Bull Math Biol (2021) 83:89
https://doi.org/10.1007/s11538-021-00922-3

Society for Mathematical Biology

ORIGINAL ARTICLE

Alternative Strategies for the Estimation of a Disease's Basic Reproduction Number: A Model-Agnostic Study

Gustavo Nicolás Páez[1] · Juan Felipe Cerón[2] · Santiago Cortés[2] · Adolfo J. Quiroz[3] · José Fernando Zea[4] · Camila Franco[3] · Érica Cruz[4] · Gina Vargas[4] · Carlos Castañeda[4]

Received: 17 November 2020 / Accepted: 8 June 2021 / Published online: 3 July 2021

Abstract

This work presents a model-agnostic evaluation of four different models that estimate a disease's basic reproduction number. The evaluation presented is twofold: first, the theory behind each of the models is reviewed and compared; then, each model is tested with eight impartial simulations. All scenarios were constructed in an experimental framework that allows each model to fulfill its assumptions and hence, obtain unbiased results for each case. Among these models is the one proposed by Thompson et al. (Epidemics 29:100356, 2019), i.e., a Bayesian estimation method well established in epidemiological practice. The other three models include a novel state-space method and two simulation-based approaches based on a Poisson infection process. The advantages and flaws of each model are discussed from both theoretical and practical standpoints. Finally, we present the evolution of Covid-19 outbreak in Colombia as a case study for computing the basic reproduction number with each one of the reviewed methods.

Keywords Basic reproduction number · Bayesian statistics · Poisson process · Kalman filter · Simulations

✉ Gustavo Nicolás Páez
gn.paez145@uniandes.edu.co

✉ Carlos Castañeda
ccastanedao@ins.gov.co

[1] Myanmar Development Institute, Naypyitaw, Myanmar

[2] Factored.Ai, Bogotá, Colombia

[3] Universidad de los Andes, Bogotá, Colombia

[4] Instituto Nacional de Salud, Bogotá, Colombia

1 Introduction

During the COVID-19 pandemic, the effective reproduction number R has become one of the main tools for public administrations to understand the local evolution of the epidemic and later to plan accordingly. The importance of this measurement lies in its ability to estimate the number of secondary cases arising from a single infectious individual during their entire infectious period. Given its potential, several governments (Zhao et al. 2020; Dickens et al. 2020; "The R number" 2020), including the Government of Colombia, have used it to mandate lockdowns and mobility restrictions to keep infection rates in check, thus avoiding the saturation of health facilities.

Despite its popularity, there is no consensus on the best algorithm for the computation of this measurement. Even the most popular models are nuanced by particular diseases' biological factors and the social contexts of the populations they affect (Adam 2020). For this reason, the purpose of this study is to develop four statistical models for the estimation of the effective reproduction number in a way that captures both the biological and social dimensions of the epidemic.

We begin by presenting the model proposed by Cori et al. (2013) and extended by Thompson et al. (2019), which has been implemented in several software packages and has remained popular during the COVID-19 pandemic. Due to its statistical techniques, we named this model as the Bayesian model. We analyze the model's assumptions, advantages, and limitations. In particular, we identify the modelling features that weaken the predictive power of this model under scenarios of social changes. For the COVID-19, for example, the most common tool implemented by the governments has been the implementation of curfews and lockdowns that significantly change the mobility patterns of people. Thus, under these circumstances, conclusions derived from the previous model become less reliable.

By acknowledging the merits and weaknesses of the Bayesian model, we then propose three alternative models that, from a theoretical standpoint, address these challenges. The first model, called state-space model, comes from a time-series approach, and introduce elements of state space models into the logic of the disease dynamic to better predict its reproductive rate. The other two models, named General Poisson and Exponential Poisson, take a discrete event simulation approach *that pay more attention to* the data generation processes and via Monte Carlo simulations estimate the evolution of the disease. While tackling the issues of the Bayesian model, there is no preference hierarchy among these four options. For example, the last two models are excellent at capturing the impact of policies but require strong computational capacity. In contrast, the state-space model is easy to implement and captures policies well, but it is more volatile than the previous models, so noise from the data can confuse it.

In order to better understand the scope and limitations of these models, we present eight simulated epidemic scenarios and the corresponding R estimations from each of the models. This allows us to empirically compare the advantages and disadvantages of the application of each model. Moreover, with the insights derived from the simulation exercise, we use the developed models in a case study

to analyze the reproduction number of the COVID-19 pandemic in Colombia during the first months of the outbreak. The results of these exercises are summarized in Appendix as a Table 1 that is intended to guide researchers and practitioners in the cases where each model should be preferred, and what are the caveats that they need to be aware of. In that way the paper brings to the academic literature novel modelling strategies to capture the dynamics of diseases under context with social changes and at the same time can be used as a quick reference guide for practitioners that want to use these techniques to develop policies to control the spread of the disease.

2 Estimating the Reproduction Number

This section describes the theory and calibration of four alternative models for the estimation of the reproduction number. The first subsection is dedicated to conceptualizing the reproduction number and the scope of its definitions. The second subsection presents two of the models, both of which are based on the concept of the "time-since-infection," and estimates reproduction numbers using time series methods. The final subsection presents two novel models based on stochastic simulation processes that describe the infection dynamics of the disease. The models are calibrated to fit the observed cases, and the reproduction number series are extracted from the models.

2.1 Different Reproduction Numbers

One of the key elements for understanding the evolution of an epidemic is the basic reproduction number. This metric estimates the number of secondary cases arising from a single infectious individual during their entire infectious period. Its value can be divided into three components (van den Driessche 2017):

1. The duration of the infectious period.
2. The probability of the infection being passed on from an infected individual to a susceptible secondary individual.
3. The mean number of times that an infectious individual comes across a susceptible individual in a single unit of time.

Such a decomposition makes it evident that this measurement is influenced by the disease's biological factors (which determine how long an individual is infectious and define how easy it is for the disease to be transmitted between individuals) as well as by social factors (which determine the types and frequencies of interactions between individuals). Because of the latter type of factors, the basic reproduction number of almost any disease depends on the social characteristics of the population, making it hard to extrapolate results across populations. Moreover, these social factors are also time-dependent, as they are affected by changing factors such as mandatory quarantines, social distancing guidelines and changes in the susceptible population.

Authors such as Fraser (2007) and Cori et al. (2013) proposed differentiating between two types of reproduction numbers. The *effective reproduction number*, $R(t)$, is defined as the mean number of individuals who an individual whose infection started on day t will infect given that all three of the aforementioned factors remain constant. On the other hand, the *case reproduction number*, R_C, accounts for changes in these factors during the individual's infectious period. Hence, the first number has a prospective and counterfactual nature, whereas the second represents a retrospective approach that allows for the visualization of the influence of social changes (usually driven by public policy) on the evolution of the epidemic. Last, both of these indicators indicate whether the epidemic is expanding (if they are greater than 1) or contracting (if they are smaller than 1).

Under these definitions, the following sections present the models used to estimate each of the indicators. Please note that, in what follows, the quantities represented by a single symbol may technically vary between models, however its conceptual purpose is the same. This was designed to highlight the correspondent parts among the models without the need of additional symbols that will complicate the reading flow.

2.2 Time-Since-Infection Models

This section presents two models based on the original estimations used by Kermack-McKendrick (1927). We choose to follow this modeling strategy because of its intuitive calculations (Fraser 2007) and its popularity in evaluating the most recent outbreaks (Thompson et al. 2019). The reference model for this section is the one developed by Cori et al. (2013), which epitomizes the current state of this family of models. Given the techniques that are used in this work, we refer to this model as the "*Bayesian model*" from this point on. We begin by describing its main features, discussing its limitations, and finally introducing a novel model that overcomes these limitations.

2.2.1 Bayesian Model

Let $I_N(t)$ be the number of individuals who become infectious during period t. This stochastic value depends on the number of infectious individuals found during previous time steps adjusted by their infectious potential during those time steps. The renewal equation for this process is defined as

$$E\left[I_N(t)\right] = \sum_{\tau=1}^{\infty} \beta(t,\tau) I_N(t-\tau). \tag{1}$$

Equation (1) is understood as the conditional expectation of $I_N(t)$ given the incidences noted in previous periods. We omit the usual conditional notation from the expected value to simplify the notation, and carry this convention throughout the rest of the work.

To illustrate Eq. (1), an individual who became infectious during period $t-10$ is expected to infect $\beta(t,\ 10)$ susceptible individuals during period t. Furthermore, the value of $\beta(t,\tau)$ is assumed to be multiplicatively decomposable into two factors:

$\phi(t)$, which reflects the social characteristics of the population, and $\omega(\tau)$, which is exclusively associated with the biological dimension of the disease. We therefore write

$$\beta(t,\tau) = \phi(t)\omega(\tau). \tag{2}$$

Based on the logic suggested by Wallinga and Teunis (2004), $\omega(\tau)$ can be interpreted as the distribution of the serial interval of the disease. In other words, $\omega(\tau)$ represents the probability that an individual begins to be infectious τ units of time after the person who infected them became infectious themselves. Under this interpretation, Wallinga and Teunis (2004) deduced that the mean number of individuals who each infected individual will infect is given by

$$R(t) = \sum_{\tau=1}^{\infty} \beta(t,\tau) = \phi(t). \tag{3}$$

Combining both equations, we obtain

$$R(t) = \frac{E\left[I_N(t)\right]}{\sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau)} \tag{4}$$

and

$$E\left[I_N(t)\right] = R(t) \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau). \tag{5}$$

Thus, following Fraser's (2007) logic:

$$R_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau) R(t-\tau). \tag{6}$$

Assuming we possess a priori knowledge about the serial interval, the calibration of this model is reduced to the estimation of $E\left[I_N(t)\right]$. To that end, we assume that the number of infectious cases during period t follows a Poisson distribution with a mean dictated by Eq. (5). Furthermore, we also assume that $R(t)$ remains constant during a time window of length w following period t. Under these assumptions,

$$\sum_{t'=t-w}^{t} I_N\left(t'\right) \sim \text{Poisson}\left(R(t) \sum_{t'=t-w}^{t} \sum_{\tau=1}^{\infty} \omega(\tau) I_N\left(t'-\tau\right) \right). \tag{7}$$

Therefore, given a prior distribution of $R \sim \text{Gamma}\left(\gamma_1, \gamma_2\right)$, where γ_1 and γ_2 are the rate and shape parameters, respectively, the posterior distribution of R is

$$R(t) \sim \text{Gamma}\left(\gamma_1 + \sum_{t'=t-w}^{t} I_N\left(t'\right), \gamma_2 + \sum_{t'=t-w}^{t} \sum_{\tau=1}^{\infty} \omega(\tau) I_N\left(t'-\tau\right) \right). \tag{8}$$

According to Fraser (2007), the main advantage of this model is the ease of its computations. However, conceptual problems arise from the estimation of the distribution of the serial interval. As the author suggested, serial intervals are influenced by social directives. Hence, the assumption of a fixed serial interval distribution can affect the ability of our estimated $R(t)$ to capture such changes. In addition to this conceptual problem, the selection of the window w poses a challenge. This meta-parameter has a smoothing effect over subsequent periods. For this reason, this estimation of $R(t)$ reflects the infection rate over a small number of subsequent days instead of a single day.

