Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Spatial and Spatio-temporal Epidemiology 39 (2021) 100455

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Robust trend estimation for COVID-19 in Brazil ☆

Fernanda Valente, Márcio P. Laurini *

FEARP-USP, Brazil

ARTICLE INFO

Keywords:
Epidemic model
Time series decomposition
Spatio-temporal count process

ABSTRACT

Estimating patterns of occurrence of cases and deaths related to the COVID-19 pandemic is a complex problem. The incidence of cases presents a great spatial and temporal heterogeneity, and the mechanisms of accounting for occurrences adopted by health departments induce a process of measurement error that alters the dependence structure of the process. In this work we propose methods to estimate the trend in the cases of COVID-19, controlling for the presence of measurement error. This decomposition is presented in Bayesian time series and spatio-temporal models for counting processes with latent components, and compared to the empirical analysis based on moving averages. We applied time series decompositions for the total number of deaths in Brazil and for the states of São Paulo and Amazonas, and a spatio-temporal analysis for all occurrences of deaths at the state level in Brazil, using two alternative specifications with global and regional components.

1. Introduction

On March 11th, 2020, the World Health Organization (WHO) declared the international public health emergency caused by the novel coronavirus (SARS-CoV-2) as a global pandemic. By December 11th, 2020, there were 69,664,639 infected people and 1,583,242 deaths in 216 countries, according to Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. As the disease propagates, healthcare systems are on the verge of collapse, and despite the unprecedented global research effort, to the present date there is no effective pharmaceutical treatments available to deal with the coronavirus disease-19 (COVID-19) (Dong et al., 2020; Sanders et al., 2020). In order to reduce the transmission, non-pharmaceutical interventions have been proposed by many countries, e.g., social distancing, self-quarantine and lockdown. However, preventing the transmission and management decisions depends on how well we can assess the real number of infected people. World Health Organization has recommended massively testing of the population (World Healh Organization, 2020) and thus, has caused a great demand for diagnostic test all over the world, but the limited availability and low number of applications has increasing the number of underreporting cases (Pedersen and Meneghini, 2020; Vaid et al., 2020; Lau et al., 2020; do Prado et al., 2020; Russell et al., 2020).

In addition to the underreporting number of COVID-19 cases and deaths, it is worth noting that there are other problems related to COVID-19 data. First, a delay between the onset of symptoms and accurate diagnosis is commonly observed, which varies from country to country, depending on the local government's strategy (Contreras et al., 2020). Plus, there are also days of delay in the report of new deaths which are counted on the day that are included in the system, instead of the actual day of the death (Russell et al., 2020). This creates a relevant measurement error problem, changing the series' dependency structure. The delays are also related to the lack of personnel available to report cases on weekends, which creates a seasonality structure in the cases and death reports, generating an additional aggregation problem on the time series (Skiera et al., 2020). Plus, the delays are related to problems in the system of accounting for confirmed cases and deaths, which suffers from some instability issues. In some days the system is not available, so deaths and cases are included late.

Given the aforementioned discussion, it is possible to note that estimating the trend in COVID-19 cases and deaths it is not a trivial problem. The estimation of long-term movements is of vital importance to draw effective strategies to reduce the transmission of COVID-19. There are several different methods for trend estimation, which differ in their complexity and interpretability, where the most prevalent trend estimation methods are model-based trend extraction, nonparametric filtering, singular spectrum analysis and wavelets (Alexandrov et al., 2012).

Although there are many studies in the literature to estimate the trend of COVID-19 for different countries (e.g., Li et al., 2020; Gupta

☆ The authors acknowledge financial support from CNPq (306023/2018-0), FAPESP (2018/04654-9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001 and Instituto Escolhas. We are grateful for the comments and suggestions of two anonymous referees, and the 40 volunteers of the Brasil.IO project, which provided the data used in our study.

* Correspondence to: Av. dos Bandeirantes 3900, 14040-905, Ribeirão Preto, SP, Brazil.
E-mail address: laurini@fearp.usp.br (M.P. Laurini).

https://doi.org/10.1016/j.sste.2021.100455
Received 14 December 2020; Received in revised form 5 July 2021; Accepted 25 August 2021
Available online 13 September 2021

and Pal, 2020; Ceylan, 2020; Perone, 2020), there are little understanding about the long-term movements of COVID-19 in Brazil. Nowadays the country draws a lot of attention since it has the world's second-most cases and deaths of COVID-19 (behind just the United States), and with probably substantial underestimates since the number of tested people are relatively low. For instance, up to August 12th, about 62 thousand tests per million inhabitants have been applied in Brazil, while the United States and United Kingdom has applied more than 200 thousand tests per million inhabitants. The difficulties in estimating the amount of infected people in Brazil are related to the absence of adequate laboratory infrastructure and qualified people, difficulty in buying tests due to the high international demand, and logistical distribution of tests in a country of continental dimensions such as Brazil (Ribeiro and ao Bernardes, 2020).

Notwithstanding the supposed underreporting of COVID-19 cases, the country also faces problems caused by contradictory and inaccurate data presented by official public portals. This occurs due to the polarized opposition between the federal government, and the state and municipal governments. While Brazilian government and, especially the President Jair Bolsonaro, are continually disqualify publicly both risks and the adoption of scientifically based prevention measures, most state and municipal governments have imposed social distancing along with other public health measures to control the spread of the virus (Ortega and Orsini, 2020). Several Brazilian official control bodies (municipal, state and federal) have created official internet portals to report the number of cases. However, there are many discrepancies in these sources, presenting contradictory data on the impact of the disease (Silva et al., 2020).

In a response to conflicting numbers provided by Brazilian Health Ministry, some data collection initiatives have collected the data by each municipality, providing more accurate information to COVID-19 research, in a parallel work to the federal government. The most notable work is provided by competing Brazilian mainstream media outlets, which have established a work in a collaborative way to gather and upload necessary COVID-19 information in the 26 states and the Federal District. In addition to the daily data, the consortium also provides the so-called moving average, in order to give a better view of the evolution of confirmed cases and deaths of COVID-19 in Brazil since this is an understandable and easy to execute method to compute trend values. The moving average filter is an example of low-pass filter because it eliminates the lower or slower frequencies from the time series by means of moving average. However, this is a limited tool for some reasons. First, it cannot be used to forecasting since the trend path does not belong to any mathematical function. Second, the extent of moving average is determined ad-hoc and may carry the effect of human judgment. Also, the method assumes that the trend is always linear.

An alternative way to estimate the long-term movements is through structural decomposition (e.g., Harvey, 1990). In this sense, we propose to decompose the temporal variability observed in the time series into trend, seasonal and cycle components, which allows us to identify long-term movements, and cyclical and seasonal effects, in the presence of measurement errors. Since we are dealing with point process, we introduce a method that allows decomposition of time series in a count data framework, based on the Poisson distribution. In particular, we propose to use a Poisson process where the intensity function is decomposed into trend, seasonal and cycle components. The inference is performed following a Bayesian approach, which is able to capture the uncertainty associated with the latent factors via Bayesian credibility interval. We also present a spatio-temporal generalization of this methodology, using a formulation of Conditional Autoregressive (CAR) models with time varying spatial random effects. In particular, the resulting Bayesian hierarchical model fits within the integrated nested Laplace approximations (INLA) framework, providing an estimation in a computationally effective way (Rue et al., 2009).

Regarding this context, our main goal is to estimate the patterns of the deaths by COVID-19 in Brazil through the trend-cycle decomposition. The contribution of this paper is to explore a Bayesian version of the structural decomposition in combination with count distributions, in the task of estimating the trend of deaths of COVID-19 in Brazil, and compare it with averages approach, which are not robust to the most common problems related to COVID-19 data. In particular, we performed inference procedures for deaths reported in Brazil, and also for the states of São Paulo and Amazonas. We choose to analyze these two states in different regions (São Paulo in Southeast and Amazonas in North) since Brazil is a country with continental dimensions and is characterized by a great heterogeneity in socioeconomic and cultural context among regions. For instance, in the Southeast region, the social distance and hygiene measures to reduce the transmission of COVID-19 are not feasible in subnormal agglomerates (also known as "favelas") and peripheries, where the problem is intensified by issues in sanitation and access to water. However, it is important to highlight the difference of this problem among Brazilian regions: while about 92% of the population in Southeast are supplied with treated water, in the North, this number is just 57%. Furthermore, the North region faces a lack of heath infrastructure. As an example, the state of Amazonas concentrates all the hospital structures able to deal with COVID-19 in a single municipality, Manaus. It represents the lack of health infrastructure for the other 61 municipalities within a state whose territorial extension is larger than those of United Kingdom, Italy and France combined (Freitas et al., 2020).

Additionally, we also extend our proposed model by including spatial information in the estimation process. The use of spatial information is important in estimating the trend and cycle components, due to the nature of transmission in an epidemic process, with spatial spread dynamics. Thus, the number of cases in a region has an important impact on neighboring regions, and the incorporation of this information as a prior information for the number of cases allows a more precise recovery of the trend of occurrences. We formulate two spatio-temporal models through a conditional autoregressive structure with time-varying spatial effects. Allowing spatial effects to change over time is essential in modeling an epidemic process, as this structure allows to identify the spatial pattern of spread of the infection. The spatio-temporal models differ in terms of the components of trend, seasonality and cycle. The first model assumes components common to all regions of the country, while in the second model we introduce specific trends, seasonality and cycles for each region of the country, allowing an identification of the regional patterns of the COVID-19 epidemic.

This article is organized as follow. Section 2 contains a description of the temporal and spatio-temporal models and presents the data. Section 3 shows the results with discussion. Section 4 concludes.

2. Data and methods

2.1. Data

To perform inference procedures in all analysis we use a data set provided by BRASIL.IO (https://brasil.io/dataset/covid19/caso_full/), where the data collection is done by the reports of the number of cases and deaths from the official epidemiological bulletins of each municipality, by a task force of 40 volunteers. This daily data is available at municipal level, containing the temporal evolution for the number of reported deaths for each municipality and state. In this work, we analyze the sample of deaths from 02/25/2020 to 12/06/2020 for Brazil and the states of São Paulo and Amazonas.

2.2. Univariate time series models

The model used in this work is based on a structural decomposition suitable for modeling counting series, using a Poisson distribution structure. In this distribution, the log intensity varies over time, and is given by the sum of components of trend, seasonality and cycle. As we are using a structure of Poisson processes with stochastic intensity, the characterization of the process is given by a Cox process. In the spatio-temporal analyzes, we use a structure of a dynamic version of Besag–York–Mollié model (Besag et al., 1991), where the log of the intensity function is given by a Gaussian Markov Random Field (Rue and Held, 2005; Illian et al., 2012).

The trend component is formulated using a second order random walk structure (Rue and Held, 2005), while the seasonal component is composed of stochastic effects in the order of periodicity of the series, in this case since we use daily data. The cyclic component is based on a structure of a second-order autoregressive process, which aims to capture the sum of stationary effects with mean reversion, including the effects induced by measurement errors in the series of deaths. This component is essential for the correct identification of trend and seasonality patterns, as we will discuss below. The time series model can be written as:

$$\begin{aligned} Y_t &= Poisson(exp(\lambda_t)),\\ \lambda_t &= \mu_t + s_t + c_t\\ \Delta^2\mu_t &= \eta_\mu\\ s_t &= s_{t-1} + s_{t-2} + \cdots + s_{t-m} + \eta_s\\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \end{aligned} \tag{1}$$

where Y_t is the number of occurrences in time t, μ_t is the long term trend, which can be seen as the accumulation of all shocks that occurred in the past with non-transitory effects, and is modeled as a second-order random walk (RW2). The s_t represents the seasonal components, c_t is a cycle component represented by a second-order autoregressive process with possibly complex roots, η_μ, η_c, and η_s are nonspatial independent innovations with $\eta_\mu \sim N(0, \sigma^2_{\eta_\mu})$, $\eta_c \sim N(0, \sigma^2_{\eta_c})$, and $\eta_s \sim N(0, \sigma^2_{\eta_s})$. The resulting Bayesian hierarchical structure allows us to perform inference procedure within the INLA framework, which provides accurate and efficient approximations on Bayesian hierarchical models that can be represented as latent Gaussian models. For reasons of space, we do not detail the INLA method here, but we show the fundamental aspects of the method in Appendix A.4.

The use of an RW2 structure for the modeling of COVID cases, replacing traditional local level processes, can be justified in several aspects. The first one is the nature of extremely fast case growth in an epidemic process, which generates a dependency structure that is best approximated by an integrated second order process (I(2)). In the stages of accelerated growth of cases, we have the presence of a trend component with a non-stationary growth rate, inducing a process that needs at least two differences to induce stationarity. The RW2 process is a parsimonious way of capturing processes with this dependency structure. A second characteristic of this process is that in these situations it also imposes a smoothness structure in the process trend, since this process can be related to formulations of smoothing splines models, as discussed for example in Green and Silverman (1994) and Lindgren and Rue (2008). A similar discussion in econometrics is in the relationship between the so-called Hodrick–Prescott filter and its formulation in state space, which corresponds to a spline model with an estimated parameter of ratio between the variances of the observation and state equations. Discussions between these properties and problems using the HP filter can be found at Harvey and Trimbur (2008) and Hamilton (2018).

