On the use of growth models for forecasting epidemic outbreaks with application to COVID-19 data

Chénangnon F. Tovissodé[1], Bruno E. Lokonon[1] and Romain Glèlè Kakaï[1]

1 Laboratoire de Biomathématiques et d'Estimations Forestières/Université d'Abomey-Calavi, Benin

Email: romain.glelekakai@fsa.uac.bj

Abstract

The initial phase dynamics of an epidemic without containment measures is commonly well modeled using exponential growth models. However, in the presence of containment measures, the exponential model becomes less appropriate. Under the implementation of an isolation measure for detected infectives, we propose to model epidemic dynamics by fitting a flexible growth model curve to reported positive cases and to infer the overall epidemic dynamics by introducing information on the detection/testing effort and recovery and death rates. The resulting modeling approach is close to the SIQR (Susceptible-Infectious-Quarantined-Recovered) model framework. We focused on predicting the peaks (time and size) in positive cases, actives cases and new infections. We applied the approach to data from the COVID-19 outbreak in Italy. Fits on limited data before the observed peaks illustrate the ability of the flexible growth model to approach the estimates from the whole data.

Key words: Epidemic dynamic; Turner's generic growth model; Positive cases; Active cases; Lost cases; COVID-19

1 Introduction

The current COVID-19 pandemic caused by the new coronavirus strain SARS-nCOV2 has emerged from Wuhan, China (Giordano et al., 2020, Velavan and Meyer, 2020). The COVID-19 outbreaks totalize 21 026 758 cases and 755 786 deaths across the world on August 15, 2020 (WHO, 2020).

The worldwide social as well as economic ravages by COVID-19 has immediately motivated the use of mathmatical models for understanding the course of the epidemic and planning effective control strategies. These include for instance the SIR (Susceptible, Infectious, Recovered), SEIR (Susceptible, Exposed, Infectious, Recovered) and its variants, SIDR (Susceptible- Infectious-Recovered-Dead) and SIQR (Susceptible-Infectious-Quarantined-Recovered) models (Anastassopoulou et al., 2020, Casella, 2020, Kucharski et al., 2020, Pedersen and Meneghini, 2020). These modeling approaches use mechanistic models which incorporate key physical laws or mechanisms involved in the dynamics of the population at risk and the pathogen (Chowell, 2017). A second class of approaches uses empirical phenomenological models which does not require specific knowledge on the physical laws or mechanisms that give rise to the observed epidemic data (Chowell et al., 2016), and has been considered for instance by Agosto and Giudici (2020) for understanding both short and long term dynamics of COVID-19.

When facing an epidemic outbreak, public health officials are mostly interested in data driven, mathematically motivated, practical and computationally efficient approaches that can for instance *i*) generate estimates of key transmission parameters, *ii*) gain insight to the contribution of different transmission pathways, *iii*) assess the impact of control interventions (*e.g.* social distancing, test + isolation, vaccination campaigns) and *iv*) optimize the impact of control strategies, and *v*) generate short and long-term forecasts (Chowell, 2017). In regard to the current COVID-19 outbreak (2020), politics and public health officials are mostly worried about the ability of the disease to induce saturation of the health system, reducing the survival of patients even consulting for reasons different from the epidemic itself. High interest is thus currently given to accurate forecasting of the epidemic peak time and size, epidemic size and duration, as well as their sensitivity to control interventions in order to optimize the impact of control strategies.

An exponential-growth model is usually assumed to characterize the early phase of epidemics. But this assumption can lead to failure to appropriately capture the profile of epidemic growth and give rise to non realistic epidemic forecasts (Chowell and Viboud, 2016, Spencer and Golinski, 2020). In an ultimate view to guide control interventions aiming to limit the spread of epidemics, with focus on the COVID-19 pandemic, this work considers a flexible growth curve fitting approach for understanding the dynamics of epidemics. We use the generic growth model of Turner et al. (1976) to model the course of reported positive cases and a binomial regression to model removals (recoveries and deaths) and then infer the overall dynamics of the epidemic, in terms of observables (reported cases, active/quarantined cases) and unobservables (new infections, lost cases), and predict interest quantities such as the peak (time and size) in reported cases, active cases and new infections. The performance of the approach is assessed through

an application to daily case reporting data from Italy which has virtually completed a whole COVID-19 outbreak wave and thus offers the possibility to compare predicted outputs to real events.

2 Methods

We consider a growth curve approach for modeling the course of an epidemic along time. We follow Chowell et al. (2020), Chowell (2017), Spencer and Golinski (2020) who among others used growth models for forecasting epidemic dynamics.

2.1 Structural model for epidemic incidence

Let C_t denote the size of the detected infected population at time t, *i.e.* the cumulative number of infected, identified and isolated individuals. We assume for convenience that C_t is continuous and denote $\dot{C}_t$ its first derivative with respect to t. Also let I_t be the true size of infectives at t, related to C_t through

$$\dot{C}_t = \delta_t I_t \tag{1}$$

where $\delta_t \in (0, 1]$ is the detection rate which is closely related to the testing effort (number of tests, tracing of contact persons of identified cases and targeting exposed people) and is assumed at least twice differentiable with respect to t. We ressort to the generic growth model of Turner et al. (1976) for the identified positive cases:

$$C_t = K(1 + u_t)^{-1/\nu} \tag{2}$$

with $u_t = [1 + \nu\omega\rho(t - \tau)]^{-1/\rho}$. In (2), $K > 0$ is the ultimate epidemic size (detected), $\omega > 0$ is the "intrinsic" growth constant, ν and ρ are powers ($\nu > 0$ and $-1 < \rho < \nu^{-1}$) characterizing respectively the rates of change with respect to the initial size $C_0 = \delta_0 I_0$ (number of cases detected at time $t = 0$) and the ultimate size K, and τ is a constant of integration, determined by the initial conditions of the epidemic and implicitly the detection rate δ_0 through $C_0 = K\left[1 + (1 - \nu\omega\rho\tau)^{-1/\rho}\right]^{-1/\nu}$ for $\rho \neq 0$ and $C_0 = K\left(1 + \mathrm{e}^{\nu\omega\tau}\right)^{-1/\nu}$ for $\rho = 0$. The growth model (2) is quite flexible to handle various shapes of epidemic dynamics. Indeed, if $K \to \infty$ and $\nu\rho \to 0$, (2) specializes to the exponential growth model

$$C_t = \mathrm{e}^{\omega(t-\tau)} \tag{3}$$

where ω is the exponential growth rate. Apart from (3), other special or limiting cases of (2) include the hyper-Gompertz ($\nu \to 0$ while $\omega\nu^{1+\rho}$ is constant) and the Gompertz ($\nu \to 0$, $\rho \to 0$ while $\omega\nu$ is constant),

the Bertalanffy-Richards ($\rho \to 0$), the hyper-logistic ($\nu = 1$) and the logistic ($\nu = 1$ and $\rho \to 0$) growth models (Turner et al., 1976). From (2), the observed epidemic incidence $\dot{C}_t$ is given by

$$\dot{C}_t = K\omega u_t^{1+\rho}(1+u_t)^{-\frac{\nu+1}{\nu}}. \tag{4}$$

In order to ensure the restriction $-1 < \rho < \nu^{-1}$, we set $\rho = \rho_0 \frac{\nu+1}{\nu} - 1$ with $\rho_0 \in (0,1)$ free of ν.

