)) < \epsilon$. Algorithm 1 shows the applied ABC strategy.

4 Numerical results

Figure 1 shows the response of the deterministic dynamic model considering the parameters: $\beta = 0.5$ (transmission rate), $1/\alpha = 9$ days (latent period), $f_E = 0.15$ (asymptomatic fraction), $1/\gamma = 14$ days (recovery period), $1/\rho = 11$ (hospitalisation period), $1/\delta = 33$ (death period), $\kappa_H = 0.25$ (hospitalisation recovery-factor). The susceptible population is $N_0 = 25000$ and the initial conditions are $D_0 = 0$, $R_0 = 0$, $H_0 = 0$, $A_0 = 10$, $I_0 = 5$, $E_0 = 1$.

The number of exposed individuals increases fast and reaches its peak on the 63rd day. The other groups also raise and then fall back to zero. The number of infectious individuals reaches its peak on the 68th day, asymptomatic, recovered and deaths on the 71th, and hospitalised on the 77th. The ordinate axis shows the normalised individuals: the number of individuals is divide by the susceptible population N_0.

Algorithm 1 ABC for parameter estimation and UQ

1: **procedure** ABC($\pi(\boldsymbol{\varphi}), \mathbf{y}_m(\boldsymbol{\varphi}^*), \epsilon$)
2:     **for** $i = 1 : n_{\text{ABC}}$ **do**
3:         Sample a candidate set of parameters $\boldsymbol{\varphi}^*$ from $\pi(\boldsymbol{\varphi})$;
4:         Compute the model prediction $\mathbf{y}_m(\boldsymbol{\varphi}^*)$;
5:         Evaluate the results using error($\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*)$);
6:         **if** error($\mathbf{y}, \mathbf{y}_m(\boldsymbol{\varphi}^*)$) $< \epsilon$ **then**
7:             Accept $\boldsymbol{\varphi}^*$;
8:             Save $\boldsymbol{\varphi}^*$ and $\mathbf{y}_m(\boldsymbol{\varphi}^*)$;
9:         **else**
10:            Reject $\boldsymbol{\varphi}^*$;
11:         **end if**
12:     **end for**
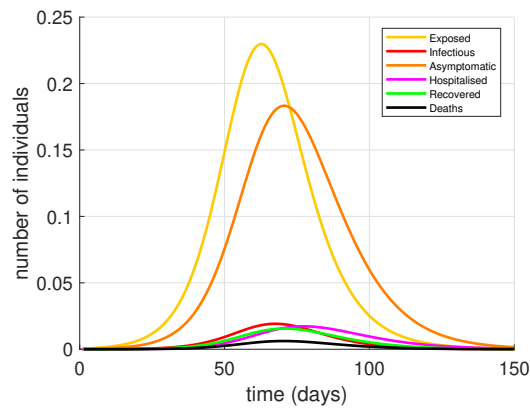13: **end procedure**

Figure 1. SEIAHRD model dynamic response.

4.1 Sensitivity analysis

Figures 2 and 3 show the sensitivity of the healthcare and death curves with respect to each parameter of the model. A Uniform probability density function with $\pm 40\%$ of a reference value is considered for each one of the eight parameters. The 95% probabilistic envelope of for healthcare and deaths number is plotted. A wide envelope indicates that the system response is sensitive to the particular parameter analysed.

Figure 2(a) shows that the transmission rate β is the most relevant parameter of the model, yielding the wider envelopes, and it is followed by the latent rate α (Fig. 2(b)). Figures 2(d), 3(b) and 3(d) show that the recovery rate γ, the death rate δ and the initial susceptible population N_0 also have a considerable impact at the healthcare and deaths curves.

The model response is less sensitive to the hospitalisation recovery-factor κ_H, for the configuration analysed (Fig. 3(c)). It is interesting to observe, Fig. 2(c) and 3(a), that the symptomatic fraction f_E and the hospitalisation rate ρ have a great impact on the hospitalisation curve, but little impact on the deaths curve.

4.2 Approximate Bayesian computation (ABC)

Data from healthcare and deaths were gathered for the period of 24th January 2020 to 28th July 2021 (156 days) from the Panel COVID-19 of the Rio de Janeiro city Government Health Department, `http://painel.saude.rj.gov.br/monitoramento/covid19.html`.

Figure 4 shows the results for the ABC procedure where $\alpha_{\text{er}} = 0.5$ and the tolerance was fixed to $\epsilon = 0.15$, after some numerical tests. Only 4% of the two thousand Monte Carlo observations were accepted.