2.3. Spatio-temporal analysis

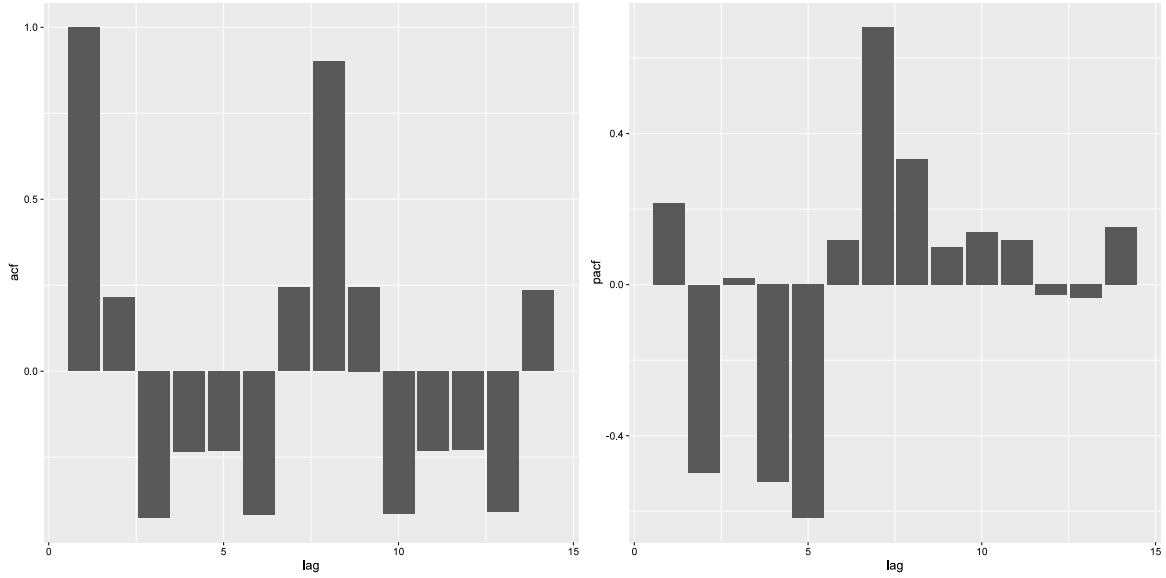
The analyzes carried out so far used only local information to carry out the inference procedures on the trend and the other components. A possible extension is the use of information on cases or deaths that occurred in some definition of neighborhood in the region. This information is important since the number of occurrences in the neighborhood can be used as prior information for the estimation in the Bayesian inference procedure, which is particularly important in the estimation of a high speed epidemic process and whose specific characteristics of transmission, latency and mortality are unknown.

In order to perform a spatio-temporal analysis of the patterns of occurrence of deaths related to COVID-19, we formulated a generalization of the decomposition of trend, seasonality and cycle components incorporating a time-varying spatial component, using a version of a model of random spatial effects with a Conditional Autoregressive (CAR) (Besag, 1974; Besag et al., 1991) structure with time dependence. This model allows to incorporate the existing information in the number of cases of the neighbor as prior information for the number of cases in the region of interest.

In relation to the model proposed in Eq. (1), the spatio-temporal version adds two modifications. The first is the use of an offset E_i to control the Exposure, which in this case is the number of inhabitants of region i. This component allows to use common components of trend, seasonality and cycle for regions with different population sizes. The second modification is the $\xi_{(i,t)}$ spatial component, which adds an additional random effect for each region i, allowing to incorporate spatial variability in the occurrence rate. The CAR structure used takes as prior for the spatial effect in the region i a Normal distribution with mean given by the average of the values of the spatial effects for regions j that are neighborhood to region i, and variance controlled by a τ precision parameter multiplied by number of neighbors. We assume that this pattern varies over time, reflecting the dynamics in the spatial distribution of the number of occurrences of COVID-19. The dynamic formulation is built through an autoregressive structure in time for this component, an additional parameter Φ controlling the time dependence for this process. The model structure can be represented as:

$$\begin{aligned} Y_{(i,t)} &= Poisson(exp(\lambda_{(i,t)}E)),\\ \lambda_{(i,t)} &= \mu_t + s_t + c_t + \xi_{(i,t)}\\ \Delta^2\mu_t &= \eta_\mu\\ s_t &= s_{t-1} + s_{t-2} + \cdots + s_{t-m} + \eta_s\\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c\\ \xi_{(i,t)} \mid \xi_{(j,t),i\neq j,\tau} &\sim \mathcal{N}\left(\frac{1}{n_i}\sum_{i\sim j}\xi_{(j,t)}, \frac{1}{\tau n_i}\right)\\ \xi_{(j,t)} &= \Phi\xi_{(j,t-1)} \end{aligned} \tag{2}$$

To estimate the above spatio-temporal model, we used COVID-19 data at the state level as a spatial unit definition. Although it is possible to work with data at the municipal level, we note that there are important problems with the detailing of cases at this level. The data are only reported in detail for the largest municipalities in each state, with data for small municipalities being aggregated in terms of regions or else as a residue in the total sum of cases within each state. Thus, the state aggregation offers greater reliability in the analysis of spatial data patterns. We used the neighborhood definition as a neighborhood structure based on queen-type contiguity using the territorial division of Brazilian states. As in the first model, the spatio-temporal model is also estimated using Integrated Nested Laplace Approximations, where the prior structure for the time series components is the same used to estimate univariate models, and we also adopted a prior log-gamma with values (1.5e−5) for the precision of this component. The prior for the coefficient Φ is a penalized complexity prior (Simpson et al., 2017) for the correlation parameter, with values (3, 0.01).



(a) ACF (b) PACF

Fig. 1. Simulated ACF and PACF functions of new daily deaths under the reporting contamination mechanism. Mean values from 10,000 replications of the SEIR model with measurement error.

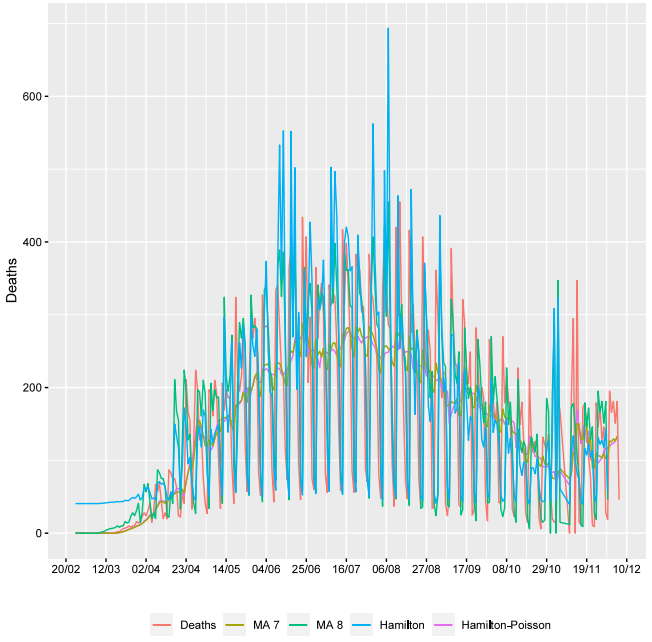


Fig. 2. Trend extraction using Moving Average and Hamilton Filters - COVID-19 related deaths for São Paulo State — 02/25/2020 to 12/06/2020.

2.4. Monte Carlo Study for SEIR model

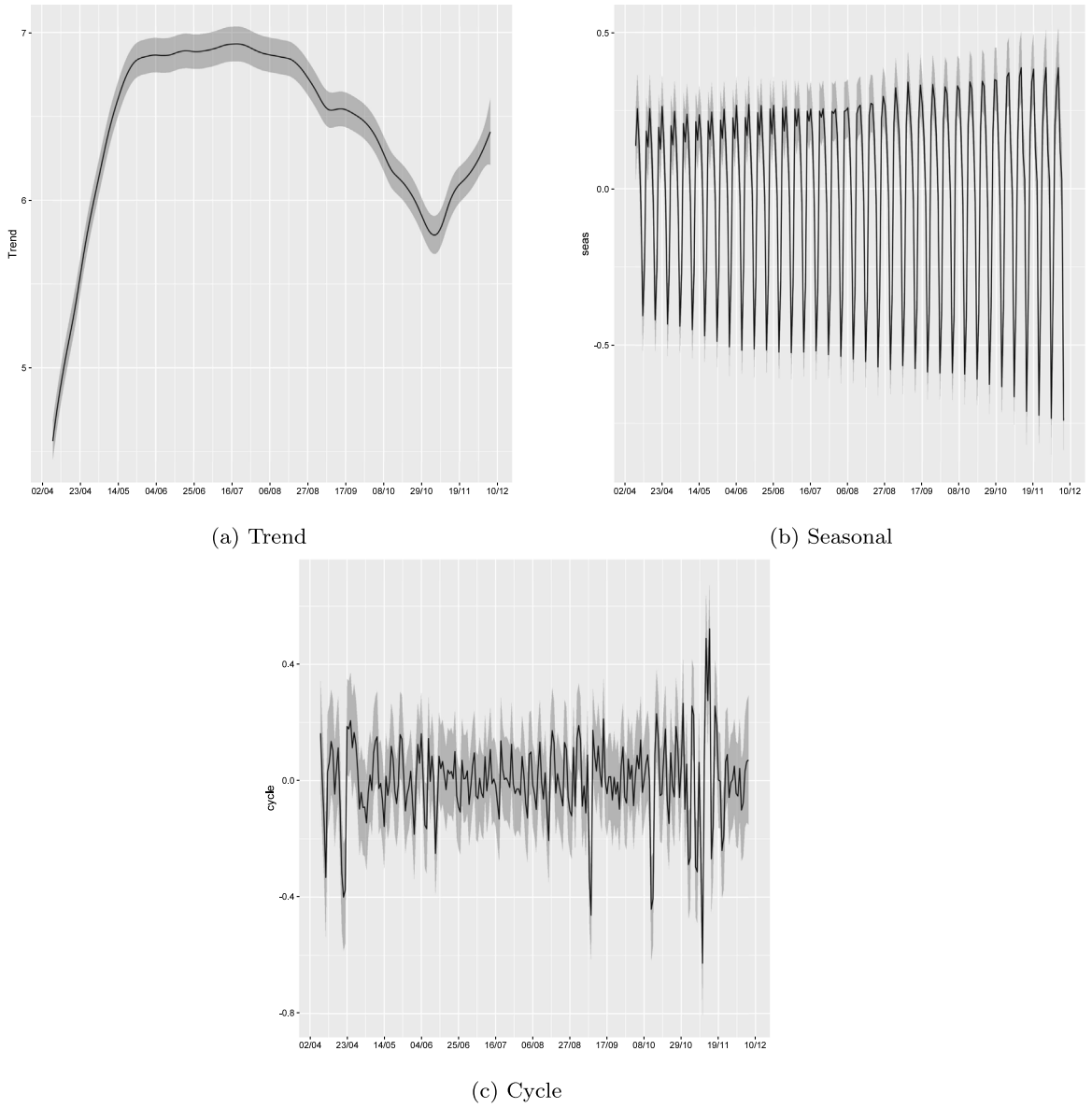
The seasonality and cycle components are introduced in order to control the effects related to the case accounting mechanism, which, as discussed in the introduction, introduces a structure of periodicity in the data, generating a seasonal pattern. It is also important to note that the dependency pattern generated by the epidemic process and this form of measurement error generates a complex contamination structure in the series of new cases or deaths, which reduces the effectiveness of simple methods of eliminating seasonality or estimating trends, in particular the use of simple methods such as a 7-day moving average in cases, which is the form used to summarize the trend of COVID-19 cases in Brazil, provided by the Brazilian mainstream media outlets.

To show the complexity of this contamination pattern and the limitation of simple trend extraction methods, we will resort to simulations of an epidemic process with a contamination structure similar to that existing for COVID-19 data, as well as some applications of simple methods of trend extraction. We performed a Monte Carlo experiment with 10,000 replications, simulating a deterministic Susceptible–Exposed–Infectious–Recovered (SEIR) epidemic process, with a stochastic structure for the number of new deaths resulting from this model. The general details of this model are presented in Appendix A.3.

The key idea is to obtain the solution of a SEIR process, which generates trajectories of the number of Susceptible, Exposed, Infectious, Recovered and Deaths over time. In our simulation, we replicated a mechanism for reporting new death cases similar to that used by health departments in Brazil. In particular, we define that each observation corresponds to a daily data in a seven-day week, and if deaths occur on Saturday or Sunday, there is a proportion of deaths that will not be counted on the same day, but only on the following Monday and Tuesday. Also, we consider that all deaths are effectively accounted for, and set that the proportion of deaths with delayed disclosure is given by a uniform distribution, with parameters (.5, .8). Indeed, this structure is a very simplified and unrealistic version of the reality due to underreporting of deaths, but it is useful to exemplify the effects of measurement error on the dynamic structure of the time series of new deaths.

As a summary of the contamination effect, we show in Fig. 1 the average values of the autocorrelation and partial autocorrelation functions for the time series of new daily deaths. We can see that the contamination mechanism induces a complex structure of temporal dependence in the series of new deaths in the SEIR process. Although there is a seasonal pattern induced by contamination, the dependency structure is more complex than a pure seasonal pattern, even in this simple contamination experiment.

The structure used in our model uses a seasonal component to control for the periodic pattern induced in the series, but also a second-order autoregressive component, to capture the remaining dynamic effects of the measurement error structure in the series. The second-order component is a parsimonious way of capturing dependency patterns and cycles with reversion to the mean, and in our analyzes it has been



(a) Trend

(b) Seasonal

(c) Cycle

Fig. 3. Posterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in Brazil — 02/25/2020 to 12/06/2020.

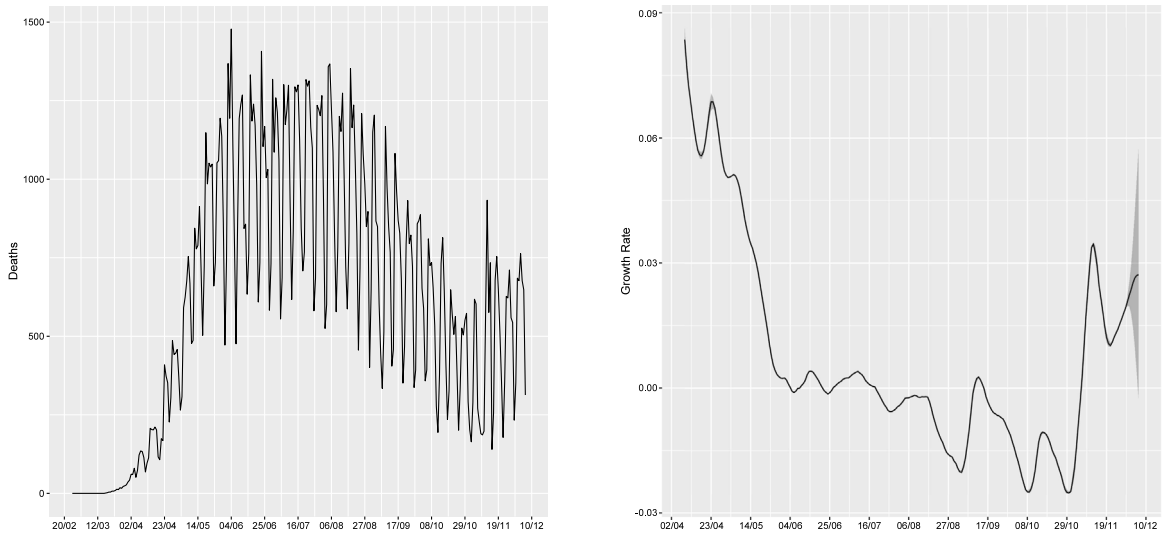
shown to be effective in allowing the recovery of the trend component in the presence of seasonality and other dynamics in the series.