2.2 Active cases and outcomes

The number A_t of detected and active cases along an epidemic outbreak is of high interest for public health officials. Indeed, A_t must be kept under the carrying capacity of the health system to avoid overload and disrupture. The derivative $\dot{A}_t$ of the detected and active cases satisfies

$$\dot{A}_t \;=\; \dot{C}_t - R_t \tag{5}$$

where $R_t = \alpha_t A_t$ denotes the number of removed and permanently immune (mortality and recovery) at time t, and α_t is the unit time removal probability, *i.e.* the odds for having an outcome (recovery or death), averaged over the active cases. Equation (5) fits in the SIQR (Susceptible, Infectious, Quarantined, Recovered) model framework (Hethcote et al., 2002) with the detected actives cases referred to as "quarantined" and the strong assumption that α_t is constant along the epidemic outbreak (see the third equation in system (6) in Hethcote et al. (2002)). The removal probability can more generally be given the logistic form $\alpha_t = \frac{e^{\eta_t}}{1+e^{\eta_t}}$ with $\eta_t = \boldsymbol{X}_t^\top \boldsymbol{\beta} + \kappa t$ where $\boldsymbol{X}_t = (X_{t1}, X_{t2}, \cdots, X_{tq})^\top$ is a vector of q covariates (known constants) and $\boldsymbol{\beta}$ is the q vector of associated effects, and κ determines the change in the log-odds ratio for having an outcome per unit time. These changes in α_t can be due to an improve in the health care system during the epidemic outbreak (increase in recovery ratio) or a deterioration of the health care system for infected individuals (increase in mortality ratio due to the outbreak). The general solution of the differential equation (5) turns to have the form

$$A_t \;=\; \begin{cases} \left[A_0 + \int_0^t \dot{C}_s e^{\alpha_s s} ds\right] e^{-\alpha_t t} & \text{if} \quad \kappa = 0 \\ \left[A_0\left(1+e^{\boldsymbol{X}_t^\top \boldsymbol{\beta}}\right)^{1/\kappa} + \int_0^t \dot{C}_s \left(1+e^{\eta_s}\right)^{1/\kappa} ds\right] \left(1+e^{\eta_t}\right)^{-1/\kappa} & \text{if} \quad \kappa \neq 0 \end{cases} \tag{6}$$

where A_0 is the number of active cases at time $t = 0$. Indeed, when $\kappa = 0$, taking the first derivative of (6) yields $\dot{A}_t = \dot{C}_t e^{\alpha_t t} e^{-\alpha_t t} + \left[A_0 + \int_0^t \dot{C}_s e^{\alpha_s s} ds\right](-\alpha_t) e^{-\alpha_t t} = \dot{C}_t - \alpha_t A_t$ which is the equation (5). For $t = 0$, the integral in (6) vanishes, resulting as expected in $A_t = A_0$ since $e^{-\alpha_t t} = 1$. When $\kappa \neq 0$, the first derivative of (6) is $\dot{A}_t = \dot{C}_t \left(1+e^{\eta_t}\right)^{1/\kappa} \left(1+e^{\eta_t}\right)^{-1/\kappa} + A_t \left(\frac{1}{-\kappa}\right)(\kappa e^{\eta_t}) \left(1+e^{\eta_t}\right)^{-1} = \dot{C}_t - \frac{e^{\eta_t}}{1+e^{\eta_t}} A_t$ which reduces to $\dot{A}_t = \dot{C}_t - \alpha_t A_t$ in accordance with (5). Here, for $t = 0$, $e^{\eta_t} = e^{\boldsymbol{X}_t^\top \boldsymbol{\beta}}$ so that $A_t = A_0$.

There are no general closed form solutions for the integrals in (6), unless $\dot{C}_t$ and α_t are purposely chosen as functions of time to simplify the integral. A_t can however be obtained in practice from (6) using a numerical integration routine such as the function *integrate* in R freeware (R Core Team, 2019) or the function *integral* of Matlab (MATLAB, 2016). To nevertheless circumvent this issue during estimation under the generic growth model (2), we discretized the active cases A_t by assuming a binomial removal process R_t conditional on the detected unit time new cases Y_t as

$$R_t|A_{t-1}, Y_t \sim \text{BIN}(A_{t-1} + Y_t, \alpha_t) \tag{7}$$

$$A_t = A_{t-1} + Y_t - R_t \tag{8}$$

where $\text{BIN}(n, \alpha)$ denotes a binomial distribution with n trials and success probability α and Y_t is a non negative process with expectation $\lambda_t = \dot{C}_t$. Clearly, the bivariate process $\{A_t, R_t\}$ defined by (8-7) is not stationary. However, since $Y_t \geq 0$ and $\dot{C}_t \to 0$ as $t \to \infty$, we have $Y_t \to 0$ in distribution as $t \to \infty$, and if the removal probability α_t does not approach zero as $t \to \infty$, then $A_t \to 0$ as $t \to \infty$.

