Article

A hybrid modelling technique of epidemic outbreaks with application to COVID-19 dynamics in West Africa

Chénangnon Frédéric Tovissodé[1] [ID], Jonas Têlé Doumatè [1,2] and Romain Glèlè Kakaï [1,*] [ID]

[1] Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Abomey-Calavi (Benin); chenangnon@gmail.com (C.F.T.); jonas.doumate@fast.uac.bj (J.T.D.)
[2] Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Abomey-Calavi (Benin)
[*] Correspondence: romain.glelekakai@fsa.uac.bj (R.G.K)

Simple Summary: The intrisinc dynamics of the propagation of a disease changes along an epidemic course, especially for long lasting epidemics such as the COVID-19. Indeed, the natural evolution of the pathogen and countermeasures such as quarantining, lockdown, social distancing and vaccination modify the transmission dynamics of the disease. With a view to match these theoretical changes to potential changes in observed epidemiological data, we designed an hybrid modelling framework where we integrated: (1) two growth curves for daily reported positive cases, differentiating the early epidemic phase and a second phase with a potentially different dynamics; (2) two logistic regression models for daily recoveries and deaths; and (3) a SIQR (Susceptible, Infective, Quarantined, Recovered) mechanistic model to provide an overview of the dynamics of the disease in the target population. This joint modelling approach allows explicit analytical expressions for the different compartments of the SIQR model, circumventing common identifability issues in such models. The changes in the disease transmission pattern can be confronted to countermeasures to assess their effectiveness along time. For illustrative purposes, we applied the approach to COVID-19 data from West Africa. It turned out that the first imported COVID-19 case(s) in West Africa likely entered the region between January 28th and February 7th, 2020. Moreover, the first measures implemented by West African authorities impacted the dynamics of the disease one month after the outbreak.

Abstract: The widely used logistic model for epidemic case reporting data may be either restrictive or unrealistic in presence of containment measures when implemented after an epidemic outbreak. For flexibility in epidemic case reporting data modelling, we combined an exponential growth curve for the early epidemic phase with a flexible growth curve to account for the potential change in growth pattern after implementation of containment measures. We also fitted logistic regression models to recoveries and deaths from the confirmed positive cases. In addition, the growth curves were integrated into a SIQR (Susceptible, Infective, Quarantined, Recovered) model framework to provide an overview on the modelled epidemic wave. We focused on the estimation of: (1) the delay between the appearance of the first infectious case in the population and the outbreak ("epidemic latency period"); (2) the duration of the exponential growth phase; (3) the basic and the time-varying reproduction numbers; and (4) the peaks (time and size) in confirmed positive cases, active cases and new infections. The application of this approach to COVID-19 data from West Africa allowed to discuss the effectiveness of some containment measures implemented across the region.

Keywords: Growth model; Epidemic latency period; Reproduction number; West Africa

1. Introduction

The ravages by the COVID-19 pandemic has deepened the need of mathematical and statistical tools for understanding the dynamics of epidemics across the world. Simple mathematical models of infectious diseases are useful for providing insight into epidemic trajectories and disease dynamics [1–3]. However, applications should target complex but

parsimonious models which make realistic assumptions and let the observed data drive estimations. Indeed, although the logistic bell curve has been widely used for various epidemic data, it lacks flexibility for epidemics whose data exibits asymmetry or varying growth patterns [4–6]. With a view to allow flexibility, [7] have considered the generic growth curve of [8]. This approach concedes the simple logistic curve when it is supported by the observed data, but offers the possibility to fit various flexible growth models such as the generalized logistic model [9,10], the hyperlogistic model [8,11], the hyper-Gompertz [8] and the Gompertz curves [12,13].

However, to be realistic, models for epidemic data should be able to account for the potential effect of containment measures when implemented after an epidemic outbreak. In a target population undergoing an epidemic wave, the number of infective individuals may be assumed to follow an exponential growth in the early epidemic phase where no containment measures were implemented or the implemented measures were not yet effective [14]. Then, the variation of the number of infective individuals is expected to shift to a sub-exponential growth resulting from negative feedbacks due to a decrease in the probability that an infectious individual meets a susceptible individual [4], or to containment measures's effects, if any.

In this paper, we focus on epidemic waves managed with at least an isolation measure for all identified infectives, as for the COVID-19 pandemic in nearly all the world. The objective of this study was to provide a quantitative framework in which epidemiologists can identify, from a large family of models, the parsimonious model that explains patterns in an observed dataset, and then assess hypotheses on the potential course of related but unobservable processes of interest. Specifically, we (1) modelled confirmed positive cases using a combination of the exponential growth curve for the initial epidemic phase and the generic growth curve [8] after this initial phase. We considered the log-normal distribution to model the error structure in the data and the maximum likelihood approach for the estimation of growth parameters. This development allows the estimation of the duration of the exponential growth phase and the theoretical time and size of the peak of new positive cases. Secondly, we (2) modelled removal (recovery and death) from identified positive cases as binary processes using two logistic regression models to monitor the evolution and the peak (time and size) of the actives among detected cases. Finally, in order to provide an overall view for a target epidemic, we (3) integrated the growth curve and the logistic regression removal rates into a mechanistic SIQR model frame [15] in which the population is structured in Susceptibles, Infectives, Quarantined (identified actives cases) and Recovered individuals. The result is a mechanistic model in which the sizes of the different states (compartments) have closed form expressions. This allows inference on various epidemiological parameters such as the delay between the appearance of the first infectious case in the population ("patient zero") and the outbreak ("epidemic latency period"), the reproduction number, the unobservable new infections per unit time as well as the proportion of the target population immunized against the pathogen of the target disease.

In addition to estimates (with quantified uncertainty) for common epidemiological parameters, the proposed hybrid modelling framework extracts from the observed data and demographic rates, the evolution along the epidemic course of the key parameter for summarizing the dynamics of an epidemic: the reproduction number. The changes in this parameter can thus be confronted to control measures promoted/enforced by public health authorities and governments. For illustrative purpose, we used the developed modelling framework to: (i) model COVID-19 case reporting data (daily PCR-confirmed positives, recoveries and deaths) from Western Africa (February 28th to August 31st, 2020); and (ii) evaluate the transmission pattern of the disease in the region during the considered period. The results were used to discuss the effectivness of some containment measures implemented by governments across the region.

2. The Hybrid Modelling Framework

In this section, we describe the three sub-models integrated into the proposed modelling framework, namely, the growth model, the logistic removal rates and the Susceptible-Infective-Quarantined-Recovered (SIQR) mechanistic model.

