



An Epidemiological Compartmental Model With Automated Parameter Estimation and Forecasting of the Spread of COVID-19 With Analysis of Data From Germany and Brazil

Adriano A. Batista^{1*} and Severino Horácio da Silva^{2*}

¹ Departamento de Física, Universidade Federal de Campina Grande, Campina Grande, Brazil, ² Departamento de Matemática, Universidade Federal de Campina Grande, Campina Grande, Brazil

OPEN ACCESS

Edited by:

Jun Ma,
Lanzhou University of Technology,
China

Reviewed by:

Dibakar Ghosh,
Indian Statistical Institute, India
Samuel Bowong,
University of Douala, Cameroon

*Correspondence:

Adriano A. Batista
adriano@df.ufcg.edu.br
Severino Horácio da Silva
horacio@mat.ufcg.edu.br

Specialty section:

This article was submitted to
Dynamical Systems,
a section of the journal
Frontiers in Applied Mathematics and
Statistics

Received: 23 December 2020

Accepted: 09 February 2022

Published: 13 April 2022

Citation:

Batista AA and da Silva SH (2022) An
Epidemiological Compartmental
Model With Automated Parameter
Estimation and Forecasting of the
Spread of COVID-19 With Analysis of
Data From Germany and Brazil.
Front. Appl. Math. Stat. 8:645614.
doi: 10.3389/fams.2022.645614

In this work, we adapt the epidemiological SIR model to study the evolution of the dissemination of COVID-19 in Germany and Brazil (nationally, in the State of Paraíba, and in the City of Campina Grande). We prove the well posedness and the continuous dependence of the model dynamics on its parameters. We also propose a simple probabilistic method for the evolution of the active cases that is instrumental for the automatic estimation of parameters of the epidemiological model. We obtained statistical estimates of the active cases based on the probabilistic method and on the confirmed cases data. From this estimated time series, we obtained a time-dependent contagion rate, which reflects a lower or higher adherence to social distancing by the involved populations. By also analyzing the data on daily deaths, we obtained the daily lethality and recovery rates. We then integrate the equations of motion of the model using these time-dependent parameters. We validate our epidemiological model by fitting the official data of confirmed, recovered, death, and active cases due to the pandemic with the theoretical predictions. We obtained very good fits of the data with this method. The automated procedure developed here could be used for basically any population with a minimum of adaptation. Finally, we also propose and validate a forecasting method based on Markov chains for the evolution of the epidemiological data for up to 2 weeks.

Keywords: epidemiological model, COVID-19, parameter estimation, forecasting, compartmental model

1. INTRODUCTION

The COVID-19 pandemic already has reached practically the whole planet. According to the World Health Organization (WHO, Situation Report-41 [1]), although around 80% of the infected people present mild symptoms (equivalent to the common flu), older people and those with a history of other diseases like diabetes, cardiovascular, and chronic respiratory syndrome can develop serious problems after being infected by the SARS-CoV-2 virus. It was first identified in December 2019 in the City of Wuhan, Province of Hubei, China. From there it spread across Asia, Europe, and the other continents [2]. On March 11th, the WHO declared it to be a pandemic [3].

In this work, we investigate the spread of the pandemic in 4 different scenarios: Germany, Brazil, the Brazilian State of Paraíba, and the City of Campina Grande. We use the data from Germany

mostly as a benchmark test for our model, since very likely their COVID-19 data is one of the most accurate in the world, which is due to the widespread testing of its population. Furthermore, the social distancing, isolation, and use of personal protection equipment (PPE) there has been far more efficient than the measures taken in Brazil in containing the spread of COVID-19 infections.

The first confirmed case in Brazil occurred in February 25th of 2020, when a 61-year-old man who had returned from Italy tested positive and the first death due to the pandemic occurred on March 12. It is very unlikely all cases of COVID-19 contaminations evolved from this patient. As reported in other countries, there were multiple imported infections, which usually drive the contagion rate to very high values in the initial stages of the epidemic. In Brazil and in the majority of countries around the world social distancing policies have been adopted in order to decrease the rate of contagion and thus allow that health systems do not collapse and have conditions to treat the gravest cases of the disease.

In the State of Paraíba, according to the Paraíba Department of Health, the first confirmed case of COVID-19 was registered on March 18th of 2020. It was a man who lived in the City of João Pessoa that had returned from a trip to Europe on February 29th. On March 31st the first death due to the pandemic was recorded in the State.

Campina Grande is the second largest city in the State of Paraíba. The social distancing policy was implanted in this city on March 20th of 2020, with the closure of universities, schools, and non-essential stores. The social distancing was implemented in a preventive form since the first case of infection by the SARS-CoV-2 virus only came to be registered 1 week afterwards, on March 27th. The first death due to COVID-19 in this city was only registered on April 19th.

In this work, we adapted the simple and well-known SIR epidemiological model, developed by Kermack and McKendrick [4], to study the evolution of the dissemination of COVID-19. The SIR model is a well-known and tested epidemiological compartmental model that has been applied to very diverse epidemics (see for example [5–10]). The model receives its name for dividing the population in three groups: susceptible, infected, and removed. Although, SIR and SIR-like models, such as the SIRD and SEIR(D), are simple models compared with far more detailed alternatives, such as proposed by Ndairu et al. [11], Tang et al. [12], or Gatto et al. [13], the simplicity of SIR-like models is a strength when it comes to parameter estimation. More complex models have more parameters to be determined, and, hence, more uncertainties that may render them non-viable for large scale applications to many populations. Usually, the available epidemiological data is not complete enough to provide estimates for all the parameters of more complex models. The model we propose here has three independent parameters that need to be estimated from the epidemiological data: contagion rate, average time duration of infection, and lethality rate. We validated the proposed theoretical model with comparisons of official data of the numbers of confirmed, recovered, death, and active cases due to COVID-19 infections in Germany, Brazil, the Brazilian State of Paraíba, and the City of Campina Grande, located in Paraíba.

We used a compartmental model called SIRD, which replaces the removed by the recovered and the deceased cases. Furthermore, in the present work we use a time-dependent contagion rate and time-dependent lethality and recovery rates.

SIR(D) or SEIR(D) models with time-dependent contagion rates are not new [14], even for the COVID-19 pandemic there are already several SIR and SIR-like models with time-dependent parameters. The most difficult part of this approach is the design of a parameter estimation method that is automated and renders the model accurate. The early work by Fang et al. [15] used a SEIR model in which they estimated the parameters based on the epidemiological data using statistical methods, but at the time of publication there were only about 40 data points and scant comparison of data with theoretical model time series. Zhong et al. [16] used a time-dependent SIR model in which they estimate the parameters from the data of active and recovered cases. The authors obtained big fluctuations in the contagion rate and in the removal rate ($1/\tau$). This likely occurred because they approximated the time derivative of the infected (di/dt) on a daily basis and also, perhaps, because recovered cases data is often less reliable than the confirmed cases data. Chen et al. [17] also used a parameter estimation technique of the contagion and the removal rates, but they did not provide information on the parameters obtained in their results explicitly. Dehning et al. [18] estimated parameters based on a Bayesian inference with the Markov chain Monte Carlo technique, but not many data points were available at the time of publication and no estimates for active cases were provided. Linka et al. [19] also used a Bayesian parameter estimate in a SEIR model. None of the above articles provide information on estimates for death cases. A comparison of models (SIR, SEIR, and a branching point process) highlighting the strengths and difficulties of each model, and stressing the importance of nonpharmaceutical public health interventions was made by Bertozzi et al. [20].

The present work introduces a simple method for estimating the parameters of the epidemiological model dynamical system. The proposed method combines dynamical systems, probability theory, and statistical analysis of data. In addition to the adapted SIRD model used, we develop a probabilistic model to obtain the recovery and the lethality probabilities. Based on this model, we make estimates for the active, death, and recovered cases based only on data of the confirmed cases. The best fit for the active cases provides us with the average time duration of infection, which is held fixed in our model. With the help of this probabilistic model and the epidemiological data, we also obtain the time-dependent parameters of the model: contagion, recovery, and lethality rates. The estimate of the time-dependent contagion rate is based on a fairly simple statistical analysis of the statistical estimate of the active cases. We estimate the time-dependent lethality and recovery rates based on a statistical analysis of the real death cases data. In this way, one avoids the time-consuming task, for the programmer, of obtaining the contagion rate function that leads to the best fit of the data. In this work, the best fit of the data is done automatically in one pass of integration. Thus, this greatly reduces the time taken to fit the available data with the theoretical curve. Furthermore, we present

and validate a forecasting method based on Markov chains and on the parameter estimation method we use to make short-term predictions of the evolution of the epidemiological variables. This allows for short-time forecasting from 1 to 2 weeks in advance. In order to measure the error made when fitting the parameters to the real data, we used the Root Mean Square Error (RMSE). More precisely, in this work, we use the normalized RMSE, which allows comparisons between the fittings of different datasets and also comparisons between the fittings of different time intervals of the same dataset. Moreover, we also take into account the fact that the counting of new epidemiological cases is usually corrupted by delays in the reporting and by missed detection of infections, deaths, and recoveries. To account for that, we add to the daily epidemiological data variation zero-mean Gaussian deviates with standard deviation of 5% of the daily number of cases variation. We run this many times to verify if the curve-fitting ability of our theoretical model is robust to such random perturbations.

Additionally, we prove the well posedness of this model (existence, uniqueness, and continuous dependence on initial data). Also, an important theoretical issue we investigate is the dependence of the solutions on the parameters present in the model. Although the continuous dependence of solutions on parameters is a classic result in abstract dynamical systems (see, for example, [21], Theorem 3.2), as far as we know, this has not been proved yet for SIR-type models. Hence, we prove the continuous dependence of the solutions on the system parameters. Here, we use techniques similar to those used in da Silva [22] and Diekmann and Getto [23] to accomplish this result.

We hope that this study of epidemiological dynamics be useful in stressing the importance of public health policies about the application, maintenance, and reinforcement of social distancing measures with the objective of avoiding the collapse of the health system.

This work is organized in the following way: In Section 2, we propose our epidemiological model and we prove results on existence and uniqueness of solution and on the continuous dependence of the solution with respect to the initial data and the parameters present in the model. We develop, in Section 3, the probabilistic model with estimates of the probabilities of recovery and death. We also describe how to make statistical estimates for the various parameters used in our epidemiological model. In Section 4, we investigate the evolution of COVID-19 in Germany as a benchmark test for our model. Afterwards, we present our results and discuss about the evolution of the disease in Brazil, in the State of Paraíba, and in the City of Campina Grande. We also present a method of forecasting for short terms the evolution of the pandemic based on Markov chains. We validate the predictions of our theoretical model by fitting the epidemiological data. In Section 5, we use the root mean square error (RMSE) to measure how well our time-dependent compartmental model fits the epidemiological data. In Section 6, we add noise to the epidemiological data to see if our model is robust enough to this sort of perturbation. Finally, in Section 7 we draw our conclusions.

2. EPIDEMIOLOGICAL MODEL

The evolution of the epidemiological model we propose is determined by the following ordinary differential equations (ODE) system

$$\begin{aligned}\frac{dS}{dt} &= \nu(S + I + R) - \mu S - \kappa(t)SI, \\ \frac{dI}{dt} &= -\left(\mu + \frac{1}{\tau}\right)I + \kappa(t)SI, \\ \frac{dR}{dt} &= \rho(t)I - \mu R, \\ \frac{dM}{dt} &= \lambda(t)I,\end{aligned}\tag{1}$$

where $S(t)$, $I(t)$, $R(t)$ are, respectively, the normalized susceptible, infected, and recovered populations at time t . We define $M(t)$ as the normalized number of accumulated deaths due to the epidemic. The introduction of $M(t)$ is mostly for convenience, so that we do not have to perform a separate integration from the numerical routine used to integrate the differential equations of our model. The normalization of the variables was achieved by dividing them by P_0 , the total initial population of the region considered. For the sake of simplification, we assume that the population is homogeneous such that all the susceptible individuals have the same probability of being contaminated and the infected individuals have the same probability of recovery or death due to the infection. We also suppose that the population evolves in such a way that the newborn babies are all susceptible and the recovered are all immune. The parameter ν is the population birth rate, μ is the death rate before the onset of the pandemic, $\kappa(t)$ is the contagion rate function, $\rho(t)$ is the recovery rate, and $\lambda(t)$ is the lethality rate due to the epidemic. Although $\rho(t)$ and $\lambda(t)$ are changing over time, $\tau^{-1} = \rho(t) + \lambda(t)$ is held constant.

It is important to point out further differences between the original model (1) and the proposed SIR model [4]. This could become relevant if the pandemic lasts for over a year. This is relevant also as a source of comparison with the death rates due to the COVID-19. One important difference we introduced is that we now allow for time variation in the contagion rate $\kappa(t)$, which reflects changes in confinement, social distancing, and mask use adopted by the population. In addition, we allow for time variation in the recovered and lethality rates, which might reflect possible improvements in the treatment efficacy of COVID-19 and/or changes in the demographics of the infected. Here, we also allow variations in the total population, taking into account the contributions of birth and death rates to the population evolution.

2.1. Well Posedness

For a given time $T > 0$, for each initial value (S_0, I_0, R_0, M_0) , with $\kappa(t)$, $\rho(t)$ and $\lambda(t)$ varying continuously in the bounded interval $[0, T]$ and with the other parameters fixed, one can show that the ODE system of Equation (1) admits existence and uniqueness of solution in the time interval $[0, T]$. Furthermore, it is possible to

prove that the solutions are continuously dependent on the initial data and on the parameters κ, ρ and λ .

To prove the existence and uniqueness of solution of Equation (1), in the Euclidean space \mathbb{R}^4 , it is sufficient to verify that the function given by the right hand side of Equation (1) is Lipschitz continuous with respect to spatial variable (see [21]).

For this, let $\xi(t) = (S(t), I(t), R(t), M(t))$ be and define $g: \mathbb{R} \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ as

$$g(t, \xi) = (g_1(t, \xi), g_2(t, \xi), g_3(t, \xi), g_4(t, \xi)), \tag{2}$$

where the $g_i: \mathbb{R} \times \mathbb{R}^4 \rightarrow \mathbb{R}$ are the coordinate functions of g given by

$$g_1(t, \xi) = \nu(S + I + R) - \mu S - \kappa(t)SI,$$

$$g_2(t, \xi) = -\left(\mu + \frac{1}{\tau}\right)I + \kappa(t)SI,$$

$$g_3(t, \xi) = \rho(t)I - \mu R,$$

and

$$g_4(t, \xi) = \lambda(t)I,$$

with $\kappa, \rho, \lambda: [0, T] \rightarrow \mathbb{R}_+$ continuous functions.

Therefore, the existence and uniqueness of solution of Equation (1) follows from the proposition below, whose proof is given in the **Appendix A**.