2.2.2 State-Space Model

Looking to overcome both of the limitations mentioned in the previous section, the present model redefines Eqs. (1–6) to produce a definition of $\omega(\tau)$ that captures only the biological dimension of the disease. Moreover, the model employs a time series method known as a state-space model instead of a Bayesian estimation framework, thus avoiding issues related to the estimation window.

Let f_{Inc} be a probability function, where $f_{\text{Inc}}(t)$ is the probability that it takes an individual t days since being exposed to the disease to reach the beginning of his infectious phase. Similarly, let F_{Inf} be a cumulative distribution function, where $F_{\text{Inf}}(t)$ is the probability that a person remains infectious for at most t time steps. We thus define

$$\omega(t) = \sum_{\tau=1}^{t} f_{\text{Inc}}(\tau)\big(1 - F_{\text{Inf}}(t-\tau)\big). \tag{9}$$

Intuitively, Eq. (9) defines $\omega(t)$ as the probability that an individual who was infected t periods ago remains infectious today. Moreover, $\Omega = \sum_{\tau=1}^{\infty} \omega(\tau)$ is the expected amount of time an individual remains infectious; this is a direct consequence of applying a convolution to the survival function in Eq. (9). Equation (1) can thus be modified to address disease incubation periods:

$$E\left[\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I_N(t+\tau)\right] = \sum_{\tau=1}^{\infty} \beta(t,\tau) I_N(t-\tau). \tag{1'}$$

Notice how this equation draws a relationship between future and past incidences. Combining these two equations, we see that

$$R(t) = \frac{E\left[\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I_N(t+\tau)\right]}{\sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau)}\, \Omega \tag{2'}$$

and

$$E\left[\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I_N(t+\tau)\right] = \frac{R(t)}{\Omega} \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau). \tag{3'}$$

The quantity $R(t)$ is thus calculated as the product of the transmission rate on a given day, $\frac{E[\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I(t+\tau)]}{\sum_{\tau=1}^{\infty} \omega(\tau) I(t-\tau)}$, and the mean number of days an individual will remain infectious, Ω.

As in the previous model, we assume that the incidences at time t follow a Poisson distribution with a mean as given by Eq. (4′):

$$\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I_N(t+\tau) \sim \text{Poisson}\left(\frac{R(t)}{\Omega} \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau) \right). \tag{4′}$$

However, instead of assuming a Bayesian context for the data, we assume a non-Gaussian state-space model. State-space analysis is well suited to time series problems in which we're interested in the hidden properties of a system given a series of associated observations (Durbin and Koopman 2012). In our case, these are the $R(t)$ series and the series of incidences, respectively.

State-space models are usually described by a *state equation*, which describes the dynamics of the unobserved system state, and an *observation equation*, which describes how that state relates to the observations. Equations 10 and 11, respectively, instantiate those equations for the problem at hand.

$$\alpha_{t+1} = T\alpha_t + R\eta_t, \tag{10}$$

$$\sum_{\tau=1}^{\infty} f_{\text{Inc}}(\tau) I_N(t+\tau) = \mathrm{e}^{\alpha_{t,1}} \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t-\tau), \tag{11}$$

where

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

$$R = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix},$$

$$\alpha_t = \begin{bmatrix} \alpha_{t,1} & \alpha_{t,2} \end{bmatrix}'$$

and

$$\eta_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Under initial conditions[1]

[1] Large variances are given to the other prior distributions to highlight our lack of knowledge about them. This allows the model to quickly adjust to the data.

$$\alpha_1 = \left[\ln\left(\frac{R_0}{\Omega}\right) \; 0 \right]', \; P_1 = \begin{bmatrix} 10^9 & 0 \\ 0 & 10^9 \end{bmatrix}.$$

Kalman filtering and smoothing are applied to approximate minimum variance estimators for the means and covariances of each state vector α_t (Durbin and Koopman 2012). From the previous equations, the values of σ_1 and σ_2 are yet to be defined. However, these can be estimated as maximum likelihood estimators on the incidences. Finally, the parameter $\alpha_{1,1}$ is adjusted to match as prior the R_0 found in the epidemiological literature; however, it can be estimated also via maximum likelihood over that neighborhood to give it the freedom to adjust to the specific context being analyzed.

From the previous definitions, the system state is thus described by a two-dimensional vector in which the first component is related to $R(t)$, and the second describes a stochastic slope for that series (Harvey 1990). More explicitly, it is possible to derive from Eq. (11) that

$$R(t) = \mathrm{e}^{\alpha_{1,t}}\Omega. \tag{12}$$

Following this line of thought, $R_c(t)$ can also be computed, analogously to Eq. (6), by

$$R_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau)\frac{R(t-\tau)}{\Omega}. \tag{6'}$$

The reader should note that this estimation procedure overcomes the two theoretical problems of the Bayesian model in the previous section. However, it poses the new challenge of obtaining the a priori estimations of the incubation and infection time distributions. For that reason, even though this model represents a theoretical improvement over the Bayesian model, its application is contingent on the availability of accurate estimations of the aforementioned distributions.

2.3 Simulation-Based Models

The two models described in the previous sections are based on fitting a Poisson process that represents the transmission dynamics. Instead of getting into the details of the process's dynamics, these models reduce the underlying process to a link equation that involves the observed sequence data. This subsection presents a different perspective: one centered around modeling the process. This shifts the problem from a statistical challenge to a probabilistic challenge. These new models also rely on simulations, which introduce a computation-time problem. The first model presented in this section describes fairly general dynamics, but its calibration is computationally demanding**.** The second model limits the first one, allowing for analytic solutions to otherwise recursive computations, thus easing some of the computational burden of the model. We name them the *general Poisson model* and *exponential Poisson model* to reflect the additional restrictions added to the second model.

2.3.1 General Poisson Model

We begin by summarizing the notation for the rest of this section:

$I(t)$:= the number of infectious individuals during period t.

$I_N(t)$:= the number of newly infectious individuals during period t (i.e., they were not infectious during period $t-1$).

$E(t)$:= the number of individuals in the disease incubation stage during period t.

$E_N(t)$:= the number of newly incubating (or exposed) individuals during period t.

Let F_{inc} and F_{inf} be the distributions of infected individuals' incubation and infectious period lengths, respectively. Both distributions are assumed to be independent of time and the characteristics of individuals.

$\beta(t)$:= the expected number of individuals who a single person will infect during period t.

We chose to denote the last variable by $\beta(t)$ because of its similarity to $\beta(t,\ \tau)$ from previous sections. However, in this case, there is no lag in the computations because this value is linked with the infectious potential of an individual during period t and not to the potential of previously infected individuals that remain infectious at time t.

2.3.2 Process Description

The process begins with the following configuration:

- $E_N(0) > 0$.
- $\forall \tau \geq 0,\ I(\tau) = I_N(\tau) = 0$.
- $\forall \tau \geq 1\ E(\tau) = 0$.
- $t = 0$.

Step 1 $\forall i \in \{1 \ldots,\ E_N(t)\}$, generate two samples $x_{inc}(i)$ and $x_{inf}(i)$ from F_{inc} and F_{inf}, respectively. Next, update the process variables according to:

$$E(h) \to E(h) + 1 \text{ for all } h \in \{t+1, \ldots,\ t + x_{\text{inc}}(i)\}$$
$$I_N\big(t + x_{\text{inc}}(i) + 1\big) \to I_N\big(t + x_{\text{inc}}(i) + 1\big) + 1$$
$$I(h) \to I(h) + 1 \text{ for all } h \in \{t + x_{\text{inc}}(i) + 1, \ldots,\ t + x_{\text{inc}}(i) + x_{\text{inf}}(i)\}$$

Step 2 Update $E_N(t+1)$ as a sample of a Poisson variable with mean $\beta(t)I(t)$.

Step 3 Update $t \to t+1$ and return to Step 1.

Finally, let Ω be the expected value of F_{inf}. The reproduction numbers for this process can be computed by

$$R(t) = \beta(t)\Omega \tag{4''}$$

and

$$R_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau) R(t-\tau). \tag{6''}$$

Model Calibration

Given the distributions F_{inc} and F_{inf} and the initial number of exposed cases $E_N(0)$, fitting the model is equivalent to extracting the series $\beta(t)$ from the incidence data. We carry out this process in two steps: the first establishes a goodness-of-fit criterion, and the second reduces overfitting of the incidence series.