2.1. Mixture of Growth Models for Detected Cases

We assume that the cumulative number C_t of reported cases, as a function of the time t, has the form

$$
C_t = \begin{cases} 0 & \text{if } t \le 0 \\ e^{\omega_0(t-\tau_0)} & \text{if } 0 < t \le t_e \\ \xi + \varphi_t & \text{if } t > t_e \end{cases} \tag{1}
$$

where $t_e > 0$ is the duration from outbreak to the end of the exponential growth phase,

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

is the generic growth model [8] with $u_t = [1 + \omega\nu\rho(t-\tau)]^{-1/\rho}$, $\Omega > 0$ is a constant such that the ultimate epidemic size (detected) is $\xi + \Omega$, $\omega > 0$ is the "intrinsic" growth rate constant for the sub-exponential growth phase, $\nu > 0$ is a growth acceleration parameter, ρ ($-1 < \rho < \nu^{-1}$) is a shape parameter controling the skewness of the growth curve during the sub-exponential epidemic phase, and τ is a constant of integration determined by the initial conditions of the epidemic. The generic growth curve φ_t specified for $t > t_e$ encompasses many special or limiting cases (Table 1) including the Bertalanffy-Richards ($\rho \to 0$), the hyper-Gompertz ($\nu \to 0$ while $\omega\nu^{1+\rho} \to \tilde{\omega}$ with $\tilde{\omega}$ constant), the Gompertz ($\nu \to 0, \rho \to 0$ while $\omega\nu \to \tilde{\omega}$), the hyper-logistic ($\nu = 1$) and the logistic ($\nu = 1$ and $\rho \to 0$) growth models [8]. The parameter $\omega_0 > 0$ in (1) is the exponential growth rate for the early epidemic phase and $\tau_0 \in \mathbb{R}$ determines the growth rate at $t = 0$. The constants ω_0 and τ_0 are set such that the first derivative $\dot{C}_t$ and the second derivative $\ddot{C}_t$ of C_t with respect to t are smooth at $t = t_e$ (i.e. at the end of the exponential growth phase). Specifically,

$$
\omega_0 = \ddot{\varphi}_e \tag{3}
$$

$$
\tau_0 = t_e + \frac{\log \omega_0 - \log \dot{\varphi}_e}{\omega_0} \tag{4}
$$

where $\dot{\varphi}_e = \dot{\varphi}_{t_e}$ and $\ddot{\varphi}_e = \ddot{\varphi}_{t_e}$ are available from Table 1, and (4) follows from setting $\omega_0 e^{\omega_0(t_e-\tau_0)} = \dot{\varphi}_e$. Furthermore, the real constant ξ in (1) ensures that C_t does not jump at $t = t_e$. In other words, ξ is given by $\xi = e^{\omega_0(t_e-\tau_0)} - \varphi_e$ (with $\varphi_e = \varphi_{t_e}$) which by (4) simplifies to

$$
\xi = \frac{\dot{\varphi}_e}{\omega_0} - \varphi_e. \tag{5}
$$

The restriction $-1 < \rho < \nu^{-1}$ makes the parameters ν and ρ dependent. This can be circumvented by introducing a free working shape parameter $\rho_0 \in (0,1)$ such that $\rho = (\rho_0(\nu+1)/\nu) - 1$ [7].

In (1), the time (in e.g. day, week, month) of the first identified cases corresponds to $t = 1$. In other words, to match (1) to observed data, C_1 is identified to the number of cases reported in the time interval $(0,1]$, C_2 is the number of cases reported in the time interval $(0,2]$, and so on. If $\Omega \to \infty$ and $\nu\rho \to 0$, the curve C_t converges to an exponential growth curve with rate ω_0. However, this scenario can be ruled out since the size of any target population is finite and so does Ω. In practice, the exponential growth forever is prevented

Table 1: **Generic growth model [8] and its limiting cases: population size (φ_t), growth rate ($\dot{\varphi}_t$), and growth acceleration ($\ddot{\varphi}_t$)**

Model	Population size	Growth rate	Growth acceleration
Gen	$\Omega(1+u_t)^{-1/\nu}$	$\Omega\omega u_t^{1+\rho}(1+u_t)^{-\frac{\nu+1}{\nu}}$	$\nu\omega u_t^{\rho}\left(\frac{\nu+1}{\nu}\frac{u_t}{1+u_t}-\rho-1\right)\dot{\varphi}_t$
BR	$\Omega(1+e^{-v_t})^{-1/\nu}$	$\Omega\omega e^{-v_t}(1+e^{-v_t})^{-\frac{\nu+1}{\nu}}$	$\nu\omega\left(\frac{\nu+1}{\nu}\frac{e^{-v_t}}{1+e^{-v_t}}-1\right)\dot{\varphi}_t$
HG	$\Omega\exp\left(-w_t^{-\frac{1}{\rho}}\right)$	$\Omega\tilde{\omega}w_t^{-\frac{1+\rho}{\rho}}\exp\left(-w_t^{-\frac{1}{\rho}}\right)$	$\tilde{\omega}w_t^{-1}\left(w_t^{-\frac{1}{\rho}}-\rho-1\right)\dot{\varphi}_t$
Gom	$\Omega\exp(-e^{-x_t})$	$\Omega\tilde{\omega}\exp(-x_t-e^{-x_t})$	$\tilde{\omega}(e^{-x_t}-1)\dot{\varphi}_t$

Table notes: Gen = Generic; BR = Bertalanffy-Richards; HG = Hyper-Gompertz; Gom = Gompertz; $u_t=[1+\omega\nu\rho(t-\tau)]^{-1/\rho}$, $v_t=\nu\omega(t-\tau)$, $w_t=\tilde{\omega}\rho(t-\tau)$, and $x_t=\tilde{\omega}(t-\tau)$.

by negative feedbacks which decrease the probability that an infectious individual and a susceptible individual meet and have an adequate contact (*i.e.* contact sufficient for transmission). For instance, the growth of infectives is naturally continuously lowered by the increasing fraction of the population constituted by individuals who recovered and become less susceptible (temporarily or permanently immune) to the infection [4]. To prevent the exponential growth of infectives, control measures such as quarantining and lockdown reduce the probability of contact between susceptible individuals and infectious individuals whereas some other measures such as social distancing and wearing a face mask reduce the likelihood of transmission whenever contacts happen.

The specification of the growth model in (1) to an epidemic thus implies that the growth rate $\dot{C}_t$, *i.e.* the number of new cases reported per unit time given by

$$\dot{C}_t=\begin{cases}\omega_0 e^{\omega_0(t-\tau_0)} & \text{if } 0\le t\le t_e\\\dot{\varphi}_t & \text{if } t>t_e\end{cases} \tag{6}$$

with $\dot{\varphi}_t$ given in Table 1, will peak and then fall toward zero case per unit time. The peak occurs at a time $t_p>t_e$ when the growth acceleration $\ddot{C}_t$ given by,

$$\ddot{C}_t=\begin{cases}\omega_0^2 e^{\omega_0(t-\tau_0)} & \text{if } 0\le t\le t_e\\\ddot{\varphi}_t & \text{if } t>t_e\end{cases} \tag{7}$$

with $\ddot{\varphi}_t$ given in Table 2, vanishes. The expressions of the time (t_p) and the size ($\dot{C}_p$) of the peak are shown in Table 2 for the general situation ($\nu\neq 0$ and $\rho\neq 0$), as well as for limiting cases.

The number of detected cases C_t is the basic data reported during an epidemic. Once this has been modelled, various epidemic related quantities can be inferred upon introduction of disease related parameters (*e.g.* detection of infectives, recoveries, deaths) and demographic parameters (*e.g.* natural mortality, births, immigration).