Proposition 2.1. *The function given in (2) is Lipschitz continuous with respect to the second variable.*

Remark 2.2. *From Proposition 2.1 and Picard-Lindelöf Theorem [21], it follows that for each initial value $\xi_0 = (S_0, I_0, R_0, M_0)$, with $\kappa(t), \rho(t)$ and $\lambda(t)$ varying continuously in the bounded intervals $[0, T]$, the ODE system of Equation (1) admits existence and uniqueness of solution in the time interval $[0, T]$, which is given, for $t \in [0, T]$, by*

$$\xi(t) = \xi_0 + \int_0^t g(s, \xi(s))ds. \tag{3}$$

Furthermore, using (3) and Gronwall inequality [21], we obtain the sensitivity with respect to the initial data.

Indeed, denoting by $\xi(t, \xi_0)$ the solution of Equation (1) that at $t = 0$ is ξ_0 , we have

$$\begin{aligned} \|\xi(t, \xi_0^1) - \xi(t, \xi_0^2)\| &\leq \|\xi_0^1 - \xi_0^2\| \\ &+ \int_0^t \|g(s, \xi(s, \xi_0^1)) - g(s, \xi(s, \xi_0^2))\| ds \\ &\leq \|\xi_0^1 - \xi_0^2\| + \int_0^t L \|\xi(s, \xi_0^1) - \xi(s, \xi_0^2)\| ds. \end{aligned}$$

Hence, from Gronwall inequality

$$\begin{aligned} \|\xi(t, \xi_0^1) - \xi(t, \xi_0^2)\| &\leq \|\xi_0^1 - \xi_0^2\| e^{\int_0^t L ds} \\ &\leq e^T \|\xi_0^1 - \xi_0^2\| \rightarrow 0, \text{ as } \|\xi_0^1 - \xi_0^2\| \rightarrow 0. \end{aligned}$$

The continuous dependence of the solutions of (1) with respect to the variation of its main parameters is obtained in the proposition below and its proof is given in the **Appendix B**.

Proposition 2.3. *Under the same assumptions and notation from Proposition 2.1, the solution $(S(t), I(t), R(t), M(t))$ is continuous with respect to parameters $\kappa(t), \rho(t)$ and $\lambda(t)$ present in Equation (1).*

Remark 2.4. *It is easy to verify the positivity and the boundedness of the state variables along the trajectories of the ODE System defined in Equation (1). See the proofs in the **Appendices C,D**.*

3. ESTIMATES FOR THE PARAMETERS

In this section, we explain the derivation of the reproduction number and develop the estimation procedures for the parameters present in the model. In addition, we describe our forecasting method.

3.1. Reproduction Number

It is of paramount importance to know if a contagious disease will become epidemic or not in a population. It is also important to know when it will be possible to control an epidemic, that is, when it will be possible to block its growth. This will happen when $\frac{dI}{dt} \leq 0$. From Equation (1), we obtain that this condition is equivalent to

$$-\left(\mu + \frac{1}{\tau}\right) + \kappa(t)S(t) \leq 0 \implies \frac{\kappa S(t)}{\mu + \frac{1}{\tau}} \leq 1. \tag{4}$$

In the beginning of the epidemic, we obtain that the value of the basic reproduction number [24]

$$R_0 = \frac{\kappa(0)S(0)}{\mu + 1/\tau}. \tag{5}$$

This is the result for our model that one obtains from the value of the spectral radius of the matrix of the linearized system at the disease free equilibrium (see [25, 26] for more details). If one knows more details about the subpopulations of the infected, such as the numbers of exposed, transmitters, and hospitalized, one could devise a more precise measure of the reproduction number as was proposed by Ghosh et al. [27]. Unfortunately, in most places this detailed epidemiological information is not publicly available. The data on the exposed is very hard or impossible to obtain, unless a very intensive, fast, and widespread program of contact tracing and testing is put in place. When $R_0 > 1$, the disease will start spreading, whereas when $R_0 < 1$ the contagion loses strength and the dissemination of the virus will be controlled. In our case $S_0 = 1$ and $S(t) \leq 1$, thus at any time after the onset of the epidemic the disease will enter remission if the dynamic reproduction number

$$R_0(t) = \frac{\kappa(t)S(t)}{\mu + 1/\tau} \tag{6}$$

remains less than 1. Although, we are over 9 months into the current epidemic in Brazil at the time of writing, $M(t) \ll 1$, $S(t) \approx 1$, the critical value of $\kappa(t)$ is $\kappa^* = \mu + 1/\tau$. If the contagion rate stays below this value, the pandemic will decrease.

As there is no efficacious treatment against COVID-19 at the time of writing this paper to the authors' knowledge, it is not yet possible to easily alter the average time of infection τ . As $1/\tau \gg \mu$, thus the only viable manner of decreasing $R_0(t) \approx \kappa(t)\tau$ is by reducing the value of $\kappa(t)$, which can be obtained with social distancing measures and the use of PPE.

3.2. Probabilistic Model

It is fundamental that we have good estimates for the recovery and the lethality rates. In order to obtain these estimates we will use a very simple probabilistic model. Suppose that a person is infected at a given instant n (which may be a day, an hour, or a minute for example), then the probability that the infected remains sick until the following instant $n + 1$ is q , the probability that the infected recovers in the next instant is p , and the probability that the infected dies is s , in such a way that $p + q + s = 1$. Here, we suppose that q and $p + s$ remain constant during the course of the disease. Hence, we have the following probability table

Note that the normalization

$$\sum_{n=0}^{\infty} q^n p + \sum_{n=0}^{\infty} q^n s = \frac{p + s}{1 - q} = 1,$$

implies that this probabilistic model is well defined.

If n is sufficiently large, only two outcomes are possible: either the infected individual recovers or dies. Hence, based on the **Table 1** we find that the probability of recovery and of death are, respectively,

$$\begin{aligned} P_\rho &= \sum_{n=0}^{\infty} p q^n = p/(1 - q), \\ P_\lambda &= \sum_{n=0}^{\infty} s q^n = s/(1 - q). \end{aligned} \tag{7}$$

Using these probabilities we find that the average number of instants (minutes, hours, days, etc.) of the infection is given by

$$\bar{n} = \sum_{n=1}^{\infty} (p + s) n q^n = (p + s) F_1(q) = (1 - q) \sum_{n=1}^{\infty} n q^n, \tag{8}$$

in which the summation $F_1(q) = \sum_{n=1}^{\infty} n q^n$ can be calculated in the following form

$$\begin{aligned} q F_1(q) &= \sum_{n=1}^{\infty} n q^{n+1} = \sum_{n=2}^{\infty} (n - 1) q^n = \sum_{n=2}^{\infty} n q^n - \sum_{n=2}^{\infty} q^n \\ &= F_1(q) - \sum_{n=1}^{\infty} q^n = F_1(q) - \frac{q}{1 - q}. \end{aligned}$$

Hence, we obtain

$$F_1(q) = \frac{q}{(1 - q)^2}. \tag{9}$$

Therefore, we find that the average number of instants of the infection is

$$\bar{n} = \frac{q}{1 - q}, \tag{10}$$

from where we obtain that $q = \bar{n}/(1 + \bar{n})$. We can also find that the average number of time intervals until recovery is given by

$$\bar{n}_\rho = p F_1(q) = \frac{p q}{(1 - q)^2} = P_\rho \bar{n}$$

and the average number of time intervals until death is

$$\bar{n}_\lambda = s F_1(q) = \frac{s q}{(1 - q)^2} = P_\lambda \bar{n}.$$

Note that $\bar{n} = \bar{n}_\rho + \bar{n}_\lambda$, that is, the average infection time span is the sum of the average time span for recovery and the average time span until death. If only these two processes were present, it would lead to the following difference equation for the number of infected

$$I(n + 1) = I(n) - (p + s)I(n) = I(n) - (1 - q)I(n) = I(n) - \frac{1}{1 + \bar{n}}I(n).$$

If we take n to indicate the n -th time interval, such as a minute, in which there is not much variation in the quantities S , I , R , and M , hence, we can approximate

$$\frac{dI}{dt} \approx \frac{\Delta I}{\Delta t} = -\frac{1}{(1 + \bar{n})\Delta t} I(t_n),$$

in which $t_n = n\Delta t$. In this work, we take $\Delta t = 1 \text{ h} = 1 \text{ day}/24$. The average time span of infection can be obtained from the following equation

$$\frac{1}{\tau} = \lambda + \rho = \frac{1}{(1 + \bar{n})\Delta t}. \tag{11}$$

Hence, we obtain the following expressions for the rates of lethality and recovery

$$\begin{aligned} \lambda &= \frac{\bar{n}_\lambda}{\bar{n}\tau} = \frac{P_\lambda}{\tau}, \\ \rho &= \frac{\bar{n}_\rho}{\bar{n}\tau} = \frac{P_\rho}{\tau} = \frac{1 - P_\lambda}{\tau}. \end{aligned} \tag{12}$$

TABLE 1 | Probabilistic model.

Situation \ Instant	0	1	2	...	n	...
Recovered	p	$q p$	$q^2 p$...	$q^n p$...
Death	s	$q s$	$q^2 s$...	$q^n s$...

3.2.1. Standard Deviation

Here, we calculate the standard deviation for this probabilistic process in the number of time intervals n of infection, from the contamination until recovery or death. In order to achieve that, we have first to calculate the sum

$$F_2(q) = \sum_{n=1}^{\infty} n^2 q^n. \tag{13}$$

We can calculate this sum by noting that

$$\begin{aligned} q^2 F_2(q) &= \sum_{n=2}^{\infty} (n-2)^2 q^n = F_2(q) + 4 \sum_{n=1}^{\infty} (1-n)q^n \\ &= F_2(q) + \frac{4q}{1-q} - q - 4F_1(q). \end{aligned}$$

Hence, we find

$$F_2(q) = \frac{q(1+q)}{(1-q)^3}. \tag{14}$$

We then obtain

$$\bar{n}^2 = (1-q)F_2(q) = \frac{q(1+q)}{(1-q)^2}$$

and

$$\bar{n}^2 = \frac{q^2}{(1-q)^2}.$$

We can now write the standard deviation of n as

$$\sigma = \sqrt{\bar{n}^2 - \bar{n}^2} = \frac{\sqrt{q}}{1-q}. \tag{15}$$

This shows that the statistical fluctuations in the time duration of the infection as q grows to 1. By reducing q one not only decreases \bar{n} but also σ .

3.3. Statistical Prediction of Active, Recovery, and Death Cases

It is very important for government officials to have estimates of the number of active (infected) cases in a given population, since these people are the ones who spread the disease. Based on knowledge of new daily infections (new confirmed cases), which we can obtain from COVID-19 databases that are updated at least every day with data from across the globe, we can make several estimates. One simple estimate of the number of active cases at time t_k is

$$A(t_k) = C(t_k) - \mathcal{R}(t_k) - D(t_k) \approx C(t_k) - C(t_k - \tau), \tag{16}$$

where $C(t_k)$ is the number confirmed cases, $\mathcal{R}(t_k)$ is the number of recovered cases, $D(t_k)$ represents the accumulated number of death cases at time t_k . Another approach is to make a statistical estimate for the number of active cases at time t_k that

follows the probabilistic process described in Section 3.2, which is the following

$$A(t_k) = - \sum_{i=-1}^k q^i P_0 \Delta S(t_{k-i}) = \sum_{i=0}^k q^i \Delta C(t_{k-i}), \tag{17}$$

where $A(t_k) = P_0 I(t_k)$, $\Delta S(t_n) = S(t_n) - S(t_{n-1})$ and, likewise, $\Delta C(t_n) = C(t_n) - C(t_{n-1})$. The index k starts at 0. We also have the following probabilistic estimates for the daily increase of recovered and death cases at the k -th day

$$\Delta \mathcal{R}(t_k) = \sum_{i=0}^{k-1} q^i p \Delta C(t_{k-1-i}), \tag{18}$$

$$\Delta D(t_k) = \sum_{i=0}^{k-1} q^i s \Delta C(t_{k-1-i}), \tag{19}$$

These probabilistic estimates complement the predictions from the epidemiological dynamical system model. The probabilistic aspect of the pandemic evolution can be most clearly seen when one considers short time variations, such as daily data on the number of new confirmed, recovered, or death cases. Also this allows us to make estimates of the evolution of the active, recovered, and death cases solely from the confirmed cases time series. This is most important where the recovered cases are not available or are less reliable, specially when one is counting the recovery of outpatients, since they may not take a second test to show they are actually free of the infection.

3.4. Estimates of the Parameters Used in the Model

We now make some estimates for the parameters ν , μ , ρ , and λ present in the dynamical system given in Equation (1).

3.4.1. Birth and Death Rates

In order to make our model more precise, we obtained the annual birth rate (ABR) and the annual mortality rate (AMR) from the most recently published data from Germany and from Brazil before the spread of the pandemic.

We also converted these annual rates into daily rates using the geometrical progression formulas

$$(1 + \nu)^{365} = 1 + ABR,$$

$$(1 - \mu)^{365} = 1 - AMR.$$

Hence, we find the daily birth and mortality rates

$$\begin{aligned} \nu &= (1 + ABR)^{1/365} - 1 = e^{\frac{1}{365} \ln(1+ABR)} - 1 \\ \mu &= 1 - (1 - AMR)^{1/365} = 1 - e^{\frac{1}{365} \ln(1-AMR)} \end{aligned} \tag{20}$$

The German data on birth and death rates were obtained from the Federal Statistical Office [28, 29]. The Brazilian data was obtained from the Brazilian Institute of Geography and Statistics (IBGE), which are from 2018. For the birth rate we divided the number of born alive infants by the population estimate for

2018 and, likewise, the number of deaths by the 2018 population estimate. The data on the born alive was collected from the site <https://sidra.ibge.gov.br/tabela/2609>, the number of deaths from the site <https://sidra.ibge.gov.br/tabela/2654>, and the population estimate from the site <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=22367&t=resultados>. The birth and death rates used are shown in the **Table 2** below. We provide **Table 3** with information on the average total daily deaths (without the born dead) before the pandemic based on annual death rates from 2018. This should be compared with the average daily deaths due to COVID-19 as another way of assessing the severity of this pandemic.

3.4.2. Contagion Rate

We use the following method to estimate the time-dependent contagion rate from the epidemiological data. From Equation (1), we can write

$$\kappa(t) = \frac{1}{S(t)} \left[\frac{1}{I(t)} \frac{dI(t)}{dt} + \mu + \frac{1}{\tau} \right]. \quad (21)$$

We would like to point out though, that this is not a definition of $\kappa(t)$. It cannot depend on $I(t)$, otherwise the ODE system of Equation (1) could not be integrated. This formal equation inspired us to write the following expression for the contagion rate as function of time

$$\kappa_k = \kappa(t_k) \approx \frac{1}{S(t_k)} \left[\frac{1}{A(t_k)} \frac{\Delta A(t_k)}{\Delta t} + \mu + \frac{1}{\tau} \right], \quad (22)$$

where $\Delta t = 1$ day. The daily approximation has too much fluctuation. A better approach is to replace the daily derivative by the slope of a linear regression of 7 consecutive active cases data points $\{A_k, A_{k+1}, \dots, A_{k+6}\}$. One rolls this week-long window over the entire set of data points calculating κ_k . For the last 6 days of data, we use a backwards window with the data points $\{A_k, A_{k-1}, \dots, A_{k-6}\}$. We chose the week interval, because all epidemiological data we have seen present weekly modulations.