Note that in the presence of these contamination mechanisms, simple seasonal filtering methods, such as the use of 7-day moving averages, will not be sufficient to extract the trend component in the series. To illustrate this problem, we will use an example with real data, in particular, the same dataset provided by BRASIL.IO, which will be also used in this work for the temporal and spatio-temporal analyzes.

We used as an example the extraction of a trend measure for COVID-related deaths for the state of São Paulo. In this example, we compare the trend obtained by a simple 7-day moving average, a moving average with the number of days determined using the modified Akaike criterion with correction for small samples (AICc) (see Svetunkov and Petropoulos, 2018 for details in this procedure), and also the trend extraction using the trend filter proposed by Hamilton (2018), assuming a periodicity of 7 days. We also used a version of the Hamilton filter based on a generalized linear model assuming that the data follow a Poisson process.

Fig. 2 shows the time series of COVID-related deaths and the trends extracted by aforementioned methods. First, an important result is that the optimal criterion for the number of lags in the moving averages procedure selected by the AICc criterion points out to 8 lags, confirming that the dependence induced by the death accounting process generates a more complex dependency structure than a seasonality of 7 days. Additionally, another interesting result is that the Hamilton filter, a trend extraction tool based on regressions against past lags and with optimally properties in general contexts, cannot effectively separate the trend component from the seasonal pattern, probably due to the stochastic and non-stationary component of the seasonality component induced by the measurement error process.

These results show that simple trend extraction methods, such as the use of simple moving averages, may not be adequate in the presence of measurement error structures such as the COVID-19 data present here. In the following sections we show the results obtained with the component extraction methods proposed in this work, using a formulation of time series applied to univariate data (Section 3.1), and also a spatio-temporal version for data at the state level in Brazil (Section 3.3). In this work we focus on the analysis of data on deaths related to COVID-19 in Brazil, since these data are less affected by the underreporting problem. Data from confirmed cases are difficult to analyze and especially to compare among regions, since testing procedures and recommendations



(a) Deaths (b) Growth Rate

Fig. 4. Number of reported deaths and the estimated growth rate of the Trend in Brazil — 02/25/2020 to 12/06/2020.

are quite heterogeneous, reducing the reliability of the data and the resulting analyzes.

3. Results

In this section we report and discuss, first, the results based on the univariate analysis, and the comparison between the estimated trend component and moving averages approach. After, we present the results based on the spatio-temporal model. Our objective is to show what are the advantages and limitations of each model used in the work, and how it can be useful to understand the dynamics of evolution of deaths by COVID-19. The pure time series model has the advantage of only depending on the data from the region itself, but it does not capture the spatial dependence of the process. The two versions of the spatio-temporal models, with global and regional components, illustrate the gains and limitations of the use of global and regional components in spatio-temporal modeling.

3.1. Univariate time series models

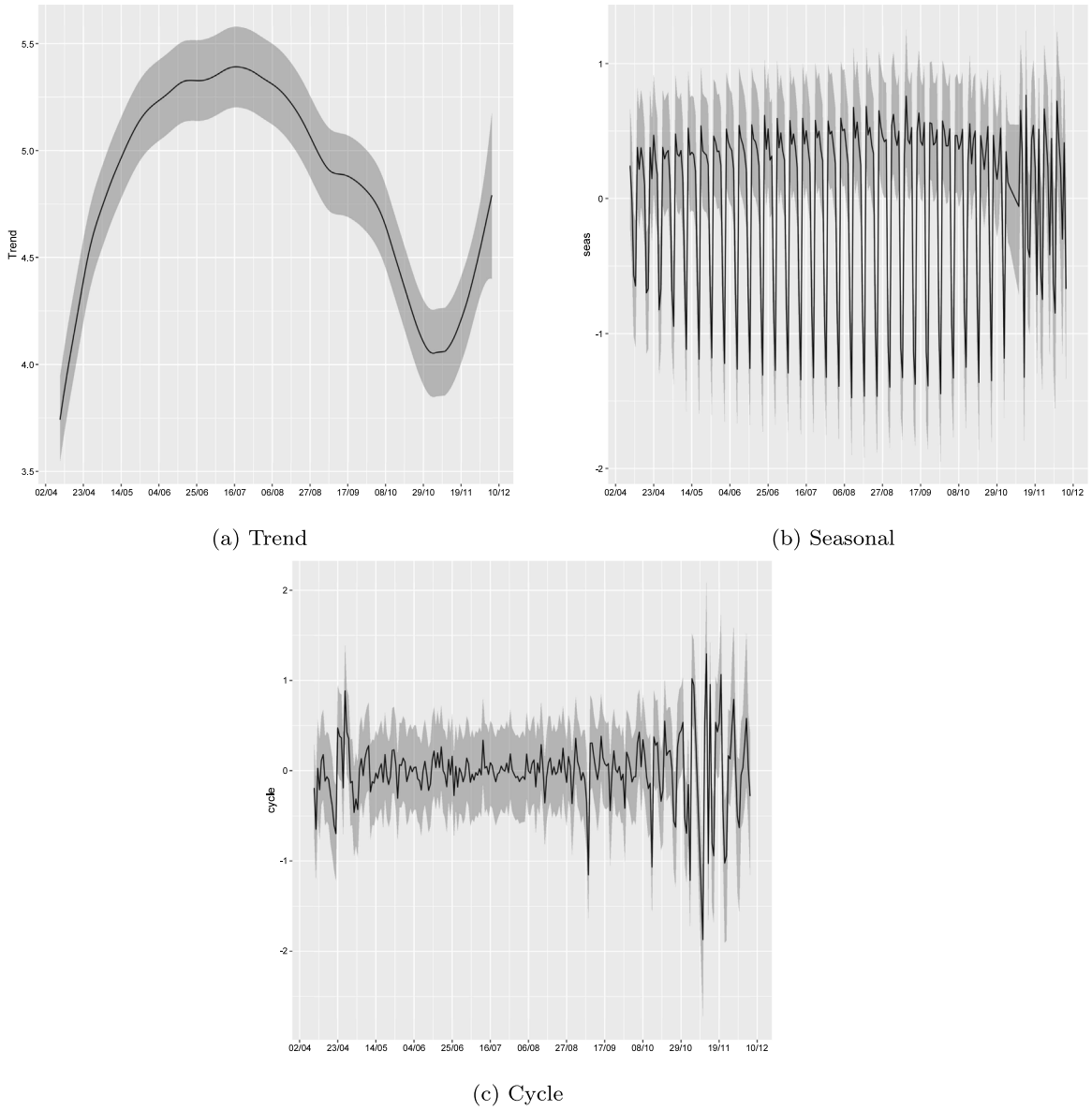
We performed inference procedures based on the model described in Section 2, thus, in this case, the estimated parameters for each location are the precision of the trend component ($1/\eta_\mu$), seasonal component ($1/\eta_s$), and cycle component ($1/\eta_c$), and the parameters of the second-order autoregressive process of the cycle, parameterized as first and second order partial correlation coefficients (PACF1 and PACF2). The precision parameters represent the variability associated with the trend, seasonal and cycle components, where high values indicate low variability, and the parameters of the second-order autoregressive process of the cycle are related to the autoregressive parameters in the AR(2) representation of the cycle.

We use a structure of log-gamma priors for the precision components, with values (1, 5e−05) for the trend and seasonal components. For the AR(2) component we use a penalized complexity prior for the precision, with values (3, 0.01), and for the first and second order autocorrelations parameters penalized complexity priors with values (0.5, 0.5) and (0.5, 0.4). The motivations for the use of penalized complexity priors is discussed in Simpson et al. (2017). These priors are invariant to reparameterizations and have excellent robustness properties, and in the assumed values they are not informative, which is adequate for a situation of a new epidemic with unknown previous behavior. The tables with the estimated parameters for Brasil, São Paulo and Amazonas states are presented in the Appendix of the article (respectively, Tables A.1–A.3).

The dissemination of the COVID-19 among Brazilian states vary due to the heterogeneity in socioeconomic and cultural context among regions. In order to assess the strong regional differences in the spread of infections, we performed inference procedures for the states of São Paulo, localized in the Southeast region of Brazil, and for the state of Amazonas, in the North region. The Southeast region includes the three largest metropolitan areas in Brazil, São Paulo, Rio de Janeiro, and Belo Horizonte. Also, it is the main industrialized area and concentrates the biggest population of the five Brazilian regions, with a high population density. It is not surprising that the Southeast region presents the highest number of cases and deaths by COVID-19 in Brazil, and were the first region to diagnoses COVID-19 cases and to step up social distancing policies to slow down the spread of coronavirus. On the other hand, the North region is the largest region of Brazil, and concentrates a large number of indigenous people and their descendants, who are part of the COVID-19 risk group, which makes the region particularly sensitive to coronavirus disease (Ferrante and Fearnside, 2020).

A primary motivation for the present study was to estimate the long-term movements for COVID-19 data in Brazil. Thus, to better understanding and discuss the results, we plotted the estimated trend, seasonal and cycle components for death cases registered in Brazil (posterior mean and 95% Bayesian credibility interval; see Fig. 3). The first case of COVID-19 in Brazil was diagnosed on February 26th, in the state of São Paulo. On March 17th, the first death in Brazil was registered in the state of Rio de Janeiro, and on March 20th the community transmission of the disease was announced by the Brazilian government. Thus, the most important result is related to the trend component, where it can be seen that since registering the first case of death by COVID-19 in Brazil, it took a steep movement upwards until the first week of April, where the death trend slowly reveals a turning point, i.e., the pace of the trend has slowed down, reaching the peak in the last week of April.

Despite the evidence of the effectiveness of the measures implemented by many countries to slow down the transmission of the COVID-19 (Kraemer et al., 2020; Gatto et al., 2020; Saez et al., 2020), the persistence in a peak plateau stage varies across countries, depending on how quickly the policies were implemented and another factors such as population density, and health systems structure (Hsiang et al., 2020; Deb et al., 2020). In particular, the persistence in the high plateau stage can be explained by the containment measures introduced late, during



(a) Trend

(b) Seasonal

(c) Cycle

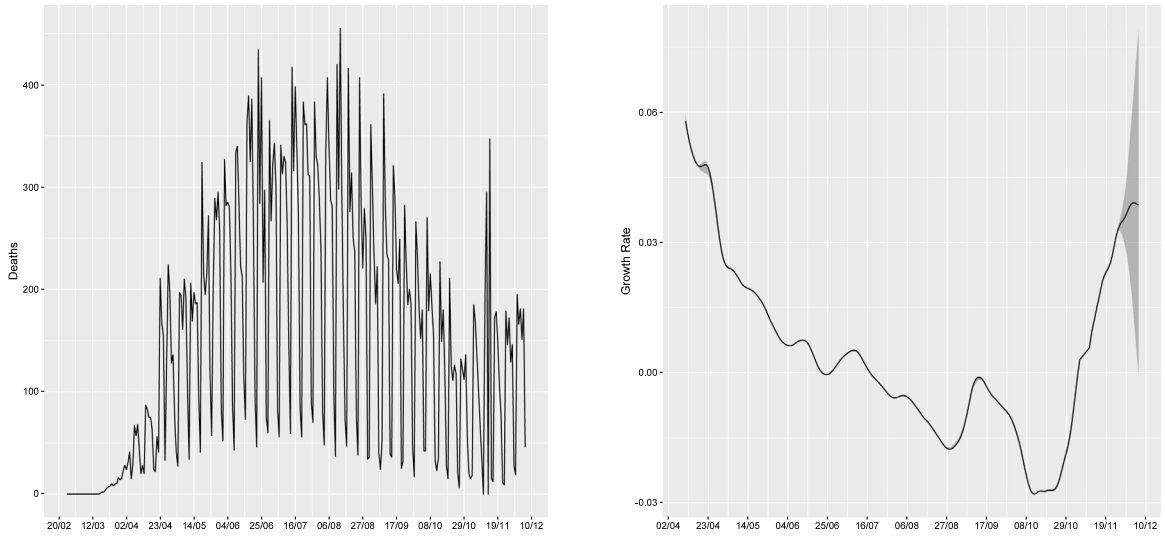
Fig. 5. Posterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in the state of São Paulo — 02/25/2020 to 12/06/2020.

the evolution of the new cases and deaths, which requires a long time to revert the death trend into a decreasing one.

In the case of Brazil, the trend component remained relatively stable in a high peak plateau from May to August, with a slight decrease just around mid-August. In particular, the sustained high plateau observed in the estimated trend component of deaths by COVID-19 in Brazil may be related to the absence of political actions at federal level and the early relaxation of the isolation measures (Canabarro et al., 2020). In addition, Fig. 4 shows the observed number of death cases by COVID-19 and the growth rate for Brazil, where it is possible to see that the days with the highest rates occurred until mid-May.