2.2. Infectives, Epidemic Latency Period and Active Cases

Since only a fraction of infectives are identified at a time t, the number I_t of infective individuals in the target population is obtained using (6) as $I_t=\delta^{-1}\dot{C}_t$ [7] which reads

$$I_t=\begin{cases}I_0 e^{\omega_0 t} & \text{if } t\le t_e\\\delta^{-1}\dot{\varphi}_t & \text{if } t>t_e\end{cases} \tag{8}$$

Table 2: **Peak time (t_p) and size ($\dot\varphi_p = \dot\varphi_{t_p}$) of the generic growth curve [8] and its limiting cases**

Model	Peak time (t_p)	Peak size ($\dot\varphi_p$)
Generic	$\tau + \frac{1}{\nu\omega\rho}\left\{\left[\frac{1-\nu\rho}{\nu(1+\rho)}\right]^\rho - 1\right\}$	$\Omega\omega\left[\nu\frac{1+\rho}{1-\rho\nu}\right]^{1+\rho}\left(\frac{\nu+1}{1-\rho\nu}\right)^{-\frac{\nu+1}{\nu}}$
BR ($\rho \to 0$)	$\tau - \frac{\log\nu}{\omega\nu}$	$\Omega\omega\nu(1+\nu)^{-\frac{\nu+1}{\nu}}$
HG ($\nu \to 0$, $\nu\omega^{(1+\rho)} \to \tilde\omega$)	$\tau + \frac{(1+\rho)^{-\rho}}{\tilde\omega\rho}$	$\Omega\tilde\omega\left[(1+\rho)e^{-1}\right]^{1+\rho}$
Gompertz ($\rho \to 0$ in HG)	τ	$\Omega\tilde\omega e^{-1}$

Table notes: BR = Bertalanffy-Richards; HG = Hyper-Gompertz; $u_p = \nu(1+\rho)/(1-\rho\nu)$, t_p is the root of $\ddot\varphi_t$; the expressions of $\dot\varphi_t$ (growth rate) and $\ddot\varphi_t$ (growth acceleration) are given in Table 1.

where $I_0 = \delta^{-1}\omega_0 e^{-\omega_0\tau_0}$ is the number of infectives at the outbreak ($t = 0$) and $\delta \in (0, 1]$ is the detection rate assumed constant along the epidemic course (after the outbreak). Note that the number of infectives before the outbreak ($t < 0$) is obtained by back extrapolation as $I_t = I_0 e^{\omega_0 t}$, *i.e.* considering an exponential growth before the outbreak [14].

We refer to the time from the appearance of the first infectious case in the population ("patient zero") to the outbreak as the "epidemic latency period". An estimate of the duration t_o of this period is obtained by setting $I_t = 1$ [14]. By (8), the duration of the epidemic latency period is estimated by $t_o = \omega_0^{-1}\log I_0$ which on using (4) simplifies to

$$t_o = \frac{\log\dot\varphi_e - \log\delta}{\omega_0} - t_e. \tag{9}$$

The number of detected and active cases, *i.e.* individuals tested positive and in isolation at an hospital or at home at time t is denoted Q_t following [15] for "Quarantined" state, although we refer to Q_t as "Actives". Given the detected cases C_t in (1), Q_t satisfies

$$\dot Q_t = \dot C_t - (\alpha_t + \epsilon_t)Q_t \tag{10}$$

where α_t is the recovery rate and ϵ_t is the death rate (natural and disease-related mortality) of actives. Indeed, following [7], we allow the removal rates α_t and ϵ_t from Q_t to be time varying. This is appropriate when recovery and death data are available aside the reported positive cases per unit time. The two rates are here given the logistic forms

$$\alpha_t = \left[1 + e^{-(\kappa_0 + \kappa t)}\right]^{-1} \tag{11}$$

$$\epsilon_t = \left[1 + e^{-(\lambda_0 + \lambda t)}\right]^{-1}. \tag{12}$$

Accordingly, at outbreak ($t = 0$), the recovery rate is $\alpha_0 = 1/(1 + e^{-\kappa_0})$ and the death rate is $\epsilon_0 = 1/(1 + e^{-\lambda_0})$. Then, along the epidemic course, κ and λ determine the changes in the log-odds ratio to have an outcome per unit time. Under constant removal rates assumption ($\kappa = \lambda = 0$), solving the differential (10), gives the actives cases as (assuming that C_w is differentiable for $0 < w < t$)

$$Q_t \;=\; \left[Q_0 + \int_0^t \dot{C}_w e^{(\alpha_0+\epsilon_0)w} dw\right] e^{-(\alpha_0+\epsilon_0)t}$$

where Q_0 is available from (21) and represents the number of persistent cases from previous epidemic waves (isolated actives) at the outbreak of the target epidemic wave (*e.g.* $Q_0 = 0$ for a new disease-related epidemic). Taking the expression of C_t in (1) into account yields for $\kappa = \lambda = 0$

$$Q_t \;=\; \begin{cases} Q_0 e^{-(\alpha_0+\epsilon_0)t} + \frac{\delta I_0}{\omega_0+\alpha_0+\epsilon_0}\left[e^{\omega_0 t} - e^{-(\alpha_0+\epsilon_0)t}\right] & \text{if } 0 < t \le t_e \\[2mm] \left[Q_e e^{(\alpha_0+\epsilon_0)t_e} + \int_{t_e}^t \dot{\varphi}_r e^{(\alpha_0+\epsilon_0)r} dr\right] e^{-(\alpha_0+\epsilon_0)t} & \text{if } t > t_e \end{cases} \tag{13}$$

where $Q_e = Q_{t_e}$ is the number of active cases at the end of the exponential growth phase. For the general situation where the rates α_t and ϵ_t may be time dependent, the number of active cases is

$$Q_t \;=\; \begin{cases} \left[Q_0 F_0 + \omega_0 \int_0^t e^{\omega_0(r-\tau_0)} F_r dr\right] F_t^{-1} & \text{if } 0 < t \le t_e \\[2mm] \left[Q_e F_{t_e} + \int_{t_e}^t \dot{\varphi}_r F_r dr\right] F_t^{-1} & \text{if } t > t_e \end{cases} \tag{14}$$

where F_t is defined as

$$F_t = \begin{cases} e^{(\alpha_0+\epsilon_0)t} & \text{if } \kappa = 0 \text{ and } \lambda = 0 \\ e^{\alpha_0 t}\left(1 + e^{\lambda_0+\lambda t}\right)^{1/\lambda} & \text{if } \kappa = 0 \text{ and } \lambda \neq 0 \\ \left(1 + e^{\kappa_0+\kappa t}\right)^{1/\kappa} e^{\epsilon_0 t} & \text{if } \kappa \neq 0 \text{ and } \lambda = 0 \\ \left(1 + e^{\kappa_0+\kappa t}\right)^{1/\kappa}\left(1 + e^{\lambda_0+\lambda t}\right)^{1/\lambda} & \text{if } \kappa \neq 0 \text{ and } \lambda \neq 0 \end{cases}. \tag{15}$$

2.3. Overall Epidemic Dynamics

The dynamics of an epidemic as expressed by the variations of the infectives I_t, is determined by the combination of the transmission rate (new infections) and the average residence time, *i.e.* the average duration from infection to isolation, recovery or death. The core parameter for summarizing this dynamics is at a moment t the reproduction number denoted $\mathcal{R}_t$, which is indeed crucial for quantifying the intensity of control measures required to control an epidemic [6].

The reproduction number is defined as the average number of secondary cases generated by a primary case. With a view to derive $\mathcal{R}_t$ under the growth model in (1), we first considered an overall picture of the target population in order to enlighten the sources (transmission and removal) of the variations of I_t as given in (8).