TABLE 2 | Birth and death rates.

Population ABR (year ⁻¹)	ν (day ⁻¹)	AMR (year ⁻¹)	μ (day ⁻¹)
Germany	0.0095	2.591×10^{-5}	0.0115
Brazil	0.0139	3.7844×10^{-5}	6.1560×10^{-3}
Paraiba	0.0148	4.0139×10^{-5}	6.5325×10^{-3}
Campina Grande	0.0158	4.2943×10^{-5}	7.1342×10^{-3}

TABLE 3 | Pre-pandemic populations and average daily deaths.

Location	P_0	μ (day ⁻¹)	average deaths (day ⁻¹)
Germany	83,149,360	3.169×10^{-5}	2635
Brazil	211,049,527	1.6918×10^{-5}	3570
Paraiba	4,018,127	1.7956×10^{-5}	72
Campina Grande	409,731	1.9616×10^{-5}	8

The most difficult part of this estimating method of $\kappa(t)$ occurs at the beginning of the data sets, when the number of active cases is very small. From Equation (22), one can see that for small values of $A(t_k)$, any errors in the derivative approximation $\frac{\Delta A(t_k)}{\Delta t}$ are amplified. Therefore, in most cases we discard the first days or weeks of epidemiological data, usually up to the day the first death occurred. In some cases, we also had to apply a cutoff to eliminate the highest values of $\kappa(t)$ right at the beginning of the time series. Apart from the numerical errors described above, notice that any new imported infected case contributes strongly to the spread of the disease near the outbreak of the epidemic. Hence, at the beginning $\kappa(t)$ tends to be very high. This also reflects the fact that at the start of the pandemic most populations did not keep social distancing nor used PPE such as masks.

3.4.3. Lethality and Recovery Rates

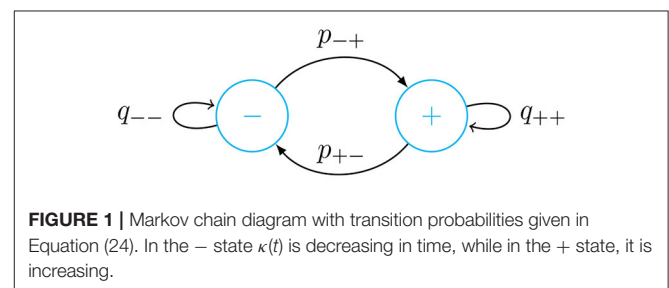
We use the following method to estimate the time-dependent lethality rate. Based on Equation (19), we can write the probability of dying for those infected with COVID-19 in a 1 day time interval as

$$s_k = \frac{\Delta D(t_k)}{\sum_{i=0}^{k-1} q^i \Delta C(t_{k-1-i})}. \quad (23)$$

From this we obtain the daily lethality probability $P_\lambda(k)$ from Equation (7). Afterwards, from Equation (12), we find the lethality rate λ_k for the k -th day. Consequently, from $\rho_k + \lambda_k = 1/\tau$, we also find the recovery rate ρ_k at the k -th day. The simpler approach of estimating s_k with $\Delta D_k/A_{k-1}$ introduces far more fluctuations.

3.5. Forecasting

We use a list with the last 3 weeks of $\kappa(t)$ data to calculate the transition probabilities of a simple two-state Markov chain as shown in **Figure 1**. From this list, we obtain a list of differences $\Delta\kappa_i = \kappa(t_i) - \kappa(t_{i-1})$. Based on this, we make two probability distributions, one for increments in $\kappa(t)$ and the other for decrements. If $\Delta\kappa_i < 0$, in the differences list, we replace it with 0, otherwise we replace it with 1. From this list of 0's and 1's, we obtain a list of lengths of the continuous sequences of 0's and 1's. We then obtain the average length of continuous stretches of 0's and 1's. We call the average length of increments of the contagion



rate \bar{n}_+ and of decrements \bar{n}_- . Based on this we find the following transition probabilities

$$\begin{aligned}
 q_{++} &= \frac{\bar{n}_+}{1 + \bar{n}_+}, & p_{+-} &= 1 - q_{++}, \\
 q_{--} &= \frac{\bar{n}_-}{1 + \bar{n}_-}, & p_{-+} &= 1 - q_{--}.
 \end{aligned}
 \tag{24}$$

Now we use the Markov chain to decide between increasing or decreasing $\kappa(t)$. The values of the increments or decrements are taken from the two probability distributions obtained as described above. We then randomly obtain predicted values of κ for 2 weeks based on a 3-week history of data. We repeat this process for 1,000 times and make an average of all these trajectories. From these trajectories we also obtain a 95% confidence interval. In the implementation of the forecasting method, we compare actual data with our predictions for the two most recent weeks. Using this same approach, we can forecast the lethality rate based on a list of daily variation of $\lambda(t)$.

4. RESULTS AND DISCUSSION

We used the Odeint function of the Python’s scientific library package SciPy [30] to integrate the ODE system of Equation (1) with the integration time-step $dt = 1.0/24$, which corresponds to an hour when the time unit is a day. In the cases investigated, we took τ 14 days. We chose the value that provided the best fit of the active cases with the delay estimate given in Equation (16) when the active cases data was available, such as in the cases of Germany and Brazil. When the active cases data was not available, we chose τ that would give the best fit between the theoretical prediction for the active cases and the corresponding estimate of Equation (16). We suppose this parameter does not change appreciably during the time scale of the outbreak of the pandemic until now, or at least until more efficient treatments are discovered. The initial values used are: $S(0) = 1 - C_0/P_0$, $I(0) = A_0/P_0$, $R(0) = \mathcal{R}_0/P_0$, and $M(0) = D_0/P_0$, with P_0 being the population just before the outbreak of the pandemic, C_0 is the initial number of confirmed cases, A_0 is the initial number of active cases, \mathcal{R}_0 is the initial number of recovered cases, and D_0 is the initial number of death cases, usually either 0 or 1. We now apply our model to the four cases of COVID-19 dissemination: in Germany, in Brazil, in the State of Paraíba, and in the City of Campina Grande.

In **Figure 2**, we show results of numerical simulation for a range of values of κ . Unlike the other results, κ is held constant during each time integration of the equations of motion given in Equation (1). This result is important in conveying the message of the paramount importance of the contagion rate on the possible outcomes of the pandemic. Not only we observe an increase in the number of deaths when the contagion rates increase, but we also see a sharp transition. When there is a growth in the contagion rate from 0.1 to 0.15, this gives rise to an extremely sharp increase in the number of deaths. This implies that there is a critical value of κ , around which there is a rapid increase in the total number of deaths due to the pandemic. Beyond the critical value, we see a saturation in the total number of deaths. The value of κ that corresponds to $R_0 = 1$ in our model is $\kappa = \mu + 1/\tau =$

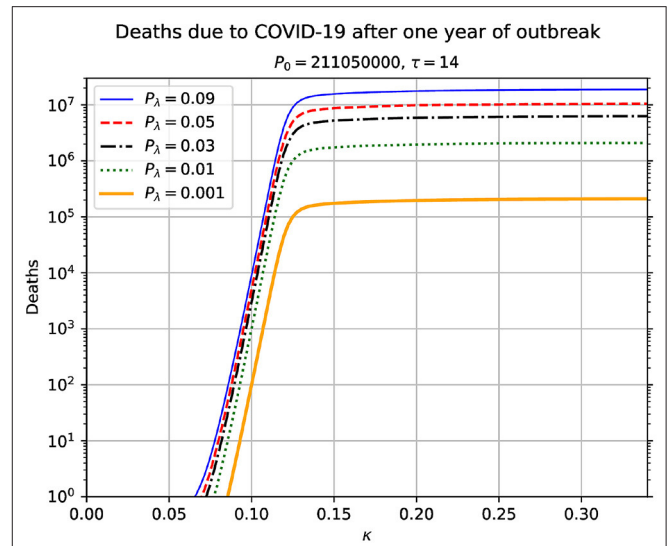


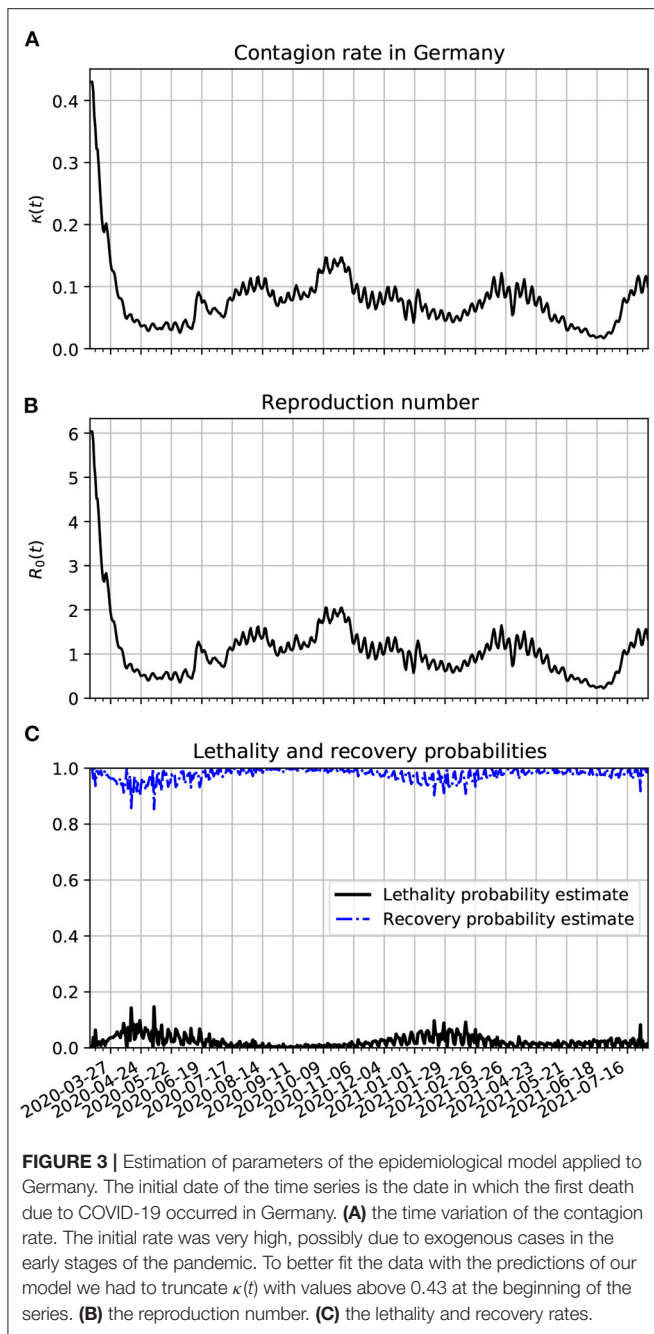
FIGURE 2 | Total number of accumulated deaths due to the epidemic as a function of the contagion rate κ . The total time duration for each value of κ is 365 days. This result shows the high sensitivity of the model time evolution on the contagion rate. The used parameters in this simulation are indicated above the figure. In each case, it is assumed that κ remains constant during the time span of the simulation. The value of κ that corresponds to $R_0 = 1$ is $\kappa^* = \mu + 1/\tau = 0.0714$.

0.0714. We believe, this reinforces the great relevance of social distancing, since increasing the average distance between people, we will be decreasing the rate of contagion and, consequently, decreasing also the number of deaths due to COVID-19.

4.1. Evolution of COVID-19 in Germany

We consider the case of Germany as a benchmark test for our epidemiological model. This is so because it is widely believed that the cases from Germany are better accounted for than in most other countries, with widespread testing of the population (<https://www.labmate-online.com/news/laboratory-products/3/breaking-news/how-germany-has-led-the-way-on-covid-19-testing/52141>). The initial population considered is $P_0 = 83.14936$ millions. The first contagion was registered on 01/27/2020 and the first death due to COVID-19 was registered on 03/09/2020. We chose this latter date as the initial point of our numerical integration.

In **Figure 3A**, we show the graph with the parameter estimation for the contagion rate function $\kappa(t)$. At the beginning of the time series, the contagion rate is very high. To achieve the best fit between theory and the data in **Figure 4**, we had a cutoff at $\kappa = 0.41$, otherwise we used the estimation method described in Section 3.4.2 to obtain this time series. Note that we observe weekly modulations, likely reflecting the fact that in weekdays the contamination is higher than in weekends. The rapid decrease of $\kappa(t)$ intensifies approximately around 03/22/2020, when strict social distancing rules were imposed by the German Government [18, 31]. This shows that these measures were very efficient in containing the spread of the epidemic. In **Figure 4B**, we show the corresponding reproduction number time series. Around April/09, the $R_0(t)$ becomes below 1 and stays there until about



June/13, when for a period of roughly 10 days it rose slightly above 1. In **Figure 4C** we show the time-dependent lethality and recovery probabilities, $P_\lambda(t)$ and $P_\rho(t)$, respectively. These estimates were obtained using the method described in Section 3.4.3. From early December/2020 to late March/2021, there was an increase in lethality, this might be due to the spread of COVID-19 variants, such as the alpha variant across Germany in that period. The decrease in $R_0(t)$ from April/2021 to June/2021 is very likely due to the increased vaccination of the population. The subsequent increase in $R_0(t)$ may be due to relaxing of preventive

