Mathematical modeling and forecasting of COVID-19: experience in Santiago de Cuba province

E. E. Ramirez-Torres
Departamento de Biomédica, Facultad de Ingeniería en Telecomunicaciones, Informática y Biomédica,
Universidad de Oriente, Santiago de Cuba, Cuba.

A. R. Selva Castañeda
Departamento de Telecomunicaciones, Facultad de Ingeniería en Telecomunicaciones, Informática y Biomédica,
Universidad de Oriente, Santiago de Cuba, Cuba

Y. Rodríguez-Aldana
Centro de Estudios de Neurociencias y Procesamiento de Señales,
Universidad de Oriente, Santiago de Cuba, Cuba

S. Sánchez Domínguez
Departamento de Matemática, Facultad de Ciencias Naturales y Exactas,
Universidad de Oriente, Santiago de Cuba, Cuba

L. E. Valdés García
Centro Provincial de Higiene, Epidemiología y Microbiología, Santiago de Cuba, Cuba

A. Palú-Orozco
Centro Provincial de Higiene, Epidemiología y Microbiología, Santiago de Cuba, Cuba

E. R. Oliveros-Domínguez
Dirección de Informatización, Universidad de Oriente, Santiago de Cuba, Cuba

L. Zamora-Matamoros
Departamento de Matemática, Facultad de Ciencias Naturales y Exactas,
Universidad de Oriente, Santiago de Cuba, Cuba

R. Labrada-Claro
Departamento de Matemática, Facultad de Ciencias Naturales y Exactas,
Universidad de Oriente, Santiago de Cuba, Cuba

M. Cobas-Batista
Departamento de Matemática, Facultad de Ciencias Naturales y Exactas,
Universidad de Oriente, Santiago de Cuba, Cuba

D. Sedal-Yanes
Facultad de Ciencias Sociales, Universidad de Oriente, Santiago de Cuba, Cuba

O. Soler-Nariño
Facultad de Ciencias Sociales, Universidad de Oriente, Santiago de Cuba, Cuba

P. A. Valdés-Sosa
Join China-Cuba Lab for Frontier Research in Translational Neurotechnology, Sichuan, China

J. I. Montijano
Instituto Universitario de Investigación de Matemáticas y Aplicaciones,
Universidad de Zaragoza, Zaragoza, Spain

L. E. Bergues Cabrales
Centro Nacional de Electromagnetismo Aplicado,
Universidad de Oriente, Santiago de Cuba, Cuba.
e-mail: berguesc@yahoo.com

In the province of Santiago de Cuba, Cuba, the COVID-19 epidemic has a limited progression that shows an early small-number peak of infections. Most published mathematical models fit data with high numbers of confirmed cases. In contrast, small numbers of cases make it difficult to predict the course of the epidemic. We present two known models adapted to capture the noisy dynamics of COVID-19 in the Santiago de Cuba province. Parameters of both models were estimated using the approximate-Bayesian-computation framework with dedicated error laws. One parameter of each model was updated on key dates of travel restrictions. Both models approximately predicted the infection peak and the end of the COVID-19 epidemic in Santiago de Cuba. The first model predicted 57 reported cases and 16 unreported cases. Additionally, it estimated six initially exposed persons. The second model forecasted 51 confirmed cases at the end of the epidemic. In conclusion, an opportune epidemiological investigation, along with the low number of initially exposed individuals, might partly explain the favorable evolution of the COVID-19 epidemic in Santiago de Cuba. With the available data, the simplest model predicted the epidemic evolution with greater precision, and the more complex model helped to explain the epidemic phenomenology.

1. Introduction

The ongoing COVID-19 epidemic represents a global challenge for health-care systems. The SARS-COV-2 coronavirus causing the COVID-19 is highly transmissible [1, 2]. Nevertheless, the epidemic propagation in a region/country is influenced by several factors, including governmental measures [3]. The Cuban government has implemented several strategies of social isolation to control this epidemic. As a result, a small number of cases is reported in the province of Santiago de Cuba. Mathematical models are useful for understanding the COVID-19 transmission and the influence of government interventions [4–7]. Currie *et al.* [4] suggest that system dynamics (differential-equation-based modeling) is suitable to model government decisions that cause social distancing and isolation.

The most common system dynamics in epidemic modeling are SIR (susceptible - infectious - removed) and SEIR (susceptible-exposed-infectious-removed) type. The classic SIR and SEIR models assume that individuals change classes (*e.g.*, from susceptible to infected) at constant rates. This assumption is not realistic in most situations [8]. Additionally, it does not explain why the effective reproduction number of COVID-19 changes over time [9]. Some authors introduce time-dependent parameters to formulate realistic models of the current pandemic [5, 10–12].

A challenge for COVID-19 modelers is to describe the difference between total cases and diagnosed cases. One approach is to exclude this difference and fit the model to reported data [16–18]. Another option is to consider undetected cases as a compartment of the differential equations system [6, 7, 12, 19, 20]. Finally, some authors avoid the complication of adding a new compartment, and they model the error with different *ad hoc* methods [5, 10, 21, 22].

Mathematical models have been used to forecast the dynamics of COVID-19 in delimited geographical areas. These models usually fit big-number data after several days of epidemic evolution [7, 10, 23–26]. The official data reported from Santiago de Cuba show a maximum value of 36 active cases reached 31 days after the first confirmed case.

This peak is followed by a decreasing trend. These small and noisy data make difficult the model fitting. A common strategy to describe COVID-19 dynamics is to update one or more model parameters based on key days of government actions [7, 10, 12].

Ordinary differential equation (ODE) based models are widely used to capture the kinetic of biological systems and to predict the behaviour of time-dependent variables. The minimization of a cost function (as a strategy for fitting a given dynamic model) does not take into account the uncertainty of parameters. Additionally, this approach can lead to unrealistic fittings [27].

Bayesian inference offers a platform that deals naturally with both problems, but it requires the calculation of likelihood functions [13–15, 28]. In many practical problems, likelihood functions become analytically intractable; nevertheless, data can be simulated from a parameter vector by applying some algorithm that depends on the model. Thus, several techniques have been developed to infer parameters without using likelihood functions. These techniques are usually called approximate Bayesian computation (ABC) or free likelihood inference [28, 29].

ABC can be used to estimate parameters when the available data of an infectious disease is coarse [30]. Besides, ABC accomplishes parameters estimation for models that include dedicated laws to describe the error between total cases and reported cases.

The aim of this paper is to adapt two models reported in the literature to capture the dynamics of the noisy and small-number data of COVID-19 in Santiago de Cuba. For this, we modify an extended SEIR model [10] by introducing a dynamic-mode gamma distribution to describe the error between total and reported cases. Additionally, we adapt the classic SIR model reported in [31] by introducing Poisson noise in the observation of removed cases. For both models, we update a parameter on critical days of government actions and estimate the parameter values with ABC. We are not aware that these modifications have been reported in the literature.

2. Methods

2.1. Data

Because no personal data of patients were used, the ethical approval and individual consent were not applicable.

We used official data of cumulative, infected, and removed cases of COVID-19 in Santiago de Cuba province. These data were reported from March 20th to May 4th, 2020. Daily data was provided by the Provincial Direction of Public Health of Santiago de Cuba and visualized in an information board from a web platform (http://www.covid19scu.uo.edu.cu). Forty-nine patients with SARS-COV-2 were confirmed using real-time polymerase chain reaction (PCR) tests. PCR tests were made in the Laboratory of Molecular Biology, Provincial Center of Hygiene, Epidemiology and Microbiology of Santiago de Cuba.