2.3.1. The SIQR Model

Following [7] and [14], we considered the Susceptible - Infectious - Quarantined - Recovered (SIQR) model of [15] to obtain a picture of the different states of individuals in a target population. We used the "quarantine-adjusted incidence" version [15] of this model since the underlying transmission mechanism explicitly recognizes the isolation of detected cases. In this framework, the size N_t of the target population (assumed finite but large) satisfies at time t

$$N_t = S_t + I_t + Q_t + R_t \tag{16}$$

where S_t is the size of the class of susceptible individuals, I_t is the class of infectives, Q_t is the size of the class of detected active cases, and R_t is the size of the class of individuals who recovered (both detected and not detected). We assume that the infection has zero latent period (susceptible individuals become infectious as soon as they become infected). The individuals in the classes R are assumed permanently immune within the period of time considered. It is also assumed that known active cases (in the class Q) do not mix with others classes, and do not infect susceptibles (*i.e.* the transmission rate from Q-class individuals is considered negligible). The corresponding SIQR model is described by the following set of nonlinear differential equations [15]

$$
\begin{align}
\dot{S}_t &= \eta - \beta_t(S_t + R_t)I_t/(N_t - Q_t) - \mu S_t \tag{17} \\
\dot{I}_t &= [\beta_t(S_t + R_t)/(N_t - Q_t) - (\gamma + \delta_t + \pi)]I_t \tag{18} \\
\dot{Q}_t &= \delta_t I_t - (\alpha_t + \epsilon_t)Q_t \tag{19} \\
\dot{R}_t &= \gamma I_t + \alpha_t Q_t - \mu R_t \tag{20}
\end{align}
$$

where η is the recruitment rate of susceptibles (births and immigration), β_t is the total number of adequate contacts (*i.e.* contacts sufficient for transmission) per unit time, μ is the per capita natural mortality rate, α_t and γ are the recovery rates from actives Q_t and infectives I_t resepectively, ϵ_t and π are the death rates (natural and disease-related) for actives Q_t and infectives I_t resepectively, and δ_t is the detection rate which is null ($\delta_t = 0$) for $t < 0$ and equals $\delta_t = \delta$ for $t \geq 0$. Note that (19) is the same as (10) for $t \geq 0$. Unlike in [15], we allow the transmission rate β_t to be time varying as a consequence of the form of the number of infectives I_t already available in (8). The transfer diagram for this SIQR model is shown in Figure 1.
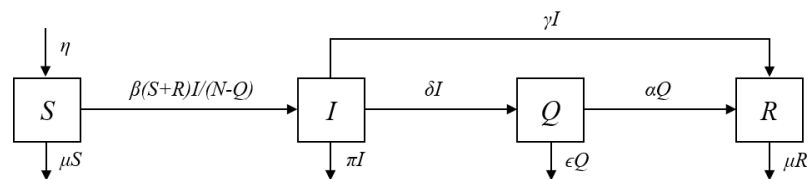
Figure 1. Transfer diagram for a SIQR model with quarantine-adjusted incidence. S is the class of susceptibles, I is the class of infectives, Q is the class of detected active cases, *i.e.* individuals tested positive and in isolation at an hospital or at home, R is the class of individuals who who contracted the disease, were detected or not, and have recovered. The individuals in the clases R are considered permanently immune.

The system (17–20) always has the disease–free equilibrium $P_0 = (S = \eta/\mu, I = 0, Q = 0, R = 0)$, *i.e.* in the absence of the disease, the population size N_t approaches the carrying capacity $N^* = \eta/\mu$. Hethcote et al. [15] showed in the case of a constant transmission rate ($\beta_t = \beta$) that the system can have an endemic equilibrium point. Furthermore, such endemic points may be either locally asymptotically stable or subject to Hopf bifurcation depending on model parameters, giving rise to unstable spiral and periodic solutions [15]. In the modelling framework considered in this work, the long term dynamics of a target disease is solely determined by the ultimate epidemic size Ω (detected). Indeed, $\lim_{t \to \infty} \varphi_t = \Omega$ so that $\lim_{t \to \infty} \dot{\varphi}_t = 0$ since Ω is finite and therefore, $\lim_{t \to \infty} I_t = 0$ by (8). Consequently, the disease always dies out in the long run and the system tends to the disease–free equilibrium P_0. This happens because the fraction of infectives in the population decreases to very near zero and the fraction of quarantined (Q) decreases to zero (through recovery and death). Eventually, over 100 or more years, the recovered people (R) slowly die off and the birth process slowly increases the susceptibles (S), until everyone is susceptible at the disease–free equilibrium P_0 [16]. Note, however, that the SIQR model described by (17–20) together with (8) is meant for a single epidemic wave, whereas it is possible to

have successive epidemic waves, or even overlapping epidemic waves [1] which would be described by a mixture of many SIQR models.

From (19), the number Q_t of known active cases before the outbreak, is given by

$$Q_t = Q_o e^{-(\alpha_0 + \epsilon_0)(t + t_o)} \quad \text{for} \quad -t_o \leq t \leq 0 \tag{21}$$

on assuming constant recovery (α_0) and death (ϵ_0) rates before the outbreak, and on denoting Q_o the number of persistent cases from previous epidemic waves (*e.g.* $Q_o = 0$ for a new disease-related epidemic).

2.3.2. Susceptibles and Recovered

In addition to infectives (I_t) and actives (Q_t) already available from the growth curve C_t, the computation of the population size in (16) requires the expressions of the sizes of the compartments of susceptibles (S_t) and immunes (R_t). From (18), we deduce the transmission rate, *i.e.* the number of adequate contacts per unit time (for $I_t > 0$)

$$\beta_t = \left(\gamma + \delta_t + \pi + \frac{\dot{I}_t}{I_t} \right) \left(1 + \frac{I_t}{S_t + R_t} \right). \tag{22}$$

Inserting this in (17) and replacing in light of (8) $\dot{I}_t / I_t = \omega_0$ for $t \leq 0$ and $\dot{I}_t / I_t = \ddot{C}_t / \dot{C}_t$ for $t > 0$ yields

$$\dot{S}_t = \begin{cases} \eta - (\gamma + \pi + \omega_0) I_0 e^{\omega_0 t} - \mu S_t & \text{if} \quad t \leq 0 \\ \eta - \delta^{-1} [(\gamma + \delta + \pi) \dot{C}_t + \ddot{C}_t] - \mu S_t & \text{if} \quad t > 0 \end{cases} . \tag{23}$$

Therefore, the number of susceptible individuals is given for $-t_o \leq t \leq 0$ by

$$S_t = \frac{\eta}{\mu} + \left(S_o - \frac{\eta}{\mu} \right) e^{-\mu(t_o + t)} - \frac{\omega_0 + \gamma + \pi}{\omega_0 + \mu} I_0 \left(e^{\omega_0 t} - e^{-(\omega_0 + \mu)t_o - \mu t} \right) \tag{24}$$

where S_o is the number of susceptibles at the begining of the epidemic, obtained from (16) with $t = -t_o$ ($I_o = 1$) as

$$S_o = N_o - Q_o - R_o - 1 \tag{25}$$

where N_o is the initial population size (*i.e.* at $t = -t_o$) and K_o is the number of known immune individuals at the begining of the target epidemic (recovered from past outbreaks if any). The number of susceptibles after the outbreak ($t > 0$) is

$$S_t = \begin{cases} \frac{\eta}{\mu} + \left(S_0 - \frac{\eta}{\mu} \right) e^{-\mu t} - \frac{\omega_0 + \gamma + \delta + \pi}{\omega_0 + \mu} I_0 \left(e^{\omega_0 t} - e^{-\mu t} \right) & \text{if} \quad 0 < t \leq t_e \\ \frac{\eta}{\mu} + \left(S_e - \frac{\eta}{\mu} \right) e^{\mu(t_e - t)} - \left\{ \int_{t_e}^{t} [1 + \delta^{-1}(\gamma + \pi + z_r)] \dot{\varphi}_r e^{\mu r} dr \right\} e^{-\mu t} & \text{if} \quad t > t_e \end{cases} \tag{26}$$

where S_0 is the number of susceptibles at the outbreak ($t = 0$) and is available from (24), $S_e = S_{t_e}$ is the number of susceptibles at the end of the exponential growth phase and $z_t = \ddot{\varphi}_t / \dot{\varphi}_t$ is the ratio of the growth acceleration $\ddot{\varphi}_t$ to the growth rate $\dot{\varphi}_t$ (Table 2).