2.3 Peak of detected cases

The epidemic peak is an important event in the disease dynamic and can be estimated for a better management of the epidemic. For an epidemic described by the exponential growth model (3) ($K \to \infty$ or $\nu\rho \to 0$) does not peak. Otherwise, the peak in the detected number of infected individuals corresponds to the maximum of the incidence rate $\dot{C}_t$. This maximum is then attained when $\ddot{C}_t = \frac{\partial \dot{C}_t}{\partial t} = 0$. We have from (4)

$$\ddot{C}_t = \nu\omega u_t^{\rho}\left[\frac{\nu+1}{\nu}\frac{u_t}{1+u_t} - (1+\rho)\right]\dot{C}_t. \tag{9}$$

Solving $\ddot{C}_t = 0$ for t using (9) yields the peak time $t_p = \tau + \left\{\left[\frac{1-\nu\rho}{\nu(1+\rho)}\right]^{\rho} - 1\right\}/(\nu\omega\rho)$ which reads,

$$t_p = \tau + \frac{1}{\omega\left[\rho_0 - (1-\rho_0)\nu\right]}\left[\left(\frac{\rho_0}{1-\rho_0}\right)^{1-\rho_0(\nu+1)/\nu} - 1\right] \tag{10}$$

on replacing $\rho = \rho_0\frac{\nu+1}{\nu} - 1$. Inserting t_p in (1) and denoting $u_p = \nu\frac{1+\rho}{1-\rho\nu}$ gives the peak

$$\dot{C}_p = K\omega u_p^{1+\rho}\left(\frac{\nu+1}{1-\rho\nu}\right)^{-\frac{\nu+1}{\nu}}. \tag{11}$$

At the peak in detected cases, the cumulative number of detected cases is $C_p = K(1+u_p)^{-1/\nu}$.

2.4 Overall epidemic dynamics

An important interest in modeling the epidemic incidence is the derivation of quantities related to the overall dynamics of the epidemic, in both detected and undetected cases.

2.4.1 Total cases: detected and losts

Let us denote S_t the cumulative number of cases from the epidemic outbreak to t, and let $\dot{S}_t$ be the first derivative of S_t. We also introduce Λ_t, the cumulative number of lost cases (with first derivative $\dot{\Lambda}_t$), *i.e.* people who were infected, undetected, and removed from infectives (mortality and recovery).

The size of the lost cases is determined by the unit time removal rate $\pi_t \in (0,1)$ from undetected infectives (π_t is an average over all infectives, *i.e.* irrespective of the time since infection onset). The lost rate π_t which is assumed at least twice differentiable with respect to t, depends on various factors like the disease related mortality, the average infection duration, the natural proportion of asymptomatics within infectives, and the existence and the use of medicines that may reduce symptoms (induced asymptomatics). It is worthwhile noticing that π_t can be estimated from the removal rate α_t in the detected cases (see section 2.2), taking into account various factors that may induce difference between the two rates. For instance, since the undetected cases include asymptomatics, desease related mortality may be lower and recovery rate higher in undetected as compared to detected cases. However, efficiency of the health care system in treating identified and isolated cases can reduce mortality thereby reducing α_t, but also improve recovery thereby increasing α_t.

With the above notations, the lost cases count Λ_t satisfies the differential equation,

$$\dot{\Lambda}_t = \pi_t(1-\delta_t)I_t \tag{12}$$

whereas the cumulative number of cases S_t is given on setting $\upsilon_t = (1-\pi_t)(1-\delta_t)$ by

$$S_t = C_t + \Lambda_t + \upsilon_t I_t. \tag{13}$$

The factor υ_t represents at time t the proportion of infectives who will potentially continue to spread the epidemic after adequate contacts (*i.e.* contacts sufficient for transmission) with susceptibles. In other words, the number of undetected currently infectives is $(1-\pi_t)(\delta_t^{-1}-1)\dot{C}_t$. From (1), the infectives I_t and its first derivative with respect to time $\dot{I}_t$ are given for $t \geq 0$ by

$$I_t = \delta_t^{-1}\dot{C}_t \tag{14}$$

$$\dot{I}_t = \delta_t^{-1}\left[\ddot{C}_t - \dot{\delta}_t\delta_t^{-1}\dot{C}_t\right] \tag{15}$$

where $\dot{\delta}_t$ is the first derivative of the detection rate δ_t with respect to t. Straightforward algebraic

operations then give the number of new cases and the cumulative number of cases as

$$\dot{S}_t = \left[\pi_t\delta_t^{-1} + (1-\pi_t)\left(1-\left(\delta_t^{-1}-1\right)\delta_t^{-1}\dot{\delta}_t\right) + \dot{\upsilon}_t\delta_t^{-1}\right]\dot{C}_t + (1-\pi_t)\left(\delta_t^{-1}-1\right)\ddot{C}_t \quad (16)$$

$$S_t = C_t + \Lambda_t + (1-\pi_t)\left(\delta_t^{-1}-1\right)\dot{C}_t \quad (17)$$

where $\dot{\upsilon}_t = -(1-\pi_t)\,\dot{\delta}_t - (1-\delta_t)\,\dot{\pi}_t$ with $\dot{\pi}_t$ the first derivative of the lost rate π_t, and the cumulative losts Λ_t is given for $t \geq 0$ by

$$\Lambda_t = S_0 + \int_0^t \pi_s(\delta_s^{-1}-1)\dot{C}_s ds \quad (18)$$

with S_0 the cumulative number of all cases until the first detection date $t = 0$. The total size of the epidemic is $S_\infty = C_\infty + \Lambda_\infty$ since $\dot{C}_t \to 0$ as $t \to \infty$. Under the Turner's growth model, $S_\infty = K + \Lambda_\infty$. Let us assume a constant detection rate $\delta_t = \delta$ closely related to detection effort but also to the average duration from infection to recovery or death of non-isolated cases. Assuming in addition a constant lost rate ($\pi_t = \pi$), we have $\dot{\delta}_t = \dot{\pi}_t = \dot{\upsilon}_t = 0$, and the new cases $\dot{S}_t$ and its accumulation S_t as well as the lost cases Λ_t simplify to

$$\dot{S}_t = \left[1+\pi\left(\delta^{-1}-1\right)\right]\dot{C}_t + (1-\pi)\left(\delta^{-1}-1\right)\ddot{C}_t \quad (19)$$

$$S_t = S_0 + \left[1+\pi\left(\delta^{-1}-1\right)\right]C_t + (1-\pi)\left(\delta^{-1}-1\right)\dot{C}_t \quad (20)$$

$$\Lambda_t = S_0 + \pi(\delta^{-1}-1)C_t. \quad (21)$$

The total epidemic size is here $S_\infty = S_0 + \left[1+\pi\left(\delta^{-1}-1\right)\right]K$.