2.2. Epidemiological evidence

The first day of the COVID-19 in Santiago de Cuba province counted when the first case was diagnosed (March 17th, 2020). The first case (day one), the second case (day four), and the third case (day seven) were contacts of different travelers from various countries with COVID-19 transmission reported. The number of diagnosed cases increased from day seven and they were contacts of three first confirmed cases. The epidemiological investigation revealed that the primary infectious focus was small in Santiago de Cuba province due to the small number of travelers that arrived in this province. All confirmed cases with COVID-19 were attended and treated in the Dr. José Joaquín Castillo Duany hospital of Santiago de Cuba. The confirmed patients remained at least 14 days admitted to the hospital. When two PCR tests were negative, patients were sent home under medical supervision. Additionally, all contacts of these confirmed patients were quickly isolated during 14 days to detect symptomatic and non-symptomatic cases. After 14 days of isolation, unconfirmed cases of COVID-19 were sent home and daily evaluated by physicians.

The government of Santiago de Cuba took different actions that restrained COVID-19 transmission, such as the urban and interprovincial travel restrictions; workers were encouraged to work from their homes; safe transportation was provided for essential people in the different prioritized activities; school activity was temporarily suspended at all educational levels; social movement was restrained; and the modified quarantine was established on the case clusters of importance for disease transmission.

2.3. Methodology

In order to understand and predict COVID-19 in Santiago de Cuba, we have chosen two deterministic models: one more explanatory and the other more parsimonious. The first model allowed to elaborate hypotheses about some possible causes of the few reported cases. The second model was used to make faster and accurate predictions of the epidemic evolution. For both models, we updated a parameter on key days of government action. Also, we included error models to explain the variability of the observed data. Details of the chosen models are shown in the following section.

2.4. Models

Lin *et al.* [10] reported an SEIR model with three additional classes: the total population size (N), the public perception of risk (D), and the cumulative cases number (C). Also, they modeled the zoonotic transmission during the first month and then only person-to-person transmission, taking into account the emigration rate inside the country.

Unlike [10], we did not consider the zoonotic transmission because infection in Cuba was not originated by animals. Also, we neglected the emigration rate for two reasons: first, the few days elapsed before the transport restriction by air, land, and sea; second, the social isolation actions taken, such as remote work, stopping school activities, among others. As a result of these two assumptions, we obtained the following model, named model-I:

$$
\begin{cases}
\dfrac{dS}{dt} = -\dfrac{\beta(t)}{N}SI \\[2mm]
\dfrac{dE}{dt} = \dfrac{\beta(t)}{N}SI - \sigma E \\[2mm]
\dfrac{dI}{dt} = \sigma E - \gamma I \\[2mm]
\dfrac{dR}{dt} = \gamma I \\[2mm]
\dfrac{dD}{dt} = d\gamma I - \lambda D \\[2mm]
\dfrac{dC}{dt} = \sigma E,
\end{cases}
\tag{1}
$$

with

$$
\beta(t) = \beta_0(1-\alpha)\left(1 - \frac{D}{N}\right)^k,
\tag{2}
$$

where σ^{-1}, γ^{-1}, d, λ^{-1}, $\beta(t)$, β_0, α and k are the mean latent period, the mean infectious period, the proportion of severe cases, the mean duration of public reaction, the dynamic transmission rate, the initial transmission rate, the governmental measure strength and the intensity of individual response, respectively. The last expression of Eq. (1) follows from $C = I + R$. As α was considered a stepwise function in [10], we used two milestone dates: the days when interprovincial transport (April 10th, 2020) and urban transport (April 25th, 2020) were restricted.

To describe the error in reported data, we assumed that the difference $C_T - C_R$ between total cases (C_T) and reported cases (C_R) is a gamma probability distribution. The assumption of a discrete random variable as a continuous one

provided some advantages in the assessment of parameter estimation. For instance, it was possible to perform a nonparametric continuous hypothesis test over residues. The shape parameter of the gamma distribution was estimated, and the scale parameter was changed dynamically. This approach allowed us to place the mode closer to zero in the most optimistic scenarios. We used an incremental error in the reported rate during the early days of the epidemic. A dedicated law was formulated for the number of unreported cases per each reported one ($U(t)$), given by

$$U(t) = \begin{cases} pt + \delta, & \text{if } t < d_T \\ pd_T + \delta, & \text{if } t \geq d_T, \end{cases} \quad (3)$$

where p and δ are tuning parameters and d_T is the time from which COVID-19 transmission showed higher regularity in the case count. We empirically set the values in $p = 0.01$ day^{-1}, $\delta = 0.04$ and $d_T = 8$ days. We assumed that $U(t)$ increased during the days when almost no new cases arose, and after that, it remained roughly constant. This assumption is consistent with a previous study [32] that reported growth in undocumented cases during the early days of the epidemic. The expression for the dynamic mode of error ($M(t)$) is

$$M(t) = \frac{C_T}{U(t) + 1}. \quad (4)$$

The difference between C_T and $M(t)$ was considered a time-dependent approximation of undetected cases.
Model-II consisted of a standard SIR model [31], adapted with the introduction of Poisson noise in the removed cases. For the parameter calibration of model-II, reported cumulative and removed cases were considered. We described $\beta(t)$ in the model-II as a stepwise function that included changes in key dates. The error was assumed as a standard normal distribution for model-II.

2.5. Parameter estimation

Parameter estimation was performed with the ABC approach [28, 33, 34]. We preferred the sequential Monte Carlo (ABC SMC) method described in the Algorithm 4.8 of [34] because of its well-documented results. Nevertheless, the ABC SMC method can perform slowly computing adaptive tolerances.

This sampler (Algorithm 1 in Appendix A) outsmarts its predecessors in automatic computing of tolerances based on the effective sample size of the previous population. We used $N = 10^4$, $h_0 = 10^8$, $T = 5$ and a Gaussian acceptance kernel function K_h. We refer readers to [34] for a deeper understanding of ABC samplers and their adjustable parameters.

The reweight step of the chosen ABC SMC sampler led to an optimization problem in the calculation of h_m. Thus, it may introduce delays in the computation process, especially if the posterior distribution is distant from the prior distribution. Therefore, we introduced the hybrid version of ABC techniques, named ABC H algorithm (Algorithm 2

in Appendix B) as a hybrid of rejection sampler (ABC RS) and Markov chain Monte Carlo (ABC MCMC) approaches. Although we have not demonstrated the ergodicity of the ABC H algorithm, we empirically ascertained that it found a maximum-a-posteriori (MAP) located within the parameter region determined by the ABC SMC algorithm, consuming less computational time. Hence, we preferred to use the MAPs obtained by the ABC H sampler for quick prediction and the ABC SMC process for more in-depth posterior Bayesian analysis. Specifically, we used the ABC SMC results to compute the credible intervals of posterior distributions. Details of the algorithms ABC SMC and ABC H are shown in A and B, respectively.