From the transfer diagram in Figure 1, the total number of individuals who were infected and then recovered, and are alive can be decomposed as

$$R_t = K_t + U_t \tag{27}$$

where K_t is the number of individuals who were tested positive, were isolated and then recovered (known), and U_t is the number of individuals who contracted the infection, but were not detected and have recovered (unknown). Equation (20) is then equivalent to the system

$$
\begin{aligned}
\dot{K}_t &= \alpha_t Q_t - \mu K_t \\
\dot{U}_t &= \gamma I_t - \mu U_t.
\end{aligned}
$$

(28)
(29)

From (28), the number of known recovered individuals K_t is given for $-t_o \leq t \leq 0$ by

$$
K_t = \begin{cases} [K_o + \alpha_0 Q_o (t_o + t)]e^{-\mu(t_o+t)} & \text{if } \mu = \alpha_0 + \epsilon_0 \\ K_o e^{-\mu(t_o+t)} + \frac{\alpha_0 Q_o}{\mu-(\alpha_0+\epsilon_0)} \left[e^{-(\alpha_0+\epsilon_0)(t_o+t)} - e^{-\mu(t_o+t)} \right] & \text{if } \mu \neq \alpha_0 + \epsilon_0 \end{cases}.
$$

(30)

After the outbreak, K_t is given by

$$
K_t = \begin{cases} \left[K_0 + \int_0^t \alpha_r Q_r e^{\mu r} dr \right] e^{-\mu t} & \text{if } 0 < t \leq t_e \\ \left[K_e e^{\mu t_e} + \int_{t_e}^t \alpha_r Q_r e^{\mu r} dr \right] e^{-\mu t} & \text{if } t > t_e \end{cases}
$$

(31)

where K_0 (available from (30)) is the number of known recovered individuals before the considered outbreak (recovered from past outbreaks if any), and $K_e = K_{t_e}$ is the number of known recovered individuals at the end of the exponential growth phase. From (29), the number of unknown recovered individuals is

$$
U_t = \begin{cases} \frac{\gamma}{\omega_0+\mu} I_0 \left[e^{\omega_0 t} - e^{-(\omega_0+\mu)t_o - \mu t} \right] & \text{if } -t_o \leq t \leq t_e \\ \left[U_e e^{\mu t_e} + \gamma \delta^{-1} \int_{t_e}^t \dot{\varphi}_r e^{\mu r} dr \right] e^{-\mu t} & \text{if } t > t_e \end{cases}
$$

(32)

where $U_e = U_{t_e}$ is the number of undetected and recovered cases at the end of the exponential growth phase.

2.3.3. The Effective Reproduction Number

From the definition of the effective reproduction number as the average number of secondary cases generated by a primary case, the threshold $\mathcal{R}_t$ corresponds to the product of the transmission rate β_t given in (22) and the average residence time $1/(\gamma + \delta_t + \pi)$ in the class of infectives, *i.e.*

$$
\mathcal{R}_t = \beta_t / (\gamma + \delta_t + \pi).
$$

This effective reproduction number is sometimes referred to as a "quarantine" reproduction number [15] or simply a "control" reproduction number to acknowledge the influence of isolation of identified infectives, and other control measures if any [17]. The basic reproduction number defined as the average number of secondary infections produced when one primary infectious individual enters a completely susceptible population ($S_o = N_o - 1$, $I_o = 1$, $Q_o = 0$, $R_o = 0$), is here given by $\mathcal{R}_o = \left(1 + \frac{\omega_0}{\gamma+\pi} \right) N_o / (N_o - 1)$. This expression is simplified, assuming $N_o / (N_o - 1) = 1$ for the sake of beauty [18] and mostly because N_o is large (recall this is a model assumption), as

$$
\mathcal{R}_o = 1 + \frac{\omega_0}{\gamma + \pi}.
$$

(33)

During the epidemic latency period ($t_o < t < 0$) where the growth is exponential ($\dot{I}_t / I_t = \omega_0$) and the detection rate is $\delta_t = 0$, the time-varying reproduction number is given by

$$\mathcal{R}_t = \left(1 + \frac{\omega_0}{\gamma + \pi}\right)\left(1 + \frac{I_t}{S_t + R_t}\right) \quad \text{for} \ -t_o \leq t < 0. \tag{34}$$

The later expression points out that the time-varying reproduction number increases as susceptibles (S_t) become infectives after adequate contact with infectious individuals (I_t), but decreases as more and more infectives recover and move to the immune class R_t, or control measures become effective. The outbreak reproduction number which accounts for the detection rate $\delta_t = \delta$ is given by

$$\mathcal{R}_0 = \left(1 + \frac{\omega_0}{\gamma + \delta + \pi}\right)\left(1 + \frac{I_0}{S_0 + R_0}\right). \tag{35}$$

indicating how the average residence time is decreased by the quarantining of some infectives. From the outbreak, the time-varying effective reproduction number during the remaining of the exponential phase has the same form

$$\mathcal{R}_t = \left(1 + \frac{\omega_0}{\gamma + \delta + \pi}\right)\left(1 + \frac{I_t}{S_t + R_t}\right) \quad \text{for} \ 0 \leq t \leq t_e. \tag{36}$$

It appears from (34) and (36) that $\mathcal{R}_t > 1$ during the whole exponential growth phase as expected. During the sub-exponential growth phase, the time-varying effective reproduction number is given by

$$\mathcal{R}_t = \left(1 + \frac{z_t}{\gamma + \delta + \pi}\right)\left(1 + \frac{I_t}{S_t + R_t}\right) \quad \text{for} \ t > t_e \tag{37}$$

where $z_t = \ddot{\varphi}_t / \dot{\varphi}_t$ (Table 2).