2.4.2 Epidemic peak

At the time t_p of the peak of reported cases ($\ddot{C}_t = 0$) under constant detection and lost rates, the new infectives is $\dot{S}_p = \left[1+\pi\left(\delta^{-1}-1\right)\right]\dot{C}_p$ with $\dot{C}_p$ given in (11). This however corresponds to the peak in the overall new cases $\dot{S}_t$ only under the unrealistic assumption $\delta = 1$. The peak of new infections occurs when the second derivative $\ddot{S}_t$ of S_t with respect to t vanishes ($\dddot{S}_t = 0$). We have from (16)

$$\ddot{S}_t = \left[\pi_t\delta_t^{-1} + (1-\pi_t)\left(1-\left(\delta_t^{-1}-1\right)\delta_t^{-1}\dot{\delta}_t\right) + \dot{\upsilon}_t\delta_t^{-1}\right]\ddot{C}_t + (1-\pi_t)\left(\delta_t^{-1}-1\right)\dddot{C}_t + \Psi_t \quad (22)$$

where $\dddot{C}_t$ (the third derivative of C_t with respect to t) and Ψ_t are given by

$$\dddot{C}_t = \nu^2\omega^2 u_t^{2\rho}\left[\frac{(\nu+1)(2\nu+\nu\rho+1)}{\nu^2}\left(\frac{u_t}{1+u_t}\right)^2 - \frac{3(\nu+1)(\rho+1)}{\nu}\frac{u_t}{1+u_t} + (1+\rho)(2\rho+1)\right]\dot{C}_t \quad (23)$$

$$\Psi_t = \left\{\dot{\pi}_t\left[\delta_t^{-1}\left(1+\left(\delta_t^{-1}-1\right)\dot{\delta}_t\right)-1\right] + (1-\pi_t)\,\delta_t^{-1}\left[\ddot{\upsilon}_t - \left(\delta_t^{-1}-1\right)\ddot{\delta}_t\right]\right.$$
$$\left. + \delta_t^{-2}\dot{\delta}_t\left[(1-\pi_t)\left[\dot{\pi}_t(1-\delta_t) + \dot{\delta}_t(2\delta_t^{-1}-\pi_t)\right]-\pi_t\right]\right\}\dot{C}_t - \left[\dot{\pi}_t\left(\delta_t^{-1}-1\right) + (1-\pi_t)\,\delta_t^{-2}\dot{\delta}_t\right]\ddot{C}_t \quad (24)$$

with $\ddot{\pi}_t$ and $\ddot{\delta}_t$ the second derivatives of respectively π_t and δ_t with respect to t. The peak time and value depend on the particular forms of δ_t and π_t as functions of time. We here restrict attention to the simple situation with constant positive detection and lost rates ($\delta_t = \delta$ with $\delta \in (0,1)$ and $\pi_t = \pi$ with $\pi \in (0,1)$) where $\dot{\delta}_t = \dot{\pi}_t = \Psi_t = 0$ and (22) reduces to

$$\ddot{S}_t \quad = \quad \left[1 + \pi\left(\delta^{-1} - 1\right)\right] \ddot{C}_t + (1-\pi)\left(\delta^{-1} - 1\right) \dddot{C}_t. \tag{25}$$

It appears that the peak of new infections occurs before the time t_p of the peak in detected cases. Indeed, at $t = t_p$, we have $\ddot{C}_t = 0$, $(1-\pi)\left(\delta^{-1} - 1\right) > 0$ and $\dddot{C}_t < 0$ so that $\ddot{S}_t < 0$, *i.e.* $\dot{S}_t$ is already in its descending phase. The expression (25) indicates that at the time t_P of the peak of new infections, $\ddot{C}_t$ is equal to $\ddot{C}_P = -\zeta \dddot{C}_P$ where $\zeta = \frac{(1-\pi)(1-\delta)}{\pi + \delta(1-\pi)}$ and $\dddot{C}_P$ is given by (23) with $t = t_P$. The lower ζ, the lower $|\ddot{C}_P|$, and the lower the difference $t_p - t_P$ (delay of the observed peak). Differentiating ζ with respect to δ gives $\frac{\partial \zeta}{\partial \delta} = -\frac{1-\pi}{[\pi + \delta(1-\pi)]^2} < 0$, hence the higher δ, the lower the delay between the observed peak time and the time of the peak in new infections. Using (9) and (23), and setting $z_t = \frac{u_t}{1+u_t}$, $\ddot{S}_t$ becomes

$$\begin{aligned} \ddot{S}_t \quad = \quad & \nu\omega u_t^{\rho} \delta^{-1} \left\{ \frac{\nu\omega\,(1-\pi)\,(1-\delta)}{1 + \nu\omega\rho(t-\tau)} \left[\frac{(\nu+1)(2\nu + \nu\rho + 1)}{\nu^2} z_t^2 - \frac{3(\nu+1)(\rho+1)}{\nu} z_t + (1+\rho)(2\rho+1) \right] \right. \\ & \left. + \left[\delta + \pi\,(1-\delta)\right] \left[\frac{\nu+1}{\nu} z_t - (1+\rho) \right] \right\} \dot{C}_t \end{aligned} \tag{26}$$

which does not have a closed form root. The root t_P can however be obtained using root finding numerical routines such as the R function *uniroot* or the Matlab function *fzero*. Afterwards, the peak $\dot{S}_p$ size (the maximum number of new infections) is obtained using (19).

2.5 Statistical models and inference

Let us consider a record of new confirmed infected cases $Y_1, Y_2, \cdots, Y_n$, active cases $A_0, A_1, \cdots, A_{n-1}$, removed cases $R_1, R_2, \cdots, R_n$ (available from (8) as $R_t = Y_t - A_t + A_{t-1}$) and the associated vectors of covariates $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ at n time points. The parameters K, ω, ν, ρ_0, τ, and κ can be estimated using maximum likelihood (ML) by assigning to each Y_t an appropriate statistical distribution with expectation $\lambda_t = \dot{C}_t$ and a dispersion parameter $\sigma > 0$, and probability density function (pdf) or probability mass function (pmf) $f(Y_t|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (K, \omega, \nu, \rho_0, \tau, \kappa, \boldsymbol{\beta}^\top, \sigma)^\top$. We subsequently consider inference under log-normal and negative binomial distributions.