As far as we know, the ABC H sampler is a new idea, and its main disadvantage may be the inaccuracy of the sampling from the posterior distribution. This issue was not critical because we aimed to give a quick prediction based on the MAP. Nevertheless, we hypothesize that for a reasonably high value assigned to the fourth sampling choice (w_3, corresponding to a conventional ABC MCMC), it is possible to obtain posterior distributions similar to ABC SMC results. Further research is required to verify this hypothesis. We used the scheme $w = \{0.2, 0.2, 0.3, 0.3\}$ for sampling choice. Other values used in ABC H sampler were $N = 10^3$, $h = 800$, and a Gaussian acceptance kernel function K_h.

Because actual data can be directly compared to simulated data for ODE models, it was not necessary to use summary statistics as criteria of divergence. We used the sum of squared errors as the distance criterion, which is closely related to the likelihood [33,35,36], and suitable for infectious-disease modeling [30]. We used a methodology for the prior-distribution choice that consisted of several trials mimicking different favorable and unfavorable scenarios, followed by feedback from epidemiologists.

Values of $d = 0.2$ and $\sigma = 1/3$ day^{-1} in model-I were reported in [10]. The remaining parameter values and the initial conditions $E(0)$ and $D(0)$ were estimated with ABC starting from locally uniform prior distributions.

In our approach, two key dates of travel restrictions divided the evolution of the epidemic into three periods. Three governmental action strengths were estimated for these periods: before travel restrictions (α_1), after interprovincial travel restriction (α_2), and after the urban public transport restriction in Santiago de Cuba (α_3). The three values β_1, β_2, and β_3 were considered in model-II taking into account the same criteria for key dates mentioned above.

We calculated the coefficient of determination R^2 for the regression analysis of MAP fit to assess the quality of parameter-estimation results for models I and II. We performed the frequentist and nonparametric Mann-Whitney U test [37,38], taking as null hypothesis that the actual residues obtained arose from the assumed gamma distribution. Furthermore, an approximation of maximum likelihood (ML) ratio between models (ML of model-I to ML of model-II) was computed with ABC output.

TABLE I. Summary table of model-I parameters and results.

Parameter	Notation	Prior or fixed value	MAP (ABC SMC)	MAP (ABC H)
Transmission rate	β_0	$U(0.1, 1.5)$	0.4811 days^{-1} $CI(0.3680, 0.5468)$	0.4340 days^{-1}
	α_1		0.5245 $CI(0.3717, 0.7534)$	0.6179
Fitting parameter	α_2	$U(0.1, 1)$	0.2855 $CI(0.1775, 0.3692)$	0.1801
	α_3		0.9665 $CI(0.8527, 1.0000)$	0.9417
Intensity of responds	k	$U(600, 4000)$	3430 $CI(3335, 3598)$	3405
Mean latent period	σ^{-1}	3 days	-	-
Mean infectious period	γ^{-1}	$U(1, 40)$	21 days $CI(1, 33)$	25 days
Proportion of severe cases	d	0.2	-	-
Mean duration of public reaction	λ^{-1}	$U(10, 30)$	20 days $CI(1, 39)$	50 days
Initial susceptible population	$S(0)$	$9.41 \cdot 10^5$	-	-
Initial exposed population	$E(0)$	$U(1, 10)$	6 $CI(4, 7)$	6
Initial infectious population	$I(0)$	1	-	-
Initial removed population	$R(0)$	0	-	-
Initial risk perception	$D(0)$	$U(1, 10^3)$	163 $CI(142, 188)$	148
Initial cumulative cases	$C(0)$	1	-	-

All simulations were made in a 256-core high-performance-computing (HPC) processor with 256 GB RAM using Python 3.6 [39]. The HPC machine was acquired by the Flemish Development Cooperation through VLIR-UOS (Flemish Interuniversity Council-University Cooperation for Development of Belgium) in the context of the Institutional University Cooperation program with Universidad de Oriente, Santiago de Cuba.

3. Results

Model-I simulation (Fig. 1) was performed with MAPs calculated with ABC SMC. All parameters and results of model-I fitting are summarized in Table I. The Mann-Whitney U test performed on the residues of model-I fit did not reject the null hypothesis based on a 5% significance level.

Figure 2 was obtained using the MAPs of posterior distributions of model-II parameters that were computed with ABC SMC. These MAPs and other values from model-II are summarized in Table II.

In all trials made, the ABC H sampler consumed less time and reached similar parameter regions to the ones obtained by ABC SMC (Figs. 3-6) for the same prior distributions.

Model-I predicted a total of 73 infected at the end of the COVID-19 epidemic in Santiago de Cuba, with 21% of undocumented cases. The final value of $M(t)$ was 57, being the model-I prediction of confirmed cases. Model-II forecasted 51 confirmed cases at the end of the epidemic. Additionally, the ML ratio calculated was 0.941; and the values of determination coefficients were $R^2 = 0.9450$ for model-I and $R^2 = 0.9781$ for model-II.
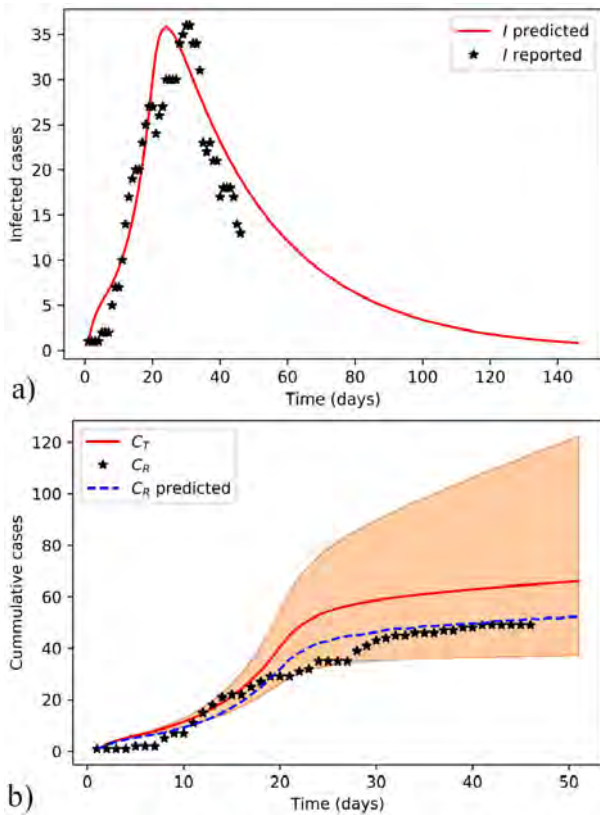
FIGURE 1. Estimated evolution of COVID-19 epidemic in Santiago de Cuba with model-I. The selected parameter values for simulation correspond to the MAP of the approximate posterior distribution. The model-predicted and reported infected individuals a), and the cumulative b), both predicted in simulation (C_T and C_R predicted) and observed (C_R) cases are plotted. The shaded area is limited by the extremes of 95% credible intervals of posterior distributions obtained with the ABC SMC sampler.
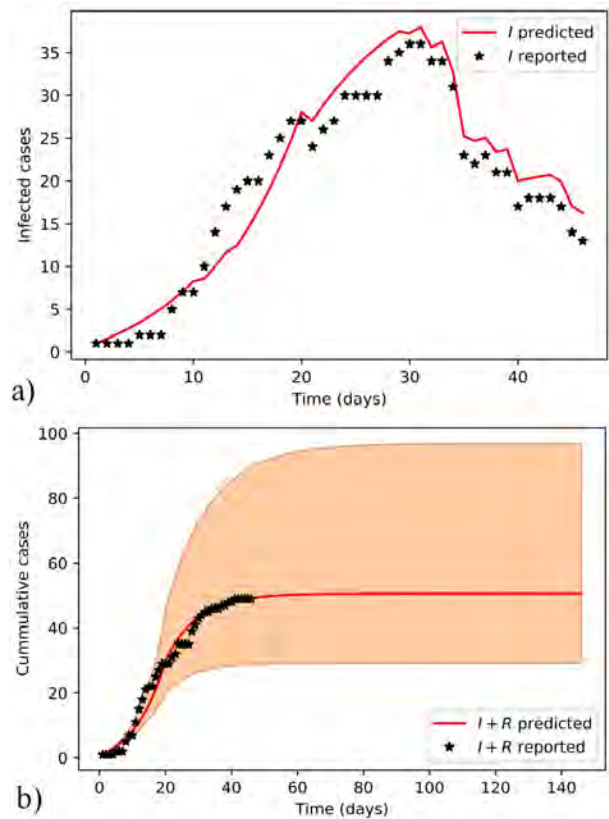