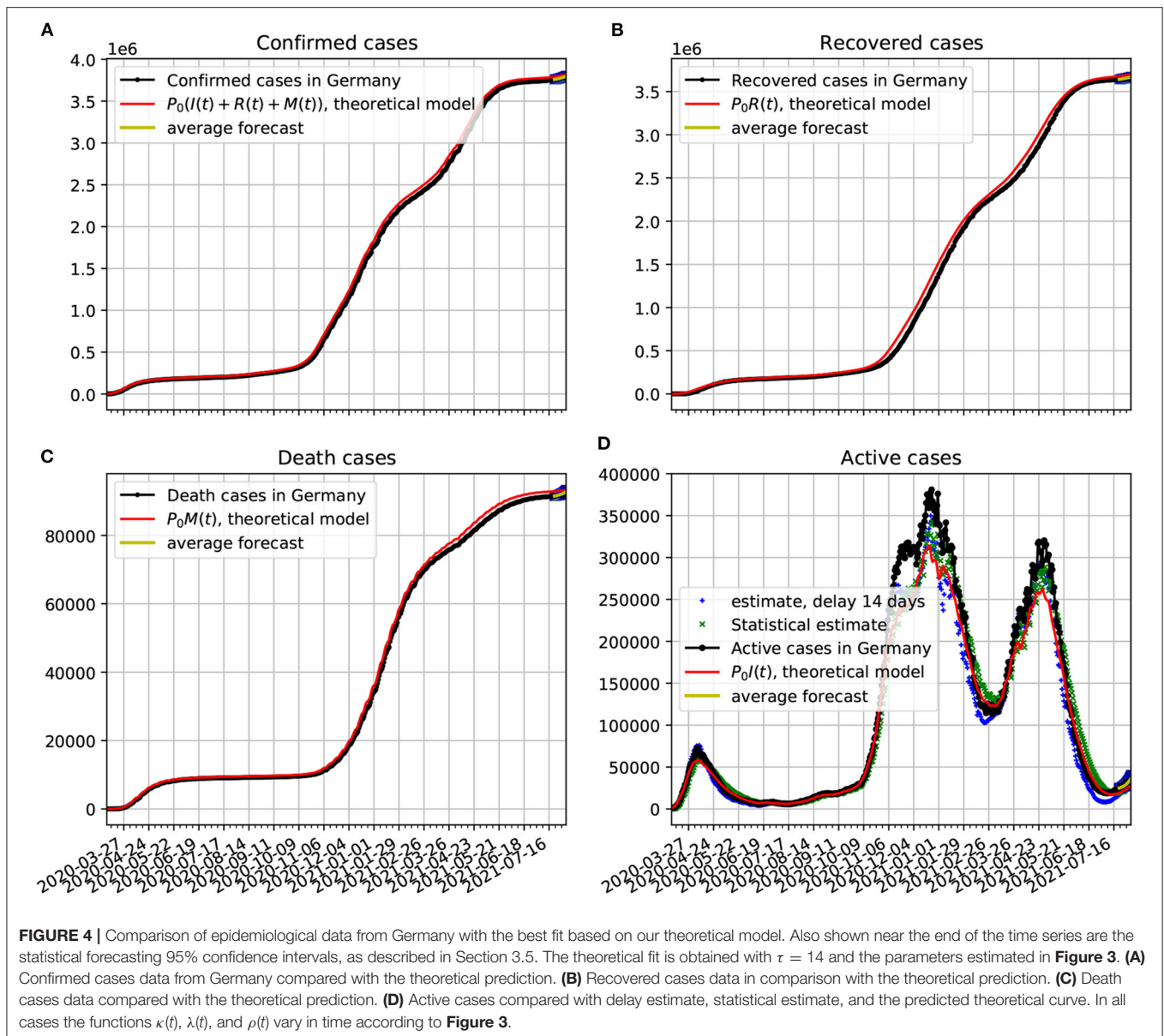
measures by the population and also due to spread of the delta variant.

In **Figure 4A**, we show the official data on the confirmed cases of COVID-19 plotted alongside the theoretical prediction, in which a very good agreement with the epidemiological data was obtained. The contagion rate function used is given in **Figure 3**. In frame **Figure 4B**, we plot the recovered cases data along with the theoretical predictions based on our model. In frame **Figure 4C**, we plot the death cases with respective theoretical curve. In frame **Figure 4D**, we plot the active cases data along with the theoretical predictions based on our model. The theoretical fit is not as good as in the previous figures, but it is still quite reasonable. We see a slow decline in the number of infected. Once the total number of confirmed cases basically saturates, the evolution of the active cases can be traced with a purely statistical model as the one we developed in Section 3.2. From the statistical point of view the slow decay of the active cases has to do with the large value of the dispersion in the duration of the infection, as shown in Section 3.2.1. We also plot two estimates of the active cases obtained solely from an analysis of the confirmed cases data. The delay estimate of active cases at a given time t is based on Equation 16. The statistical estimate is obtained from the probabilistic process given in Equation 17. This estimate came closer to the theoretical dynamical system model, what shows its consistency with the probabilistic model of Section 3.2. Both estimates came fairly close to the real data. Note also, that at the last 2 weeks of the time series we validate the forecasting model based on the Markov chain. Here, we show a 95% confidence band along 2 weeks. In this case the all epidemiological data fell within the predictions.

4.2. Evolution of COVID-19 in Brazil

We consider the initial time the day of the first death case in Brazil. We take Brazil's population to be approximately $P_0 = 211.050 \times 10^6$. The initial date of the time series of the data sets used is 03/20/2020, 1 day after the first official death due to COVID-19 was recorded. We used the pre-pandemic birth and death rates shown in **Table 2**. The estimates of contagion, lethality and recovery rates are obtained according to the methods described in Sections 3.4.2 and 3.4.3, respectively. The time variation of the contagion rate reflects the fact that the population slowly took heed of the gravity of the pandemic and started adopting social distancing measures and using PPE.

In **Figure 5A**, we show the time evolution of the contagion rate $\kappa(t)$. The sharp drop of this rate is likely due to the increase of isolation and social distancing that grew at the second half of March in Brazil. This is certainly due to better precautions by the population (isolation, social distancing, hand washing, and the increased use of PPE). In frame **Figure 5B**, we plot the reproduction number as a function of time. It is basically a scaled version of $\kappa(t)$. In frame **Figure 5C**, we show the time evolution of the lethality and recovery probabilities, $P_\lambda(t)$ and $P_\rho(t)$, respectively. One sees more fluctuations near the beginning of the time series likely because there were less active cases then. Also, one sees that the lethality probability was very high on average in the early stages of the pandemic. This might be related



with the small amount of testing in Brazil, specially in the early stages of the pandemic. Another possibility is that the medical treatment and procedures for the more severe cases of COVID-19 are being better treated since May/2020. In March/2021, there was an increase in lethality, this might be due to the gamma variants that spread from the city of Manaus across Brazil in that period.

In **Figure 6A**, we compare the official data (black dotted line) of confirmed cases with the corresponding time series obtained from our proposed model (red line). In frame **Figure 6B**, we show a comparison between the number of recovered cases and the predicted number of recovered cases predicted by the theoretical model. The discrepancies in the fitting may have to do with delays in the confirmation of the recovered cases, as we can see in the jumps that occurred from around October/2020. One possible source of systematic error, toward sub-notification of recovered cases, could occur in mild cases, as recovered outpatients may

fail to take another test to confirm their recoveries. In frame **Figure 6C**, we show a comparison between the number of death cases due to COVID-19 and the number of deaths predicted by the theoretical model. In frame **Figure 6D**, we plot the active cases data along with the theoretical predictions based on our model. The theoretical fit is not as good as in the previous figures, but it is still quite reasonable. We again validate our forecasting model with a 95% confidence interval based on Markov chain. In all cases the epidemiological data fell within the prediction range.

4.3. Evolution of COVID-19 in Paraíba

The initial population of Paraíba is $P_0 = 4,018,127$. The first case of COVID-19 contamination was registered on 03/18/2020 and the first recorded death on 04/06/2020. We did not plot the number of recovered cases because we have not been able, so far, to obtain this data for Paraíba.

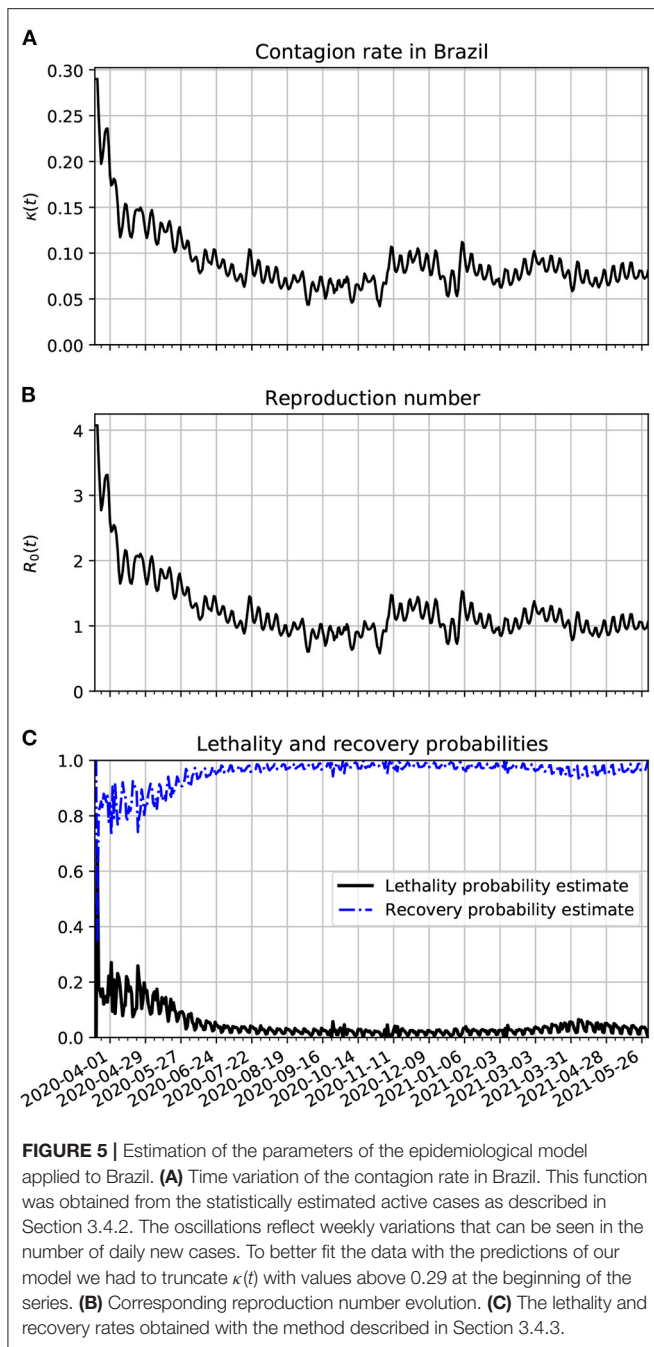


FIGURE 5 | Estimation of the parameters of the epidemiological model applied to Brazil. **(A)** Time variation of the contagion rate in Brazil. This function was obtained from the statistically estimated active cases as described in Section 3.4.2. The oscillations reflect weekly variations that can be seen in the number of daily new cases. To better fit the data with the predictions of our model we had to truncate $\kappa(t)$ with values above 0.29 at the beginning of the series. **(B)** Corresponding reproduction number evolution. **(C)** The lethality and recovery rates obtained with the method described in Section 3.4.3.

In **Figure 7A**, we show the time evolution of the contagion rate $\kappa(t)$. Initially, the contagion rate is not as high as in Brazil's case. From about 04/06/20, the rate of contagion starts decreasing on average, although with a higher amplitude of modulation as in the case of Brazil. One sees also a decrease in the amplitude of the modulations from February to May 2021. The weekly modulations may be due to the use of public transportation during the work days of the week. In frame **Figure 7B**, we plot the reproduction number as a function of time. It is basically a scaled version of $\kappa(t)$. In frame **Figure 7C**, we show the time evolution of the lethality and recovery probabilities, $P_\lambda(t)$ and

$P_\rho(t)$, respectively. Again, one sees more fluctuations near the beginning of the time series likely because there were very few active cases then. Also, one sees that the lethality rate started decreasing on average after the first week of April. This might be related with the increased amount of testing in Paraíba after 2 months after the onset of the pandemic. The increase in lethality around March and April of 2021 may also be due to the spread of the gamma variants. The decreased lethality from June/2021 is likely due the vaccinations, specially of the older population.

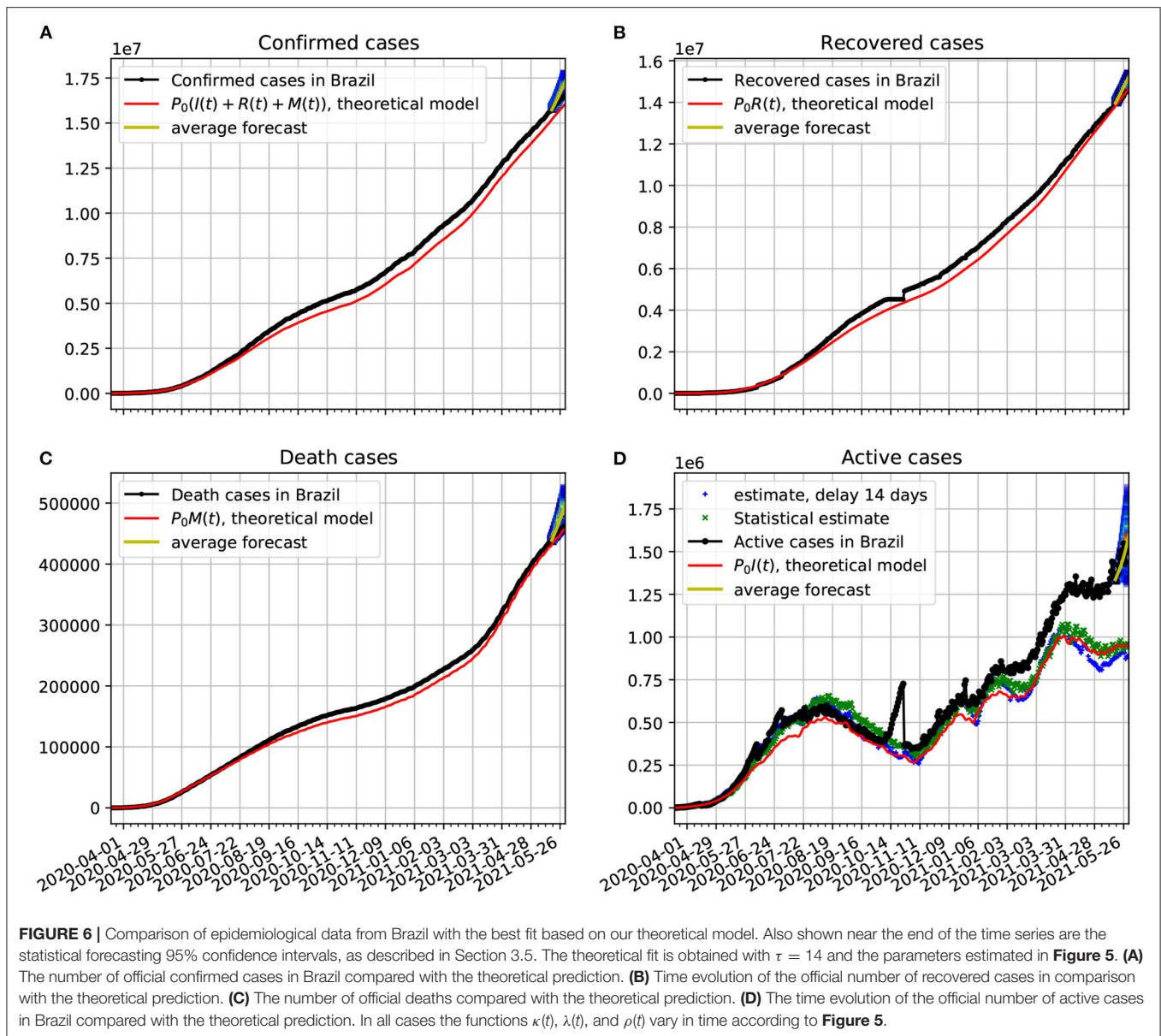
In **Figure 8A**, we compare the official data (black dotted line) of confirmed cases with the corresponding time series obtained from our model (red line). In frame **Figure 8B**, we show a comparison between the number of death cases due to COVID-19 and the number of deaths predicted by the theoretical model. In frame **Figure 8C**, we show the estimated active cases. One result is based on delay and the other on the statistical method. Both methods are described in Section 3.3. We also show epidemiological model estimate (solid red line). Note also, that at the last 2 weeks of the time series we validate the forecasting model based on the Markov chain. Here, we show a 95% confidence band along 2 weeks. In this case the all epidemiological data fell well within the predictions.