2.3.4. Total and Lost Cases

The total number of persons infected during an epidemic wave is indicative of the overall cost of the epidemic in terms of its overall impact on the society (in regard to *e.g.* health, work, communication). The total number of new infections denoted $\dot{T}_t$ is given by

$$\dot{T}_t = (\gamma + \delta_t + \pi)I_t + \dot{I}_t. \tag{38}$$

The total number of cases is thus given by

$$T_t = \begin{cases} 1 + \frac{\omega_0 + \gamma + \pi}{\omega_0}\left(I_0 e^{\omega_0 t} - 1\right) & \text{if} \ -t_o \leq t \leq 0 \\ T_0 + \frac{\omega_0 + \gamma + \delta + \pi}{\omega_0} I_0\left(e^{\omega_0 t} - 1\right) & \text{if} \ 0 < t \leq t_e \\ T_e + \delta^{-1}[(\gamma + \delta + \pi)(\varphi_t - \varphi_e) + \dot{\varphi}_t - \dot{\varphi}_e] & \text{if} \ t > t_e \end{cases} \tag{39}$$

where $T_e = T_{t_e}$, $\varphi_e = \varphi_{t_e}$ and $\dot{\varphi}_e = \dot{\varphi}_{t_e}$. The increase in lost cases is $\dot{\Lambda}_t = (\gamma + \pi)I_t$ per unit time so that the number of lost cases Λ_t is given by

$$\Lambda_t = \begin{cases} \frac{\gamma + \pi}{\omega_0}\left(I_0 e^{\omega_0 t} - 1\right) & \text{if} \ -t_o \leq t \leq t_e \\ \Lambda_e + \frac{\gamma + \pi}{\delta}(\varphi_t - \varphi_e) & \text{if} \ t > t_e \end{cases}. \tag{40}$$

with $\Lambda_e = \Lambda_{t_e}$. In particular, the number of lost cases during the entire epidemic latency period is $\Lambda_0 = (\gamma + \pi)(I_0 - 1)/\omega_0$.

2.3.5. Epidemic Peak

The peak of new infections occurs when $\ddot{T}_t$ vanishes. We have from (38)

$$\ddot{T}_t = (\gamma + \delta_t + \pi)\dot{I}_t + \ddot{I}_t. \tag{41}$$

During the exponential growth phase, both $\dot{I}_t$ and $\ddot{I}_t$ are increasing functions of time so that the peak of new infections occurs after t_e, *i.e.* the peak time t_{new} satisfies $t_{new} > t_e$. Hence the peak time is the solution of

$$(\gamma + \delta + \pi)\ddot{\varphi}_t + \dddot{\varphi}_t = 0 \tag{42}$$

which can be solved for t using a numerical root finding routine such as the R [19] function *uniroot* or the Matlab [20] function *fzero*. Afterwards, the peak size $\dot{T}_{new}$ (the maximum number of new infections per unit time) is obtained by inserting t_{new} in (38).

2.4. Long Term Epidemic Dynamics

The specification of the growth model in (1) to an epidemic implicitly assumes that the number of infectives in (8) peaks at time t_p and then tends to zero. The decay of infectives after the peak can happen at various rates, depending on the growth pattern (determined by contacts between infectives and susceptibles or intermediate hosts), the response of infected individuals's organism (natural or induced with medicines or a vaccin) to the disease (recovery and death process) and the testing efforts (detection followed by isolation). There are actually two alternative paths from a disease related state (*i.e.* $I_t > 0$) toward the unique (disease–free) equilibrium P_0: transmissions either stop ($\mathcal{R}_t$ reaches zero) or continue longtime at a rate which cannot sustain an epidemic ($0 < \mathcal{R}_t \leq 1$). We discuss these two scenarii in this section. Because the behaviour of $\mathcal{R}_t$ for $t > t_e$ depends on $z_t = \ddot{\varphi}_t / \dot{\varphi}_t$ (see (37)), we shall make use of the minimum of z_t (over $t > t_e$) and the limit $\lim_{t \to \infty} z_t$ given in Table 3 for the general and limiting expressions of z_t.

Table 3: **Minimum point ($t_{z_{min}} = \arg\min_{t>t_e}\{z_t\}$), minimum value $z_{min} = \min_{t>t_e}\{z_t\}$ and limit $z_{lim} = \lim_{t \to \infty} z_t$ of the ratio $z_t = \ddot{\varphi}_t / \dot{\varphi}_t$ of the growth acceleration $\ddot{\varphi}_t$ to the growth rate $\dot{\varphi}_t$ of the generic growth curve (φ_t) [8] and its limiting cases**

Model	$t_{z_{min}}$	z_{min}	z_{lim}
Generic	$\tau + \frac{1}{\nu\omega\rho}\left(u_z^{-\rho} - 1\right)$	$\nu\omega u_z^{\rho}\left(\frac{\nu+1}{\nu}\frac{u_z}{1+u_z} - (1+\rho)\right)$	0
BR ($\rho \to 0$)	∞	$-\nu\omega$	$-\nu\omega$
HG ($\nu \to 0$, $\nu\omega^{(1+\rho)} \to \tilde{\omega}$)	$\tau + \tilde{\omega}^{-1}\rho^{-(1+\rho)}$	$-\tilde{\omega}\rho^{\rho}$	0
Gompertz ($\rho \to 0$ in HG)	∞	$-\tilde{\omega}$	$-\tilde{\omega}$

Table notes: BR = Bertalanffy-Richards; HG = Hyper-Gompertz; φ_t is as defined in (2) and z_t is available from Table 2, $u_z = (\sqrt{1 - \rho_0} - 1)/\sqrt{1 - \rho_0}$ with $\rho_0 = \nu(\rho + 1)/(\nu + 1)$.

2.4.1. Straight End of Transmissions

The transmission of a target disease ends when the transmission rate β_t and accordingly the number of new infections ($\dot{T}_t$) drops to zero at a finite time point which is solution of the equation

$$z_t + (\gamma + \delta + \pi) = 0. \tag{43}$$

Actually, because the transmission rate per capita per unit time $\beta_t(S_t + R_t)/(N_t - Q_t)$ is a non negative quantity, (22) implicitly assumes that $\dot{I}_t/I_t \geq -(\gamma + \delta_t + \pi)$. This condition holds for $t \leq t_e$ since $\dot{I}_t/I_t = \omega_0 > 0$. For the sub-exponential growth phase ($t > t_e$), the assumption is equivalent to

$$z_t + (\gamma + \delta + \pi) \geq 0. \tag{44}$$

The importance of the inequality in (44) becomes more apparent when considering the reproduction number given in (37): the restriction ensures that $\mathcal{R}_t \geq 0$. Therefore, if (43) has a solution $t_z \in (t_{new}, \infty)$, then the transmission of the infection (from the infectives already present in the population to the susceptibles) ends at $t = t_z$ and $\mathcal{R}_t = 0$ for $t \geq t_z$. The existence of a solution t_z of (43) can be checked by comparing the minimum value z_{min} of z_t (Table 3) to the total rate $(\gamma + \delta + \pi)$ of removals from I_t. Indeed, if we have $z_{min} = -(\gamma + \delta + \pi)$, then $t_z = t_{min}$. Furthermore, if $z_{min} < -(\gamma + \delta + \pi)$, there exists a solution $t_z \in (t_{new}, t_{min})$ which can be found using a numerical routine. In either of these two cases, the number of susceptibles afterwards stays at $S_z = S_{t_z}$ and the number of infectives follows an exponential decay as

$$I_t = I_z \, e^{-(\gamma + \delta + \pi)(t - t_z)} \quad \text{for } t > t_z \tag{45}$$

where $I_z = I_{t_z}$ is given by (8). The number of new detected cases is $\dot{C}_t = \delta I_t$ as before, but the number of known active cases becomes

$$Q_t = \left[Q_z F_z + \delta I_z \int_{t_z}^{t} e^{-(\gamma + \delta + \pi)(r - t)} F_r dr \right] F_t^{-1}. \tag{46}$$