FIGURE 2. Estimated evolution of COVID-19 epidemic in Santiago de Cuba with model-II. Predicted infected individuals follow the dynamics of the reported ones a). A long-term scenario of both reported and predicted cumulative cases, and the area limited by the extremes of the 95% credible intervals of posterior distributions obtained with the ABC SMC sampler b) are plotted.

4. Discussion

COVID-19 dynamics in Santiago de Cuba is difficult to capture empirically without the introduction of stepwise functions for parameter α in model-I and parameter β in model-II. Results show that model-I (Fig. 1) and model-II (Fig. 2), fitted with ABC algorithms, can predict an early small-number peak of infected cases and the approximate ending of the COVID-19 in Santiago de Cuba. The non-rejection of the null hypothesis on the residuals allows us to continue considering the validity of the error law for model-I.

All MAPs calculated with ABC H are within credible intervals obtained with ABC SMC, except parameter λ of model-I. Even for this parameter, the histogram generated by ABC H partially overlapped the one generated by ABC SMC. We deduce that both algorithms find the same parameter region for the chosen prior. A visual comparison of Fig. 3 with Fig. 4, and Fig. 5 with Fig. 6, reveals that ABC H does not sample from the same distribution as ABC SMC. Although this does not affect the use of ABC H in this work, a modification of the ABC H sampler is required to make it a fully

functional ABC algorithm. The scheme $w = \{1, 0, 0, 0\}$ for sampling choice corresponds to a ABC RS process [33, 40]. The scheme $w = \{0, 0, 0, 1\}$ resembles an ABC MCMC method [41]. ABC RS and ABC MCMC are algorithms that sample from adequate posterior distributions but in an inefficient time, so other intermediate schemes may be evaluated in further works.

The computed percentage of undocumented cases is in contrast with the 86% of unreported infected cases estimated for China [43]. This result suggests that the limited progression of the COVID-19 epidemic in Santiago de Cuba may be due to opportune epidemiological investigations, effective control measures in each source of infection, and a low number of initially exposed individuals ($E(0) = 6$). Additionally, case clusters prevail over transmission clusters for COVID-19 in Santiago de Cuba province. The policymakers used these estimates to complement an investigation into the initial infectious load of SARS-COV-2 in Santiago de Cuba.

The MAPs of α posterior distributions were estimated at $\alpha_1 = 0.5245$, $\alpha_2 = 0.2855$ and $\alpha_3 = 0.9665$ (Table I). Value of α is expected to grow as government measures increase but, surprisingly, the value after the closure of transportation

TABLE II. Summary table of model-II parameters.

Parameter	Notation	Prior or fixed value	MAP (ABC SMC)	MAP (ABC H)
	β_1		0.5674 days^{-1}	0.5149 days^{-1}
			$CI(0.3067, 0.7534)$	
Transmission rate	β_2	$U(0.1, 0.9)$	0.5855 days^{-1}	0.6056 days^{-1}
			$CI(0.4794, 0.7088)$	
	β_3		0.3555 days^{-1}	0.3548 days^{-1}
			$CI(0.2443, 0.4525)$	
Mean infectious period	γ^{-1}	$U(1, 10)$	2 days	2 days
			$CI(1, 3)$	
Interval expectation of removed	λ	$U(1, 20)$	18 days	18 days
			$CI(12, 23)$	
Initial susceptible population	$S(0)$	9.41×10^5	-	-
Initial infectious population	$I(0)$	1	-	-
Initial removed population	$R(0)$	0	-	-

between provinces is lower than before ($\alpha_2 < \alpha_1$). We consider two possible explanations for this result:

First, the interpretation of α as government action is somewhat forced. In a previous study [42], α was interpreted as seasonality of transmission associated with the school calendar for a similar formulation of $\beta(t)$. Consequently, the interpretation of α remains open.

Second, the increase in PCR tests coincides with the period corresponding to α_2. Neither the deterministic model nor the error law explicitly takes into account the variability of PCR tests. This variability influences the number of detected cases, and therefore the estimation of all transmission-rate parameters.

Thus, in our results, we do not interpret α as government action. We consider α as a multifactorial parameter that works as a calibrator of $\beta(t)$ dynamics. The only dynamic variable the explicitly influences $\beta(t)$ in Eq. (2) is D [10,42]. This influence is in agreement with the epidemiological and sociological researches carried out in Santiago de Cuba, which confirm the reduction of the transmission rate by the high-risk perception of both decision-makers and population. Nevertheless, D depends explicitly on I, which in turn depends on E; so the variable D is affected by changes in these state variables; and therefore, they indirectly influence the dynamics of $\beta(t)$.

The estimated value $k = 3430$ ($k = 1117$ in [10]) may indicate a good response from individuals in Santiago de Cuba to COVID-19 epidemic.

The mean infection period for model-I may be explained by the inclusion of unreported cases and the delay in the re-

ports of those recovered. The value of this parameter for model-II is plausible because it is the number of days expected for a diagnosed patient to recover, and the diagnosis is usually confirmed several days after the exposition. The interval expectation of removed, $\lambda_{Poisson} = 18$ days with $CI\,[12, 23]$, seems consistent with the patient-discharge policy implemented by the Ministry of Public Health of the Republic of Cuba.

One possible explanation for the ML ratio is that the error law for model-I (Eq. (3)) can only partially describe the complex variability of the reporting rate. As model-II is more parsimonious than model-I, we did not perform some further analysis like Akaike's information criterion (AIC) or Bayes factor, which should favour model-II over model-I. On the other hand, the model-I allowed us to estimate the number of individuals initially exposed, and to compute the percentage of unreported cases. We argue that simpler models like model-II should be preferred for accurate prediction, whereas more explanatory models like the model-I are to be used for phenomenological analysis.