4.4. Evolution of COVID-19 in Campina Grande

The first case of COVID-19 contamination was registered on 03/27/20 and the first recorded death on 04/16/2020. We did not plot the number of recovered cases because we have not been able, so far, to obtain this data for Campina Grande.

In **Figure 9A**, we show the time evolution of the contagion rate $\kappa(t)$. Initially, the contagion rate is not as high as in Brazil's case, but takes longer to drop. Only from about 06/07/20, the rate of contagion starts decreasing on average, although with a higher amplitude of modulation as in the case of Brazil or Paraíba. As time passes, one sees also a decrease in the amplitude of the modulations. This is certainly due to better precautions by the population (isolation, social distancing, hand washing, and the increased use of PPE). In frame **Figure 9B**, we plot the reproduction number as a function of time. It is basically a scaled version of $\kappa(t)$. Since about the beginning of July, $R_0(t)$ has been modulating around 1. Although, this is not enough to end the pandemic since the number of estimated active cases is still very high. In frame **Figure 9C**, we show the time evolution of the lethality and recovery probabilities, $P_\lambda(t)$ and $P_\rho(t)$, respectively. Again, one sees more fluctuations near the beginning of the time series likely because there were very few active cases then. Also, one sees that the lethality probability started decreasing on average after the first week of April. This might be related with the increased amount of testing in Campina Grande. The decreased lethality from June/2021 is also likely due the vaccinations, specially of the older population.

In **Figure 10A**, we compare the epidemiological data (black dotted line) of confirmed cases with the corresponding time series obtained from our model (red line). In frame **Figure 10B**, we show a comparison between the number of death cases due to



COVID-19 and the number of deaths predicted by the theoretical model. In frame Figure 10C, we show the estimated active cases. One result is based on delay (black x) and the other on the statistical method (green +). Both methods are described in Section 3.3. We also show the epidemiological model estimate (solid red line). Note also, that at the last 2 weeks of the time series we again validate the forecasting model based on the Markov chain. Here, we show a 95% confidence band along 2 weeks. In this case the all epidemiological data fell well within the predictions.

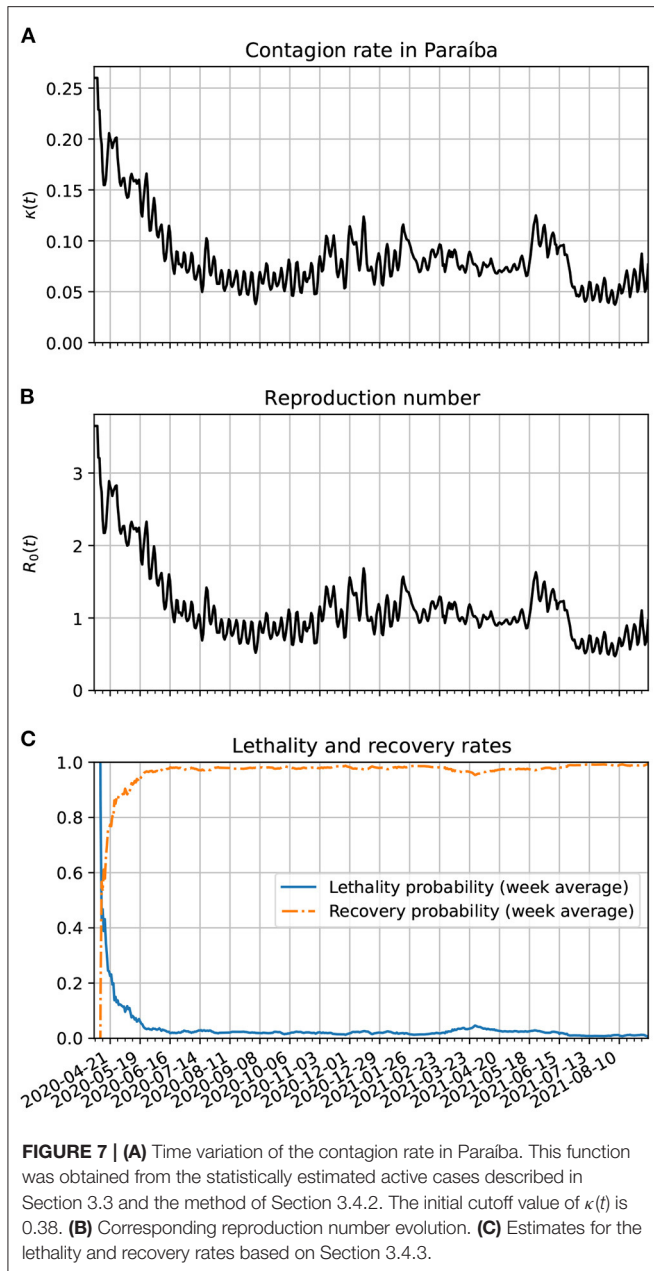
5. ANALYSIS OF CURVE FITTING OF THE EPIDEMIOLOGICAL DATA

Although there are several papers that use root-mean-square error (RMSE) and RMSE-like measures of goodness-of-fit such

as Fernandes [32], Kirbaş et al. [33], Uba et al. [34], Al-qaness et al. [35], and Gupta et al. [36], the populations are different to the ones we investigated or, as in the case of Germany, the datasets are very short compared with our datasets. The closest reference we found for comparison with our results is the work of Uba et al. [34] that only uses statistical fitting of confirmed cases data for the first 120 days of the Covid-19 epidemic in Brazil. And even then, our model is on par or better than their best fits for this time period. We also found that better measures of goodness-of-fit than RMSE are the normalized RMSE. The normalized RMSE allows comparisons of fits of different datasets and also different periods along the same dataset.

Below we provide results for the normalized RMSE, defined as

$$NRMSE = \frac{RMSE}{\bar{y}}$$



where

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N_p} (f(t_n) - y(t_n))^2}{N_p}}$$

where N_p is the number of data points, $f(t_n)$ is the estimate from the model, $y(t_n)$ represents the data, and \bar{y} is the average of y on the same time interval of the measurement of RMSE. We show results for both the confirmed and the death cases. We used two ways to calculate the NRMSE, one is the running window NRMSE with a window length of N_w days. In the cases shown, we use $N_w = 90$. As a comparison, we also calculated the cumulative

NRMSE starting from the first 90 days of pandemic data until a recent date.

In **Figure 11**, we show two different ways of measuring the NRMSE of our epidemiological data. Both the running-window and the cumulative NRMSE showed very similar results. Overall, the cumulative NRMSE values are slightly smaller for the cumulative case than the running-window case, assuming the window size of the running-window case is the initial length of the cumulative case dataset. We also notice that the cumulative case results are smoother (with broader peaks) than the running-window case. The best fitting was obtained with the data from Germany. We also notice a strong correlation between the Paraíba and Campina Grande data. This is to be unexpected because nearly the same non-farmaceutical measures were taken in both populations. In addition to that, Campina Grande is the second-largest city in Paraíba.

6. NOISE IN THE EPIDEMIOLOGICAL DATA

Here, we add noise to the epidemiological data to see if our model estimates are robust enough to the presence of uncertainties in the recording of epidemiological data. These uncertainties can arise from delays in reporting cases and also missed detection due to insufficient testing of the populations. One way to reproduce these uncertainties is to add a Gaussian noise to the epidemiological data. At each day, we add a random value obtained from a normal distribution with zero mean and standard deviation of 5% of the day's cases, which could be confirmed, death, and recovered cases. We add these statistically independent deviates to each one of these datasets, noticing that for the case of confirmed, death, and recovered cases the Gaussian deviate with zero mean and standard deviation of 5% of the day's confirmed, death, and recovered cases, respectively. From these modified epidemiological data, we obtain new time-dependent contagion, lethality, and recovery rates and integrate the equations of motion of the corresponding model. From Equation (22), we estimate that the noise effect on the contagion rate should be small. This is confirmed by our numerical results as one can see in **Figure 12A**, specifically in the inset.

What is not usually small is the effect on the populations as can be seen in **Figure 12B**. In this figure, we generated 100 datasets (not plotted) with added Gaussian noise with zero mean and standard deviation on the n -th day given by $\sigma_n = 0.05\Delta C(t_n)$. To each of these datasets we plot the corresponding theoretical estimate (colored solid lines) For comparison, we also plot the confirmed cases data without noise (black dots).

In **Figure 12C**, we investigate the effect of added Gaussian noise to the curve fitting of the epidemiological data by our model via the cumulative normalized for confirmed cases. Observe that a large fraction of the time-series of the NRMSE have higher values than the the NRMSE without added noise. This likely reflects the fact that the contagion process of COVID-19 is not really random and requires contact between infected and susceptible people as assumed by compartmental epidemiological models such as the one we use. Nevertheless, even in the worst cases, the curve fitting of the data with noise is still very good,

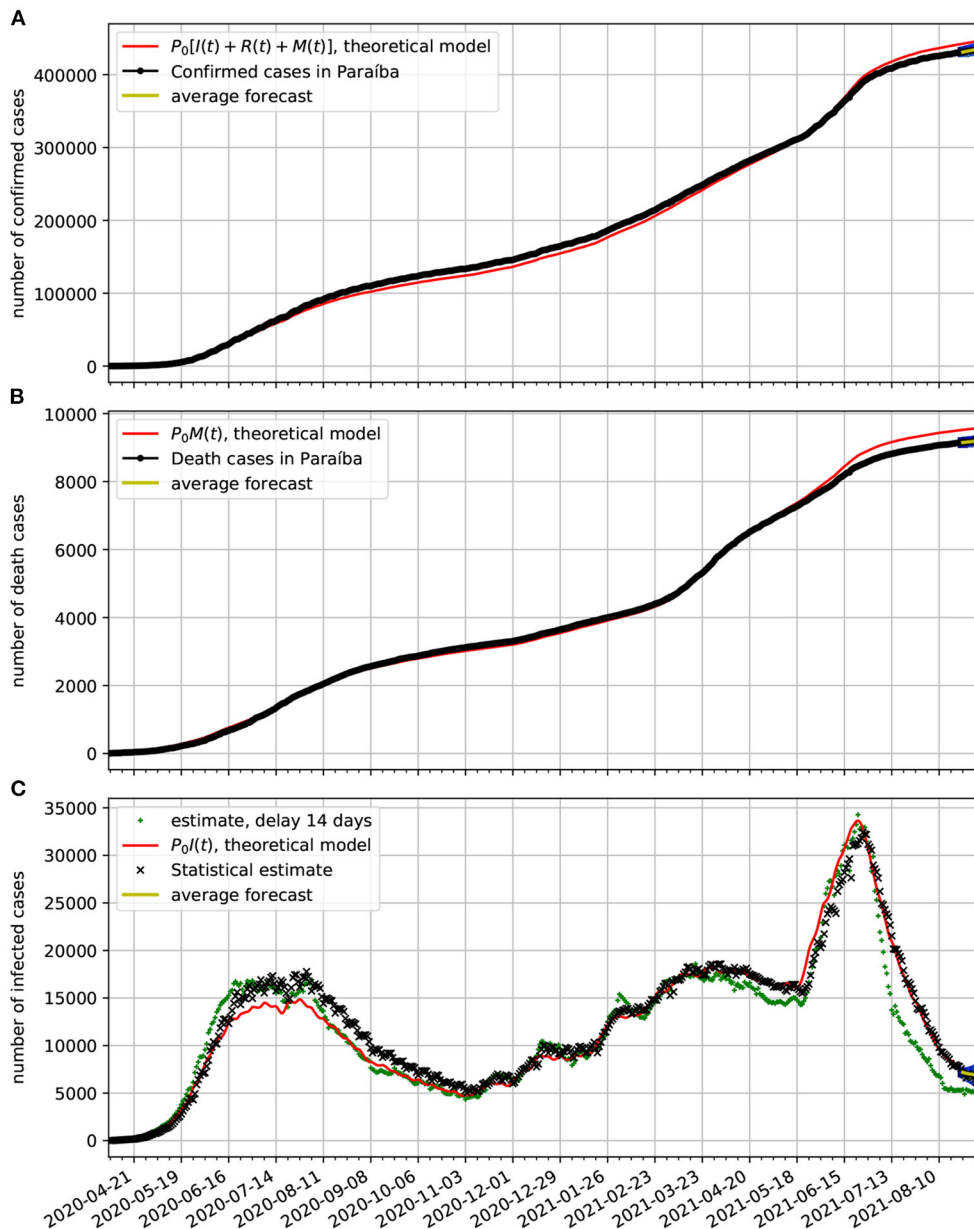


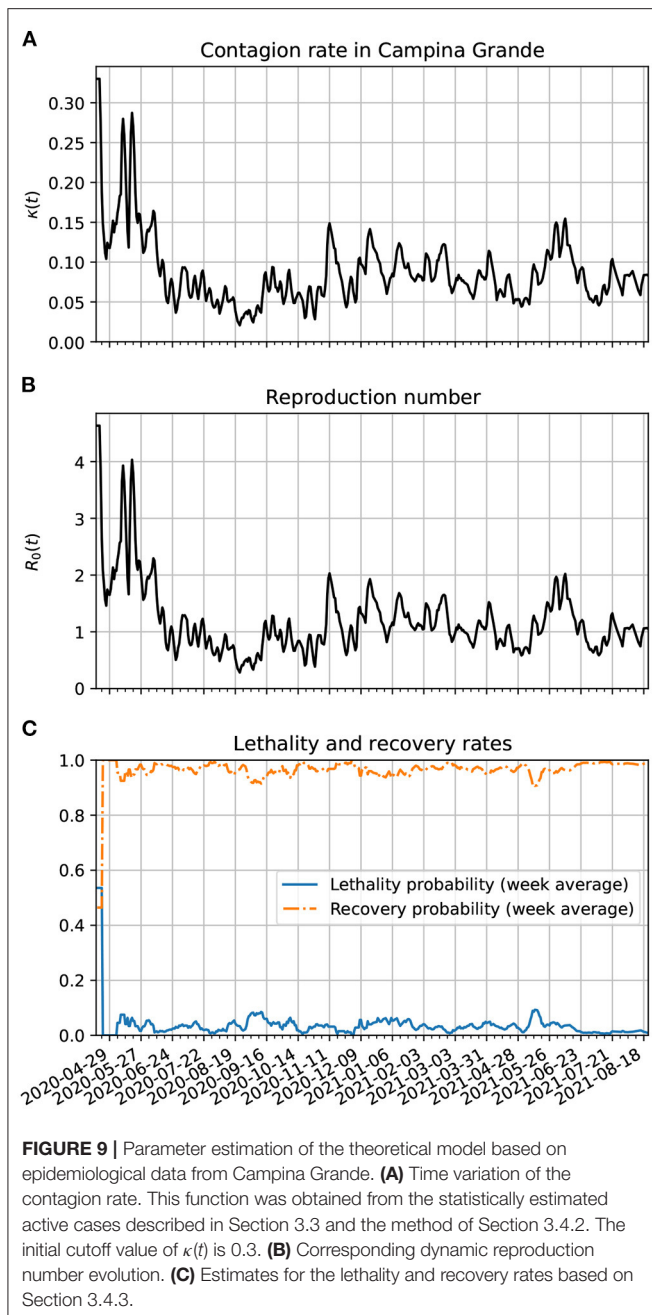
FIGURE 8 | Comparison of epidemiological data from Paraíba with the best fit based on our theoretical model. Also shown near the end of the time series are the statistical forecasting 95% confidence intervals, as described in Section 3.5. The theoretical fit is obtained with $\tau = 13$ and the parameters estimated in **Figure 7**. **(A)** Comparison of the number of official confirmed cases with the theoretical prediction. **(B)** Comparison of the number of official deaths due to COVID-19 infection with the theoretical prediction. **(C)** Comparison of the number of the estimated number of active cases with the theoretical prediction.

since the cumulative NRMSE is small, demonstrating that our model is robust to this sort of perturbation.