A more accelerated downward trend is observed from the end of August, with a significant reduction in the trend until the beginning of November. From this point on, we observed a further acceleration of the trend, going to the end of the analyzed sample. This acceleration was interpreted as a reflection of the accelerated reduction in social distance measures taken in various locations, and may also be related to agglomerations caused by the electoral campaign for the election of mayors and councilors in this period. This acceleration is being called a "second wave", although there was in fact no control over the pandemic in the period before this acceleration in the number of deaths. The estimated parameters for Brazil are shown in Table A.1 in Appendix, where it is possible to note a high precision associated with the trend, seasonality and cycle components.

Fig. 5 presents the posterior mean of the estimated trend, seasonality and cycle components, and the associated 95% Bayesian credibility interval for deaths by COVID-19 in the state of São Paulo, and Table A.2, in the Appendix, reports the estimated parameters. The first case of COVID-19 in the state of São Paulo was on February 26th and the first death was registered on March 18th. The state government imposed the first isolation measures between March 17th and March 22th (Decree 64.881), which established the mandatory closure of non-emergency services, as well as educational institutions. The estimated trend of death (see Fig. 5) presents a rapidly increase from the beginning of the sample until the end of March, where a turning point is slowly revealed. This change in the rising trend can be attributed to the first isolation measures adopted by the state of São Paulo authorities, as previously discussed in the literature (e.g., Cruz, 2020). However, the abrupt change in the evolution of the death trend observed in the mid-April suggests that the level of adherence of the citizens to isolation measures dropped down over time, or the measures taken by state government were not enough to cope with the outbreak. After reaching



(a) Deaths (b) Growth Rate

Fig. 6. Number of reported deaths and the estimated growth rate of the Trend in the state of São Paulo — 02/25/2020 to 12/06/2020.

the peak on mid-May, the death trend remained in a long persistence peak plateau stage. Also, Fig. 6 shows that the highest growth rate of deaths occurred around mid-April, whereas the lowest growth rate was registered around mid-August, suggesting a decreasing trend from mid-August. Similar to the pattern observed in the data from Brazil, in the state of São Paulo we observed a significant drop in the trend from June onwards, and a strong acceleration in the cases from November.

The estimated trend, seasonality and cycle components, and the associated 95% Bayesian credibility interval for the state of Amazonas are presented in Fig. 7, and Table A.3, in the Appendix, reports the estimated parameters. In the state of Amazonas, the first death by COVID-19 was recorded on March 24th, on same day that isolation measures were imposed by local authorities. Despite the North region being the last Brazilian region to register the first case of COVID-19, the state of Amazonas was one of the first to collapse the health system. Indeed, through Fig. 7 it is possible to see that the death trend in the state of Amazonas reached the peak faster than the state of São Paulo, where the social measures were able at least in part to flatten the epidemic curve and postpone the peak of death cases. This result may be explained by some reasons. First, the state capital, Manaus, concentrates all the intensive care units able to deal with the coronavirus disease in the state. With few roads and transport mainly by rivers, many patients were not able to access proper health care, especially indigenous people, who are particularly sensitive to COVID-19 and usually live in isolated areas, far from doctors and access to medicines. In addition, the state of Amazonas presents a high social vulnerability index, and one of the worst rates of hospital beds per population over country (Guerra-Shinohara et al., 2020). Since the beginning of June, the state capital, and most populous city in the state, established some relaxation strategies in the containment measures, including reopening schools, which could have led to the observed increasing pattern in the estimated trend component in the end of July (Ferrante et al., 2020). A second local peak is observed in August, followed by a reduction in the trend until mid-October, and thereafter a further acceleration that persists until the end of October, and after that period a further reduction in the trend, showing a behavior different from the pattern observed for Brazil.

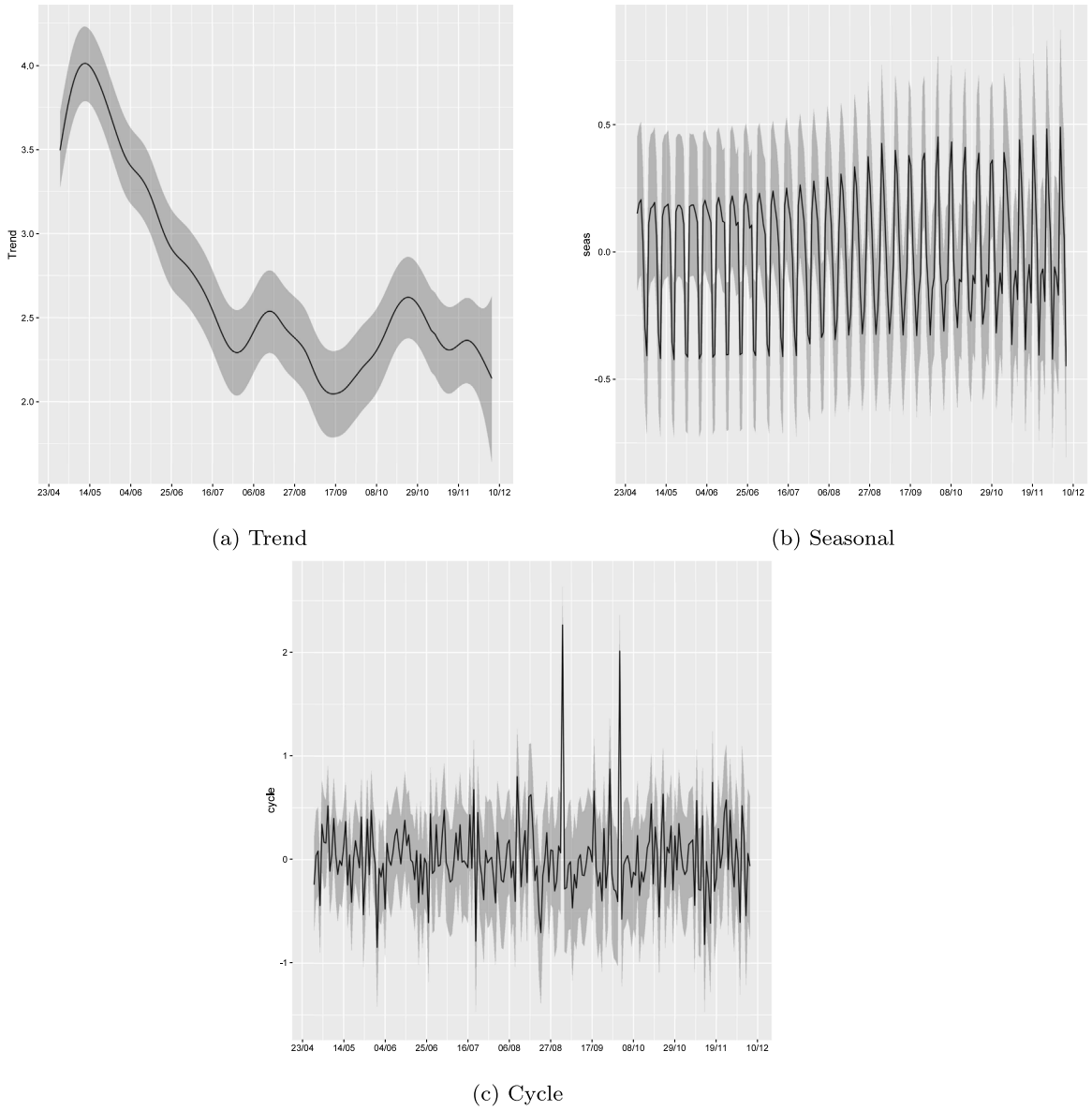
The state of Amazonas is particularly interesting as it shows the difficulty in controlling the COVID epidemic. According to the estimates presented in Buss et al. (2020), about 76% of the population of Manaus, the capital of Amazonas, had already been infected until October 2020, and even so the epidemic had not been fully controlled, showing that the proportion number of infected people required the existence of a herd immunity is quite high. The general trend of declines observed since November may indicate the start of herd immunity, since the trend contrasts with the acceleration pattern in the rest of Brazil. In addition, in Fig. 8, it is possible to observe the great heterogeneity in the rates of infection growth in this state, showing the complexity in the transmission patterns of COVID-19.

In order to show the model's ability to fit the death cases, we plotted the observed number of deaths by COVID-19 and the predicted value of death count in each day given by the sum of the estimated trend, seasonal, and cycle components, along with the 95% credibility interval of this sum (see Fig. 9), for Brazil and the states of São Paulo and Amazonas. The results visually suggest that the model has a good fit.

3.2. Comparison between moving averages approach and the proposed model

We compared the performance of our proposed model in monitoring the long-term movements of mortality of COVID-19 with the so-called moving average filter. The importance of the results interpretation is evident, and this is why we need to properly highlight an important point about the estimated values for the trend component. It is worth noting that the moving average filter is a way to extract a trend from time series count data. On the other hand, in our model the trend component can be interpreted as the mean of the log-transformed count data or the mean of the log intensity. Due to the non-linearity of our model, the anti-log (exponential) transformation is not suitable to obtain the original data. Therefore, in order to provide comparable results, we plotted (see Fig. 10) the moving average of the log-transformed data (green line) along with the estimated trend component (red line; posterior mean and 95% Bayesian credibility interval) and the observed log-transformed count data (blue line).

As discussed above, one of the most advantages of our proposed model is the ability to extract a smoother trend component with reduced residual autocorrelations due to the inclusion of the seasonal and cycle components. Based on the comparison between the moving average filter and the estimated trend component from our proposed model, it is possible to see why is important to add the seasonality and cycle components in estimating a trend that really reflects the permanent patterns in COVID-related deaths. The cycle component, which corresponds to a second-order autoregressive structure, is a parsimonious way of incorporating the transient effects induced by the



(a) Trend

(b) Seasonal

(c) Cycle

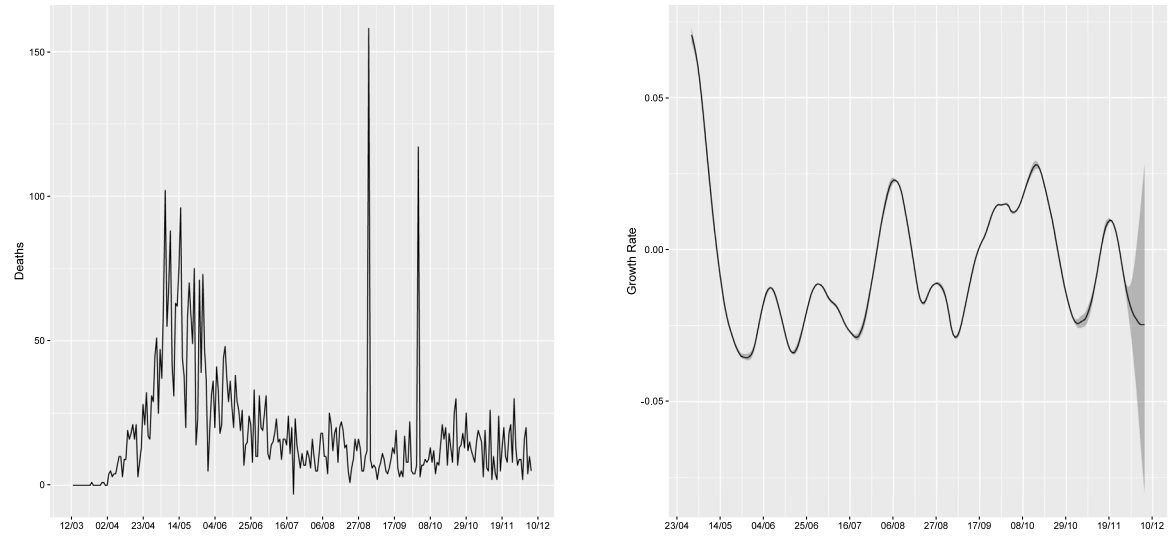
Fig. 7. osterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in the state of Amazonas — 02/25/2020 to 12/06/2020.

contamination mechanism generated by the case accounting structure. Also, it is useful to capture local, non-persistent patterns in the data, isolating long-term effects. These two components allow a more robust estimation of the trend component, avoiding the greatest fluctuation observed in the trend estimation given by the moving average.

Additionally, it is important to note that the cycle and seasonal components play important roles by adding additional robustness to days with large atypical patterns in the reported data, which normally correspond to the date when adjustments are made for unreported deaths in past periods. For example, looking at Fig. 8 containing deaths for the state of Amazonas, we can see two dates with aberrant numbers of deaths (09/02/2020 with 158 deaths and 01/10/2020 with 117 deaths). These dates correspond to adjustments where all unreported deaths in the previous months were added, and thus are observations from much earlier periods that do not correspond to current death patterns. The model captures these atypical days through the cycle component, without changing the trend estimation, as opposed to what happens with the moving average estimator, which is quite sensitive to this form of contamination, as seen in subfigure (c) of Fig. 10. This highlights the robustness properties of the model proposed in this work to recover the trend in death patterns related to COVID-19.

3.3. Spatio-temporal analysis

For the spatio-temporal analysis we used the same data source and the same sample period as Section 3.1, using a state level aggregation of the death data related to COVID-19. The results of the estimation of this model are shown in Table A.4 in the Appendix, and the estimated trend, seasonality and cycle components are shown in Fig. 11. Note that the components are now estimated with an adjustment for the size of the population in each region, which corresponds to exposure in the Poisson process. To recover the specific effect of the component for each state, it is necessary to multiply by the size of the population. We can observe that the spatio-temporal model allows to recover a smoother trend component compared to those obtained in univariate models, indicating that the spatial component captures part of the irregularity observed in the data on death by COVID-19. The trend estimated by this model indicate a general peak about the end of July, and also indicates the existence of a new acceleration (the “second wave”) from the second week of November. The seasonality and cycle components are consistent with the patterns observed in the aggregated data for the whole country.



(a) Deaths (b) Growth Rate

Fig. 8. Number of reported deaths and the estimated growth rate of the Trend in the state of Amazonas — 02/25/2020 to 12/06/2020.