The chat on the left in Fig. 2 shows a remarkable tracking of the data variation by the fitted curve. A conventional SIR model cannot achieve this. Model-II accomplishes this tracking by dividing the adjustable parameters of the SIR model into three time periods. Furthermore, the cases recovered on the day estimated by the SIR model are not subtracted, but on the day expected by the Poisson error. In other words, model-II is advantageous over a conventional SIR as it is a piecewise SIR with a Poisson noise in the number of removed cases.
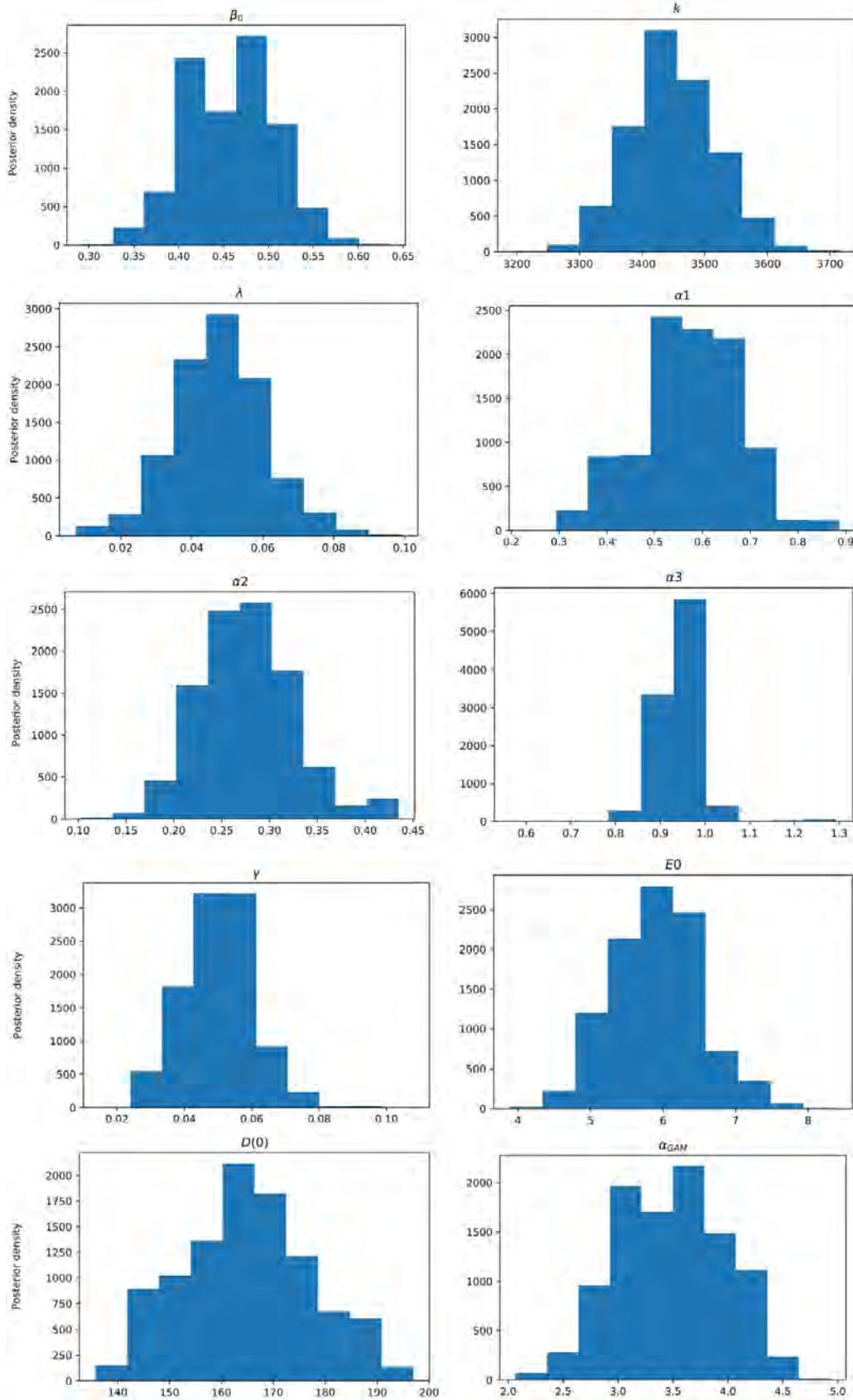
FIGURE 3. Posterior density estimation of model-I parameters using ABC SMC sampler and COVID-19 data from Santiago de Cuba. The estimated parameters are β_0, k, λ, α_1, α_2, α_3, γ, $E(0)$, and the scale parameter of the gamma distribution (α_{GAM}).