7. CONCLUSION

Here we summarize the main contributions of our epidemiological study of the evolution of the COVID-19 pandemic in four populations. The proposed model is an adaptation of the SIR model [4], a SIRD model [37], with some notable differences. In the SIRD model we replace the removed

population by the recovered and the deceased. We allow that the contagion rate varies in time so that it reflects the fact that social distancing and isolation changes over time. We developed two ways to obtain the active population from the data on confirmed cases only. In the first approach, the number of active cases at time t is estimated simply by the difference between the confirmed population at time t minus the confirmed population at time $t - \tau$, where τ denotes the average time span of infection. In the second approach, the number of active cases is estimated by the probabilistic model proposed here. Both approaches result



in fairly accurate predictions of the active cases when the official data is available. In cases in which the data on recovered cases is not available, this approach could be the only way to estimate the number of active cases. Furthermore, we propose a simple method to track automatically the contagion rate based on the statistical estimate of active cases. We divided this data in weekly intervals and for each interval we made a linear regression to obtain the slope, and based on the equation in our model for the infected, we could obtain a daily contagion rate. This contagion rate was fed back into the equations of motion and we were able to fit the data with a minimum of or no *ad hoc* interventions. We noticed that the contagion rate could reach very high values at

the initial stages of the spread of the epidemic in all populations investigated. This likely happens because of inherent numerical errors, as described in Section 3.4.2, and multiple infections that are imported from contaminated visitors or people returning from trips abroad. This type of dynamics is more relevant at the beginning, when the local contaminated population is small and before barriers on traveling were imposed by the governments. Here, we also take into account the contribution from the pre-pandemic birth and death rates to the evolution of the populations investigated. This could become relevant if the pandemic lasts for over a year and also it is important as a comparison for the lethality of the pandemic. According to the results exposed in the previous section, with our model we could fit the official case data from Germany, Brazil, the State of Paraíba, and city of Campina Grande quite well.

The modeling of the spread of the pandemic in Germany is very emblematic, since it clearly shows that the strict social distancing measures imposed by the government on 03/22/2020 were very effective in containing the disease, reducing the $R_0(t)$ from 2.8 down to roughly 1 about 2 weeks later, and further down to approximately 0.5 in more 2 weeks, according to our model. The increase in lethality from December/2020 through March/2021 was probably due to the spread of COVID-19 variants, such as the alpha variant. The decrease in $R_0(t)$ from April/2021 to June/2021 is very likely due to the increased vaccination of the population. The subsequent increase in $R_0(t)$ may be due to relaxing of preventive measures by the population and also due to spread of the delta variant.

In the case of Brazil, we conclude that, based on the fit of the proposed model, the decrease of the reproduction number has been far more difficult and bumpier. This means the implanted social distancing measures are having an effect in thwarting the spread of the disease, but it has not been enough, since to control the pandemic $R_0(t)$ has to remain consistently below 1 for several weeks, whereas in reality it has been oscillating around 1 for most of the time and the number of active cases is still very high. The increase in lethality after March/2021 has likely to do with the spread of gamma variants across the country.

The evolution of COVID-19 in the State of Paraíba has been similar to the national case. The amplitude of the weekly modulations of the contagion rate is considerably higher than in Brazil. This may be due to the smaller population involved. The higher lethality from February/2021 through May/2021 is probably due to the spread of the gamma variants, while the lower lethality after the last week of June/2021 is likely due to the increased anti-COVID-19 vaccination of the population. The decrease in the effective reproduction number after June/2021 may also be due to the vaccination.

As was commented in the Introduction, Campina Grande adopted a social distancing policy 1 week earlier than the report of the first confirmed case. Despite of this, the rate of contagion did not decrease as fast as it did in Germany, Brazil or Paraíba. It presents two major peaks in $\kappa(t)$ spaced apart by a week since the outbreak of the disease here. Probably these could be linked to agglomeration events such as in-branch governmental relief payments to unemployed people. Also, one sees higher amplitude of the modulation of the contagion rate

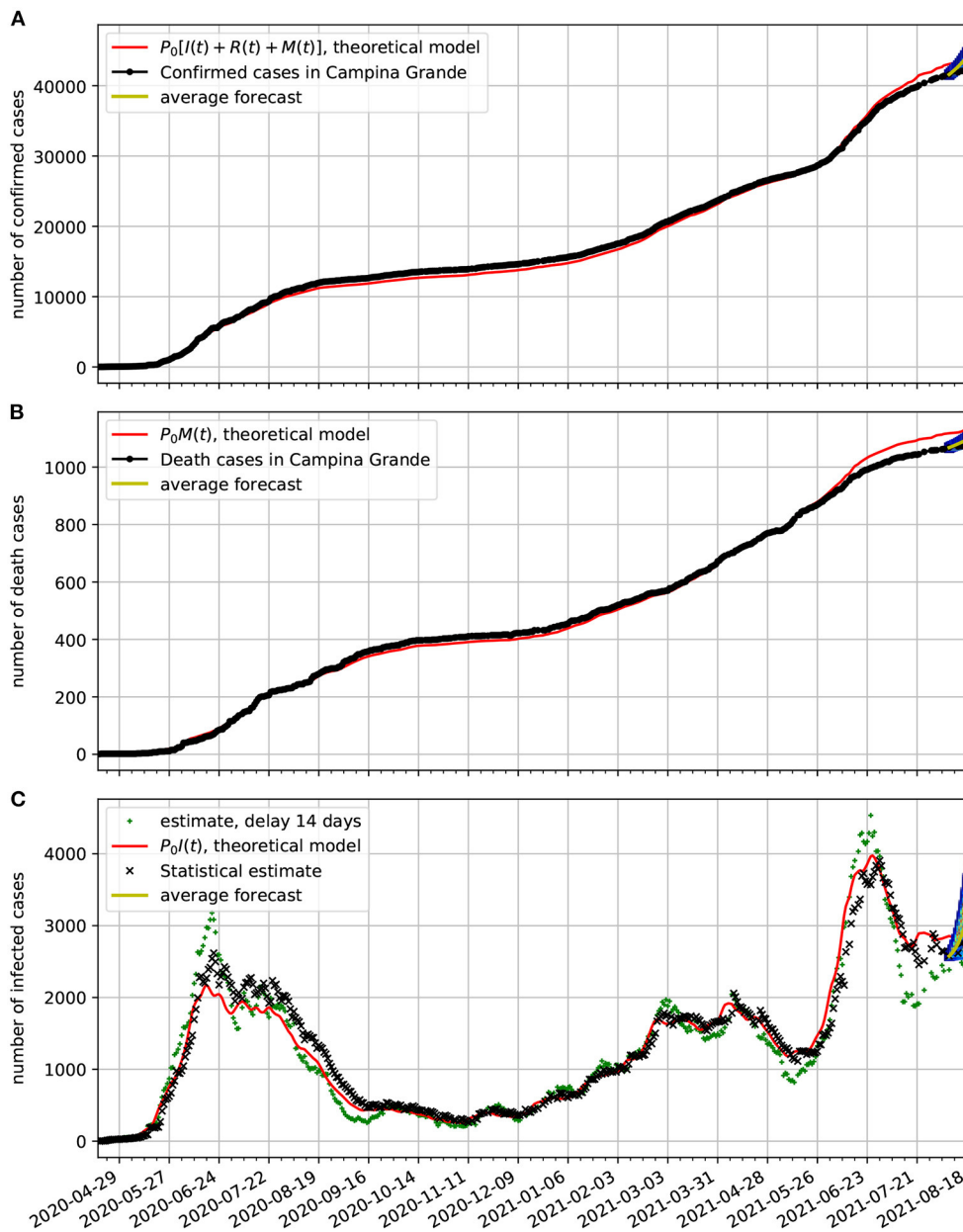


FIGURE 10 | Comparison of epidemiological data from Campina Grande with the theoretical fit based on our model. Also shown near the end of the time series are the statistical forecasting 95% confidence intervals, as described in Section 3.5. In all cases, we used $\tau = 14$ day and the parameters estimated in **Figure 9**. **(A)** Comparison of the number of official confirmed cases with the theoretical fit. **(B)** Comparison of the number of official deaths due to COVID-19 infection with the theoretical fit. **(C)** Comparison of the number of the estimated number of active cases with the theoretical fit.

than in the other populations studied, this might be linked to the smaller size of the population involved. In a similar way to Paraíba, from mid June/2021 onward there is a substantial decrease in the lethality probably caused by the anti-COVID-19 vaccination campaign.

We also proposed and validated a simple forecasting method based on Markov chains and on our parameter estimation method for the evolution of the epidemiological data for up to 2 weeks. For the populations we investigated, the epidemiological

data fell well within the 95% confidence interval of our predictions.

Furthermore, we calculated the NRMSE, which measures the fitting error of the differences between the epidemiological data and the estimates obtained from the evolution provided by our compartmental model with the time-dependent parameters. We use the NRMSE since it allows comparisons between curve fittings to data points at different time intervals or even between different datasets. In addition to that, we also analyzed the

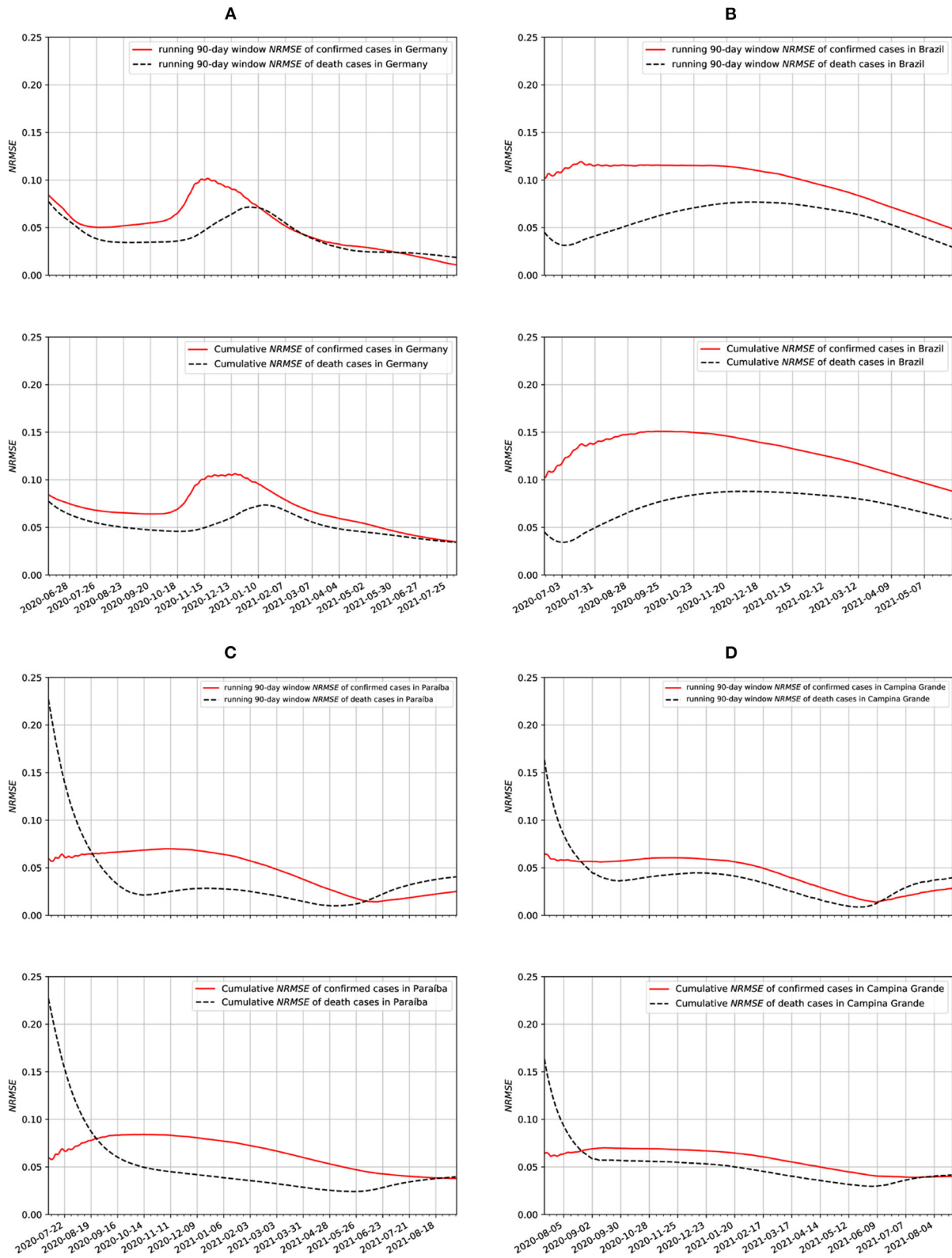
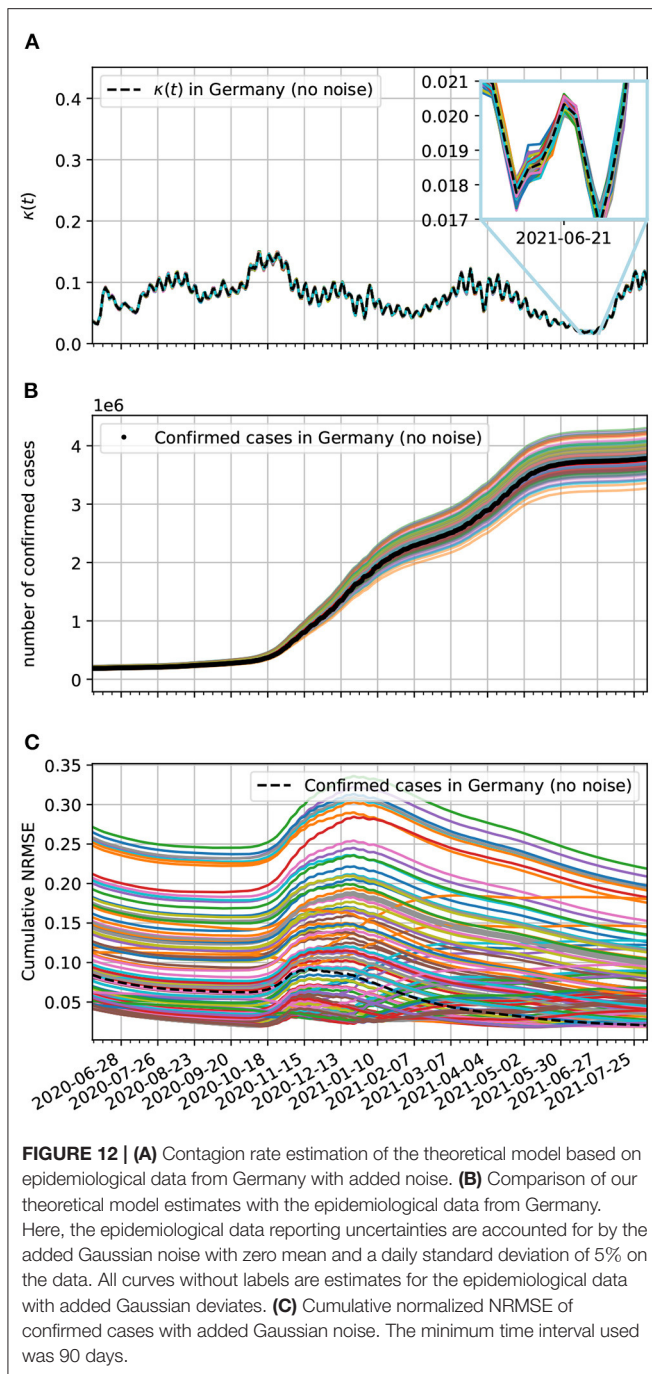