For all $t \in 1, \ldots, T$, let $\widehat{I_N(t)}$ be the observed number of individuals who start being infectious during period t. On the other hand, let $f(x : \beta)$ be the fraction of the process simulations in which $I_N(t) = x$ given the parameters B. Then, the log-likelihood[2] of the observed incidence series is given by:

$$ll(\beta) = \sum_{t=1}^{T} \ln\left(f\left(\widehat{I_N(t)} : \beta\right)\right). \tag{13}$$

The maximum likelihood estimator B can be obtained from Eq. (11). Nevertheless, this method produces significant overfitting of the incidence series. We avert this defect by introducing a regularization factor λ on the series $\beta(t)$ in the following cost function:

$$C(\lambda;\beta) = -ll(\beta) + \frac{\lambda}{T-2} \sum_{t=3}^{T} \left(\frac{\beta(t-1)}{\beta(t)} - \frac{\beta(t-2)}{\beta(t-1)}\right)^2. \tag{14}$$

A higher value of λ will thus impose a higher cost on the nonlinear changes along the $\beta(t)$ series; these nonlinear changes are likely due to overfitting. The final estimation can then be rewritten as

$$\beta_{\text{opt}}(t;\lambda) = (C(\lambda;\beta(t))). \tag{15}$$

We end this subsection by providing a selection criterion for the meta-parameter λ. Assuming the process's true parameters are $\beta_{\text{opt}}(t;\lambda)$ for some $\lambda \in R^+$, it holds that

$$2\left(ll\left(\beta_{\text{opt}}(t;0)\right) - ll\left(\beta_{\text{opt}}(t;\lambda)\right)\right) \sim \chi^2_T, \tag{16}$$

where T is the length of the vector $\beta(t)$ as well as that of the incidence time series. To arrive at this statement, suppose vector $\beta_{\text{opt}}(t;\lambda)$ is indeed the true series $\beta(t)$ describing the process. $\beta_{\text{opt}}(t;0)$ is optimized freely with T parameters, whereas $\beta_{\text{opt}}(t;\lambda)$ is said to be optimized over a subset of the previous optimization space. Equation 14 is then the null hypothesis of a Chi-squared likelihood-ratio test.

A reasonable meta-parameter λ is one for which the p value of the test is not too small (we recommend a value above 5%), ensuring that the regularized optimum is not

[2] The presented log-likelihood function is an approximation based on the number of simulations completed. Nevertheless, by the law of large numbers, it is well known that this estimation converges to that of the likelihood process.

significantly different from the unrestrained optimum. Our algorithm then proceeds by calculating the p values associated with increasingly high values of λ and then by selecting the highest value of λ for which the p value remains above the predetermined threshold.

2.3.3 Exponential Poisson Model

An advantage of the general Poisson model is its flexibility in terms of the distributions F_{inc} and F_{inf}. However, its reliance on simulations results in the requirements of extensive computations as well as a large sample of data in order for its results to be statistically meaningful. Given those limitations, we now propose a less flexible model than the general Poisson model, but one that allows for the analytical computation of previously recursive calculations while preserving most of the general model's power. We modify the general model as follows:

τ_1 := Let $e^{-\tau_1}$ be the probability that an individual incubating the infection during period t becomes infectious during period $t+1$.

τ_2 := Let $e^{-\tau_2}$ be the probability that a previously infectious individual becomes noninfectious during period t.

Notice how these new definitions limit F_{inc} and F_{inf} to memoryless processes. We thus derive

$$E[I(t)] = e^{-\tau_2}E[I(t-1)] + (1-e^{-\tau_1})E[E(t-1)], \tag{17}$$

$$E[E(t)] = \beta(t-1)E[I(t-1)] + e^{-\tau_1}E[E(t-1)], \tag{18}$$

$$E\left[I_N(t)\right] = (1-e^{-\tau_1})E[E(t-1)], \tag{19}$$

$$E\left[E_N(t)\right] = \beta(t-1)E[I(t-1)]. \tag{20}$$

These equations allow for a very efficient calculation of the expected state of the system during any period given only its initial state. Last, note that Eqs. (4″) and (6″) remain unchanged since this model is a particular instance of the general model.

Model Calibration

Under the new assumptions, the likelihood of the parameter vector β can be rewritten as

$$ll(\beta) = \sum_{t=1}^{T}\left(E[I(t)]\ln\left(\widehat{I_N(t)}\right) - E[I(t)] - \ln\left(\widehat{I_N(t)}!\right)\right). \tag{4′}$$

Equations (12–14) remain unchanged from those of the general case.

2.3.4 Construction of Confidence Intervals

Unlike for the time-since-infection models, extracting confidence intervals for $R(t)$ from the Poisson models is not entirely straightforward. We thus propose a bootstrap procedure to this end, based on the technique proposed by Davison and Hinckley (1999).

Let $I_N(t)$ be the simulated incidence series obtained by minimizing the loss function described in Eq. (12). The residual of this series with respect to the observed series is defined by

$$e(t) = I_N(t) - \widehat{I_N(t)} \tag{21}$$

For an integer $v > 0$, $v \leq t \leq T - v$ defines the local moving average

$$\bar{I}_v(t) = \frac{1}{2v+1} \sum_{\tau=t-v}^{t+v} \widehat{I_N(\tau)} \tag{22}$$

and the local variance estimator

$$s^2(t) = \frac{1}{2v} \sum_{\tau=t-v}^{t+v} \left(\widehat{I_N(\tau)} - \bar{I}_v(t) \right)^2. \tag{23}$$

In all of the simulations in this study, v is set to 2. Nevertheless, experiments have shown this process to be robust to changes in this meta-parameter.

For extreme values of t, Eqs. (20) and (21) are adapted by omitting the necessary terms. Given the likely heteroscedasticity of series $\widehat{I_N(t)}$, we standardize the above residuals by

$$r(t) = \frac{e(t)}{s(t)}. \tag{24}$$

Resampling is then performed as follows: let integer B denote the number of resampling iterations (typically, $B = 1000$), and let T be the length of the observed incidence series. For each b in $\{1, \ldots, B\}$, sample $\{r_b^*(1), \ldots, r_b^*(T)\}$ with replacement from $\{r(1), \ldots, r(T)\}$. The b-th bootstrapped series is then obtained as

$$\widehat{I_b^*(t)} = I_N(t) + r_b^*(t)s(t). \tag{25}$$

The model in question is then fit to the series $\widehat{I_b^*(t)}$ following the procedure given in the previous subsections, yielding a bootstrapped series $R_b^*(t)$. For each t, the collection of estimators $R_1^*(t), \ldots, R_B^*(t)$ allows for the extraction of quantiles that define a confidence interval for $R(t)$.

Finally, notice that every step of the calibration procedure is continuous. Thus, by the envelope theorem, the bootstrapped $\beta(t)$ are suitable for defining confidence intervals for $R(t)$ and $R_c(t)$.

3 Evaluation of Simulated Data

In the present section, the models described above are compared in terms of the quality of their fit to simulated data generated for eight theoretical settings that correspond to different functional forms of $R(t)$; these data were simulated following the logic of the disease infection process. For the simulations, a discrete model is used where individuals in an incubation state turn infectious in the next stage with probability F_{inc}, while infectious individuals turn noninfectious in the next stage with probability F_{inf}. Following the settings of the general Poisson model, for these simulations, we have

$$E_N(t+1) \sim \text{Poisson}(\beta(t)I(t)).$$

In this way, from a given series of $\beta(t)$ coefficients or, equivalently, from a series of reproduction numbers $R(t)$, a series of new infected cases, $I_N(t)$, can be simulated.

Visual inspection is used as the method of smoothing the hyperparameters (time windows for the Bayesian model and λ for the general Poisson model). We discuss this selection in Sect. 3.5.

The parameters used for the simulations are as follows:

1. Bayesian model window size: 7
2. Maximum number of lag periods: 7 in Eqs. (7) and (7′)
3. F_{inc} and F_{inf} are generated from geometric distributions with parameters 2 and 7, respectively.
    a. For the time series models, given the truncation in the number of lags allowed, only the first seven values are considered, and their probabilities are normalized so that their sum is 1 while keeping the mean unchanged.
    b. The serial interval is computed from the original F_{inc} and F_{inf} distributions. Then, the truncation and normalization of the probabilities are performed.
    c. Finally, the general Poisson model is not included in the simulations. The high computational cost for this model implies that its calibration can take several hours or even days for a sufficiently long data series, limiting its use in public policy decisions. Thus, the models included are the Bayesian model, the state-space model and the exponential Poisson model.

3.1 Simulation Scenarios

To attain a complete evaluation of the models studied, the simulations considered represent different disease transmission dynamics. In practice, the $R(t)$ series corresponds to a combination of some of these basic dynamics as well as other dynamics. Nevertheless, studying these dynamics separately helps in identifying the strengths and limitations of each model. The scenarios considered are the following:

1. $R(t)$ remains constant.
2. $R(t)$ linearly increases.
3. $R(t)$ linearly decreases.

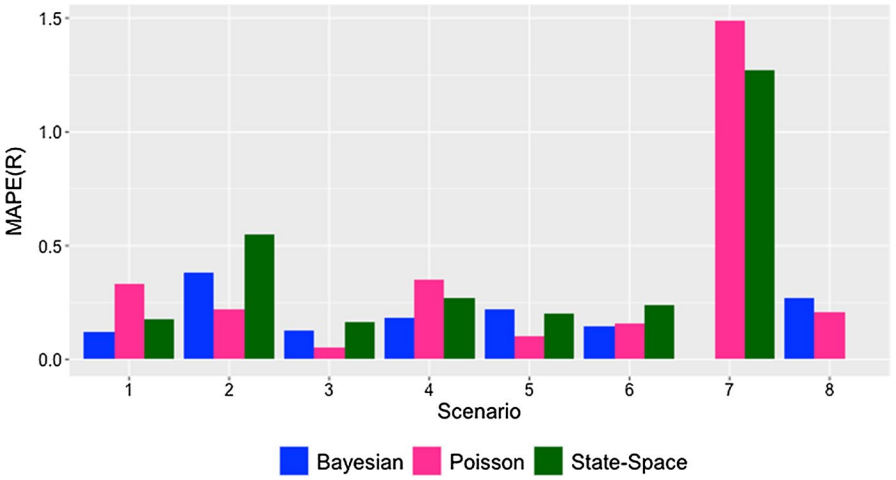

Fig. 1 MAPE between the ground truth and every model basic reproduction number obtained from each simulation (Color figure online)

4. $R(t)$ increases in steps.
5. $R(t)$ decreases in steps. This scenario resembles the results obtained by different models fitted to the COVID-19 epidemic data in Colombia up to April 2020.
6. $R(t)$ remains constant except for two large infectious peaks.
7. $R(t)$ suddenly vanishes.
8. The observed new case (incidence) series presents small values with respect to its expected value.