where $Q_z = Q_{t_z}$ is given by (14) and $F_z = F_{t_z}$ is given by (15). Whereas the number K_t of known immunes has the same expression given in (31) with Q_t given by (46), the number U_t of unknown immunes becomes

$$U_t = U_z \, e^{-\mu(t - t_z)} - \frac{\gamma I_z}{\gamma + \delta + \pi - \mu} \left[e^{-(\gamma + \delta + \pi)(t - t_z)} - e^{-\mu(t - t_z)} \right] \tag{47}$$

where $U_z = U_{t_z}$ is given by (32). From (45), the number of infectives falls to 1 at time

$$t_f = t_z + \frac{\log I_z}{\gamma + \delta + \pi}. \tag{48}$$

Finally, since the removal rate of infectives is $\gamma + \delta + \pi$ per unit time, the probability that the number of infectives drops to zero at a time $t_{end} = t_f + r$ with r a non negative integer is $(\gamma + \delta + \pi)(1 - \gamma - \delta - \pi)^r$. Under this scenario, the system (17–20) will tend to the disease free equilibrium P_0 at which the size of the population stabilizes at $N^* = \eta/\mu$.

2.4.2. Asymptotic End of Transmissions

When the shape of the curve of infectives has growth parameters such that $z_t = \ddot{\varphi}_t / \dot{\varphi}_t > -(\gamma + \delta + \pi)$ for $t > t_e$, the transmission of the disease does not stop straightly, but continues at a low rate. Indeed, under this scenario, inserting the limit $\lim_{t\to\infty} I_t = 0$ in (37) yields

$$\mathcal{R}_\infty = \lim_{t\to\infty} \mathcal{R}_t = \left(1 + \frac{z_{lim}}{\gamma + \delta + \pi}\right) \tag{49}$$

where $z_{lim} = \lim_{t\to\infty} z_t \leq 0$ is available from Table 3 (note that $z_{lim} = -\nu\omega$ when $\rho \to 0$ and $z_{lim} = 0$ otherwise). Therefore, $\mathcal{R}_\infty \leq 1$ and the population asymptotically tends to the disease–free equilibrium P_0 [15]. However, if $z_{lim} > -(\gamma + \delta + \pi)$, then we also have $\mathcal{R}_\infty > 0$. For instance, under the simple logistic growth model ($\nu = 1$, $\rho \to 0$), z_t decreases and tends to $-\omega$ as $t \to \infty$ (Table 3) and $\mathcal{R}_\infty = 1 - \omega/(\gamma + \delta + \pi)$ which satisfies $0 < \mathcal{R}_\infty < 1$ (from $\omega < \gamma + \delta + \pi$). In general, when $\rho \neq 0$, the shape of φ_t may allow z_t to properly decrease for $t > t_e$ and become negative from $t > t_{new}$ so that $\mathcal{R}_t < 1$. However, whence z_t reaches its limit $z_{lim} > -(\gamma + \delta + \pi)$, it bounces and tends to 0 (Table 3) so that $R_\infty = 1$.

The limit (49) shows that when $\mathcal{R}_t$ does not sharply reach zero but $\rho \to 0$, the asymptotic reproduction number depends on rate parameters (γ, δ and π) that can be controlled to hasten the disease to die out. In the situation where $\rho \neq 0$, R_∞ is independent of model parameters, so that the long run dynamics is less likely to respond to changes in the rate parameters.

2.5. Statistical Model and Inference

To allow likelihood inference in the growth models in (1) using observed epidemiological data, we followed [7] and assigned to new reported cases Y_t ($t = 1, 2, \cdots, n$) a log-normal distribution with probability density function (pdf)

$$f_Y(Y_t|\boldsymbol{\theta}) = \frac{1}{\sigma(Y_t + 1)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\log(Y_t + 1) - \log(\dot{C}_t + 1)}{\sigma} + \frac{\sigma}{2}\right]^2\right) \tag{50}$$

where $\sigma > 0$ is a dispersion parameter (standard deviation at logarithmic scale). This specification yields the mean $E[Y_t] = \dot{C}_t$ and the variance $Var[Y_t] = (\dot{C}_t + 1)^2 \left(e^{\sigma^2} - 1\right)$ while allowing null values of Y_t. In addition, the numbers of new recoveries G_t and new deaths M_t from known active cases Q_t ($t = 1, 2, \cdots, n$) were modelled using logistic regression models with probability mass functions (pmf)

$$f_G(G_t|\boldsymbol{\theta}, Q_{t-1}, Y_t) = \binom{Q_{t-1} + Y_t}{G_t} \alpha_t^{G_t}(1 - \alpha_t)^{Q_{t-1} + Y_t - G_t} \tag{51}$$

$$f_M(M_t|\boldsymbol{\theta}, Q_{t-1}, Y_t) = \binom{Q_{t-1} + Y_t}{M_t} \epsilon_t^{M_t}(1 - \epsilon_t)^{Q_{t-1} + Y_t - M_t} \tag{52}$$

where $\alpha_t = \left[1 + e^{\kappa_0 + \kappa t}\right]^{-1}$ and $\epsilon_t = \left[1 + e^{\lambda_0 + \lambda t}\right]^{-1}$. The parameter vector indexing the pdf in (50) and the conditional pmf in (51) and (52) is $\boldsymbol{\theta} = (\Omega, \omega, \nu, \rho, \tau, t_e, \sigma, \kappa_0, \kappa, \lambda_0, \lambda)^\top$ when the generic growth curve was considered for the sub-exponential growth phase. If a special case of the generic growth curve is desired, the corresponding restricted parameters must be withdrawn from $\boldsymbol{\theta}$. For instance, the use of a hyper-logistic growth curve ($\nu = 1$) implies $\boldsymbol{\theta} = (\Omega, \omega, \rho, \tau, t_e, \sigma, \kappa_0, \kappa, \lambda_0, \lambda)^\top$. Given Q_0, the conditional log-likelihood of an observed series $\{Y_t, G_t, M_t\}$ with $t = 1, 2, \cdots, n$, as a function of the parameter $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}) \;=\; \ell_Y(\boldsymbol{\theta}) + \ell_G(\boldsymbol{\theta}) + \ell_M(\boldsymbol{\theta}) \tag{53}$$

$$\text{where } \ell_Y(\boldsymbol{\theta}) \;=\; \sum_{t=1}^{n} \log f_Y(Y_t|\boldsymbol{\theta}) \tag{54}$$

$$\ell_G(\boldsymbol{\theta}) \;=\; \sum_{t=1}^{n} \log f_G(G_t|\boldsymbol{\theta}, Q_{t-1}, Y_t) \tag{55}$$

$$\ell_M(\boldsymbol{\theta}) \;=\; \sum_{t=1}^{n} \log f_M(M_t|\boldsymbol{\theta}, Q_{t-1}, Y_t). \tag{56}$$

The conditional maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained using an optimization algorithm to maximize the log-likelihood function ℓ. Note that the three components of $\ell(\boldsymbol{\theta})$ are separable and can be maximized independently. In other words, the parameter vector $\boldsymbol{\theta}$ has the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_Y^\top, \boldsymbol{\theta}_G^\top, \boldsymbol{\theta}_M^\top)^\top$ and the maximum likelihood estimates of the components $\boldsymbol{\theta}_Y = (\Omega, \omega, \nu, \rho, \tau_0, t_e, \sigma)^\top$, $\boldsymbol{\theta}_G = (\kappa_0, \kappa)^\top$ and $\boldsymbol{\theta}_M = (\lambda_0, \lambda)^\top$ can be obtained by maximizing ℓ_Y, ℓ_G and ℓ_M respectively.