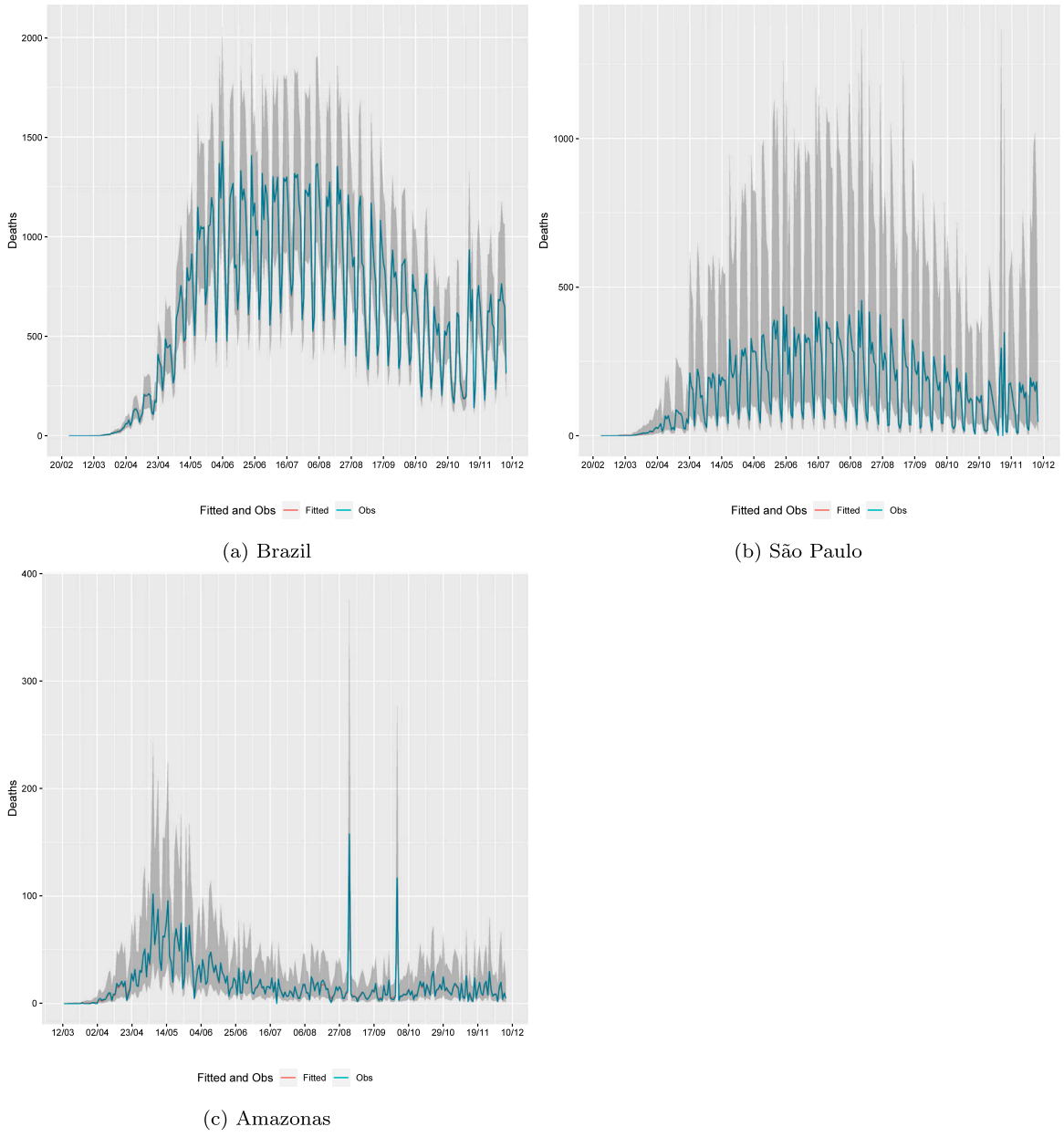
Table 1
Estimated parameters of deaths reported in Brazil - Spatio-temporal model with region specific trend, seasonal and cycle components.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend southeast	1.68e4	1257.446	1.46e4	1.67e4	1.95e4	1.65e4
Precision for trend south	4.94e3	336.875	4.29e3	4.94e3	5.61e3	4.95e3
Precision for trend north	1.65e4	1186.588	1.41e4	1.65e4	1.88e4	1.66e4
Precision for trend northeast	4.80e4	4829.782	4.10e4	4.72e4	5.95e4	4.46e4
Precision for trend center-west	3.05e4	3986.805	2.45e4	2.98e4	3.99e4	2.80e4
Precision for seas southeast	2.43e2	18.723	2.09e2	2.42e2	2.82e2	2.39e2
Precision for seas south	2.28e3	163.394	1.96e3	2.28e3	2.60e3	2.29e3
Precision for seas north	1.11e4	1284.404	8.37e3	1.12e4	1.32e4	1.18e4
Precision for seas northeast	1.27e5	9207.029	1.12e5	1.26e5	1.48e5	1.24e5
Precision for seas center-west	5.62e3	421.451	4.92e3	5.58e3	6.56e3	5.45e3
Precision for cycle southeast	3.21e1	2.927	2.65e1	3.20e1	3.80e1	3.20e1
PACF1 for cycle southeast	1.47e−1	0.037	7.30e−2	1.47e−1	2.20e−1	1.47e−1
PACF2 for cycle southeast	−1.24e−1	0.034	−1.92e−1	−1.23e−1	−6.00e−2	−1.20e−1
Precision for cycle south	3.45e1	2.481	3.02e1	3.43e1	3.99e1	3.36e1
PACF1 for cycle south	1.16e−1	0.034	5.10e−2	1.16e−1	1.84e−1	1.13e−1
PACF2 for cycle south	4.50e−2	0.036	−2.00e−2	4.30e−2	1.19e−1	3.60e−2
Precision for cycle north	1.54e1	1.449	1.31e1	1.52e1	1.88e1	1.47e1
PACF1 for cycle north	4.83e−1	0.034	4.23e−1	4.80e−1	5.55e−1	4.69e−1
PACF2 for cycle north	−3.30e−1	0.030	−3.91e−1	−3.29e−1	−2.72e−1	−3.25e−1
Precision for cycle northeast	4.07e2	28.659	3.57e2	4.05e2	4.69e2	3.99e2
PACF1 for cycle northeast	7.49e−1	0.027	7.06e−1	7.45e−1	8.07e−1	7.31e−1
PACF2 for cycle northeast	−2.70e−2	0.098	−2.32e−1	−2.00e−2	1.51e−1	5.00e−3
Precision for cycle center-west	2.37e1	2.481	1.99e1	2.33e1	2.95e1	2.22e1
PACF1 for cycle center-west	6.10e−2	0.042	−2.90e−2	6.40e−2	1.34e−1	7.60e−2
PACF2 for cycle center-west	−2.41e−1	0.035	−3.04e−1	−2.43e−1	−1.69e−1	−2.50e−1
Precision for CAR	4.81e−1	0.013	4.57e−1	4.81e−1	5.08e−1	4.79e−1
Group Φ	8.00e−1	0.006	7.88e−1	8.00e−1	8.12e−1	8.00e−1
Deviance Information Criterion (DIC)	40003.80					
Watanabe–Akaike information criterion (WAIC)	40265.38					

Although the spatio-temporal model presented above allows a useful recovery of the general patterns in the dynamics of deaths related to COVID-19 in Brazil, the assumption of common components for all Brazilian states is a very restricted assumption. The Brazilian regions have different dates of onset of contagion, and also very different dynamics over time. To avoid these restrictions, and still maintain a relatively parsimonious model, we performed a new estimation of the spatial model allowing individual components of trend, seasonality and cycle for each region of the country, that is, components for the North, Northeast, Midwest regions, Southeast and South, using the regional division of the country used by *Instituto Brasileiro de Geografia e Estatística* (IBGE).

The parameters estimated in this version with individual components for the Brazilian regions are shown in Table 1. Figs. 12, 14 and 15 show the components of trend, seasonality and estimated cycles for each region of Brazil in this model. In Fig. A.1 in the Appendix we also show the posterior mean of the spatial effects for two different dates in the sample, illustrating the evolution of the spatial dynamics in deaths related to COVID. Again, the spatio-temporal model makes it possible to recover trend components with greater smoothness than univariate models, indicating the importance of the process of spatial dependence in the dynamics of COVID-19. The regional components of trend also seem to confirm the patterns observed for the number of deaths in Brazil. There seems to be a reduction in the trend of deaths for the Southeast, South and Northeast regions from the end of July, and also a new acceleration from November.

As the latent components in this formulation are formulated in terms of exposure, that is, the number of inhabitants in each region,



(a) Brazil

(b) São Paulo

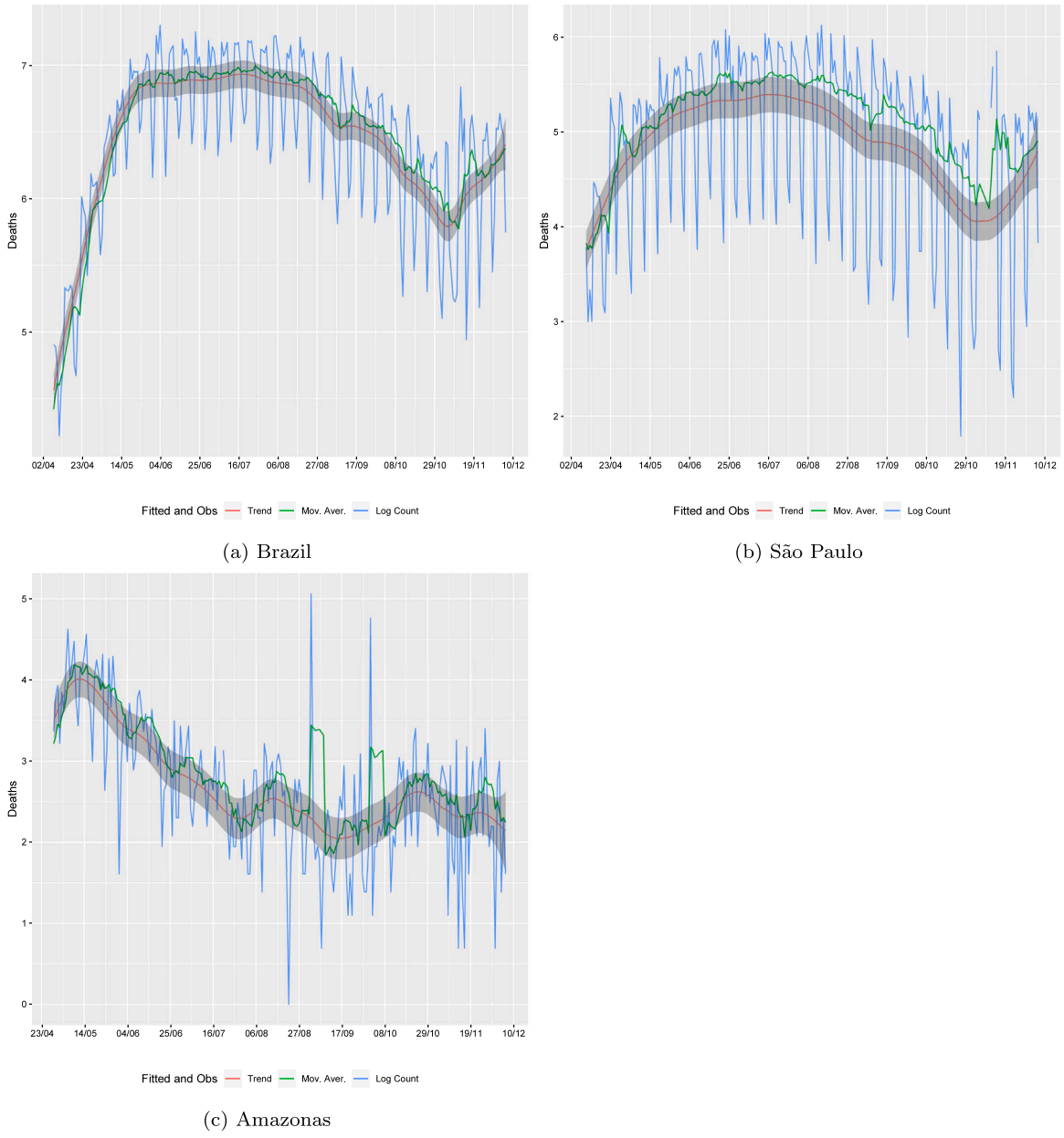
(c) Amazonas

Fig. 9. Predicted values and observed deaths - Brazil, São Paulo and Amazonas — 02/25/2020 to 12/06/2020.

it is not so easy to directly interpret the estimated trend value. For a more direct interpretation, we carried out a transformation of the trend towards the number of daily deaths per million inhabitants, that is, assuming an exposure of 1,000,000 for each trend. The trend in daily deaths per million inhabitants per region is shown in Fig. 13. We can observe the rapid peak that was reached in the North, which reflects the high number of deaths observed in Amazonas in May, and then the rapid reduction and the new acceleration observed at the end of the analyzed sample. Another notable pattern is the high rate of deaths in the Center-West region, which reflects the number and acceleration of cases and deaths in the Federal District, a region with a higher population density. For the Southeast, South and Northeast regions, the pattern is more similar, indicating a acceleration in the death rate at the end of the analyzed sample.

The seasonal components (Fig. 14) are quite different across regions, which is probably reflecting different patterns of occurrence and also methods of accounting for cases and deaths. The cycle (Fig. 15) components show a little present cycle at the beginning of the sample, which is expected due to a limited number of cases, but a greater relative importance at the end of the sample, after the acceleration in the cases. We can also analyze these components by the estimated autoregressive coefficients. In Table 1 we have the first and second order partial correlation coefficients for each region, and we can observe a variety of patterns in these estimated coefficients. The second-order autoregressive component seems to have a clear interpretation of the cyclical component for the Southeast, North and Northeast regions, while in terms of magnitude it seems less important for the South and Midwest regions.

We can compare the models with common components and regional components using information criteria. In Tables A.4 and 1 we report two common information criteria for Bayesian estimates. Deviance Information Criterion (DIC) and Widely Applicable Information Criterion (WAIC), which is a generalized version of AIC. By the two criteria the model with regional components is selected as the most suitable model. This result indicates the existence of a relevant heterogeneity in COVID-19 standards in Brazil.