3.2 Evaluation Metric: Mean Absolute Percentage Error (MAPE)

As a quantitative measure, this paper uses the mean absolute percentage error to compare the theoretical value of $R(t)$ against its estimated value produced by each of the models. This metric is chosen due to its intuitive interpretability and its robustness with respect to changes in the magnitude of $R(t)$ during the observation period.

$$\text{MAPE}\left(R,\ \hat{R}\right) = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{R(t)-\widehat{R(t)}}{R(t)}\right|$$

3.3 Results

Figure 1 summarizes the adjustment quality of each model for each of the different scenarios. The graphs omit the cases in which MAPE of $R(t)$ is larger than 200%, as this allows for an optimal visualization of the cases with reasonably small errors. The following section analyses each of the scenarios; still Table 1 in Appendix summarizes

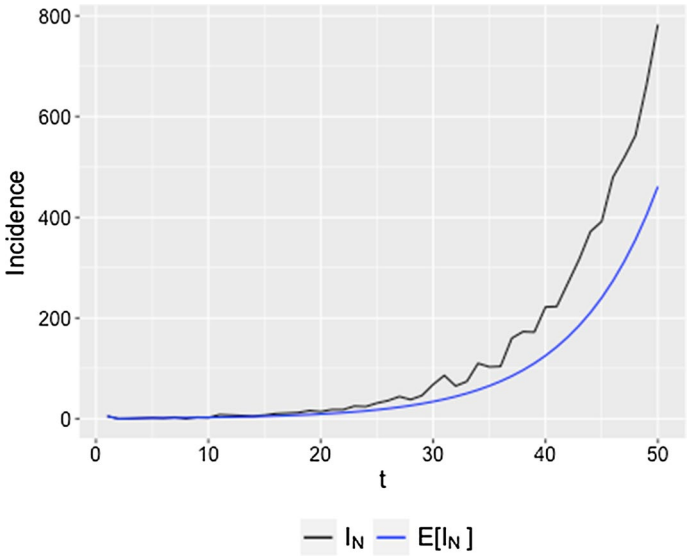

Fig. 2 Simulated new incidences and expected new incidences for scenario 1 (authors' simulations) (Color figure online)

the main conclusions that practitioners and researchers need to be aware of in case they want to use these models for their analyses.

We now present the detailed results for each simulation. For the following figures, the new case series for each simulation and the corresponding expected incidences are shown in the figures on the left. In contrast, the right side figures consist of the $R(t)$ series estimated by the three models in each of the scenarios.

Scenario 1. $R(t)$ remains constant (Figs. 2, 3).

In this first scenario, one can observe that the exponential Poisson model is the only one that captures the nature of the series from the start. The other models take longer to adjust to the theoretical series. In the long run, all three models underestimate the real $R(t)$, with the exponential Poisson model presenting the largest deviation from the real value.

Scenario 2. $R(t)$ increases linearly (Figs. 4, 5).

In the second scenario, one can notice the sensitivity of the state-space model, which reacts quickly and strongly to changes in the number of cases presented. The Bayesian model displays certain "resistance to change," which leads to its systematic underestimation of the real value. In this case, the best estimation of the real $R(t)$ value is attained by the exponential Poisson model, although it underestimates the real value for most of the time interval considered.

Scenario 3. $R(t)$ linearly decreases (Fig. 6, 7).

Scenario 3 begins with a large infection potential that produces a large number of cases from the start. All three models adjust rapidly to this setting. The best estimation is achieved by the exponential Poisson model, which does not underestimate $R(t)$ except at the beginning of the interval. The state-space had a similar performance to the Bayesian, yet at the end it captured the value better.

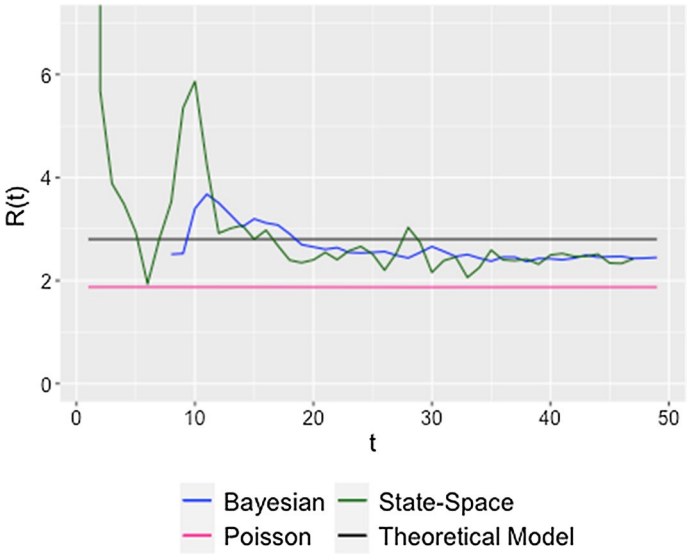

Fig. 3 $R(t)$ computed by each model on scenario 1 versus ground truth (authors' simulation) (Color figure online)

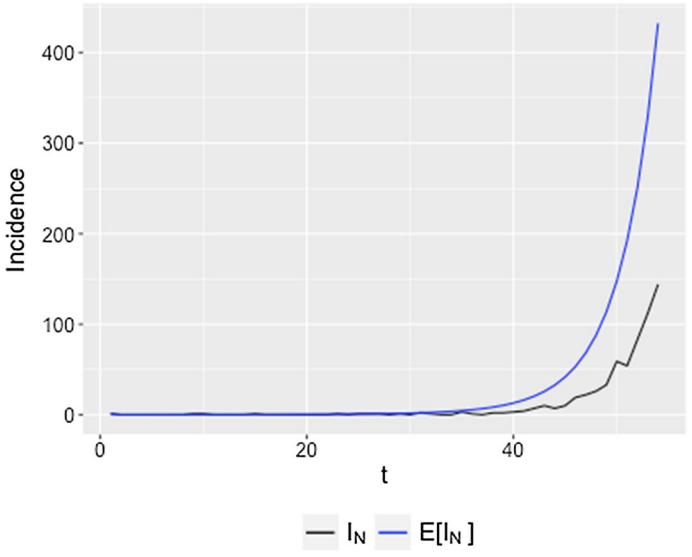

Fig. 4 Simulated new incidences and expected new incidences for scenario 2 (authors' simulations) (Color figure online)

Scenario 4. $R(t)$ increases by steps (Fig. 8, 9).

Scenario 4 can be associated with the end of a social distancing and quarantine period since at the beginning, there is a low contagion rate that suddenly increases due to the change in social behavior. In this case, the state-space model is the first

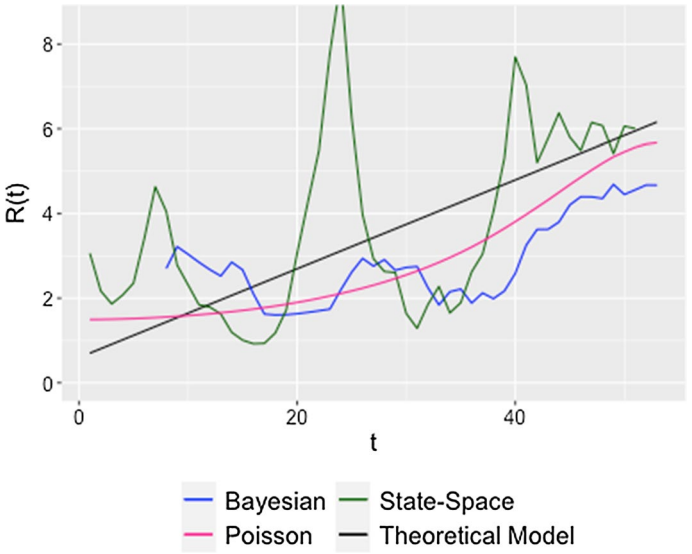

Fig. 5 $R(t)$ computed by each model on scenario 2 versus ground truth (authors' simulation) (Color figure online)

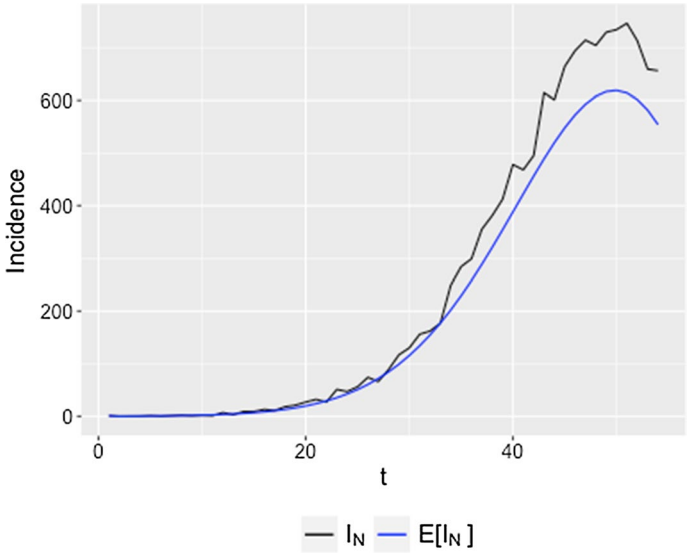

Fig. 6 Simulated new incidences and expected new incidences for scenario 3 (authors' simulations) (Color figure online)