Figures 4(a) and 4(b) shows the ABC results, where the 95% probabilistic envelope (black lines) is plotted together with data from Rio de Janeiro city. These envelopes are a consequence of the uncertain parameters (posterior probability distribution). The model prediction could envelope most of the available noisy data. However, the end of the healthcare curve, Fig. 4(a), escapes the model prediction.

Table 1 shows the parameter bounds, the maximum a posterior (MAP) estimate obtained applying ABC.
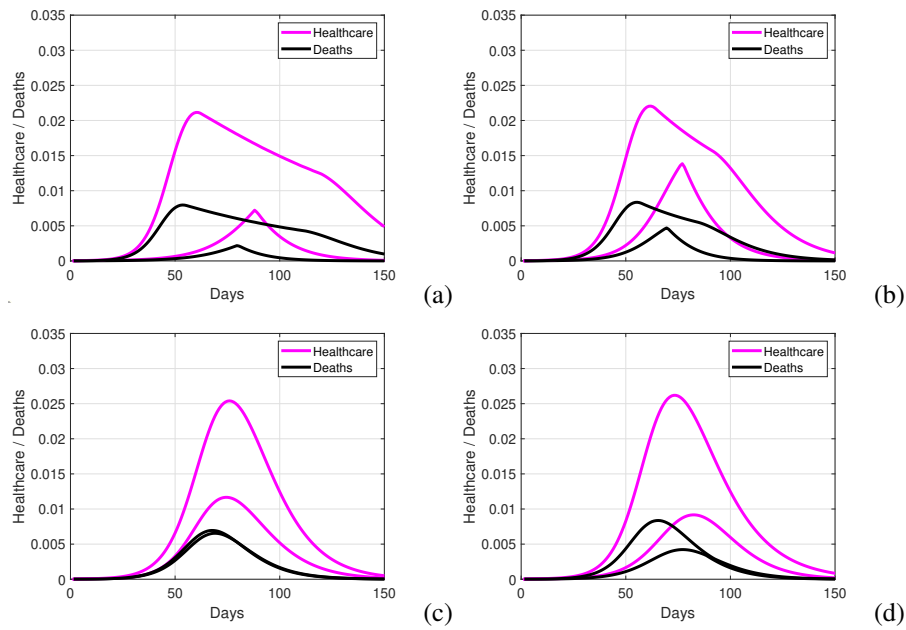

Figure 2. Probabilistic envelope (95%) of the system response for a Uniform input of the parameters $\beta \in [0.3, 0.7]$ (a), $1/\alpha \in [5.4, 12.6]$ (b), $f_E \in [0.09, 0.21]$ (c) and $1/\gamma \in [8.4, 19.6]$ (d). Two thousand Monte Carlo samples.
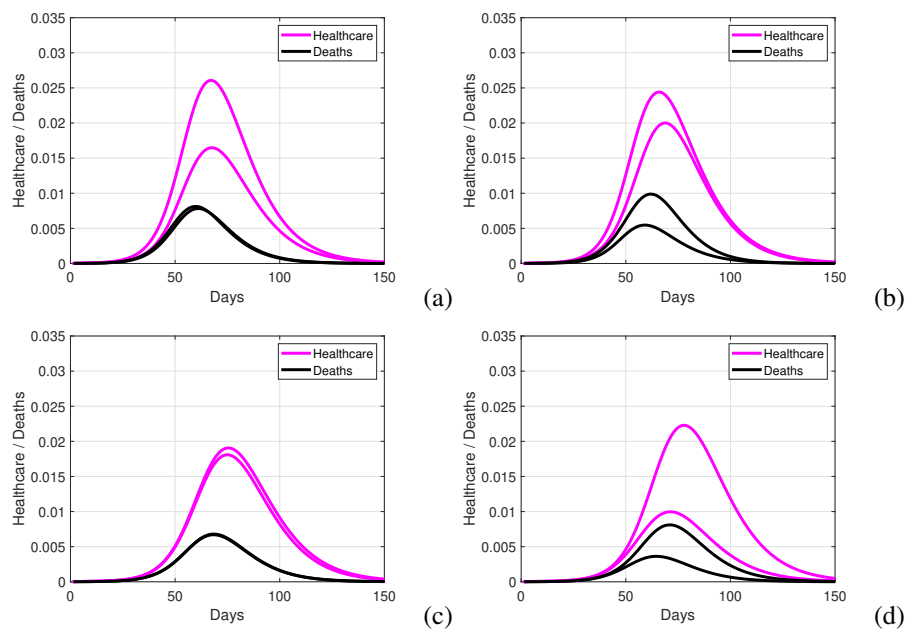

Figure 3. Probabilistic envelope (95%) of the system response for a Uniform input of the parameters $1/\rho \in [8.3, 20.0]$ (a), $1/\delta \in [23.8, 55.5]$ (b), $\kappa_H \in [0.15, 0.35]$ (c) and $N_0 \in [15000, 35000]$ (d). Two thousand Monte Carlo samples.