FIGURE 11 | Running window and cumulative NRMSE data for: **(A)** Germany, **(B)** Brazil, **(C)** Paraíba, and **(D)** Campina Grande.



robustness of our model when one introduces random errors in the epidemiological data. This mimics the fact that in real life, a fraction of the new cases of confirmed, deceased, and recovered, are reported with delays or are missed entirely.

Based on the results shown here, we conclude that the public health officials should look into the local dynamics of the spread of the disease as they compare with the theoretical predictions of models such as the one developed here. In this way, they will know where the social distancing and isolation is being more efficiently implemented. The models will be more relevant

and accurate if there is more testing of the population. Even random testing should be considered, as one gains statistical information on the spread of the disease and discovers where there is more under-notification. Also, cellphone data of the motion of the population, as used by Peixoto et al. [38] and Linka et al. [19], should be considered as a means of predicting the contagion and of identifying hot spots of COVID-19. This could help identify where there are more contagions: bars, churches, supermarkets, offices, pharmacies, hospitals, bakeries, restaurants, public transportation, delivery services, or family visits, etc. Furthermore, by comparing different local strategies one could gain insight on what works better for slowing the spread of the disease. It would be interesting to see the amount of correlation in population mobility and contagion rate, this could be specially relevant at the city level. In a more refined work, one could couple several nearby cities into a network of populations.

8. CODE AVAILABILITY

The code used in this paper is publicly available at the gitHub repository https://github.com/aabatista/covid19_parameter_estimation. See also the **Supplementary Material** for a description of the code.

DATA AVAILABILITY STATEMENT

All raw data supporting the conclusions of this article is publicly available. Covid-19 epidemiological data for Germany and Brazil was obtained from the site <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases> (accessed on 10/09/2021, with data collected until 09/09). The Covid-19 epidemiological data for the State of Paraíba and the City of Campina Grande was obtained from the site: <https://brasil.io/dataset/covid19/files/> (accessed on 10/09/2021, with data collected until 09/09).

AUTHOR CONTRIBUTIONS

AB designed research, developed compartmental model, statistical modeling, automated parameter estimation and forecasting, wrote and executed python code, wrote manuscript text, and made figures. SS designed research, developed compartmental model, wrote manuscript text, and wrote mathematical proofs. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Francisco A. Brito, Antônio A. Lisboa de Souza, Michelli Barros, Joelson Campos, Pammela Queiroz, and Aldo Trajano for suggestions during the development of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fams.2022.645614/full#supplementary-material>

REFERENCES

- Coronavirus disease 2019. (COVID-19) *Situation Report-41* (2020). Available online at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2 (accessed on June 15, 2020).
- Hui SD, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health: the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis.* (2020) 91:264–6. doi: 10.1016/j.ijid.2020.01.009
- WHO. *Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020* (2020). Available online at: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020> (accessed on May 27, 2020).
- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A.* (1927) 115:700–21. doi: 10.1098/rspa.1927.0118
- Alcaraz GG, Vargas-De-León C. Modeling control strategies for influenza A H1N1 epidemics: SIR models. *Revista Mexicana de Física.* (2012) 58:37–43. Available online at: <https://www.redalyc.org/articulo.oa?id=57030391007>
- Bastos SB, Cajueiro DO. Modeling and forecasting the COVID-19 pandemic in Brazil. *arXiv preprint arXiv:200314288.* (2020) doi: 10.1038/s41598-020-76257-1
- Isea R, Lonngren KE. On the mathematical interpretation of epidemics by Kermack and McKendrick. *Gen Math Notes.* (2013) 19:83–7. Available online at: https://www.emis.de/journals/GMN/yahoo_site_admin/assets/docs/6_GMN-3602-V19N2.32210220.pdf
- Crokidakis N. Data analysis and modeling of the evolution of COVID-19 in Brazil. *arXiv preprint arXiv:200312150.* (2020).
- Castilho C, Gondim JA, Marchesin M, Sabeti M. Assessing the efficiency of Differential control strategies for the COVID-19 epidemic. *Electron J Diff Equat.* (2020) 2020:1–17. Available online at: <https://ejde.math.txstate.edu/Volumes/2020/64/castilho.pdf>
- Kolokolnikov T, Iron D. Law of mass action and saturation in SIR model with application to Coronavirus modelling. *Infect Dis Model.* (2021) 6:91–7. doi: 10.1016/j.idm.2020.11.002
- Ndairou F, Area I, Nieto J, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals.* (2020) 135:1–6. doi: 10.1016/j.chaos.2020.109846
- Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med.* (2020) 9:462. doi: 10.3390/jcm9020462
- Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc Natl Acad Sci USA.* (2020) 117:10484–10491. doi: 10.1073/pnas.2004978117
- Smirnova A, deCamp L, Chowell G. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model. *Bull Math Biol.* (2019) 81:4343–65. doi: 10.1007/s11538-017-0284-3
- Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data driven analysis. *J Med Virol.* (2020) 92:645–59. doi: 10.1002/jmv.25750
- Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D. Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model. *IEEE Access.* (2020) 8:51761–9. doi: 10.1109/ACCESS.2020.2979599
- Chen YC, Lu PE, Chang CS. A Time-dependent SIR model for COVID-19. *arXiv preprint arXiv:200300122.* (2020).
- Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science.* (2020) 369. doi: 10.1126/science.aab9789
- Linka K, Goriely A, Kuhl E. Global and local mobility as a barometer for COVID-19 dynamics. *Biomech Model Mechanobiol.* (2021) 20:651–69. doi: 10.1007/s10237-020-01408-2
- Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. *Proc Natl Acad Sci USA.* (2020) 117:16732–16738. doi: 10.1073/pnas.2006520117
- Hale JK. *Ordinary Differential Equations.* Dover,VA: Dover Corporation (2009).
- Da Silva SH. Asymptotic behavior for a non-autonomous model of neural fields with variable external stimuli. *J Differ Equ.* (2020) 2020:1–16. Available online at: <https://ejde.math.txstate.edu/Volumes/2020/92/silva.pdf>
- Diekmann O, Getto P. Boundedness, global existence and continuous dependence for nonlinear dynamical systems describing physiologically structured populations. *J Diff Equat.* (2005) 215:268–319. doi: 10.1016/j.jde.2004.10.025
- Murray JD. *Mathematical Biology: I. An Introduction.* vol. 17. Springer Science Business Media; (2007).
- Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J Math Biol.* (1990) 28:365–82. doi: 10.1007/BF00178324
- van den Driessche PJW. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci.* (2002) 180:29–48. doi: 10.1016/S0025-5564(02)00108-6
- Ghosh S, Senapati A, Chattopadhyay J, Hens C, Ghosh D. Optimal test-kit-based intervention strategy of epidemic spreading in heterogeneous complex networks. *Chaos.* (2021) 31:071101. doi: 10.1063/5.0053262
- Bundesamt S. *Daten der Lebendgeborenen, Totgeborenen, Gestorbenen und der Gestorbenen im 1. Lebensjahr.* Available online at: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Geburten/Tabellen/lebensgeborene-gestorbene.html> (accessed on June 15, 2020).
- Bundesamt S. *Sterbefälle und Lebenserwartung.* Available online at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/_inhalt.html;jsessionid=CE8082CB6507F3C8C4A7C93E1AE3B3FC.internet8721 (accessed on June 15, 2020).
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods.* (2020) 17:261–72. doi: 10.1038/s41592-020-0772-5
- Welle D. *What are Germany's New Coronavirus Social Distancing Rules?* (2020). Available online at: <https://www.dw.com/en/what-are-germanys-new-coronavirus-social-distancing-rules/a-52881742> (accessed on July 23, 2020).
- Fernandes R. Compartmental epidemiological models for COVID-19: sources of uncertainty, goodness-of-fit and goodness-of-projections. *IEEE Latin Am Trans.* (2021) 19:1024–32. doi: 10.1109/TLA.2021.9451248
- Kırbaş I, Sözen A, Tuncer AD, Kazancıoğlu FŞ. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fractals.* (2020) 138:110015. doi: 10.1016/j.chaos.2020.110015
- Uba G, Yakasai H, Abubakar A, Abd Shukor MY. Predictive mathematical modelling of the total number of COVID-19 cases for Brazil. *J Environ Microbiol Toxicol.* (2020) 8:16–20. doi: 10.1038/s41598-021-95815-9
- Al-qaness MAA, Saba AI, Elsheikh AH, Elaziz MA, Ibrahim RA, Lu S, et al. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Safety Environm Protect.* (2021) 149:399–409. doi: 10.1016/j.psep.2020.11.007
- Gupta M, Jain R, Taneja S, Chaudhary G, Khari M, Verdú E. Real-time measurement of the uncertain epidemiological appearances of COVID-19 infections. *Appl Soft Comput.* (2021) 101:107039. doi: 10.1016/j.asoc.2020.107039

37. Hethcote HW. The mathematics of infectious diseases. *SIAM Rev.* (2000) 42:599–653. doi: 10.1137/S0036144500371907
38. Peixoto PS, Marcondes D, Peixoto C, Queiroz L, Gouveia R, Delgado A, et al. Potential dissemination of epidemics based on Brazilian mobile geolocation data. Part I: Population dynamics and future spreading of infection in the states of São Paulo and Rio de Janeiro during the pandemic of COVID-19. *medRxiv.* (2020) doi: 10.1101/2020.04.07.20056739

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Batista and da Silva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Here we summarize the proofs of proposition II.1, Proposition II.3 and Remark II.4.

A. Proof of Proposition II.1

Proof: Initially we will denote

$$\kappa_{max} = \sup_{t \in [0, T]} \kappa(t) \text{ and } \rho_{max} = \sup_{t \in [0, T]} \rho(t)$$

and we will consider \mathbb{R}^4 with the sum norm, that is, for $\xi = (S, I, R, M)$,

$$\|\xi\| = |S| + |I| + |R| + |M|,$$

where $|S| \leq S_{max}$, $|I| \leq I_{max}$, $|R| \leq R_{max}$ and $|M| \leq M_{max}$.

Hence, it is easy to see that

$$\begin{aligned} |g_1(t, \xi_1) - g_1(t, \xi_2)| &\leq |v - \mu||S_1 - S_2| + v|I_1 - I_2| + v|R_1 - R_2| + \kappa(t)|S_1I_1 - S_2I_2| \\ &\leq |v - \mu||S_1 - S_2| + v|I_1 - I_2| + v|R_1 - R_2| + \kappa(t)|S_1||I_1 - I_2| \\ &\quad + \kappa(t)|I_2||S_2 - S_1| \\ &\leq |v - \mu||S_1 - S_2| + v|I_1 - I_2| + v|R_1 - R_2| + \kappa_{max}S_{max}|I_1 - I_2| \\ &\quad + \kappa_{max}I_{max}|S_1 - S_2| \\ &\leq (|v - \mu| + \kappa_{max}I_{max})|S_1 - S_2| + (v + \kappa_{max}S_{max})|I_1 - I_2| \\ &\quad + v|R_1 - R_2|. \end{aligned}$$

Hence, by writing $L_1 = \sup\{|v - \mu| + \kappa_{max}I_{max}, v + \kappa_{max}S_{max}\}$, we obtain

$$\begin{aligned} |g_1(t, \xi_1) - g_1(t, \xi_2)| &\leq L_1 (|S_1 - S_2| + |I_1 - I_2| + |R_1 - R_2|) \\ &\leq L_1 \|\xi_1 - \xi_2\|. \end{aligned}$$

Similarly,

$$\begin{aligned} |g_2(t, \xi_1) - g_2(t, \xi_2)| &\leq \left(\mu + \frac{1}{\tau}\right) |I_2 - I_1| + \kappa(t)|S_1I_1 - S_2I_2| \\ &\leq \left(\mu + \frac{1}{\tau}\right) |I_2 - I_1| + \kappa(t)|S_1||I_2 - I_1| + \kappa(t)|I_2||S_1 - S_2| \\ &\leq \kappa_{max}I_{max}|S_1 - S_2| + \kappa_{max}S_{max}|I_1 - I_2| + \left(\mu + \frac{1}{\tau}\right)|I_2 - I_1|; \end{aligned}$$

thus, writing $L_2 = \max\{\kappa_{max}I_{max}, \kappa_{max}S_{max} + \mu + \frac{1}{\tau}\}$ we have

$$\begin{aligned} |g_2(t, \xi_1) - g_2(t, \xi_2)| &\leq L_2 (|S_1 - S_2| + |I_1 - I_2|) \\ &\leq L_2 \|\xi_1 - \xi_2\|; \end{aligned}$$

$$\begin{aligned} |g_3(t, \xi_1) - g_3(t, \xi_2)| &\leq \rho(t)|I_1 - I_2| + \mu|R_2 - R_1| \\ &\leq \rho_{max}|I_1 - I_2| + \mu|R_1 - R_2| \\ &\leq L_3 (|I_1 - I_2| + |R_1 - R_2|) \\ &\leq L_3 \|\xi_1 - \xi_2\|, \end{aligned}$$

where $L_3 = \max\{\rho_{max}, \mu\}$ and

$$|g_4(t, \xi_1) - g_4(t, \xi_2)| \leq \lambda(t)|I_1 - I_2|.$$

Since $\frac{1}{\tau} = \rho(t) + \lambda(t)$, it follows that $\lambda(t) = \frac{1}{\tau} - \rho(t)$. Thus

$$\lambda_{max} = \sup_{t \in [0, T]} \left(\frac{1}{\tau} - \rho(t) \right) < \tau^{-1}.$$

Hence,

$$|g_4(t, \xi_1) - g_4(t, \xi_2)| \leq \frac{1}{\tau} |I_1 - I_2| \leq \frac{1}{\tau} \|\xi_1 - \xi_2\|.$$

Therefore,

$$\|g(t, \xi_1) - g(t, \xi_2)\| \leq L \|\xi_1 - \xi_2\|,$$

where $L = \max\{L_1, L_2, L_3, \frac{1}{\tau}\}$. This concludes the proof.