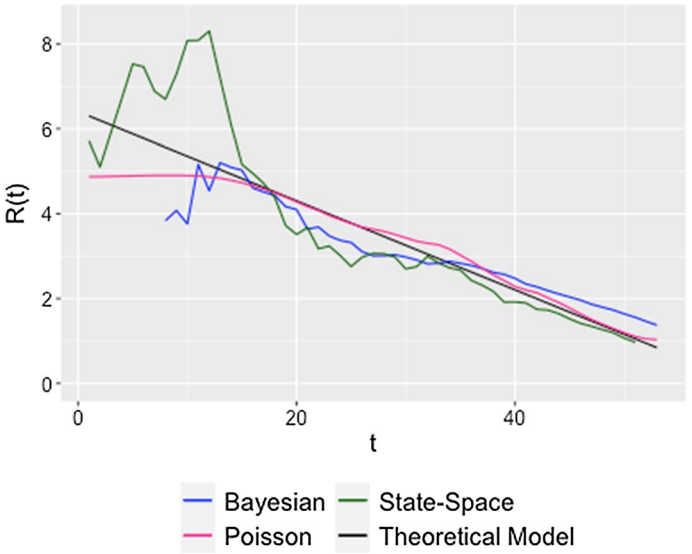

Fig. 7 $R(t)$ computed by each model on scenario 3 versus ground truth (authors' simulation) (Color figure online)

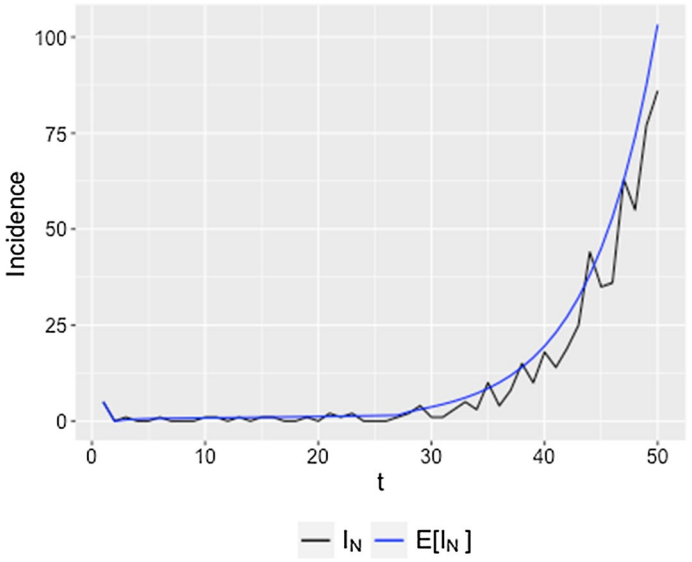

Fig. 8 Simulated new incidences and expected new incidences for scenario 4 (authors' simulations) (Color figure online)

to adjust to the change. The exponential Poisson model also adjusts rapidly, but the amount of adjustment is insufficient, probably due to the regularization that this model requires. The Bayesian model adjusts well to the new level of $R(t)$ but requires a relatively long time to reach the correct level. After the structural change,

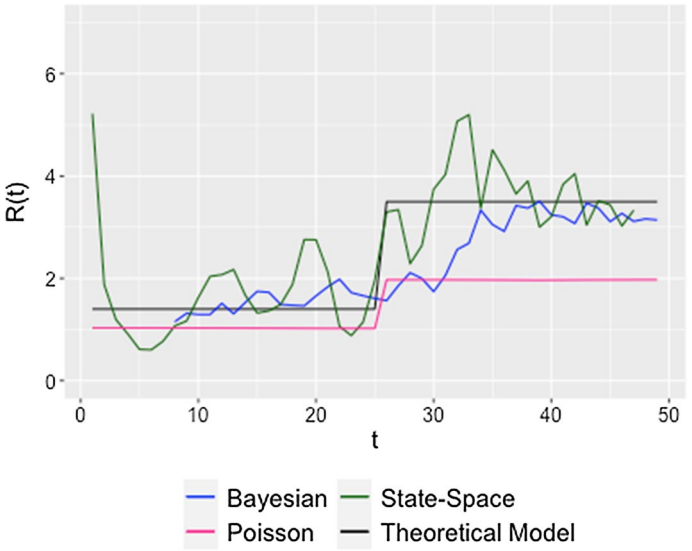

Fig. 9 R(t) computed by each model on scenario 4 versus ground truth (authors' simulation) (Color figure online)

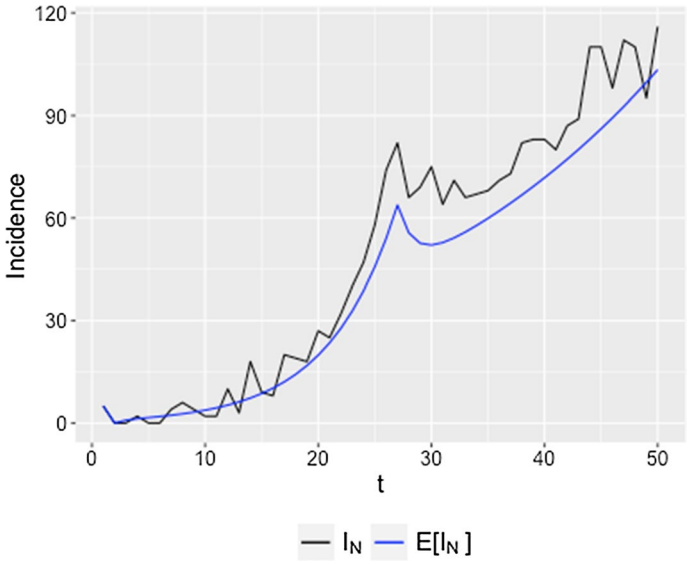

Fig. 10 Simulated new incidences and expected new incidences for scenario 5 (authors' simulations) (Color figure online)

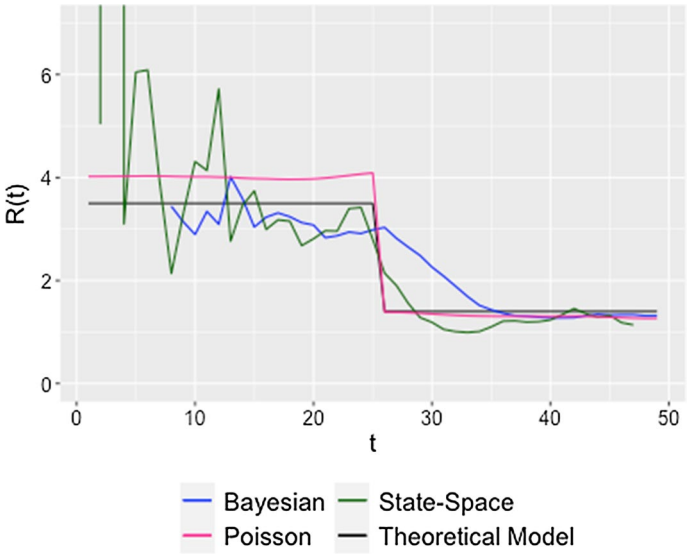

Fig. 11 R(t) computed by each model on scenario 5 versus ground truth (authors' simulation) (Color figure online)

both the Bayesian and the state-space models are closer to the theoretical value than the exponential Poisson model.

Scenario 5. R(t) decreases by steps (Figs. 10, 11).

In contrast to what happened in the previous case, in scenario 5, the exponential Poisson model adjusts rapidly and precisely to the level change in $R(t)$, followed (in terms of their response times) by the state-space model and the Bayesian model (in that order). After the structural change, during period 25, the exponential Poisson model is closest to the new level of the reproduction number. In the scenarios where the $R(t)$ values are initially high, the adjustments of the models are systematically better than the adjustments for low initial values, especially that of the exponential Poisson model. This could be attributed to the fact that as the number of cases reported increases, the exponential Poisson distribution moves away from near-zero values, thereby producing difficulties in the iteration of the simulation procedure.

Scenario 6. R(t) remains constant except for two large infectious peaks (Figs. 12, 13).

Scenario 6 is included to represent the situations where due to an unexpected event, the $R(t)$ value increases significantly, returning later to its base state. In this case, the Bayesian model performs an oversmoothing of the data and fails to detect the infectious peaks. The state-space model presents a large initial overestimation and displays other estimated peaks in the wrong places. In turn, the exponential Poisson model adjusts best to the observed levels and does the best job of identifying the locations of the peaks.

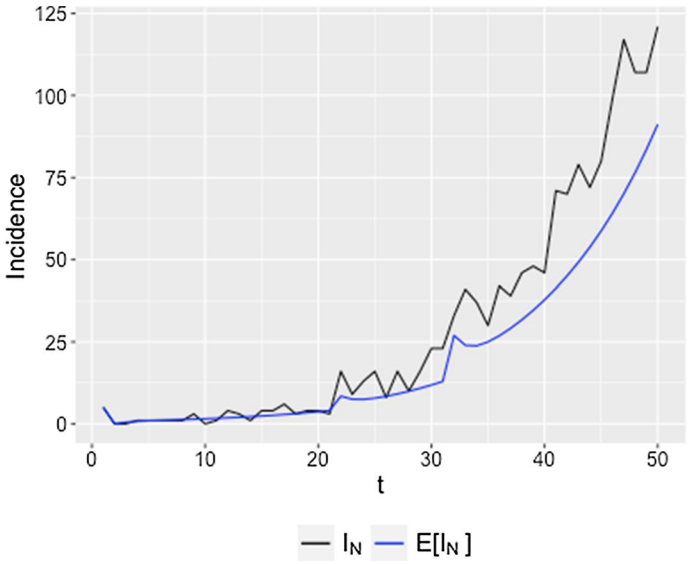

Fig. 12 Simulated new incidences and expected new incidences for scenario 6 (authors' simulations) (Color figure online)

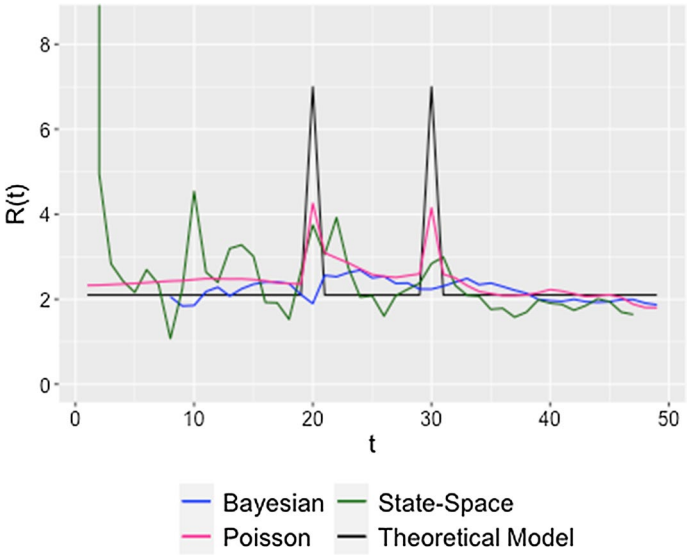

Fig. 13 R(t) computed by each model on scenario 6 versus ground truth (authors' simulation) (Color figure online)

Scenario 7. $R(t)$ suddenly vanishes (Figs. 14, 15).