FIGURE 4. Posterior density estimation of model-I parameters using ABC H sampler and COVID-19 data from Santiago de Cuba. The estimated parameters are β_0, k, λ, α_1, α_2, α_3, γ, $E(0)$, and the scale parameter of the gamma distribution α_{GAM}.
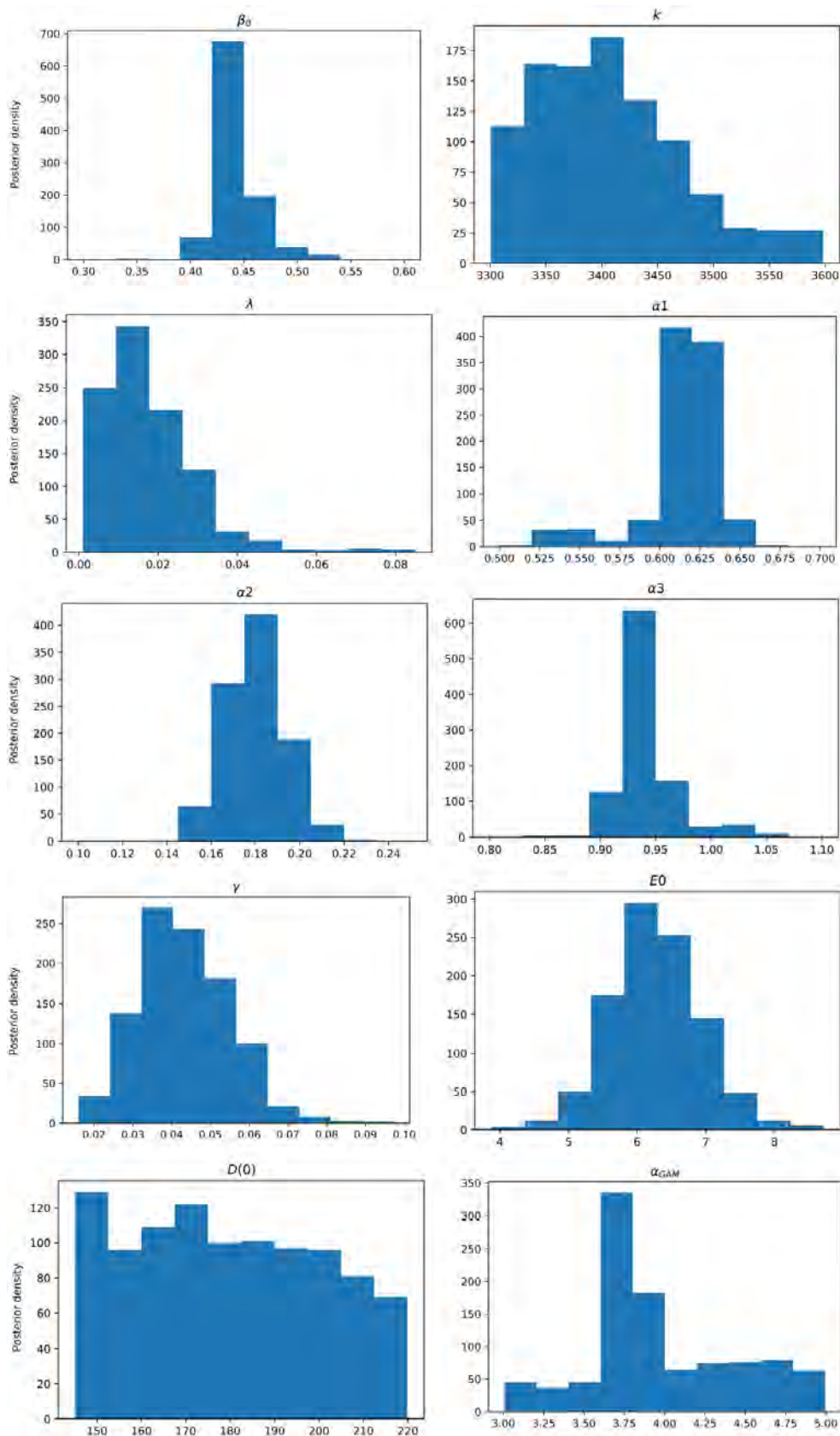
FIGURE 5. Posterior density estimation of model-II parameters using ABC SMC sampler and COVID-19 data from Santiago de Cuba. The estimated parameters are β_1, β_2, β_3, γ, and the interval expectation of the Poisson distribution (λ_{Poisson}).
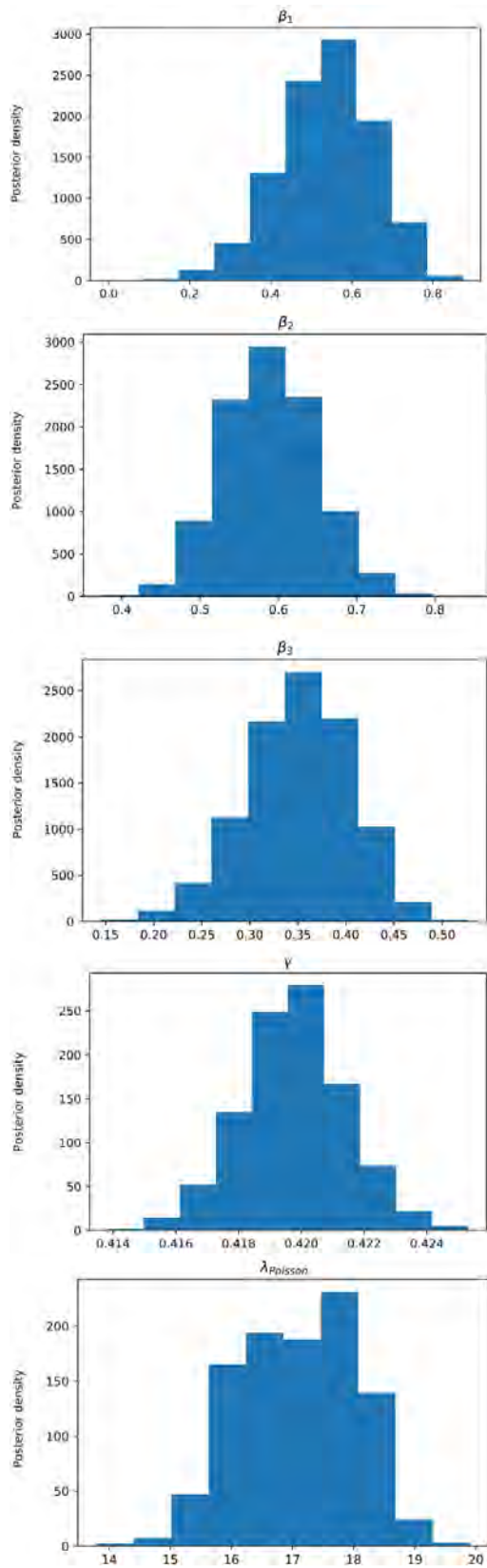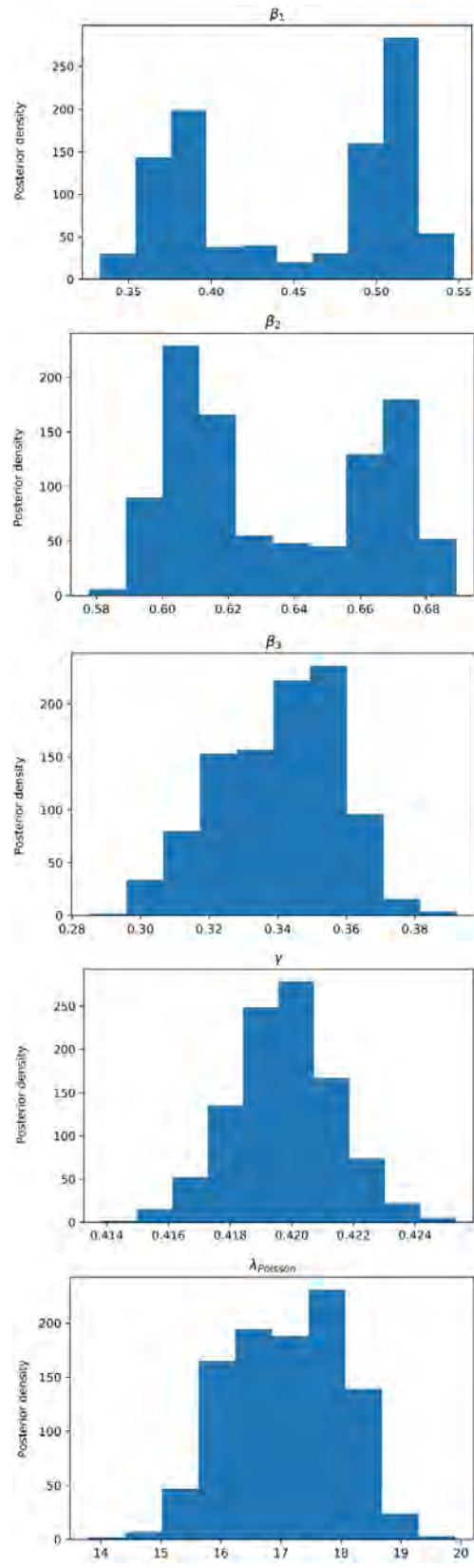
FIGURE 6. Posterior density estimation of model-II parameters using ABC H sampler and COVID-19 data from Santiago de Cuba. The estimated parameters are β_1, β_2, β_3, γ, and the interval expectation of the Poisson distribution (λ_{Poisson}).

We hold that the difference between the predicted and reported data in both models is acceptable because of reporting delays, which are in agreement with previous studies [10, 32, 44].

One advantage of using a Bayesian approach in parameter estimation is to use different prior distributions in order to set favourable and unfavourable scenarios for models. When the data do not show a peak yet, most models become highly unidentifiable, in such a way that one can find several parameter combinations leading to reasonable fits. Therefore, we deem it essential to complement the mathematical analysis with an evaluation of epidemiological investigation, looking for the most plausible prediction.

The error law for the model-I (Eq. (3)) and the assumption of the gamma distribution cannot fully capture the variability of the reporting rate. This approach is a baseline for future improvements.

The influences of β_0, α, D and k in $\beta(t)$ should not be discussed separately. The variables β_0, α, D, and k are dynamically interrelated, taking into account that the society (complex system) is constituted by the interaction and coupling of its elements as a global unit, in agreement with [45].

Further studies will evaluate the influence of other variables on the favourable evolution of COVID-19 in Santiago de Cuba: government-induced social isolation, the social response of individuals, and the environmental factors (*e.g.*, temperature and relative humidity). Additionally, future works should study the ABC H performance for intermediate values of w, looking for an acceptable compromise of efficiency and accuracy. Another possibility is to study adaptive schemes of w, in order to accept a group of candidates in a short time, and then ensure sampling of the posterior distribution.