2.5.1 Log-normal model

Epidemic incidence case data are generally fitted through non linear least squares applied at logarithmic scale (Chowell and Viboud, 2016, Chowell et al., 2007, Viboud et al., 2016). To deal with zero incidence

cases, the logarithmic transform is usually applied on the shifted cases $Y_t + 1$. Mimicking this procedure in a likelihood inference framework, we consider a log-normal distribution assumption for the shifted incidence cases, *i.e.* $Y_t + 1 \sim \text{LN}(\lambda_t + 1, \sigma)$. The pdf of Y_t, adapted from Limpert et al. (2001), reads

$$f(Y_t|\boldsymbol{\theta}) = \frac{1}{\sigma(Y_t+1)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log(Y_t+1) - \log(\lambda_t+1)}{\sigma} + \frac{\sigma}{2}\right)^2\right\} \tag{27}$$

so that Y_t has expectation $E[Y_t] = \lambda_t$ and variance $Var[Y_t] = (\lambda_t + 1)^2 \left(\mathrm{e}^{\sigma^2} - 1\right)$.

2.5.2 Negative binomial model

Since incidence cases are counts, Y_t can be assumed to follow the negative binomial distribution, *i.e.* $Y_t \sim \text{NB}(\lambda_t, \sigma)$ with pmf

$$f(Y_t|\boldsymbol{\theta}) = \frac{\Gamma(Y_t + 1/\sigma)}{\Gamma(Y_t+1)\Gamma(1/\sigma)} \left(\frac{\sigma\lambda_t}{\sigma\lambda_t + 1}\right)^{1/\sigma} \left(\frac{1}{\sigma\lambda_t + 1}\right)^{Y_t}. \tag{28}$$

The incidence case Y_t then has expectation $E[Y_t] = \lambda_t$ and variance $Var[Y_t] = \lambda_t(1 + \sigma\lambda_t)$.

2.5.3 Likelihood inference

Based on the information $\{Y_t, R_t\}$ for $t = 1, 2, \cdots, n$, the conditional log-likelihood of the parameter $\boldsymbol{\theta}$ given A_0 is

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[\log f(Y_t|\boldsymbol{\theta}) + \log f_B(R_t|\boldsymbol{\theta})\right] \tag{29}$$

where $f_B(R_t|\boldsymbol{\theta}) = \binom{A_{t-1}+N_t}{R_t} \alpha_t^{R_t} (1-\alpha_t)^{A_{t-1}+N_t-R_t}$ is the binomial probability mass function for R_t. The function $\ell(\cdot)$ can be maximized to obtain the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ using an optimization routine such as the function *optim* in R or the function *fminsearch* of Matlab. Let $H(\boldsymbol{\theta})$ the hessian matrix of $\ell(\boldsymbol{\theta})$ and define the covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = -\left[H(\boldsymbol{\theta})\right]^{-1}$. The large sample distribution (*i.e.* for $n \to \infty$) of the maximum likelihood estimator is multivariate normal with mean $\widehat{\boldsymbol{\theta}}$ and covariance matrix $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\widehat{\boldsymbol{\theta}}\right)$.

2.6 Application to reported COVID-19 new cases in Italy

2.6.1 The data

In order to test the reliability of the Turner's growth model in predicting the dynamics of an epidemic, we use data from one of the countries which has completed a whole COVID-19 outbreak wave. The daily case reporting data in Italy has been obtained from https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series. We use only the confirmed data

(2020-02-20 to 2020-07-11) available on 2020-07-28, discarding the latest data subject to possible reporting delay, as indicated by the Istituto Superiore di Sanità (ISS) at https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-dashboard.

2.6.2 Data analysis

All analyses were performed in R (R Core Team, 2019). We have fitted the Turner's growth model curve to the whole italian data. Both the log-normal and the negative binomial distributions were used and the fit with the lowest root mean square error (RMSE) computed for the daily new positive cases was selected as the best. We next derived peak statistics (time and size) for daily new reported cases and actives cases. We also inferred the daily new infections from assuming constant detection and lost rates and estimated its peak (time and size). The detection ($\delta = 0.033$/day) and the lost rates ($\pi = 0.1$/day) for Italy were obtained from Pedersen and Meneghini (2020). These rates follows from assuming an average time of duration from infection to recovery or death of non-isolated cases of ten days (hence $\pi = 0.1$/day) and that during this detection window, 1/3 of infectives are tested positives (hence $\delta = 0.033$/day).
In order to assess the ability of the model to predict the peak of the new positive cases in countries which has not yet reached the peak, we retrospectively fitted the model to the italian data before the observed peak (day 29 after the first case notification), using data of the first two weeks, and then data of the first three weeks. For these analyses with limited data, we fitted the full Turner's growth model to the positives cases, but also its special cases, namely the hyper-Gompertz ($\nu \to 0$ while $\omega\nu^{1+\rho}$ is constant), the Gompertz ($\nu \to 0$, $\rho \to 0$ while $\omega\nu$ is constant), the Bertalanffy-Richards ($\rho \to 0$), the hyper-logistic ($\nu = 1$) and the logistic ($\nu = 1$ and $\rho \to 0$) models using the log-normal distribution for the daily counts. We then computed the Akaike's Information Criterion (AIC) defined as AIC $= -2\widehat{\ell} + 2N_p$ with $\widehat{\ell}$ the maximized log-likelihood and N_p the number of parameters in a fitted model. We finally retained and presented the best fit (lowest AIC value).

3 Results

3.1 Modeling the whole italian data

Table 1 shows parameter estimates using the whole italian COVID-19 daily case reporting data from 2020-02-20 to 2020-07-11, with standard errors and 95% confidence intervals. It appears that the log-normal distribution based fit recorded the lowest RMSE and is thus retained for subsequent analyses. The confidence bounds for the parameter ρ ($\hat{\rho} = 0.32$ with $CI(\rho) = [0.29, 0.35]$) indicate that neither

the logistic growth model ($\rho \to 0$ and $\nu = 1$) nor the Bertalanffy-Richards growth model ($\rho \to 0$) are appropriate for this dataset. It can be observed that ν is not significantly different from 1 ($\hat{\nu} = 0.85$ with $CI(\nu) = [0.70, 1.05]$), hence the hyper-logistic model ($\nu = 1$) is compatible with the data. The fitted equation is for $t \geq 0$

$$\widehat{C}_t = \frac{253124.1}{\left\{1 + \left[1 + 0.0242\,(t - 39.3877)\right]^{-3.1659}\right\}^{1.1691}} \tag{30}$$

with a coefficient of determination of $R^2 = 99.97\%$. The curves fitted to the new positive cases and the cumulative number of positive cases are shown on figure 1 (A-B). It can be observed on figure 1 (A) that the peak of new positive cases occurred 29 days after the first case notification whereas the maximum likelihood estimate of the theoretical peak time is five days later as shown in table 2 ($\widehat{t_p} = 34.10$, $CI(t_p) = [31.94, 36.41]$ days). The theoretical peak size is on average 5298 new positive cases ($\widehat{\dot{C}_p} = 5298.96$, $CI(\dot{C}_p) = [4609.72, 6091.25]$ new cases) against a maximum of 6248 observed new positive cases.