Figure 5 shows the scatter plot of the ABC posterior (accepted) samples. The charts in the diagonal shows the posterior marginal histograms. The original Uniform distributions are updated to distributions with different shapes. Correlations between pairs of random variables can also be observed for the posterior distribution. The prior distribution does not consider any correlation. The second chart of the first line (and the first chart of the second line) shows that there is a positive correlation (0.72) between the transmission rate β and the latent period $(1/\alpha)$.

Figure 6 shows the identified transmission rate function. Its value starts at 0.42 (high value) and then decreases to 0.37. It has been reported that transmission rate can be high in the beginning of the disease spread and then decreases. In most of the analysed models found in the literature this parameter is considered constant. Allowing it to vary over time, permitted the model to better represent the problem and improve its fit to data.
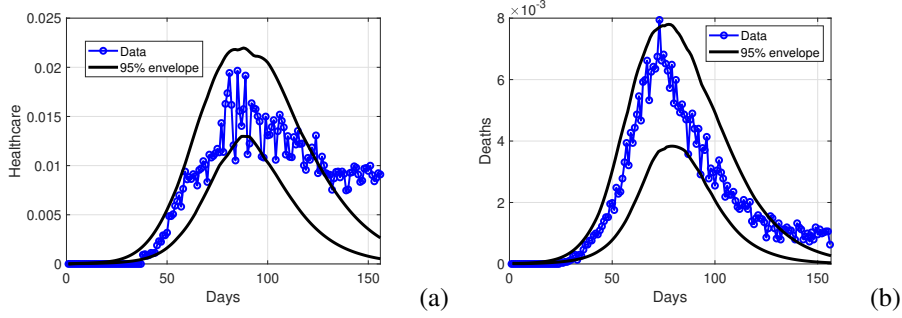
(a) (b)

Figure 4. Probabilistic envelope (95%) of the system response for the identified model together with hospitalisation and deaths data of Rio de Janeiro city; period of 24th January 2020 to 28th July 2021. Charts (a) and (b) are the hospitalisation and deaths curves obtained with ABC.

Parameter	Bounds	ABC
β_1 (-)	[0.30,0.70]	0.50
β_2 (-)	[0.24,0.56]	0.44
ρ_β (-)	[0.06,0.14]	0.10
t_β (days)	[30,70]	43
$1/\alpha$ (days)	[4.2,9.8]	12.5
f_E (-)	[0.09,0.21]	0.17
$1/\gamma$ (days)	[9,21]	19.6
$1/\rho$ (days)	[8.3,20.0]	16.7
$1/\delta$ (days)	[23.8,55.5]	28.5
κ_H (-)	[0.15,0.35]	0.21
N_0 (individuals)	[15,35]$\times10^3$	27.1$\times10^3$

Table 1. Parameters, bounds and maximum a posterior estimate (MAP) obtained with ABC. Prior distribution follows a Uniform distribution with bound $\pm40\%$ of a reference value; only t_β considers a larger percent range.
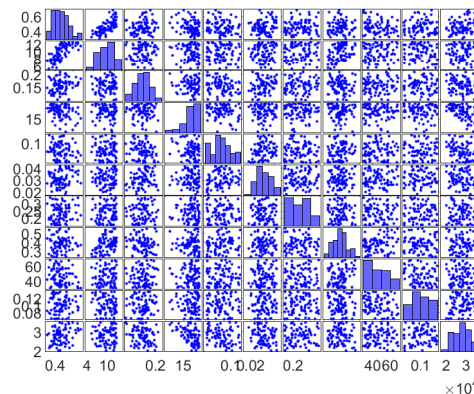

Figure 5. Scatter plot of the posterior PDF of the parameters: β_1, $1/\alpha$, f_E, $1/\gamma$, ρ, δ κ_H, β_2, t_β, ρ_β, and N_0.