(a) Brazil

(b) São Paulo

(c) Amazonas

Fig. 10. Predicted Values - Estimated trend and moving average filter for the log-transformed count data — 02/25/2020 to 12/06/2020.

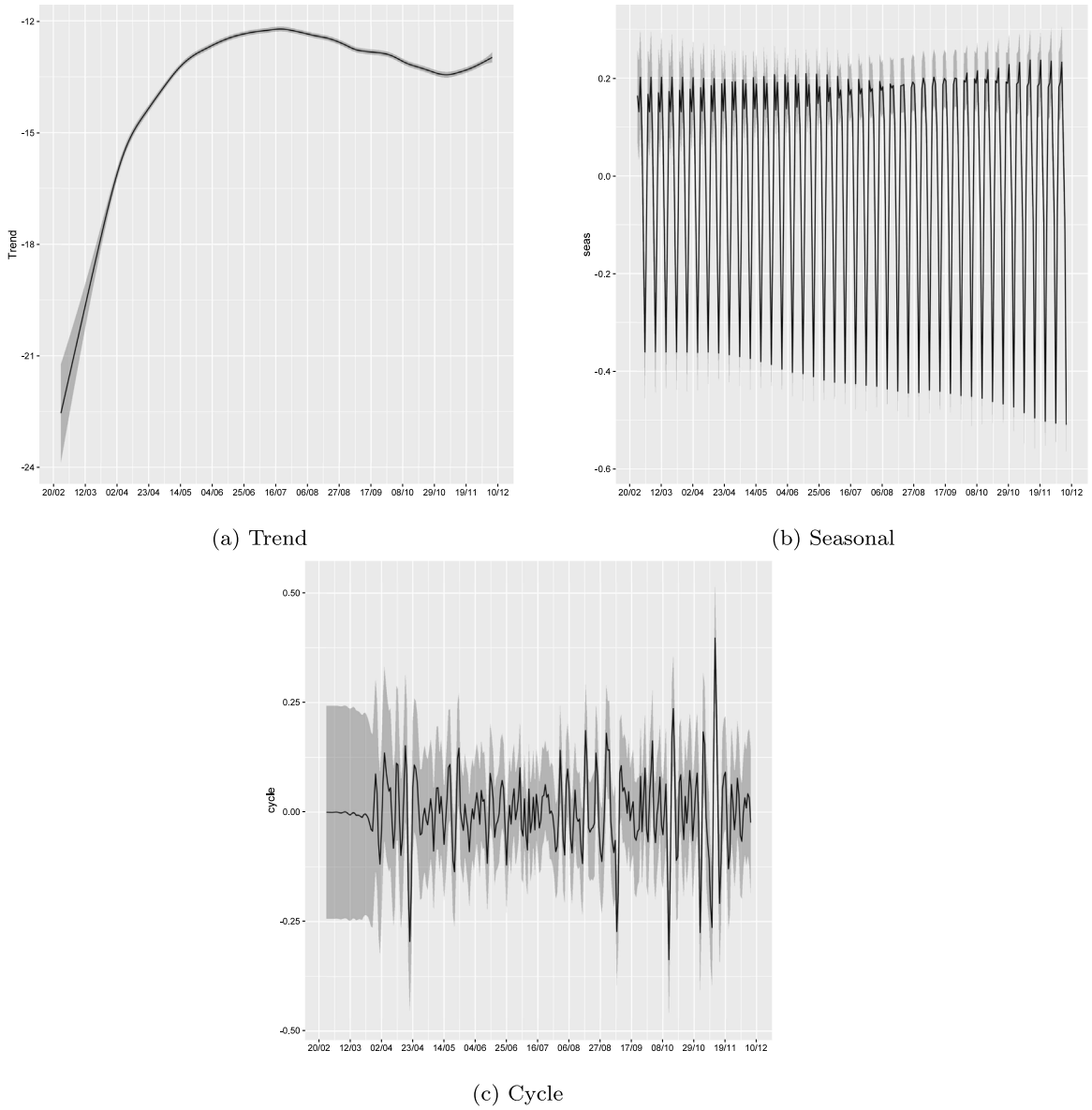
We compared some model fit measures for all models analyzed in this work, which allows to analyze especially the importance of the spatial component in the parameterization of the models, as well as the effects in the adoption of specific components of trend, seasonality and cycle for each region. Table 2 shows the mean error (ME), root mean squared error (RMSE) and the mean absolute error (MAE) of univariate models and space–time models, using the posterior median as the point measure of model fit. We also built the specific adjustment measures for the states of São Paulo and Amazonas, obtained by the models with spatial components, for comparison with the adjustment of the time series models.

An important first result is that the fit of spatio-temporal models is clearly superior to that of univariate models of time series in all metrics. Note that this gain reflects the importance of inter-state transmission patterns in the COVID-19 dynamics. Regarding the inclusion of specific components for regions, we can see that in general these specific components lead to a general reduction in the mean error term, but are slightly worse in terms of RMSE and MAE for some regions, which indicates a possible overparameterization of the model and the presence of possible common trends.

4. Conclusions

The methods proposed in this work are based on hierarchical formulations using Bayesian inference methods for time series and spatio-temporal processes. In particular, we proposed methods to estimate the trend in the deaths by COVID-19, through trend-cycle decomposition for counting processes with latent components, which were applied for the total number of deaths in Brazil and for the states of São Paulo and Amazonas. In addition, the estimated trend component was compared to the empirical analysis based on moving averages. In order to achieve the main goal, we used the BRASIL.IO daily data series on deaths by COVID-19 in Brazil and in the states of São Paulo and Amazonas, from 02/25/2020 to 12/06/2020.

The use of Bayesian inference methods is especially useful in situations such as the spreading a new epidemic with transmission, latency



(a) Trend

(b) Seasonal

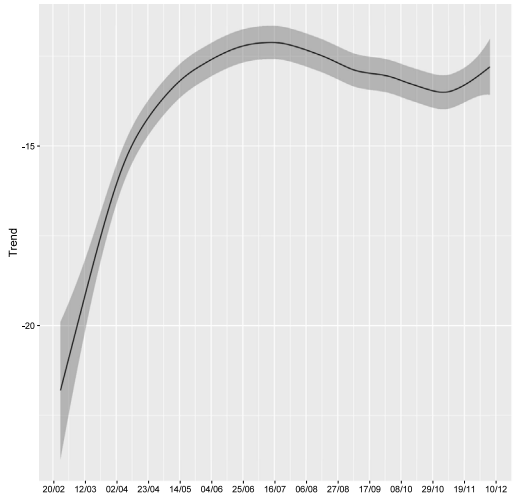
(c) Cycle

Fig. 11. Posterior mean and 95% credibility interval of Trend, Seasonal and Cycle decomposition - Spatio-Temporal model with common trends, seasonal and cycle components — 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

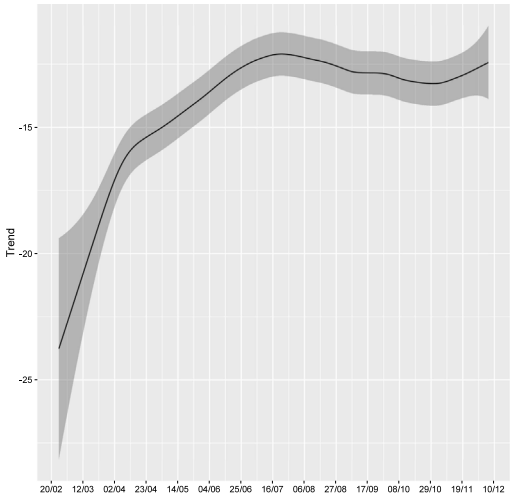
and mortality characteristics that are not well known. The hierarchical formulation reflects the process of information accumulation in time and space, where the posterior distribution in a period t serves as a prior for the next period, and also the use of information from neighbors as a prior structure to make inference about the parameters in a certain region. In a situation where there is little prior information on essential aspects of the problem, the incorporation of learning in time and space through a Bayesian mechanism allows for the efficient use of the new information available, which is essential for the emergency formulation of health and prevention policies, in a general environment of information of dubious quality or propagation of false information or without a scientific basis. Bayesian learning is essential in this context as a basis for evidence and data-based policies.

The results for Brazil showed that our proposed models were able to capture the long-term movements in death cases, showing an accelerated growth pattern until early April, where the pace of the trend has slowed down, reaching the peak in July, suggesting that the isolation measures taken by authorities might have been partly effective to change the growth of the death trend and postpone the peak. However, the observed sustainable high peak between May and August provided evidence that the social distancing measures might have not been enough to slow down the spread of the COVID-19, which can be related to the absence of political actions at federal level. The model also captured a negative trend in the number of new deaths until the beginning of November 2020, and a possible "second wave" after this period, with a new accelerated growth in the trend, which may be related to the electoral period and an excessive relaxation of social distance measures by local authorities. The results of the estimated trend component for the states of São Paulo are similar to those observed for Brazil. For the state of Amazonas, the estimated trend reached the peak faster than the previous analyzed states, which may be due to the lack of health care infrastructure and the difficult of some patients to access proper heath care, concentrated in the state capital, Manaus.

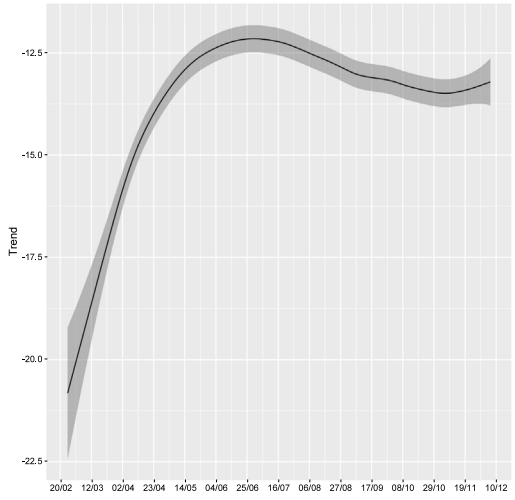
Additionally, the results obtained indicate that the trend component estimated by the univariate and spatio-temporal models is a more robust indication of the general patterns in the occurrence of deaths related to COVID-19, overcoming the existing limitations in simple smoothing measures such as the use of moving averages, which are not robust to the measurement errors introduced by the case accounting



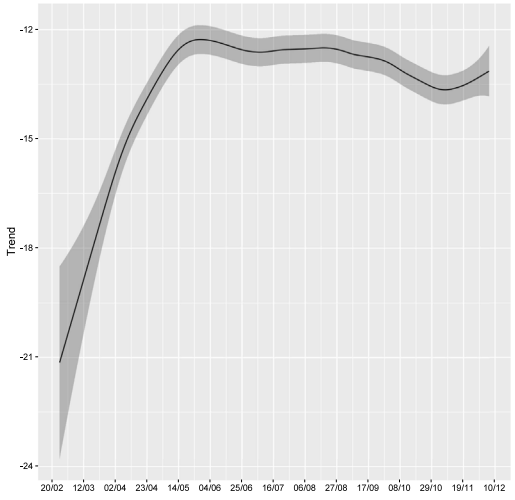
(a) Trend Southeast



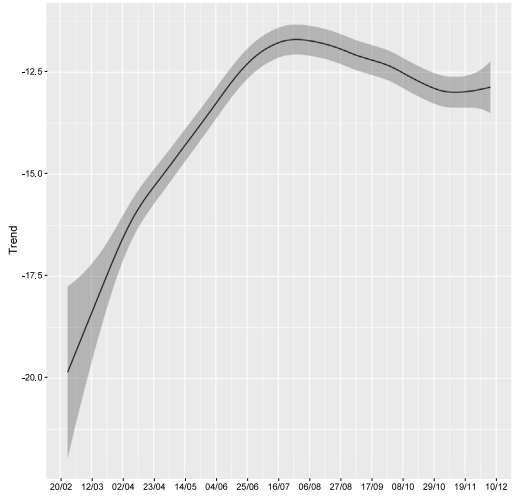
(b) Trend South



(c) Trend Northeast



(d) Trend North



(e) Trend Center-West

Fig. 12. Posterior mean and 95% credibility interval of Trend - Spatio-temporal model with region specific trends, seasonal and cycle components — 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

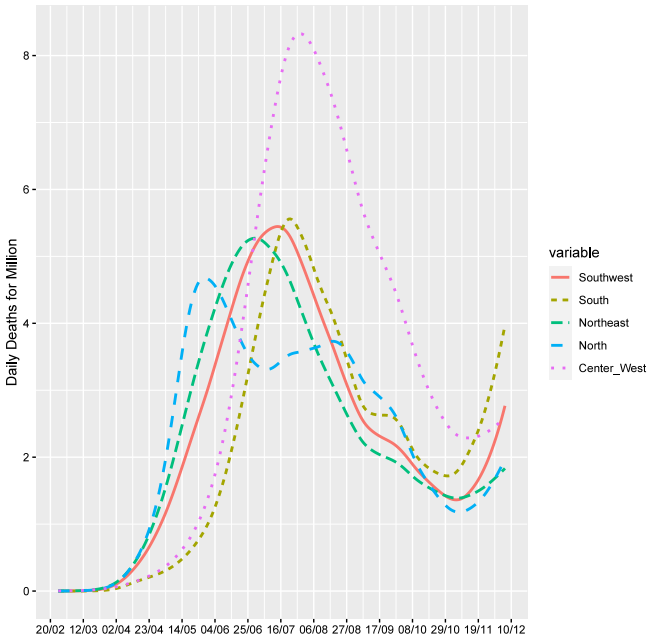


Fig. 13. Region-specific Trends in Deaths by day by Million inhabitants — 02/25/2020 to 12/06/2020.

Table 2
Model fit statistics.