Table 1. Estimate, standard error (SE) and 95% confidence interval ($CI_{95\%}$) of Turner's growth model parameters fitted to the italian COVID-19 daily case reporting data from 2020-02-20 to 2020-07-11, using the log-normal distribution (RMSE = 514.24, $R^2 = 99.97\%$) and the negative binomial distribution (RMSE = 530.93, $R^2 = 99.93\%$)

Model	Log-normal fit			Negative binomial fit		
parameter	Estimate	SE	$CI_{95\%}$	Estimate	SE	$CI_{95\%}$
K	253124.1	12623.0	[242620.4, 243285.2]	242952.6	169.6	[242951.2, 242954.0]
ω	0.0896	0.0113	[0.0700, 0.1146]	0.0902	0.0098	[0.0729, 0.1117]
ν	0.8553	0.0906	[0.6951, 1.0526]	0.8300	0.0771	[0.6918, 0.9959]
ρ	0.3159	0.0142	[0.2892, 0.3451]	0.3231	0.0124	[0.2996, 0.3484]
τ	39.3877	2.5181	[34.7491, 44.6456]	39.3457	2.2393	[35.1927, 43.9888]
β	-4.0229	0.0060	[-4.0348, -4.0111]	-4.0229	0.0060	[-4.0348, -4.0111]
κ	0.0076	0.0001	[0.0075, 0.0078]	0.0076	0.0001	[0.0075, 0.0078]
σ	0.4332	0.0257	[0.3857, 0.4867]	0.1466	0.0175	[0.1160, 0.1853]

Notes: β and κ define the daily removal rate from detected cases as $\alpha_t = \frac{e^{\beta+\kappa t}}{1+e^{\beta+\kappa t}}$; σ is the log-normal/negative binomial distribution scale parameter (see pdf (27) and pmf (28)).

From the estimate of the parameter β given in table 1 ($\widehat{\beta} = -4.02$ with $CI(\beta) = [-4.03, -4.01]$), it comes that the daily removal rate (recoveries and deaths) averaged $\widehat{\alpha}_0 = 1.8\%$ in the very early phase of the epidemic ($t \approx 0$ day). Then, from the estimate of κ ($\widehat{\kappa} = 0.0076$ with $CI(\kappa) = [0.0075, 0.0076]$), it appears that the removal rate increased with time, *i.e.* the probability for an active case to recover or die during a one day period increased on average by 5.5% over a week. Figure 1 (C) displays the actives cases and the corresponding fitted curve using the removal probability along with the fitted equation (30). The actives cases is predicted to peak on day 56 ($\widehat{t_a} = 55.71$ days, $CI(t_a) = [53.87, 57.62]$ days) to

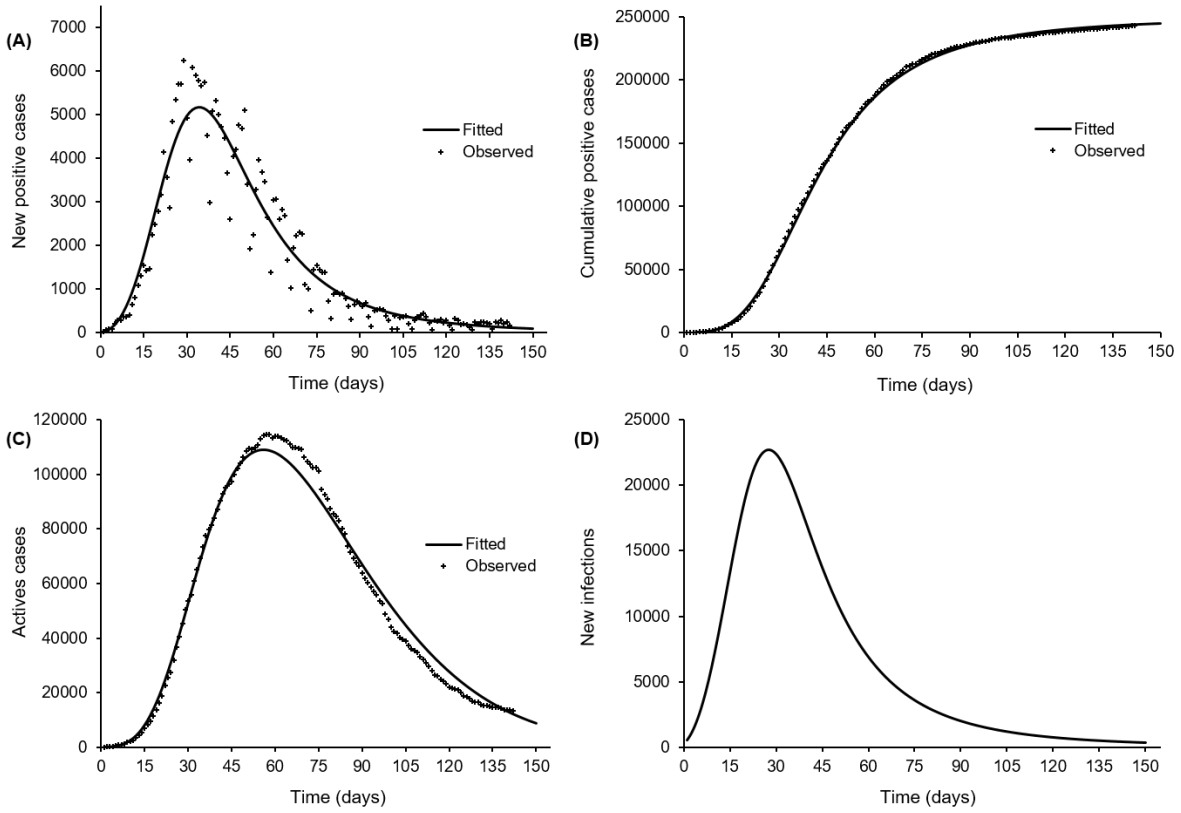

Fig 1. Log-normal fit of Turne's generic growth curve to the italian COVID-19 daily case reporting data from 2020-02-20 to 2020-07-11 (A), cumul of positive cases (B), active (quarantined) cases (C) and estimated (average) daily new infections based on a detection rate of $\delta = 0.033$/day and a lost rate (recovery or death) of non detected cases of $\pi = 0.1$/day (D)

111070 actives ($\widehat{A}_a = 111069.88$ cases, $CI(A_a) = [98580.39, 125141.70]$ cases) whereas the observed peak amounted 114683 cases and occurred 58 days after the first case notification.