In scenario 7, we observe the same behaviors exhibited by the three methods as those seen in scenario 5. The drop in $R(t)$ is detected fastest and most precisely by the exponential Poisson model.

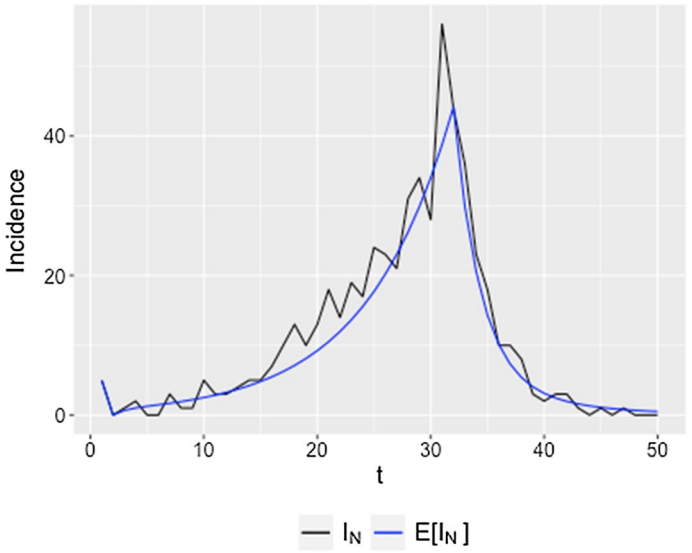

Fig. 14 Simulated new incidences and expected new incidences for scenario 7 (authors' simulations) (Color figure online)

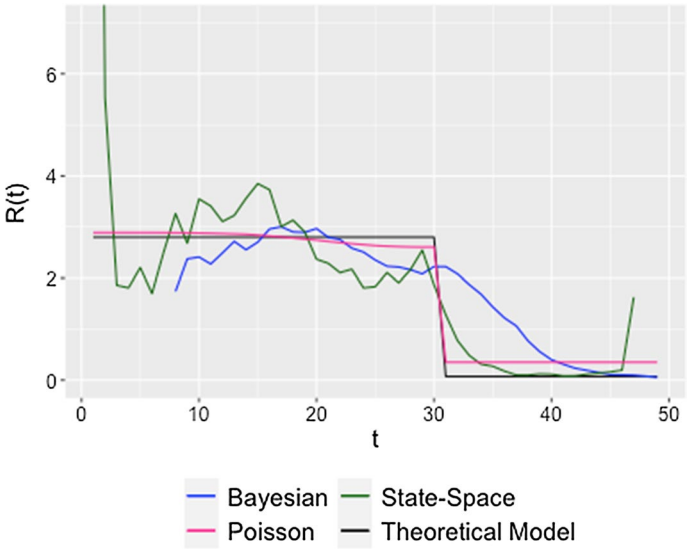

Fig. 15 R(t) computed by each model on scenario 7 versus ground truth (authors' simulation) (Color figure online)

Scenario 8 (Figs. 16, 17).

Scenario 8 shows that for realizations with case numbers that are markedly below the expected value, both the Bayesian model and the exponential Poisson model

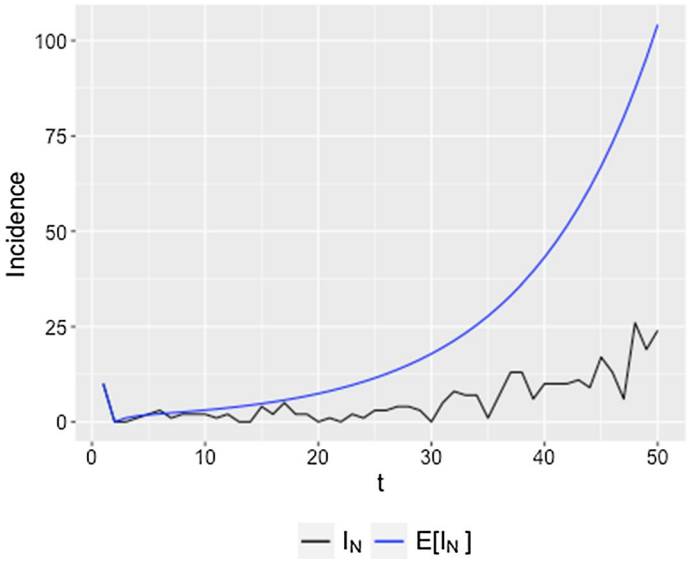

Fig. 16 Simulated new incidences and expected new incidences for scenario 8 (authors' simulations) (Color figure online)

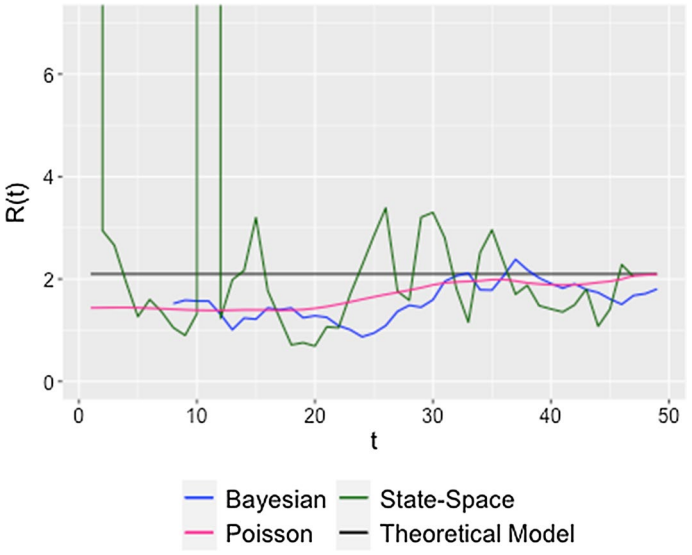

Fig. 17 R(t) computed by each model on scenario 8 versus ground truth (authors' simulation) (Color figure online)

underestimate $R(t)$; however, their estimations improve toward the end of the interval, while the state-space model displays oscillations above and below the theoretical level with some largely overestimated peaks. This behavior of the state-space model can be attributed to the low counts of disease cases.

3.4 Discussion

The Bayesian model is highly stable. For this reason, it adjusts well under scenario 1, in which $R(t)$ is constant. However, this quality is also the reason why it achieves poor performances under scenarios 2 and 3, where the infection rate increases and decreases, respectively. Moreover, this model requires the calculation of a constant serial interval with a fixed time window, which requires regularization hyperparameters. However, these values also make the model react slowly to fast changes in the infection patterns, as can be visualized for scenarios 4 and 5.

The state-space model reacts quickly to changes in the behavior of the infection. However, it also exhibits a small delay in noticing these changes, as can be observed for scenarios 5 to 7. This is because the dependent variable requires the estimation of future incubation times, and this creates this delay in the reaction of the model to the trend. In a similar way, this method reports high instability at the beginning of every simulation. We thus recommend starting analyses based on this method only after using some periods for the initial calibration of the model. Moreover, it is important to highlight that due to the lack of a regularization parameter, this method is more sensitive to changes in the data than the other models.

Finally, from the results obtained from the exponential Poisson model, it can be inferred that its main advantages are the following:

1. The visual adjustment of the visualization parameter allows the researchers to tune the smoothness of the estimator according to their expert knowledge of epidemiology and historical performance.
2. The model has a strong capacity to identify fast changes in data trends but is also capable of identifying smooth areas and reducing noise generated by the data collection process (scenarios 2 and 3).
3. Although it has a reaction lag, its lag is significantly lower than those of the other models. Moreover, it also presents highly accurate results under low incidence scenarios (scenario 8).

Finally, given that the optimization of the model includes a heuristic process, its computational complexity is greater than that of the other two. This issue becomes relevant due to the need to produce models for different regularization standards to evaluate their fitness.

3.5 Comments Regarding the Regularization Parameter

This section expands upon the previous discussion about the regularization parameter of the exponential Poisson model λ that controls the smoothness of the estimation of

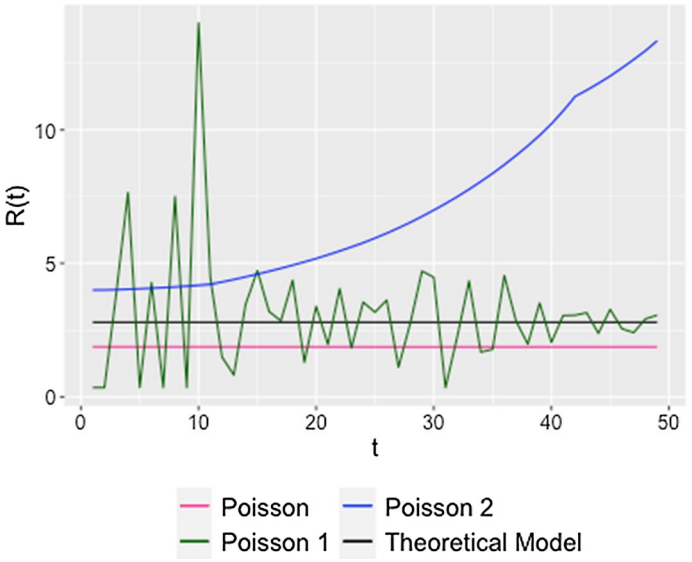

Fig. 18 Lambda regularization for constant $R(t)$ (Color figure online)