5. Conclusions

Our study indicates that modified SIR and SEIR models, combined with ABC for parameter estimation, can follow the small-number dynamics of the COVID-19 epidemic. More precisely, models follow dynamics with no explosions of new cases and with small-number peaks by including some parameters as stepwise functions of critical days. An opportune epidemiological investigation, along with a low number of initially exposed individuals, may partly explain the favourable evolution of the COVID-19 epidemic in Santiago de Cuba.

Acknowledgements

Appendix

A. Algoritm ABC Sequential Monte Carlo Algorithm

Algorithm 1 ABC Sequential Monte Carlo Algorithm

Require:

- A target posterior density $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y|\theta)$.

- A kernel function $K_h(u)$, and an integer $N > 0$.

- An initial sampling density $g(\theta)$ and sequence of proposal densities
  $g_m(\theta, \theta'), m = 1, \ldots, M$.

- A value $\alpha \in [0, 1]$ to control the effective sample size.

- A low dimensional vector of summary statistics $s = S(y)$.

Ensure: A   set   of   weighted   parameter   vectors $(\theta_M^{(1)}, w_M^{(1)}, \ldots, \theta_M^{(N)}, w_M^{(N)})$ drawn from $\pi_{ABC}(\theta|s_{obs})$

Initialise:

  For $i = 1, \ldots, N$ **Do**

- Generate $\theta_0^{(i)} \sim g(\theta)$ from initial distribution g.

- Generate $y_0^{(i)}(t) \sim p(y|\theta_0^{(i)})$ and compute summary statistics
  $s_0^{(i)}(t) = S(y_0^{(i)})$ for $t = 1, \ldots, T$.

- Compute weights $w_0^{(i)} = \pi(\theta_0^{(i)})/g(\theta_0^{(i)})$, and set $m = 1$.

  End For

Sampling:

1. Reweight: Determine h_m such that

$$ESS(w_m^{(1)}, \ldots, w_m^{(N)}) = ESS(w_{m_1}^{(1)}, \ldots, w_{m_1}^{(N)}) \text{ where}$$

$$w_m^{(i)} = w_{m_1}^{(i)} \frac{\sum_{t=1}^{T} K_{hm}(||s_m - s_{obs})||)p(s|\theta)\pi(\theta_m)}{\sum_{t=1}^{T} K_{hm-1}(||s_{m-1} - s_{obs})||)p(s|\theta)\pi(\theta_{m-1})}$$

2. Resample: if $ESS(w_{m_1}^{(1)}, ..., w_{m_1}^{(N)}) < E$ then resample N particles from the empirical distribution function $\{\theta_m^{(i)}, s_m^{(i)}(1), ..., s_m^{(i)}(T), W_m^{(i)}\}$ where

$W_m^{(i)} = w_m^{(i)} / \sum_{j=1}^{N} w_m^{(j)}$ and set $w_m^{(i)} = 1/N$.

3. Move:

For $i = 1, \ldots, N$ **Do**

  if $w_m^{(i)} > 0$ **then**

  • Generate $\theta' \propto g_m(\theta_m^{(i)}, \theta)$, $y'(t) \propto p(y|\theta_m^{(i)})$ and $s'(t) = S(y'(t))$
  for $t = 1, \ldots, T$.

  • Accept θ' with probability

$$\min \left(1, \frac{\displaystyle\sum_{t=1}^{T} K_{hm}(||s'(t) - s_{\text{obs}}||)\pi(\theta')g(\theta'|\theta_m^{(i)})}{\displaystyle\sum_{t=1}^{T} K_{hm}(||s_m^{(i)}(t) - s_{\text{obs}}||)\pi(\theta_m^{(i)})g(\theta_m^{(i)}|\theta')} \right)$$

  end if

end for

B. Algorithm ABC Hybrid Algorithm

Algorithm 2 ABC Hybrid Algorithm

Require:

• A target posterior density $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y|\theta)$.

• A kernel function $K_h(u)$, an integer $N > 0$ and a tolerance h.

• Sampling densities $g_m(\theta)$, $g_c(\theta)$ and $g_{MC}(\theta)$.

• A normalized vector $w = \{w_0, w_1, w_2, w_3\}$ to control the sampling choice and $i = 0$ to count the accepted candidates.

• A low dimensional vector of summary statistics $s = S(y)$.

Ensure A set of parameter vectors $(\theta^{(1)}, ..., \theta^{(N)})$ drawn from

$$\pi_{ABC}(\theta|s_{\text{obs}}) \propto \int K_{h M}(||s - s_{\text{obs}}||)p(s|\theta)\pi(\theta)ds$$

Initialise:

• Sample three initial points $\{\theta_m, \theta_c, \theta_{MC}\}_{i=0}$ from $\pi(\theta)$.

• Generate $y_{\{m,c,MC\}} \sim p(y|\theta_{\{m,c,MC\}})$ and compute summary statistics $s_{\{m,c,MC\}} = S(y_{\{m,c,MC\}})$.

Sampling:

while $i < N$ **do**

1. Choose sampling mode: sample w' from w according to weights.

2. Sample candidate:

  if $w' = w_0$ **then** Draw θ' from $\pi(\theta)$

  else if $w' = w_1$ **then** Propose θ' according to $g_m(\theta|\theta_m)$

  else if $w' = w_2$ **then** Propose θ' according to $g_c(\theta|\theta_c)$

  else if $w' = w_3$ **then** Propose θ' according to $g_{MC}(\theta|\theta_{MC})$

  end if

3. Accept or not:

  • Generate $y' \propto p(y|\theta')$ and $s' = S(y')$.

  • Accept θ' with probability $K_h(||s' - s_{\text{obs}}||)$

  • If θ' is accepted, do $\theta_{MC} = \theta'$ with probability

  $$\min \left(1, \frac{\pi(\theta')g_{MC}(\theta_{MC}|\theta')}{\pi(\theta_{MC})g_{MC}(\theta'|\theta_{MC})} \right)$$ and increment i.

4. Update:

  • Do $\theta_m = \theta'$ if $K_h(||s' - s_{\text{obs}}||)$ is the highest computed so far.

  • Propose θ_c' according to $g_c(\theta|\theta', \theta_c)$, generate $y_c' \sim p(y|\theta_c')$, compute summary statistics $s_c' = S(y_c')$ and do $\theta_c = \theta_c'$ if $K_h(||s_c' - s_{obs}||)$ is the highest computed so far.

end while

---

1. S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. W. Hengartner, and R. Ke, The novel coronavirus, 2019-nCoV, is highly contagious and more infectious than initially estimated, *arXiv preprint arXiv*:2002.03268 (2020), Available from: `https://arxiv.org/ftp/arxiv/papers/2002/2002.03268.pdf`.

2. L. Ferretti *et al.*, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science* **368** (2020) eabb6936, https://dx.doi.org/10.1126/science.abb6936.

3. R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, How will country-based mitigation measures influence the course of the COVID-19 epidemic?, *The Lancet* **395** (2020) 931, https://doi.org/10.1016/S0140-6736(20)30567-5.