B. Proof of Proposition II.3

Proof: Define $\theta : [0, T] \rightarrow \mathbb{R}^3$ by

$$\theta(t) = (\kappa(t), \rho(t), \lambda(t)).$$

Consider the metric defined on space of the continuous functions of $[0, T]$ in $\mathbb{R}^3, \mathbb{C}([0, T], \mathbb{R}^3)$, given by

$$dist(\theta, \theta_0) = \sup_{t \in [0, T]} |\kappa(t) - \kappa_0(t)| + \sup_{t \in [0, T]} |\rho(t) - \rho_0(t)| + \sup_{t \in [0, T]} |\lambda(t) - \lambda_0(t)|.$$

Now, we denote by $\xi_\theta(t)$ the solutions of Equation (1) with respect to parameter $\theta(t) = (\kappa(t), \rho(t), \lambda(t))$ such that $\xi_\theta(0) = \xi_0$ and we denote by $\xi_{\theta_0}(t)$ the solutions with respect to parameter $\theta_0(t) = (\kappa_0(t), \rho_0(t), \lambda_0(t))$ such that $\xi_{\theta_0}(0) = \xi_0$.

We need to show that

$$\|\xi_\theta(t) - \xi_{\theta_0}(t)\| \rightarrow 0, \text{ as } dist(\theta, \theta_0) \rightarrow 0.$$

For this, note that

$$\|\xi_\theta(t) - \xi_{\theta_0}(t)\| \leq \int_0^t \|g(s, \xi_\theta(s)) - g(s, \xi_{\theta_0}(s))\| ds. \quad (25)$$

Now

$$\begin{aligned} |g_1(s, \xi_\theta(s)) - g_1(s, \xi_{\theta_0}(s))| &= |v[S(s) - S_0(s)] + v[I(s) - I_0(s)] + v[R(s) - R_0(s)] \\ &\quad + \mu[S_0(s) - S(s)] + \kappa(s)S(s)I(s) - \kappa_0(s)S_0(s)I_0(s)| \\ &\leq (v + \mu)|S(s) - S_0(s)| + v|I(s) - I_0(s)| + v|R(s) - R_0(s)| \\ &\quad + |\kappa(s)S(s)I(s) - \kappa_0(s)S(s)I(s)| \\ &\quad + |\kappa_0(s)S(s)I(s) - \kappa_0(s)S_0(s)I_0(s)|. \end{aligned}$$

But

$$|\kappa(s)S(s)I(s) - \kappa_0(s)S(s)I(s)| = |S(s)I(s)||\kappa(s) - \kappa_0(s)|$$

and

$$|\kappa_0(s)S(s)I(s) - \kappa_0(s)S_0(s)I_0(s)| = |\kappa_0(s)||S(s)I(s) - S_0(s)I_0(s)| \leq |\kappa_0(s)| (|S(s)||I(s) - I_0(s)| + |I_0(s)||S(s) - S_0(s)|).$$

Hence,

$$\begin{aligned} |g_1(s, \xi_\theta(s)) - g_1(s, \xi_{\theta_0}(s))| &\leq (v + \mu)|S(s) - S_0(s)| + v|I(s) - I_0(s)| + v|R(s) - R_0(s) \\ &+ S_{max} I_{max} \sup_{s \in [0, T]} |\kappa(s) - \kappa_0(s)| \\ &+ \kappa_{0max} (S_{max} |I(s) - I_0(s)| + I_{0max} |S(s) - S_0(s)|) \\ &\leq (v + \mu + \kappa_{0max} I_{0max})|S(s) - S_0(s)| \\ &+ (v + \kappa_{0max} S_{max})|I(s) - I_0(s)| + v|R(s) - R_0(s)| \\ &+ S_{max} I_{max} \text{dist}(\theta, \theta_0). \end{aligned}$$

Thus, by writing $G_1 = \{v + \mu + \kappa_{0max} I_{0max}, v + \kappa_{0max} S_{max}\}$, we obtain

$$|g_1(s, \xi_\theta(s)) - g_1(s, \xi_{\theta_0}(s))| \leq G_1 \|\xi_\theta(s) - \xi_{\theta_0}(s)\| + S_{max} I_{max} \text{dist}(\theta, \theta_0). \tag{26}$$

$$\begin{aligned} |g_2(s, \xi_\theta(s)) - g_2(s, \xi_{\theta_0}(s))| &= \left| \left(\mu + \frac{1}{\tau} \right) (I_0 - I) + \kappa(s)S(s)I(s) - \kappa_0(s)S_0(s)I_0(s) \right| \\ &\leq \left(\mu + \frac{1}{\tau} \right) |I_0 - I| + |S(s)I(s)||\kappa(s) - \kappa_0(s)| \\ &+ |\kappa_0(s)||S(s)I(s) - S_0(s)I_0(s)| \\ &\leq \left(\mu + \frac{1}{\tau} \right) |I_0 - I| + S_{max} I_{max} |\kappa(s) - \kappa_0(s)| \\ &+ \kappa_{0max} (|S(s)||I(s) - I_0(s)| + |I_0(s)||S(s) - S_0(s)|) \\ &\leq \kappa_{0max} I_{0max} |S(s) - S_0(s)| + \left(\mu + \frac{1}{\tau} + \kappa_{0max} S_{max} \right) |I_0 - I| \\ &+ S_{max} I_{max} \text{dist}(\theta, \theta_0) \\ &\leq G_2 \|\xi_\theta - \xi_{\theta_0}\| + S_{max} I_{max} \text{dist}(\theta, \theta_0), \end{aligned} \tag{27}$$

where $G_2 = \max\{\kappa_{0max} I_{0max}, \mu + \frac{1}{\tau} + \kappa_{0max} S_{max}\}$.

$$\begin{aligned} |g_3(s, \xi_\theta(s)) - g_3(s, \xi_{\theta_0}(s))| &= |\rho(s)I(s) - \rho_0(s)I_0(s) + \mu[R_0(s) - R(s)]| \\ &\leq |I(s)||\rho(s) - \rho_0(s)| + |\rho_0(s)||I(s) - I_0(s)| \\ &+ \mu|R_0(s) - R(s)| \\ &\leq I_{max} \sup_{s \in [0, T]} |\rho(s) - \rho_0(s)| + \left(\sup_{s \in [0, T]} \rho_0(s) \right) |I(s) - I_0(s)| \\ &+ \mu|R_0(s) - R(s)| \\ &\leq I_{max} \text{dist}(\theta, \theta_0) + \rho_{0max} |I(s) - I_0(s)| \\ &+ \mu|R_0(s) - R(s)| \\ &\leq G_3 \|\xi_\theta(s) - \xi_{\theta_0}(s)\| + I_{max} \text{dist}(\theta, \theta_0), \end{aligned} \tag{28}$$

where $G_3 = \max\{\rho_{0max}, \mu\}$. Now,

$$\begin{aligned} |g_4(s, \xi_\theta(s)) - g_4(s, \xi_{\theta_0}(s))| &= |\lambda(s)I(s) - \lambda_0(s)I_0(s)| \\ &\leq |I(s)||\lambda(s) - \lambda_0(s)| + \lambda_0(s)|I(s) - I_0(s)| \\ &\leq I_{max} \left(\sup_{s \in [0, T]} |\lambda(s) - \lambda_0(s)| \right) + \left(\frac{1}{\tau} - \rho_0(s) \right) |I(s) - I_0(s)| \\ &\leq \frac{1}{\tau} \|\xi_\theta(s) - \xi_{\theta_0}(s)\| + I_{max} \text{dist}(\theta, \theta_0). \end{aligned} \tag{29}$$

Thus, by using Equations (26), (27), (28) and (29) in (25), we find

$$\begin{aligned} \|\xi_\theta(t) - \xi_{\theta_0}(t)\| &\leq \int_0^t \left\{ 2(S_{max} I_{max} + I_{max}) \text{dist}(\theta, \theta_0) + (G_1 + G_2 + G_3 + \frac{1}{\tau}) \|\xi_\theta(s) - \xi_{\theta_0}(s)\| \right\} ds \\ &\leq 2T(S_{max} I_{max} + I_{max}) \text{dist}(\theta, \theta_0) + \int_0^t \left(G_1 + G_2 + G_3 + \frac{1}{\tau} \right) \|\xi_\theta(s) - \xi_{\theta_0}(s)\| ds. \end{aligned}$$

Therefore, from Gronwall Lemma, it follows that

$$\|\xi_\theta(t) - \xi_{\theta_0}(t)\| \leq 2T(S_{max} I_{max} + I_{max}) \text{dist}(\theta, \theta_0) e^{(G_1 + G_2 + G_3 + \frac{1}{\tau})T} \rightarrow 0,$$

as $\text{dist}(\theta, \theta_0) \rightarrow 0$.

C. Positivity of Solutions

Here we develop a constructive proof. For a more abstract alternative proof, see the supplementary material. Notice that, in the system of Equation (1), κ , ρ , and λ are non-negative continuous functions and μ and v are positive constants. Furthermore, we have the following initial data $S(0) = 1 - \frac{1}{P_0} > 0$, $I(0) = \frac{1}{P_0} > 0$, $R(0) = 0$ and $M(0) = 0$.

From the second line of the ODE system of Equation (1), we have

$$I'(t) = \left[- \left(\mu + \frac{1}{\tau} \right) + \kappa(t)S(t) \right] I(t).$$

Then

$$\frac{I'(t)}{I(t)} = - \left(\mu + \frac{1}{\tau} \right) + \kappa(t)S(t).$$

Integrating from 0 to t , we obtain

$$\ln \frac{I(t)}{I(t_0)} = \int_{t_0}^t \left[- \left(\mu + \frac{1}{\tau} \right) + \kappa(s)S(s) \right] ds$$

Thus, since $I_0 > 0$, we have

$$I(t) = I_0 e^{\int_{t_0}^t [-(\mu + \frac{1}{\tau}) + \kappa(s)S(s)] ds} > 0. \tag{30}$$

From third equation in (1), we have

$$R'(t) + \mu R(t) = \rho(t)I(t).$$

Multiplying both members of the last equation by $e^{\mu t}$, it follows that

$$e^{\mu t} R'(t) + \mu e^{\mu t} R(t) = e^{\mu t} \rho(t)I(t)$$

or equivalently

$$\frac{d}{dt} (e^{\mu t} R(t)) = e^{\mu t} \rho(t)I(t)$$

Integrating from 0 to t , we obtain

$$e^{\mu t} R(t) - e^{\mu t} R(t_0) = \int_0^t e^{\mu s} \rho(s)I(s) ds.$$

Hence

$$R(t) = e^{-\mu(t-t_0)}R_0 + \int_0^t e^{\mu s} \rho(s)I(s)ds.$$

Since $R(t_0) = R_0 = 0$

$$R(t) = \int_0^t e^{\mu s} \rho(s)I(s)ds > 0, \tag{31}$$

because $\rho(s) > 0$ and, from (30), $I > 0$.

Now, from fourth equation in (1), we have

$$M'(t) = \lambda(t)I(t)$$

Thus

$$M(t) = M(t_0) + \int_{t_0}^t \lambda(s)I(s)ds$$

Since $M(t_0) = 0$, $\lambda > 0$ and from (30), $I > 0$, it follows that $M(t) > 0$ for all $t \in [0, T]$.

Finally, from first equation in (1), it follows that

$$S'(t) + (\kappa(t)I(t) + \mu - \nu)S(t) = \nu I(t) + \nu R(t)$$

Let $w(t) = \int_{t_0}^t \kappa(s)I(s)ds$ be. Then, multiplying both members by $e^{w(t)+(\mu-\nu)t}$, we obtain

$$\frac{d}{dt} \left[e^{w(t)+(\mu-\nu)t} S(t) \right] = \nu(I(t) + R(t))e^{w(t)+(\mu-\nu)t}.$$

Integrating from 0 to t , we obtain

$$e^{w(t)+(\mu-\nu)t} S(t) - e^{w(t_0)+(\mu-\nu)t_0} S(t_0) = \int_{t_0}^t e^{w(s)+(\mu-\nu)s} \nu(I(s) + R(s))ds.$$

Since $w(t_0) = 0$, it follows that

$$S(t) = e^{-w(t)} e^{-(\mu-\nu)(t-t_0)} S(t_0) + \int_{t_0}^t e^{-(w(t)-w(s))-(\mu-\nu)(t-s)} \nu(I(s) + R(s))ds.$$

As $S(t_0) = \frac{1}{P_0} > 0$, $\nu > 0$ and from (30) and (31) we have that $I > 0$ and $R > 0$, it follows that $S(t) > 0$ for all $t \in [0, T]$.

D. Boundedness of Trajectories

If $\mu = \nu$, then, from Equation (1),

$$\frac{d}{dt}(S + I + R + M) = 0. \tag{32}$$

Hence, $S(t) + I(t) + R(t) + M(t) = S_0 + I_0 + R_0 + M_0 = 1$. As the initial condition is $S_0 = 1 - 1/P_0$ and $I_0 = 1/P_0$, $R_0 = M_0 = 0$ and state variables are all non-negative, as shown above, we obtain that $0 \leq S(t), I(t), R(t), M(t) \leq 1$. In the case of $\mu \neq \nu$, from Equation (1), we can write

$$\frac{d}{dt}(S + I + R + M) = (\nu - \mu)(S + I + R) \leq (\nu - \mu)(S + I + R + M), \tag{33}$$

assuming $\nu > \mu$. Hence, we obtain $S(t) + I(t) + R(t) + M(t) \leq e^{(\nu-\mu)t}$. Therefore, as $S(t), I(t), R(t)$, and $M(t)$ are positive, they are all bounded in any given finite time interval. If $\mu > \nu$, we obtain $S(t) + I(t) + R(t) + M(t) \leq 1$.