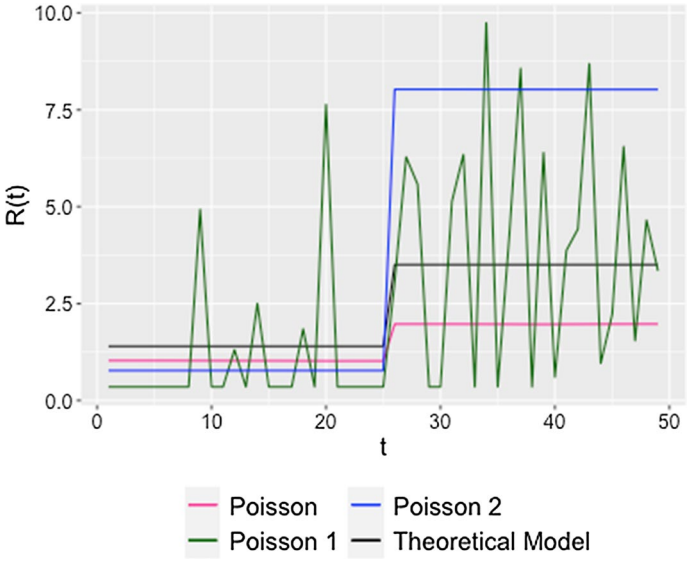

Fig. 19 Lambda regularization for changing $R(t)$ (Color figure online)

$R(t)$. To motivate this point, several regularization parameters are used to estimate scenarios 1 and 4. Figures 18 and 19 present the results for three values, where the first option ("Poisson") has an expert adjust the criterion, the second option has low

regularization, i.e., $\lambda = 0$ ("Poisson 1"), and finally, the last option has a computationally high value $\lambda = 10^{32}$ ("Poisson 2").

Scenarios 1 and 4 (Figs. 18, 19)

These figures highlight two important elements of this method to consider. On the one hand, the calibration parameter can change the tendency of the series. As seen in scenario 1, if the optimizers start in an inappropriate local optimum, the high regularization makes the model retain this mistake and produce trends that are not realistic. On the other hand, it is important that λ is high enough as it protects the series from data noise, as clearly demonstrated by Poisson 1. For this regularization parameter, the method tends to be highly sensitive to changes in the values of the data and transfers the noise to the prediction. However, the parameter can induce under- or overestimation of the series, as is visually clear from Fig. 19.

For these reasons, it is important to use this third method with a high level of caution. If the researcher has robust knowledge of the disease patterns and is capable of finding a reasonably good λ, then the method is promising. However, if the researcher is merely exploring the data, this method can induce significant errors.

4 Case Study: COVID-19 Epidemic in Colombia

To provide an illustration of the application of the previous models to a case study, we employed the three described models (Bayesian, state-space, and exponential Poisson) to describe the evolution of the COVID-19 epidemic in Colombia. The models were fit to the epidemiological surveillance data published by Colombia's National Healthcare Institute (Instituto Nacional de Salud) starting with the country's first reported case on March 6, 2020, through August 15, 2020. The meta-parameters used in this application were as follows:

1. The constant $R(t)$ window w in the Bayesian model was set to $w = 5$.
2. For ease of implementation, the supports of the distributions in the different models were limited to
   a. $\text{supp}(\omega) = [0,\ 7]$
   b. $\text{supp}\left(f_{\text{inc}}\right) = [0,\ 5]$
   c. $\text{supp}\left(f_{\text{inf}}\right) = [0,\ 15]$
3. $F_{\text{inc}} \sim \text{Gamma}(3.16,\ 5.16)$ (Gao et al. 2020) $F_{\text{inf}} \sim \text{Weibull}(24.20,\ 2.98)$ (Ling et al. 2020), and the serial interval $\omega \sim \text{Weibull}(2.23, 5.42)$(Nishiura et al. 2020).
4. In the exponential Poisson model, F_{inc} and F_{inf} were taken to be exponential distributions with the same means as those of the above distributions.
5. The regularization parameter λ in the exponential Poisson model was selected using the procedure described in Sect. 2.3.3 with a p value of 0.001.
6. 95% confidence intervals were obtained for all cases.

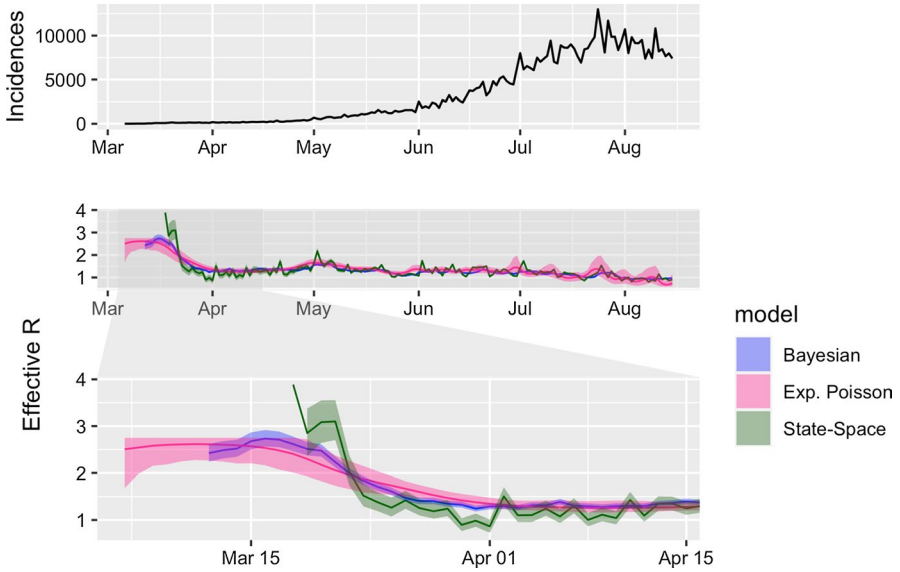

Fig. 20 R(t) estimations for the COVID-19 epidemic in Colombia from March to April 2020 (Color figure online)

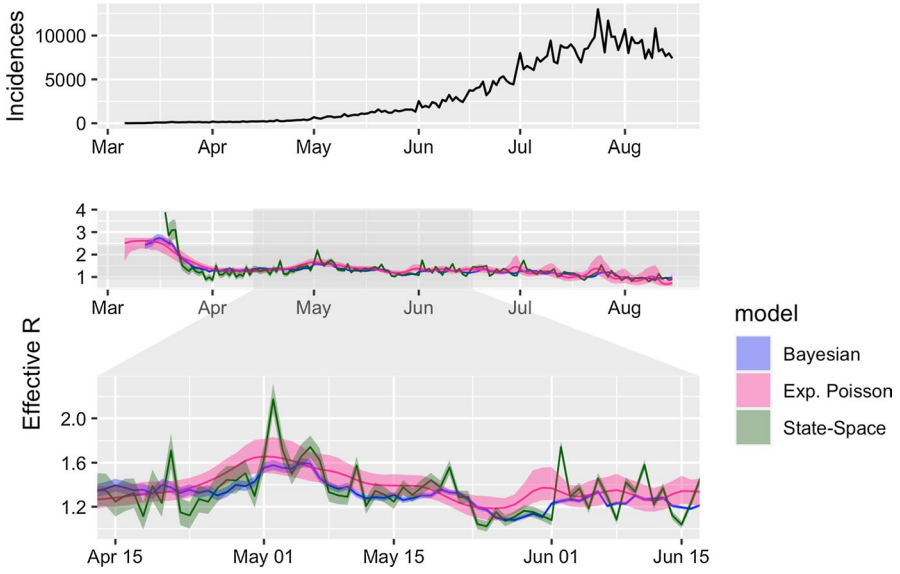

Fig. 21 R(t) estimations for the COVID-19 epidemic in Colombia from April to June 2020 (Color figure online)

These parameters were agreed by the expert committee of the Institute based on the literature review and their context-specific knowledge on the reproduction of the virus.

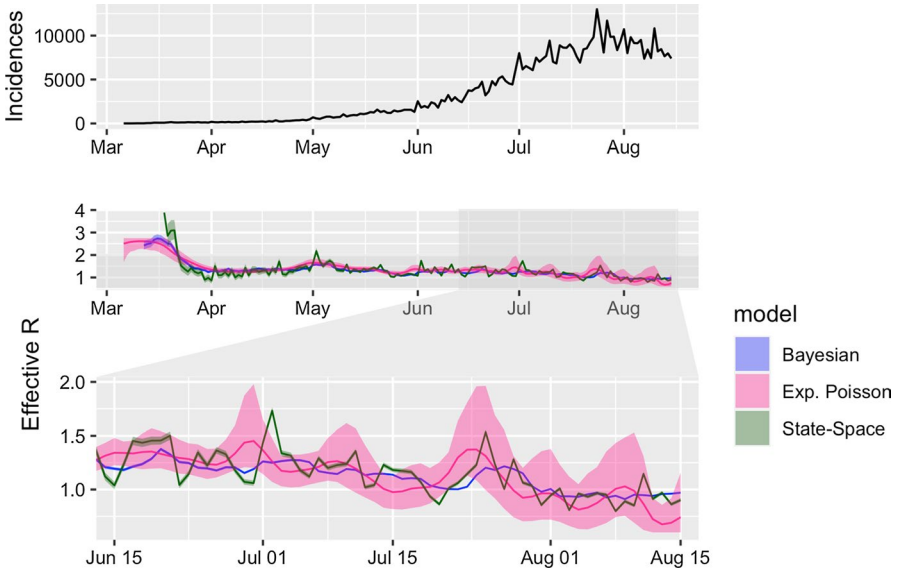

Fig. 22 R(t) estimations for the COVID-19 epidemic in Colombia from June to August 2020 (Color figure online)

4.1 Results

Figure 20 zooms into the rate $R(t)$ around the onset of the epidemic. All three models agree on a declining rate during the second half of March. This phenomenon can be explained by the first wave of quarantines in the country, including a nation-wide quarantine that started on March 24. By mid-April, the Bayesian and state-space models converged to a rate of approximately 1. Unfortunately, at the earliest stages the models have some differences while they finish the warmup period. Still all align as the sample increases.