Figure 7 shows the cumulative normalised individuals evolution. Note that there is no fluctuations in the data since the accumulation attenuates them. Healthcare and deaths data are compared with the model results considering the identified parameters using ABC. The overall trend are similar, but the model cannot capture the healthcare data tail of the healthcare data.
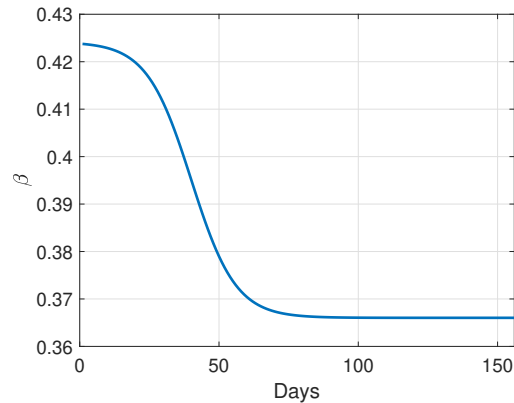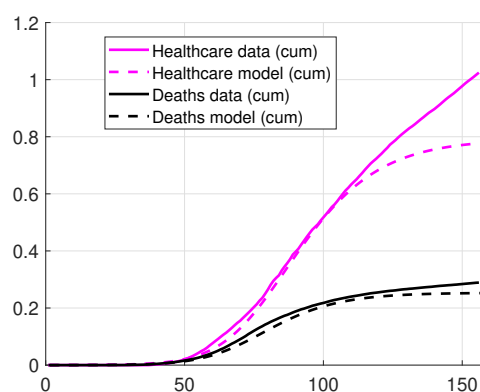
Figure 6. Identified $\beta(t)$.

Figure 7. Cumulative individuals with the identified model, comparing data and ABC results.

5 Concluding remarks

In this paper, we apply the approximate Bayesian computation (ABC) to a SEIR-type model for the COVID disease. The analysed model considers eight compartments: susceptible, exposed, infectious, asymptomatic, hospitalised, recovered and deceased (SEIAHRD). Data from the city of Rio de Janeiro (healthcare and deaths) are used in the calibration procedure. The results were consistent with the death data, but the tail of the healthcare data could not be properly represented.

The next steps of this research are to refine the identification procedure to obtain more representative parameter values and analyse different scenariostr with the calibrated model.

References

[1] WHO. Coronavirus disease 2019 (COVID-19). Situation report 24. Geneva: World Health Organization, vol. , 2020.

[2] E. Vynnycky and R. White. An Introduction To Infectious Disease Modelling. Oxford University Press, 2010.

[3] P. M. C. L. Pacheco, M. A. Savi, and P. V. Savi. COVID-19 dynamics considering the influence of hospital infrastructure: an investigation into Brazilian scenarios. Nonlinear Dynamics, vol. in press, 2021.

[4] C. P. Vyasarayani and A. Chatterjee. New approximations, and policy implications, from a delayed dynamic model of a fast pandemic. Physica D: Nonlinear Phenomena, vol. 414, pp. 132701, 2021.

[5] X. Yu, L. Lu, J. Shen, J. Li, W. Xiao, and Y. Chen. A fractional-order SEIHDR model for COVID-19 with inter-city networked coupling effects. Nonlinear dynamics, vol. 101, n. 3, pp. 1717–1730, 2021.

[6] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, E. J., S. Funk, and R. M. Eggo. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet Infectious Diseases, vol. 20, n. 5, pp. 553–558, 2020.

[7] S. He, Y. Peng, and K. Sun. SEIR modeling of the COVID-19 and its dynamics. Nonlinear dynamics, vol. 101, n. 3, pp. 1667–1680, 2021.

[8] P. Jha, L. Cao, and J. Oden. Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models. Computational Mechanics, vol. 66, n. 5, pp. 1055–1068, 2020.

[9] E. T. Jaynes. Probability Theory: the logic of science. Cambridge University Press, 2003.

[10] J. Kaipio and E. Somersalo. Statistical and Computational Inverse Problems. Springer, 2004.

[11] D. S. Sivia. Data analysis – a Bayesian tutorial. Oxford Science, 2006.

[12] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, vol. 6, n. 31, pp. 187–202, 2009.

[13] H. W. Hethcote. Mathematics of infectious diseases. SIAM Review, vol. 42, n. 4, pp. 599–653, 2000.

[14] B. Pavlack et al. EPIDEMIC - Epidemiology Educational Code. `www.EpidemicCode.org`, 2021.

[15] W. Lyra, do J.-D. Nascimento, J. Belkhiria, de Almeida L., P. P. M. Chrispim, and de I. Andrade. COVID-19 pandemics modeling with modified determinist SEIR, social distancing, and age stratification. the effect of vertical confinement and release in Brazil. PLoS ONE, vol. 15, n. 9, pp. e0237627, 2020.

[16] G. L. Vasconcelos, A. A. Brum, F. A. G. Almeida, A. M. S. Macêdo, G. C. Duarte-Filho, and R. Ospina. Standard and anomalous waves of COVID-19: A multiple-wave growth model for epidemics. medrxiv, vol. , 2021.