	ME	RMSE	MAE
Univariate Models			
Brazil	−0.5236	5.7872	4.0368
SP	−0.6094	1.9461	1.4112
AM	−0.2586	1.7920	1.2175
Spatio-temporal Model Common Components			
Brazil	−1.18e−6	2.3204	1.4613
Southeast	−0.0004	3.5379	2.3078
South	0.0124	1.5199	1.0590
Northeast	0.0008	2.0400	1.3224
North	0.0026	1.8968	1.2172
Center-West	−0.0152	2.4802	1.6213
SP	0.0040	2.2366	1.4146
AM	0.0044	2.2926	1.4384
Spatio-temporal Region-Specific Components			
Brazil	−9.79e−7	2.4051	1.5297
Southeast	1.48e−7	3.5985	2.4151
South	−2.03e−6	1.5589	1.1067
Northeast	−1.41e−6	2.2406	1.4285
North	−6.09e−7	2.0050	1.2895
Center-West	−1.00e−6	2.3820	1.5754
SP	0.0036	2.3341	1.4881
AM	0.0047	2.3669	1.5019

mechanisms. The models also obtained a very precise adjustment for the number of occurrences, especially in spatio-temporal models with the incorporation of spatial propagation patterns. These models also provided a smoother component of trend, when compared to univariate models, which can be explained by the greater information available in the estimation process, using all the information available to Brazilian states, which helps to mitigate the measurement error problem.

We believe that our work contributes to the enormous effort of studying and analyzing the impacts of COVID-19 in Brazil, complementing several other initiatives of epidemiological, statistical and computational modeling, such as projects Brasil.IO (https://brasil.io/dataset/covid19/caso_full/), Covid 19 Analytics (https://covid19analytics.com.br/), COVID-19 Brasil (https://ciis.fmrp.usp.br/covid19/), MonitoraCovid-19 (https://bigdata-covid19.icict.fiocruz.br/) and several other initiatives of great value for Brazilian society.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A.1. Estimated posterior distribution of parameters

See Tables A.1–A.4.

A.2. Spatial random effects

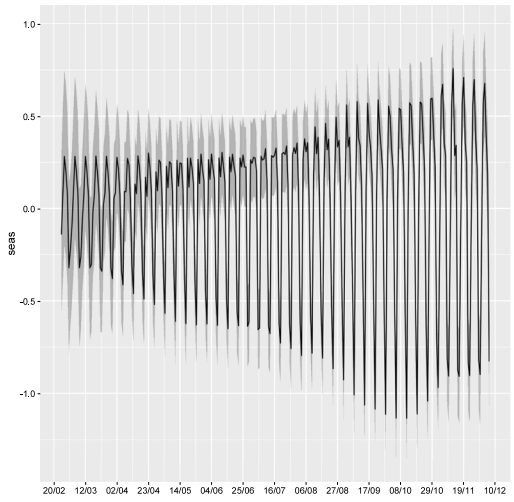
As an illustration of the dynamics of spatial random effects, we show in Fig. A.1 the effects estimated by the model with the regions specific components for the days 04/26/2020 and 08/14/2020. The variation observed in the two dates shows the importance of using a dynamic structure for the spatial effects in the modeling of epidemic processes.

A.3. SEIR model with measurement error

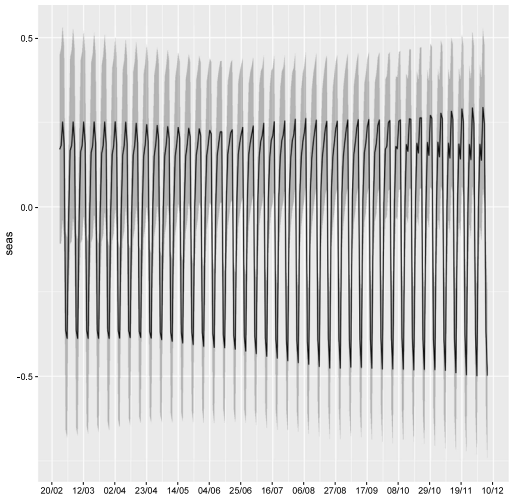
We use a (Susceptible–Exposed-Infective–Resistant) SEIR model to generate realizations of an epidemic model calibrated to reproduce some essential aspects of the COVID-19 epidemic. We use the following structure, based on Liu et al. (2020), to generate a solution for the SEIR model:

$$\begin{aligned} dS(t)/dt &= -\beta I(t)S(t)/N \\ dE(t)/dt &= \beta S(t)I(t)/N - \delta(t)E(t-7) \\ dI(t)/dt &= \delta(t)E(t-7) - \gamma(t)I(t-10) - \eta(t)I(t-10) \\ dR(t)/dt &= \gamma(t)I(t-10) \\ dF(t)/dt &= \eta(t)I(t-10) \end{aligned} \tag{3}$$

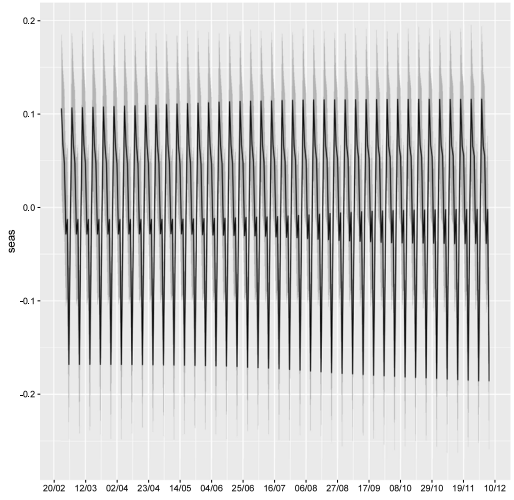
$S(t)$ is the number of susceptible individuals in the population in period t, $E(t)$ is the number of exposed individuals, $I(t)$ the number of infected, $R(t)$ the resistant individuals and $F(t)$ the number of fatalities. The total number of individuals is N, given by the sum of the individuals in all states. β is a parameter which controls the average number of



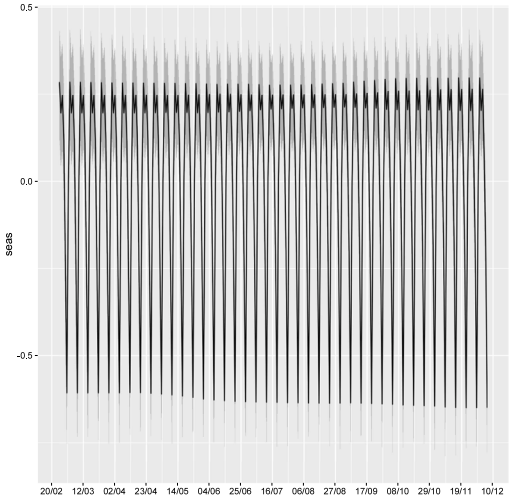
(a) Seasonal Southeast



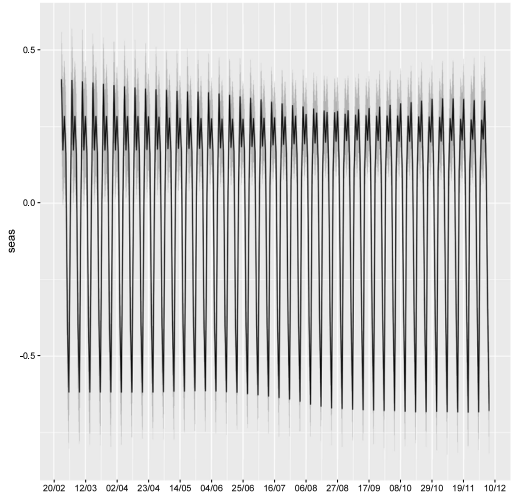
(b) Seasonal South



(c) Seasonal Northeast

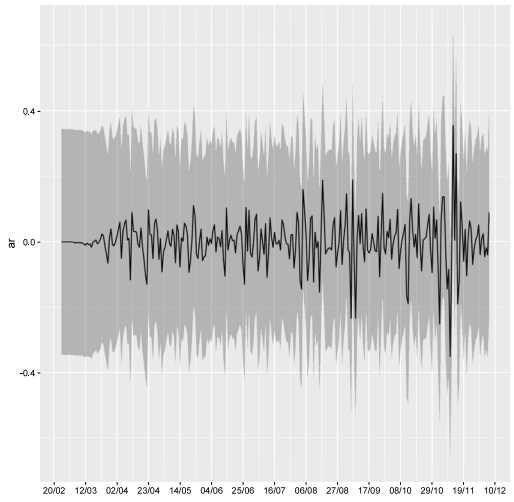


(d) Seasonal North

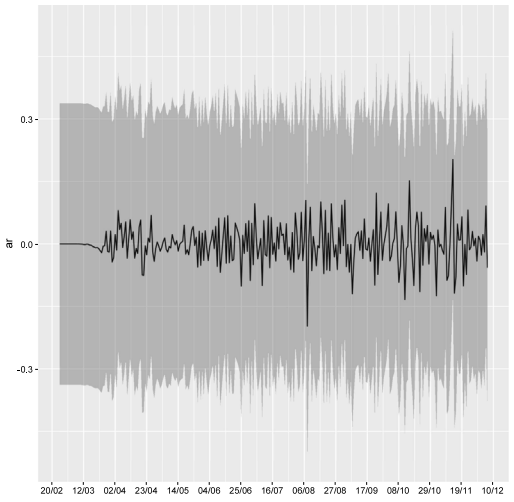


(e) Seasonal Center-West

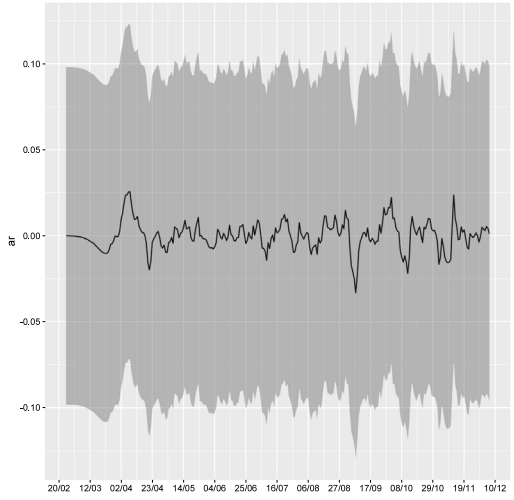
Fig. 14. Posterior mean and 95% credibility interval of Seasonal - Spatio-temporal model with region specific trends, seasonal and cycle components — 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).



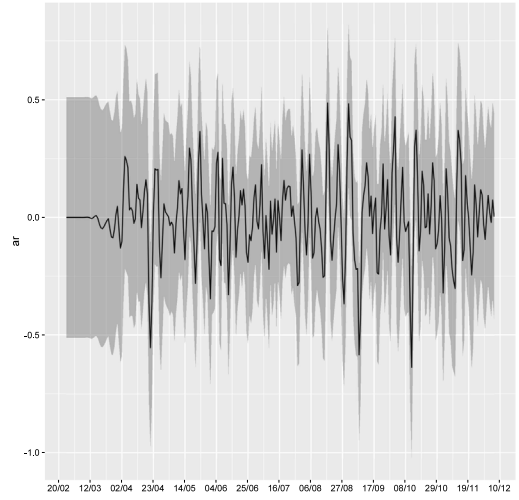
(a) Cycle Southeast



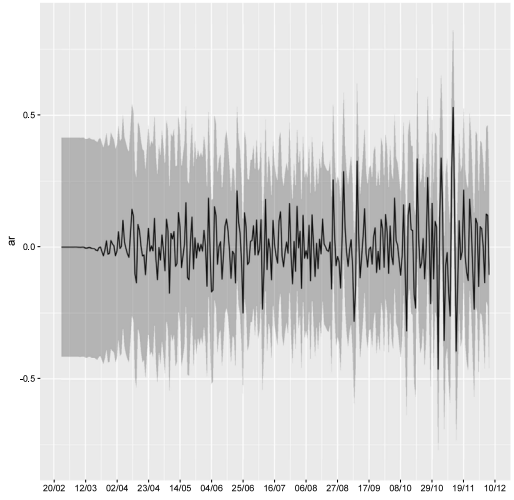
(b) Cycle South



(c) Cycle Northeast



(d) Cycle North



(e) Cycle Center-West

Fig. 15. Posterior mean and 95% credibility interval of Cycle - Spatio-temporal model with region specific trends, seasonal and cycle components — 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

Table A.1
Estimated parameters of deaths reported in Brazil.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	10147.060	3996.089	4447.125	9454.153	19872.796	8201.470
Precision for seasonality	2017.824	803.563	881.916	1874.032	3987.542	1618.580
Precision for cycle	38.444	4.970	29.407	38.212	48.877	37.840
PACF1 for cycle	0.290	0.067	0.156	0.291	0.421	0.290
PACF2 for cycle	−0.204	0.078	−0.351	−0.206	−0.047	−0.210

Table A.2
Estimated parameters of deaths reported in the state of São Paulo.