Figure 21 highlights a period of transition from total quarantines to spatially and occupationally split quarantines around the country. This period saw oscillatory patterns in $R(t)$ as well as a sharply increased rate at the beginning of May. This may be associated with changes to the restrictions for citizens under quarantine (República 2020a; b,c).

Last, Fig. 22 shows a relatively stable period. The peaks around June 19 and July 3 are likely associated with nation-wide tax exemptions on several products on those two days, which set off a shopping spree in several large cities around the country. The third peak around July 20 might be related to the independence of the country and related events around this date. The state-space model detected these phenomena the best, while the Bayesian model resulted in soft estimations that hid them away. Moreover, the state-space model identified other peaks in $R(t)$ as well; they might be associated with other singular events or simply numerical noise. This calls for a deeper investigation of the political, economic, and social events on those dates. Finally, the exponential Poisson does a good job identifying the peaks, but depending on the

period there is a lag with the state-space model. Thus, among these two models, it is possible to identify the range of dates where the events took place.

Overall, it is noteworthy that the models were aligned on the main trends of the pandemic in Colombia. However, in their differences, significant elements appear that describe the effects of public policy. First, at the beginning, due to the small number of cases, the state-space model was very noisy. In that moment, the Bayesian and the exponential Poisson model provide a more stable description of the way in which the first round of lockdowns reduced the transmission of the disease. However, the Bayesian had a higher smoothing which does not allow the visualization of the changes in the lockdown schedule. Consistently with the simulation results, these issues suggest that the other two models will provide more useful information regarding the impact of policies. In that topic, the Poisson model begins with a lag, probably due to the low number of initial cases that affect the initial performance of the model. Yet, once the cases increase, the Poisson tends to agree with the state-space model fairly well. Therefore, the Poisson model is providing a good combination of smoothness and identification of policy changes. However, it is important to realize that the exponential Poisson relies in strong assumptions about the distribution of the infection and incubation rates.

For the case of COVID-19, the literature presented at the beginning of the section, suggested that these distributions are indeed from gamma and Weibull families. In that sense, the shape of the distribution is not that different to an exponential, yet it is not the same. In this case, having the state-space model as a back-up suggests that the distribution simplification is not that bad for some policy decisions. The other important element to recall is the importance of the smoothing parameter in the Poisson model. In this case, to choose the smoothing parameter different options were considered and the selection was chosen on the capacity of the model to match the initial data (after a warming period). Still, other researchers might have chosen other parameters based on different criteria (such as the one presented in the previous sections). This adds a level of subjectivity to the model that is undesirable. Hence, the recommended way to use these models is to evaluate all of them. The Bayesian will match data at the start as policies are not yet implemented, so the social component of the transmission rate is constant. Then, once public policies start to take place, the combination of the state space model and the Poisson model (either exponential or general) can help with the analysis of policies. In that moment, the recommendation will be to evaluate reasonable smoothing parameters and check if both models are consistent, which becomes a robustness test in itself. In that last element, it would be ideal to use the Generalized Poisson as it can be a better fit for the dynamics of the disease. Yet, as it comes with a significant computation burden, the previous results show that the exponential version can provide a suitable simplification.

5 Conclusions

The effective reproduction number is a key instrument for the development of policies associated with the prevention and mitigation of an epidemic. However, its estimation requires the consideration of different sets of assumptions, which may vary widely across diseases and social contexts. For this reason, this study focused on

understanding the implications, reaches, and limitations of those assumptions and thereafter proposed three novel models that improve the estimation of $R(t)$. Our simulation studies showed none of them to be a priori preferable to the others. We thus recommend the simultaneous use of several of them (making their assumptions explicit) to provide policy-makers with complete information. This information should then be interpreted by experts, so they can make appropriate decisions. Finally, even though the results from our case study allowed for reasonable analysis of the COVID-19 epidemic in Colombia, it is important to note that there were several meta-parameters involved in the different models, such as the distributions for the infectious and incubation periods. It is thus paramount to advance studies on the biology of diseases that will enable statistical analyses such as the ones shown here by providing accurate estimations of said meta-parameters.

Appendix

Table 1 summarizes features, limitations, and elements to consider when adjusting each one of the models presented in this work to a real dataset.

Table 1 Summary of the simulation results

Model	Features	Limitations	Cautions
Bayesian	Stable predictions Doesn't require high computational power The Bayesian framework naturally gives a confidence interval for the predictions	Requires a Fixed-length serial interval Requires a hyperparameter tuning for the length of the fixed window for the Serial interval	It adjusts slowly to sudden changes in R(t) driven by social changes, for example, a lock down The smoothness of the predictions is very dependent on the window size of the serial interval
State-Space	Reacts quickly to sudden changes in the R(t). This is because the underlying process does not depend on data of a fixed window of time For a particular day it *looks ahead* to people that started incubating on time t Doesn't require high computational power	Requires two priors: An incubation and infectious distribution Is very unstable on the first days predictions (but it becomes more stable as the sample grows)	Can produce noisy estimations and lacks from a regularization parameter
Poisson	Reacts quickly to sudden changes in R(t). This is because the social component of the underlying process does not depend on data of a fixed window of time Has a regularization parameter Has a way to fit the model based on a Hypothesis testing, thus producing a statistically significant result	Requires high computational power Requires hyperparameter tuning for the regularization parameter Requires two priors: An incubation and infectious distribution	The flexibility of this model is an advantage as well as a disadvantage. This is because it requires more inputs from the researcher: the incubation and infection distribution and visual tuning for adjusting the right amount of regularization The reliance on simulations for the fitting process makes this model very expensive in terms of computations

Acknowledgements The authors of this study acknowledge the support and advice provided by Secretaría de Salud de Bogotá, researchers from Imperial College London, including Zulma Cucunubá and Juan Vesga, and members of the project Ciencia y Vida in Chile. In that same line, the authors acknowledge the different divisions of the Instituto Nacional de Salud for their support in data collection, editing and outreach channels. Finally, the authors want to recognize the useful advice provided by the anonymous referees that help improving the quality and assertiveness of the paper.

Authors' contribution Project coordinador: Gustavo Nicolás Páez, Statistical modellers: Gustavo Nicolás Páez, Juan Felipe Cerón, Santiago Cortés, and Adolfo J. Quiroz; Algorithm optimizer: José Fernando Zea; Conceptual analysists: Érica Cruz, Gina Vargas, and Camila Franco, Gustavo Nicolás Páez, Carlos Castañeda; Data Provider: Gina Vargas, José Fernando Zea, Carlos Castañeda.

Funding The current research was done under the joint effort of researchers from Instituto Nacional de Salud, Universidad de los Andes, and independent researchers. Only public available data was used, and therefore, no specific funding was used.

Declarations

Conflict of interest The authors declare that there is no financial or personal interest or belief that can affect the objectivity of the methodology and results presented in the following study.

Code availability The overall code and data are present in the webpage: https://www.ins.gov.co/Direcciones/ONS/reportes-de-modelo-para-capitales. If specific additional details are needed, the authors welcome emails from the readers.

References

Adam D (2020) A guide to R-the pandemic's misunderstood metric. Nature 583(7816):346–348

Cori A, Ferguson NM, Fraser C, Cauchemez S (2013) A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 178(9):1505–1512

Davison AC, Hinkley DV (1999) Bootstrap methods and their applications. Cambridge University Press, Cambridge

Dickens BL, Koo JR, Lim JT, Park M, Quaye S, Sun H, Lee VJ (2020) Modelling lockdown and exit strategies for COVID-19 in Singapore. Lancet Reg Health Western Pacific 1:100004

Durbin J, Koopman SJ (2012) Time series analysis by state space methods. Oxford University Press

Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE 2(8):e758. https://doi.org/10.1371/journal.pone.0000758

Gao Q, Hu Y, Dai Z, Xiao F, Wang J, Wu J (2020) The epidemiological characteristics of 2019 novel coronavirus diseases (COVID-19) in Jingmen, Hubei China. Medicine 99(23):e20605

Harvey A (1990) Forecasting. Cambridge University Press, Structural Time Series Models and the Kalman Filter

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc Royal Soc London A 115:700–721

Ling Y, Xu SB, Lin YX, Tian D, Zhu ZQ, Dai FH, Hu BJ (2020) Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. Chin Med J 133:1039

Nishiura H, Linton NM, Akhmetzhanov AR (2020) Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 93:284

República P (2020a) Gobierno expide el Decreto 847 mediante el cual dicta nuevas disposiciones para el Aislamiento Preventivo Obligatorio. Retrieved 18 August 2020, from https://id.presidencia.gov.co/Paginas/prensa/2020/Gobierno-expide-Decreto-847-mediante-el-cual-dicta-nuevas-disposiciones-para-el-Aislamiento-Preventivo-Obligatorio-200614.aspx

República P (2020b) Estas son las 43 actividades exceptuadas durante el Aislamiento Preventivo Obligatorio que regirá en Colombia desde el 1° de junio, según Decreto expedido por el Gobierno Nacional. Retrieved 18 August 2020, from https://id.presidencia.gov.co/Paginas/prensa/2020/Estas-son-43-activ

idades-exceptuadas-durante-Aislamiento-Preventivo-Obligatorio-que-regira-Colombia-desde-1-junio-200528.aspx

República P (2020c) Abecé del Decreto 593, que amplía de 35 a 41 las actividades exceptuadas del Aislamiento Preventivo Obligatorio. Retrieved 18 August 2020, from https://id.presidencia.gov.co/Paginas/prensa/2020/Abece-del-Decreto-593-que-amplia-de-35-a-41-las-actividades-exceptuadas-del-Aislamiento-Preventivo-Obligatorio-200425.aspx

The R number and growth rate in the UK. (2020). Retrieved 18 August 2020, from https://www.gov.uk/guidance/the-r-number-in-the-uk

Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, Lessler J (2019) Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics 29:100356

van den Driessche P (2017) Reproduction numbers of infectious disease models. Infectious Disease Modelling 2(3):288–303. https://doi.org/10.1016/j.idm.2017.06.002

Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 160(6):509–516

Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Wang MH (2020) Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 92:214–217