4. C. S. M. Currie *et al.*, How simulation modelling can help reduce the impact of COVID-19, *J. Simulat.* **14** (2020) 83, https://doi.org/10.1080/17477778.2020.1751570.

5. R. F. Reis *et al.*, Characterization of the COVID-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in South Korea, Italy, and Brazil, *Chaos Solitons Fract.* **136** (2020) 109888, https://doi.org/10.1016/j.chaos.2020.109888.

6. S. Zhao and H. Chen, Modeling the epidemic dynamics and control of COVID-19 outbreak in China, *Quant. Biol.* **8** (2020) 11 https://doi.org/10.1007/s40484-020-0199-0.

7. Z. Liu, P. Magal, O. Seydi, and G. Webb, Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *Biology* **9** (2020) 50, https://doi.org/10.3390/biology9030050.

8. A. L. Lloyd, Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods, *P. Roy. Soc. B-Biol. Sci.* **268** (2001) 985. https://doi.org/10.1098/rspb.2001.1599.

9. C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, *PloS One* **15** (2020) e0230405. https://doi.org/10.1371/journal.pone.0230405.

10. Q. Lin *et al.*, A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action, *Int. J. Infect. Dis.* **93** (2020) 211. https://doi.org/10.1016/j.ijid.2020.02.058.

11. J. Wangping *et al.*, Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China, *Front. Med.* **7** (2020) 169. https://doi.org/10.3389/fmed.2020.00169.

12. G. Giordano *et al.*, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, *Nat. Med.* **26** (2020) 855. https://doi.org/10.1038/s41591-020-0883-7.

13. A. Gelman, Bayes, Jeffreys, prior distributions and the philosophy of statistics, *Stat. Sci.* **24** (2009) 176. https://doi.org/10.1214/09-STS284D.

14. J. K. Kruschke, *Doing Bayesian Data Analysis* (Academic Press, Cambridge, 2015).

15. V. Amrhein, T. Roth, and F. Korner-Nievergelt, Identical twins and Bayes' theorem in the 21st century, *F1000Research* **2** (2015) 278. https://doi.org/10.12688/f1000research.2-278.v2

16. E. B. Postnikov, Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions?, *Chaos Solitons Fract.* **135** (2020) 109841. https://doi.org/10.1016/j.chaos.2020.109841.

17. D. Fanelli and F. Piazza, Analysis and forecast of COVID-19 spreading in China, Italy and France, *Chaos Solitons Fract.* **134** (2020) 109761. https://doi.org/10.1016/j.chaos.2020.109761.

18. M. D'Arienzo and A. Coniglio, Assessment of the SARS-CoV-2 basic reproduction number, R_0, based on the early phase of COVID-19 outbreak in Italy, *Biosaf. Health* **2** (2020) 57, https://doi.org/10.1016/j.bsheal.2020.03.004.

19. C. Reno *et al.*, Forecasting COVID-19-Associated Hospitalizations under Different Levels of Social Distancing in Lombardy and Emilia-Romagna, Northern Italy: Results from an Extended SEIR Compartmental Model, *J. Clin. Med.* **9** (2020) 1492. https://doi.org/10.3390/jcm9051492.

20. B. Ivorra, M. R. Ferrández, M. Vela-Pérez, and A. Ramos, Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. *The case of China, Commun. Nonlinear Sci.* **88** (2020) 105303 https://doi.org/10.1016/j.cnsns.2020.105303.

21. G. C. Calafiore, C. Novara, and C. Possieri, A modified SIR model for the COVID-19 Contagion in Italy, https://arxiv.org/pdf/2003.14391.pdf.

22. A. Lachmann, K. M. Jagodnik, F. M. Giorgi, and F. Ray, *Correcting under-reported COVID-19 case numbers: estimating the true scale of the pandemic*, https://doi.org/10.1101/2020.03.14.20036178.

23. E. L. Piccolomini and F. Zama, *Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model,* https://arxiv.org/abs/2003.09909.

24. G. Guzzetta *et al.*, Potential short-term outcome of an uncontrolled COVID-19 epidemic in Lombardy, Italy, February to March 2020, *Eurosurveillance* **25** (2020) 2000293. https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000293.

25. A. J. Kucharski *et al.*, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet. Infect. Dis.* **20** (2020) 553. https://doi.org/10.1016/S1473-3099(20)30144-4.

26. W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic?, *Infect. Dis. Model.* **5** (2020) 271. https://doi.org/10.1016/j.idm.2020.03.001.

27. D. Fernández Slezak *et al.*, When the optimal is not the best: parameter estimation in complex biological models, *PLOS ONE* **5** (2020) e13283. https://doi.org/10.1371/journal.pone.0013283.

28. M. A. Beaumont, Approximate Bayesian computation in evolution and ecology, *Annu. Rev. Ecol. Evol. Syst.* **41** (2010) 379. https://doi.org/10.1146/annurev-ecolsys-102209-144621.

29. J. Lintusaari *et al.*, Fundamentals and recent developments in approximate Bayesian computation, *Syst. Biol.* **66** (2017) e66. https://doi.org/10.1093/sysbio/syw077.

30. A. Minter and R. Retkute, Approximate Bayesian Computation for infectious disease modelling, *Epidemics* **29** (2019) 100368. https://doi.org/10.1016/j.epidem.2019.100368.

31. H. H. Weiss, The SIR model and the foundations of public health, *Mat. Mat.* **2013** (2013) 17.

32. S. Zhao *et al.*, Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak, *J. Clin. Med.* **9** (2020) 388. https://doi.org/10.3390/jcm9020388.

33. T. Toni *et al.*, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *J. R. Soc. Interface* **6** (2009) 187. https://doi.org/10.1098/rsif.2008.0172.

34. S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of Approximate Bayesian Computation* (Chapman and Hall/CRC, 2018). https://doi.org/10.1201/9781315117195.

35. H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting* (Oxford University Press, 2004).

36. A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005), https://doi.org/10.1137/1.9780898717921.

37. P. E. McKnight and J. Najab, Mann-Whitney U Test, in *The Corsini Encyclopedia of Psychology*, edited by I. B. Weiner and W. E. Craighead (Wiley, New York, 2009), https://doi.org/10.1002/9780470479216.corpsy0524.

38. G. W. Corder and D. I. Foreman, Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach (Wiley, New York, 2009), https://doi.org/10.1002/9781118165881.

39. G. V. Rossum, *The Python Library Reference* (12th Media Services, Suwanee, 2018).

40. J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, Population growth of human Y chromosomes: a study of Y chromosome microsatellites, *Mol. Biol. Evol.* **16** (1999) 1791. https://doi.org/10.1093/oxfordjournals.molbev.a026091.

41. P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, Markov chain Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci. USA* **100** (2003) 15324. https://doi.org/10.1073/pnas.0306899100.

42. D. He *et al.*, Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales, *P. Roy. Soc. B-Biol. Sci.* **280** (2013) 20131345. https://doi.org/10.1098/rspb.2013.1345.

43. R. Li *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* **368** (2020) 489. https://doi.org/10.1126/science.abb3221.

44. A. R. Tuite and D. N. Fisman, Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic, *Ann. Intern. Med.* **172** (2020) 567. https://doi.org/10.7326/M20-0358.

45. S. Galea, M. Riddle, and G. A. Kaplan, Causal thinking and complex system approaches in epidemiology, *Int. J. Epidemiol.* **39** (2010) 97. https://doi.org/10.1093/ije/dyp296.