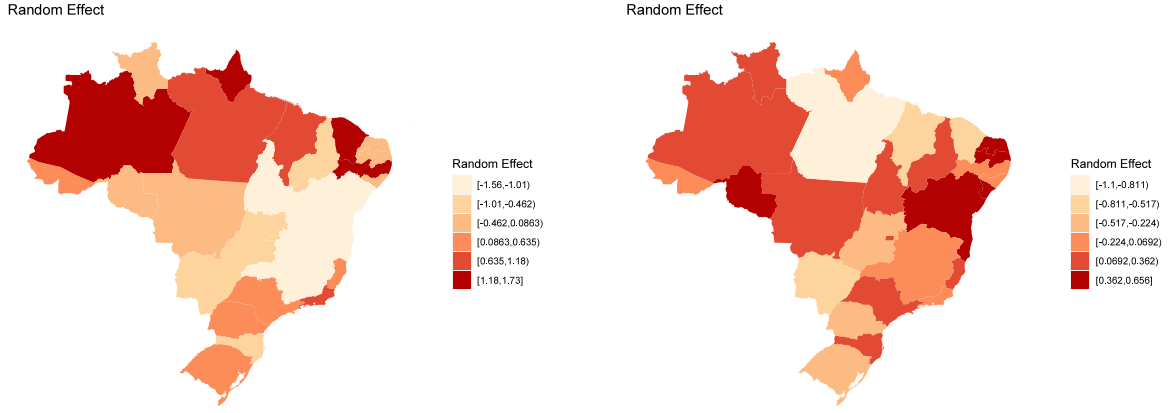
	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	10959.712	4710.471	4245.146	10164.461	2.24e4	8622.798
Precision for seasonality	6.625	30.858	15.651	37.878	1.28e2	27.062
Precision for cycle	5.058	1.085	3.119	5.007	7.350	4.933
PACF1 for cycle	−0.002	0.086	−0.170	−0.003	1.67e−1	−0.004
PACF2 for cycle	−0.126	0.085	−0.292	−0.127	4.20e−2	−0.126

Table A.3
Estimated parameters of deaths reported in the state of Amazonas.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	18777.970	2.28e4	3222.621	11039.803	8.43e4	5523.388
Precision for seasonality	590.342	5.47e2	150.322	421.364	2.04e3	259.600
Precision for cycle	4.812	6.52e−1	3.655	4.771	6.210	4.692
PACF1 for cycle	−0.069	1.05e−1	−0.269	−0.071	1.41e−1	−0.076
PACF2 for cycle	0.020	8.70e−2	−0.146	0.018	1.94e−1	0.010

Table A.4
Estimated parameters of deaths reported in Brazil - spatio-temporal model with common trend, seasonal and cycle components.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	30562.848	8148.802	16735.991	29929.161	48350.874	28689.914
Precision for seasonality	12034.725	3292.864	7366.350	11413.624	20093.088	10209.154
Precision for cycle	67.156	9.515	51.093	66.207	88.328	64.095
PACF1 for cycle	0.4308	0.052	0.334	0.439	0.537	0.440
PACF2 for cycle	−0.559	0.057	−0.657	−0.563	−0.436	−0.575
Precision for CAR	0.439	0.019	0.405	0.437	0.481	0.431
Group Φ	0.785	0.010	0.763	0.785	0.803	0.787
Deviance Information Criterion (DIC)	40015.16					
Watanabe–Akaike information criterion (WAIC)	40188.55					



(a) 26/04/2020 (b) 26/04/2020

Fig. A.1. Posterior mean of Spatial Random Effects - Spatio-temporal model with region specific trends, seasonal and cycle components — 26/04/2020 and 26/04/2020.

exposed cases that are generated by one infected person. The parameter δ is the probability of an exposed individual migrate to the infected state. γ is the parameter for the probability of transition of a infected individual to the recovered state, and η the probability of the infected individual to migrate to fatality state. We assume the values 1.3629, 0.0262, 0.01, 0.004 for $\beta, \gamma, \delta, \eta$, and a total population size of 10 million of inhabitants. The model assumes a period of 7 days between exposure and infection, and 10 days between recovery or fatality after infection. The model is simulated for 180 periods, and we focus on the number of fatalities generated by the model. As commented in the main text, we generated a measurement error process assuming that fatalities on Saturday and Monday has a proportion, drawn from a uniform

distribution with parameters (.5, .8), of data with delayed disclosure, on the following Monday and Tuesday. We generated 10,000 replications of the measurement error process, and calculated the ACF and PACF functions for the difference between the true number of fatalities and the reported daily number of fatalities, which is the seasonal effect induced by the measurement error process. For the simulations of the SEIR model we use the deSolve package from the r-project.org software. In particular, we use the dede function, which is a solver for differential equations with delay.

A.4. Integrated nested Laplace approximations

The INLA method proposed by Rue et al. (2009) is a methodology based on deterministic Laplace approximations to perform accurate and efficient approximations on the class of Bayesian hierarchical models that can be represented as Gaussian Markov random fields (GMRF). See Rue and Held (2005) for a detailed discussion on this class of models.

A latent GMRF model is a hierarchical model with the first stage/level defining a conditional distribution for the observed variable y, usually assumed to be conditionally independent given the latent factors x and some additional (hyper)-parameter θ, in the form:

$$\pi(y|x,\theta) = \prod_j \pi(y_j|x_j,\theta), \quad j \in J \tag{4}$$

with y_j for $j \in J$ observed values and J a subset of the latent factors; and $\pi(y|x,\theta)$ defining the likelihood function of observed variables. The latent (hyper)-parameters constitutes the second stage in the hierarchical formulation:

$$x_i = \text{Offset}_i + \sum_{k=0}^{\eta_f - 1} \omega_{ki} f_k(c_{ki}) + z_i^T \beta + \epsilon_i, \quad i = 0, \ldots, \eta_x - 1 \tag{5}$$

The offset term is a prior known component to be included in the linear prediction; for example, in the Poisson likelihoods the offset is the exposure effect. ω_k are known weights for each observed data point in the sample, and $f_k(c_{ki})$ represents the effect of covariates with value c_{ki} for each observation i; β are the regression parameters of linear covariates z_i. Finally, the third and last stage of the model consists of the prior distribution for the hyperparameters θ.

The INLA approach obtains accurate approximations using sequential Laplace approximations in the mode of the posterior distributions of the latent factors, written as:

$$\pi(x_i|Y) = \int \pi(x_i|\theta,Y)\pi(\Theta,Y)d\theta \tag{6}$$

and for the marginal posterior distribution of (hyper)parameters:

$$\pi(\theta_j|Y) = \int \pi(\theta|Y)d\theta_{-j} \tag{7}$$

The element θ_{-j} denotes the vector θ with its jth element omitted. The INLA method is realized in three main steps. The first is an approximation to the full posterior distribution $\pi(\theta|y)$ by a Laplace approximation in the mode of the distribution, where the mode is found using a numerical optimization algorithm. The second step is an approximation to the full conditional distributions $\pi(x_i|\theta,y)$ for specific values of θ. The last step of the approximation gets an approximation for the marginal posterior distributions in (6) and (7) by combining the two approximations in the previous steps and integrating out the irrelevant factors. This method was introduced in Rue et al. (2009), and extended to several classes of models. Surveys of recent developments in this methodology for spatial modeling can be found at Bakka et al. (2018), a discussion on the use on log Gaussian Cox processes for discrete domains in Illian et al. (2012), and textbook references in Blangiardo and Cameletti (2015) and Gomez-Rubio (2020).

References

Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T.S., 2012. A review of some modern approaches to the problem of trend extraction. Econometric Rev. 31 (6), 593–624.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F., 2018. Spatial modeling with R-INLA: A review. WIREs Comput. Stat. 10 (6), e1443.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (2), 192–236.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Statist. Math. (43), 1–59.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Models with R-INLA. Wiley.

Buss, L.F., Prete, C.A., Abrahim, C.M., Mendrone, A., Salomon, T., de Almeida-Neto, C., França, R.F., Belotti, M.C., Carvalho, M.P., Costa, A.G., Crispim, M.A., Ferreira, S.C., Fraiji, N.A., Gurzenda, S., Whittaker, C., Kamaura, L.T., Takecian, P.L., da Silva Peixoto, P., Oikawa, M.K., Nishiya, A.S., Rocha, V., Salles, N.A., de Souza Santos, A.A., da Silva, M.A., Custer, B., Parag, K.V., Barral-Netto, M., Kraemer, M.U., Pereira, R.H., Pybus, O.G., Busch, M.P., Castro, M.C., Dye, C., Nascimento, V.H., Faria, N.R., Sabino, E.C., 2020. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science 371 (6526), 288–292.

Canabarro, A., Tenório, E., Martins, R., Martins, L.s., Brito, S., Chaves, R., 2020. Data-driven study of the COVID-19 pandemic via age-structured modelling and prediction of the health system failure in Brazil amid diverse intervention strategies. PLoS One 15, 1–13, URL https://doi.org/10.1371/journal.pone.0236310.

Ceylan, Z., 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci. Total Environ. (ISSN: 0048-9697) 729, 138817.

Contreras, S., Biron-Lattes, J.P., Villavicencio, H.A., Medina-Ortiz, D., Llanovarced-Kawles, N., Olivera-Nappa, A., 2020. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. Chaos Solitons Fractals 139, 110087.

Cruz, C.H.d.B., 2020. Social distancing in São Paulo State: demonstrating the reduction in cases using time series analysis of deaths due to COVID-19. Rev. Bras. Epidemiol. 23, e200056.

Deb, P., Furceri, D., Ostry, J.D., Tawk, N., 2020. The Effect of Containment Measures on the COVID-19 Pandemic. CEPR Discussion Paper No. DP15086.

do Prado, M.F., de Paula Antunes, B.B.a., Bastos, L.d.S.L., Peres, I.T., da Silva, A.d.A.B., Dantas, L.F., Baião, F.A., Maçaira, P., Hamacher, S., Bozza, F.A., 2020. Analysis of COVID-19 under-reporting in Brazil. Rev. Bras. Terapia Intensiva 32 (2), 224.

Dong, L., Hu, S., Gao, J., 2020. Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov. Therapeut. 14 (1), 58–60.

Ferrante, L., Fearnside, P.M., 2020. Protect Indigenous peoples from COVID-19. Science 368 (6488), 251.

Ferrante, L., Steinmetz, W.A., Almeida, A.C.L., Leão, J., Vassão, R.C., Tupinambás, U., Fearnside, P.M., Duczmal, L.H., 2020. Brazil's policies condemn amazonia to a second wave of COVID-19. Nature Med. (26), 1315.

Freitas, C.M.d., Cidade, N.d.C., et al., 2020. COVID-19 AS a GLOBAL DISASTER: Challenges to risk governance and social vulnerability in Brazil. Amb. Soc. 23, e0115.

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., Rinaldo, A., 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc. Natl. Acad. Sci. 117 (19), 10484–10491.

Gomez-Rubio, V., 2020. Bayesian Inference with INLA. CRC Press.

Green, P., Silverman, B., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. CRC Press.

Guerra-Shinohara, E.M., Weber, S.S., Paniz, C., Gomes, G.W., Shinohara, E.J., Gandra, T.B.R., Pereira, I.C., Jarcem, K.G., Zanre, R.F.d.A., Barreto, A.G., De Carli, A.D., 2020. Overview on COVID-19 outbreak indicators across Brazilian federative units. MedRxiv 2020.06.02.20120220.

Gupta, R., Pal, S.K., 2020. Trend analysis and forecasting of COVID-19 outbreak in India. MedRxiv 2020.03.26.20044511.

Hamilton, J.D., 2018. Why you should never use the hodrick-prescott filter. Rev. Econ. Stat. 100 (5), 831–843.

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

Harvey, A., Trimbur, T., 2008. Trend estimation and the Hodrick-Prescott Filter. J. Japan Stat. Soc. 38 (1), 41–49.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L.Y., Hultgren, A., Krasovich, E., et al., 2020. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 584 (7820), 262–267.

Illian, J.B., Sorbye, S., Rue, H., 2012. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann. Appl. Stat. 4 (12), 1499–1530.

Kraemer, M.U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D.M., Du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., et al., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368 (6490), 493–497.

Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J., Khosrawipour, V., 2020. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. Pulmonology 15 (2), 110–115.

Li, Q., Feng, W., Quan, Y.-H., 2020. Trend and forecasting of the COVID-19 outbreak in China. J. Infect. 80 (4), 469–496.

Lindgren, F., Rue, H., 2008. On the second-order random walk model for irregular locations. Scand. J. Stat. 35 (4), 691–700.

Liu, X., Hewings, G.J., Wang, S., Qin, M., Xiang, X., Zheng, S., Li, X., 2020. Modeling the situation of COVID-19 and effects of different containment strategies in China with dynamic differential equations and parameters estimation. MedRxiv 2020.03.09.20033498.

Ortega, F., Orsini, M., 2020. Governing COVID-19 without government in Brazil: Ignorance, neoliberal authoritarianism, and the collapse of public health leadership. Global Publ. Health 15 (9), 1257–1277.

Pedersen, M.G., Meneghini, M., 2020. Quantifying undetected COVID-19 cases and effects of containment measures in Italy. ResearchGate Preprint (Online 21 March 2020) DOI 10.

Perone, G., 2020. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. MedRxiv 2020.04.27.2008153.

Ribeiro, L.C., ao Bernardes, A.T., 2020. Estimate of Underreporting of COVID-19 in Brazil by Acute Respiratory Syndrome Hospitalization Reports. Notas Técnicas Cedeplar-UFMG 010, Cedeplar, Universidade Federal de Minas Gerais.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. CRC Press.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.

Russell, T.W., Hellewell, J., Abbott, S., Jarvis, C., van Zandvoort, K., nCov working group, C., Flasche, S., Kucharski, A., et al., 2020. Using a delay-adjusted case fatality ratio to estimate under-reporting. Centre for Mathematical Modeling of Infectious Diseases Repository.

Saez, M., Tobias, A., Varga, D., Barceló, M.A., 2020. Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain. Sci. Total Environ. 138761.

Sanders, J.M., Monogue, M.L., Jodlowski, T.Z., Cutrell, J.B., 2020. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. JAMA 323 (18), 1824–1836.

Silva, L.V.E., de Andrade Abi, M.D.P., Dos Santos, A.M.T.B., de Mattos Teixeira, C.A., Gomes, V.H.M., Cardoso, E.H.S., da Silva, M.S., Vijaykumar, N.L., Carvalho, S.V., Frances, C.R.L., et al., 2020. An analysis of COVID-19 mortality underreporting based on data available from official Brazilian government internet portals. J. Med. Internet Res. 22 (8), e21413.

Simpson, D., Rue, H., Riebler, A., Martins, T., Sørbye, S., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32 (1), 1–28.

Skiera, B., Jürgensmeier, L., Stowe, K., Gurevych, I., 2020. How to best predict the daily number of new infections of covid-19. arXiv preprint arXiv:2004.03937.

Svetunkov, I., Petropoulos, F., 2018. Old dog, new tricks: a modelling view of simple moving averages. Int. J. Prod. Res. 56 (18), 6034–6047.

Vaid, S., Cakan, C., Bhandari, M., 2020. Using machine learning to estimate unobserved COVID-19 infections in North America. J. Bone Joint Surg. Am. 102 (13), e70.

World Healh Organization, 2020. Who director-general's opening remarks at the media briefing on COVID-19-11 march 2020.