The daily new infections inferred from assuming a constant detection rate ($\delta = 0.033$/day) and a constant lost rate ($\pi = 0.1$/day) is depicted on figure 1 (D). The peak in new infections likely occurred about 28 days ($\widehat{t_P} = 27.52$ days, $CI(t_P) = [25.62, 29.56]$ days) after the first case notification, and averaged 22748 new infections ($\widehat{\dot{S}}_P = 22748.38$, $CI(\dot{S}_P) = [19726.30, 26233.44]$ new infections) (table 2). The ratio of the number of infectives to the number of active cases decreased from 44.70 at the first notification day to 11.41 one week later (averaging 22.95, $CI = [22.01, 23.93]$ over this period) and to 2.99 at peak time, 22 days later.

3.2 Restrospective fits

The AICs of the retrospective fits of Tuners's growth model and its special cases to the italian COVID-19 data of the first two weeks and the first three weeks are presented in table 3. It can be observed that

Table 2. Estimate, standard error (SE) and 95% confidence interval of peak statistics using the italian COVID-19 daily case reporting data from 2020-02-20 to 2020-07-11

Quantity	Peak statistic	Estimate	SE	$CI_{95\%}$	Observed
Detected	Time (day)	34.10	1.14	[31.94, 36.41]	29
	New positive cases	5298.96	376.73	[4609.72, 6091.25]	6248
Actives (isolated)	Time (day)	55.71	0.96	[53.87, 57.62]	58
	Active cases	111069.88	6759.93	[98580.39, 125141.70]	114683
New infections	Time (day)	27.52	1.02	[25.62, 29.56]	-
	New infections	22748.38	1351.44	[19726.30, 26233.44]	-

Notes: - = not available

the best fits correspond to the hyper-logistic growth model for both data of the first two weeks (AIC = 483.03) and data of the first three weeks (AIC = 863.58). Although parsimony indicates the hyper-logistic model fits as the best, the differences Δ_{AIC} in AIC with respect to the full Turner's growth model fit are mild ($|\Delta_{AIC}| < 2$).

Table 3. AIC of Turner's growth model (2) fitted to the italian COVID-19 daily case reporting data of the first two weeks and the first three weeks from 2020-02-20, with a log-normal distribution for the positive cases

Dataset	Growth model	Restrictions	NFGMP	AIC	ΔAIC
Data of the first two weeks	Full Turner	-	5	484.49	0
	Bertalanffy-Richards	$\rho \to 0$	4	504.45	19.97
	Hyper-logistic	$\nu = 1$	4	483.03	-1.46
	Logistic	$\nu = 1$ and $\rho \to 0$	3	530.22	45.73
	Hyper-Gompertz	$\nu \to 0$ and $\omega\nu^{1+\rho}$ is constant	3	542.92	58.43
	Gompertz	$\nu \to 0$, $\rho \to 0$ and $\omega\nu$ is constant	2	499.50	15.02
Data of the first three weeks	Full Turner	-	5	864.21	0
	Bertalanffy-Richards	$\rho \to 0$	4	901.78	37.57
	Hyper-logistic	$\nu = 1$	4	863.58	-0.63
	Logistic	$\nu = 1$ and $\rho \to 0$	3	945.16	80.95
	Hyper-Gompertz	$\nu \to 0$ and $\omega\nu^{1+\rho}$ is constant	3	966.71	102.49
	Gompertz	$\nu \to 0$, $\rho \to 0$ and $\omega\nu$ is constant	2	896.52	32.30

Notes: - = not applicable; NFGMP = Number of free growth parameters; ΔAIC = difference between AICs of a special growth model fit and the AIC of the full Turner's growth model fit

Table 4 shows the estimate of the hyper-logistic growth model parameters for the two shorted datasets. It appears that the estimates of the intrinsic growth parameter ω increases slightly with data availability from $\widehat{\omega} = 0.05$ ($CI_\omega = [0.04, 0.07]$) using the data of the first two weeks, to $\widehat{\omega} = 0.07$ ($CI_\omega = [0.06, 0.08]$) using the data of the first three weeks and to $\widehat{\omega} = 0.09$ ($CI_\omega = [0.07, 0.11]$) using the whole dataset from Italy.

The estimates of the peak time and size from the two shorted datasets are shown in table 4. It can be observed that the forecast of the peak time from the data of the first two weeks is day 44 ($\widehat{t_p} = 43.38$,

Table 4. Estimate, standard error (SE) and 95% confidence interval of peak statistics using the italian COVID-19 daily case reporting data from 2020-02-20 to 2020-07-11

Model	First two weeks data			First three weeks data		
parameter	Estimate	SE	$CI_{95\%}$	Estimate	SE	$CI_{95\%}$
K	260124.1	930.9	[258305.9, 261955.1]	260122.6	633.1	[258884.8, 261366.3]
ω	0.0518	0.0066	[0.0404, 0.0665]	0.0661	0.0050	[0.0569, 0.0768]
ρ	0.3401	0.0183	[0.3061, .3779]	0.3075	0.0121	[0.2846, 0.3322]
τ	55.5287	4.3775	[47.5790, 64.8068]	47.7007	2.0405	[43.8645, 51.8724]
β	-4.7679	0.2379	[-5.2341, -4.3017]	-3.4678	0.1013	[-3.6663, -3.2693]
κ	0.1144	0.0196	[0.0761, 0.1528]	-0.0100	0.0058	[-0.0214, 0.0013]
σ	0.2081	0.0393	[0.1437, 0.3014]	0.2165	0.0334	[0.1600, 0.2930]

Notes: β and κ define the daily removal rate from detected cases as $\alpha_t = \frac{e^{\beta+\kappa t}}{1+e^{\beta+\kappa t}}$; σ is the log-normal distribution scale parameter (see pdf (27))

$CI(t_p) = [39.04, 48.22]$ days) which overestimates the observed peak time (day 29). The estimate from the data of the first three weeks reduces the delay, with $\widehat{t_p} = 38.97$ ($CI(t_p) = [36.80, 41.27]$) days. The forecast of the peak size from the data of the first two weeks is 3794 ($\widehat{\dot{C}_p} = 3793.60$, $CI(t_p) = [3032.63, 4745.53]$) new positive cases, which underestimates the observed peak (6248 new positive cases). The forecast from the data of the first three weeks also underestimates the peak but is less biased, with $\widehat{\dot{C}_p} = 4733.35$ ($CI(t_p) = [4136.58, 5416.22]$) new positive cases.

Table 5. Estimate, standard error (SE) and 95% confidence interval ($CI_{95\%}$) of the parameters of the hyper-logistic growth model fitted using the log-normal distribution to the italian COVID-19 daily case reporting data of the first two weeks (RMSE = 92.16, $R^2 = 99.68\%$) and the first three weeks (RMSE = 224.41, $R^2 = 99.87\%$) from 2020-02-20

Peak	Data of the first two weeks			Data of the first three weeks		
statistic	Estimate	SE	$CI_{95\%}$	Estimate	SE	$CI_{95\%}$
Time (day)	43.38	2.34	[39.04, 48.22]	38.97	1.14	[36.80, 41.27]
New positive cases	3793.60	433.34	[3032.63, 4745.53]	4733.35	325.46	[4136.58, 5416.22]

4 Summary and perspectives

This work proposes the use of a flexible growth model for modeling case reporting data from an epidemic outbreak with containment measures including at least isolation of individuals tested positive. The generic growth model of Turner et al. (1976) offers a flexible framework with the possibility to recover many special growth models such as the common exponential and the logistic growth models, the hyper-logistic, the hyper-Gompertz, the Gompertz and the Bertalanffy-Richards growth models. Since the special models are all nested within the generic model framework, the most appropriate model can be identified

using information criteria such as the Akaike's An Information Criterion (AIC), but a likelihood ratio test (Wilks, 1938) can also be conducted for models with different number of free parameters. When additional information can be obtained on the ability to detect infective individuals (number of tests, tracing of contact persons), the proposed framework allows to include this information so as to infer on the dynamics of the epidemic beyond the identified (positive) cases, without ressorting to mechanistic/compartmental models.

From our application to the COVID-19 outbreak data in Italy, the hyper-logistic model is the most appropriate model for the dataset. It appears that the modeling approach can predict the dynamics of an epidemic using data from first few days of an outbreak, at least in this example. Indeed, the predicted peak time (and size) for the positives cases (using only the first two/three weeks data) overestimates (and underestimates) the observed peak time (and size). However, the biases can be attributed for instance to the increase in the testing effort and isolation (and the subsequent decrease in the growth rate) in Italy where only about 3762 tests/day were performed in the first three weeks from 2020-02-20, and about 21248 tests/day were performed in the next three weeks. Our estimate of the ratio of the number of infectives to the number of active cases averaged 22.95 in the first week of the outbreak, in the range $[5, 25]$ obtained by Pedersen and Meneghini (2020) using the SIQR model. Our proposal thus offers a valid alternative to mechanistic models, for instance the picewise exponential growth used by Pedersen and Meneghini (2020) withing the SIQR model framework on the italian early outbreak data.

In a very limited data situation, we suggest a further reduction of the number of model parameters to be estimated. Indeed, since the parameter τ in the growth model (2) is a constant of integration determined by the initial conditions of the epidemic, it can be expressed in terms of other parameters and the number of cases C_0 detected at time $t = 0$ as $\tau = \frac{1}{\nu\omega\rho}\left\{1 - \left[\left(\frac{K}{C_0}\right)^{\nu} - 1\right]^{-\rho}\right\}$ for $\rho \neq 0$ and $\tau = \log\left(\left(K/C_0\right)^{\nu} - 1\right)/(\nu\omega)$ for $\rho = 0$. Consideration of a procedure where τ is not estimated as a free parameter may lead to parsimony, with inference conditional on the number of individuals tested positive at time $t = 0$. Inference on the effective reproduction number and the sensitivity of the epidemic dynamics to containment measures under the generic growth model framework is considered for a future work.

References

Arianna Agosto and Paolo Giudici. A poisson autoregressive model to understand covid-19 contagion dynamics. *SSRN ePrint*, 2020. URL https://ssrn.com/abstract=3551626.

Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, and Constantinos Siettos. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PloS one*, 15(3):e0230405, 2020.

Francesco Casella. Can the covid-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5(3):1079–1084, 2020.

G Chowell, R Luo, K Sun, K Roosa, A Tariq, and C Viboud. Real-time forecasting of epidemic trajectories using computational dynamic ensembles. *Epidemics*, 30:100379, 2020.

Gerardo Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017.

Gerardo Chowell and Cécile Viboud. Is it growing exponentially fast?–impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics. *Infectious disease modelling*, 1(1):71–78, 2016.

Gerardo Chowell, Hiroshi Nishiura, and Luis MA Bettencourt. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface*, 4(12):155–166, 2007.

Gerardo Chowell, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud. Mathematical models to characterize early epidemic growth: A review. *Physics of life reviews*, 18:66–97, 2016.

Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6, 2020.

Herbert Hethcote, Ma Zhien, and Liao Shengbing. Effects of quarantine in six endemic models for infectious diseases. *Mathematical biosciences*, 180(1-2):141–160, 2002.

Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.

Eckhard Limpert, Werner A. Stahel, and Markus Abbt. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, 341(5), 2001.

MATLAB. *version 9.0.0 (R2016a)*. The MathWorks Inc., Natick, Massachusetts, 2016.

Morten Gram Pedersen and Matteo Meneghini. Quantifying undetected covid-19 cases and effects of containment measures in italy. *ResearchGate Preprint (online 21 March 2020) DOI*, 10, 2020.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

Peter D Spencer and Adam Golinski. Modeling the covid-19 epidemic using time series econometrics. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.06.01.20118612.

Malcolm E. Turner, Edwin L. Bradley, Katherine A. Kirk, and Kenneth M. Pruitt. A theory of growth. *Mathematical Biosciences*, 29(3):367 – 373, 1976. ISSN 0025-5564. doi: https://doi.org/10.1016/0025-5564(76)90112-7. URL http://www.sciencedirect.com/science/article/pii/0025556476901127.

Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3):278, 2020.

Cécile Viboud, Lone Simonsen, and Gerardo Chowell. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics*, 15:27–37, 2016.

WHO. Coronavirus disease 2019 (covid-19): situation report, 208. Technical documents, 2020.

Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics*, 9(1):60–62, 1938.