Since both the binomial and the log-normal distributions belong to the exponential family, we have considered the common deviance statistic used in Generalized Linear Models [21] for checking the goodness-of-fit of the log-normal model associated to Y_t and the binomial models associated to G_t and M_t. We define the likelihood ℓ_s of the saturated model by replacing $\dot{C}_t$ in (50) by the observed values Y_t, α_t in (51) by the observed daily recovery probabilities $G_t/(Q_{t-1} + Y_t)$ and ϵ_t in (52) by the observed daily death probabilities $M_t/(Q_{t-1} + Y_t)$. Similarly, we define the likelihood ℓ_n of the null model by replacing each $\dot{C}_t$ by the daily mean count $\overline{Y} = n^{-1}\sum_{t=1}^{n} Y_t$, each α_t by the overall daily recovery probability $\bar{\alpha}$ (obtained assuming $\kappa = 0$) and each ϵ_t by the overall daily death probability $\bar{\epsilon}$ (obtained assuming $\lambda = 0$). The residual deviance of the maximum likelihood fit is then given by $DEV_{res} = 2\left(\ell_s - \ell(\widehat{\boldsymbol{\theta}})\right)$ and the null deviance of the null model fit is given by $DEV_{null} = 2(\ell_s - \ell_n)$. The quantity DEV_{res} is a statistic to test the null hypothesis H_0: the assumed model is not significantly different from the unknown model that generated the data. If H_0 is true, then the large sample distribution (*i.e.* as $n \to \infty$) of DEV_{res} is the χ_k^2 distribution with $k = n - m$ degrees of freedom where $m = 12$ is the number of individual model parameters in $\boldsymbol{\theta}$ [22]. If the overall goodness of fit test based on DEV_{res} rejects H_0, then the corresponding statistics (residual deviances) can be computed for the three sub-models (*i.e.* considering the log-likelihoods in (53-56)) to identify the sub-models lacking goodness-of-fit. The percentage of information explained by the maximum likelihood fit for the cumulative data can be evaluated using the common adjusted-coefficient of determination

$$r_a^2 = 1 - \left(1 - r^2\right)\frac{n-1}{n-m_Y} \tag{57}$$

where $r = cor(C_t, Y_{.t})$ is the Pearson's correlation coefficient between C_t and $Y_{.t} = \sum_{j=1}^{t} Y_j$, and m_Y is the number of individual model parameters in $\boldsymbol{\theta}_Y$. The explanative power of the overall fit can be assessed via the adjusted-deviance reduction ratio [23]

$$r_{dev}^2 = 1 - \frac{DEV_{res}}{DEV_{null}}\frac{n-1}{n-m}. \tag{58}$$

Let $H(\boldsymbol{\theta})$ be the hessian matrix of ℓ and define the asymptotic covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = -[H(\boldsymbol{\theta})]^{-1}$. In a large sample, the covariance matrix of the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ is estimated by $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\widehat{\boldsymbol{\theta}}\right)$ and square roots of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}$ provide standard errors for individual parameters in $\boldsymbol{\theta}$. For the selection of the parsimonious

model aggreeing with the observed data, the likelihood ratio statistic can be used. To test a null hypothesis H_0 against an alternative H_1 with $q > 0$ restrictions fewer than H_0, the likelihood ratio (LR) statistic is given by [24]

$$LR \quad = \quad 2\Big[\ell(\widehat{\boldsymbol{\theta}}^{(1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(0)})\Big] \qquad (59)$$

where $\widehat{\boldsymbol{\theta}}^{(0)}$ is the estimate under H_0 and $\widehat{\boldsymbol{\theta}}^{(1)}$ is the estimate under H_1. If the null hypothesis H_0 is true, the test statistic LR converges in distribution to the $\chi^2_{(q)}$ distribution with q degrees of freedom as $n \to \infty$ [25]. There are however distinct special cases of model (2) leading to the same number of parameters (m). For example, we have the Bertalanffy-Richards ($m = 7$), the hyper-logistic ($m = 7$) and the hyper-Gompertz ($m = 7$). We also have the Gompertz ($m = 6$) and the logistic ($m = 6$). In these situations, $q = 0$ and the likelihood ratio test cannot be used. Thereafter, we suggest to consider information criteria such as the Akaike's Information Criterion (AIC) (the lower the better): AIC $= -2\ell(\widehat{\boldsymbol{\theta}}) + 2m$ [26].

3. Application to COVID-19 Data of Western Africa

3.1. Context and Objectives

The region of West Africa has sixteen countries (Benin, Burkina-Faso, Cape Verde, Ghana, Guinea, Côte d'Ivoire, Mali, Mauritania, Niger, Nigeria, Senegal, Togo, Sierra Leone, Liberia, Guinea-Bissau and Gambia), covering 407,816,097 km^2 for a population of about 407,816,097 inhabitant [27]. The first COVID-19 patient was formaly identified in Western Africa in late February (27th) 2020. We considered COVID-19 daily infection (PCR-confirmed cases on the day of reporting), recovery and death data, from February 28th to August 31st, 2020, obtained from the Global Rise of Education plateforme [28]. Although the region is heterogeneous, we treated it as if it were homogeneous. Indeed, it must be keept in mind that the reported COVID-19 cases occurred in small clusters concentrated in the main cities of each country. Hence, the sparsity of the data for the whole region actually reflects data sparcity at national and city levels.

The purpose of this analysis was to demonstrate, by example, the use of the proposed modelling framework. The specific aims were to: (i) model COVID-19 case reporting data (daily PCR-confirmed positives, recoveries and deaths) from Western Africa (February 28th to August 31st, 2020); and (ii) evaluate the transmission pattern of the disease. Most West African governments have planned and subsequently implemented, several control measures, either before or overlapping with the time of diagnosis of the first national cases [29]. The modelling results were thus used to discuss the effectivness of the containment measures implemented by governments across the region, and the implications for the control of the further spread of COVID-19 in West African countries.

3.2. Data Analysis

All computations and statistical analyses were performed in R [19]. The significance level of statistical tests was set to 5%.

3.2.1. Model Fitting

We fitted the generic growth curve to the daily new infections Y_t. We used the *optim* routine of R to maximize the log-likelihood (54). We also fitted three of its special cases (Bertalanffy-Richards, hyper-logistic, hyper-Gompertz), which were compared to the generic model fit using likelihood ratio tests. Instead of directly maximizing the log-likelihoods (55) for $\widehat{\boldsymbol{\theta}}_G$ and (56) for $\widehat{\boldsymbol{\theta}}_M$ with the *optim* routine, we used the *glm* routine of R with the family specification "family = binomial(logit)". Since COVID-19 was a new disease in 2020, we considered the number of known active cases $Q_0 = 0$ at $t = 0$ in (51) and (52). We plotted the daily new positives, recoveries, deaths and actives to provide graphical insights in the fitted models.

3.2.2. Overall Epidemic Dynamics

We analysed the overall dynamics of the COVID-19 epidemic in West Africa using the mechanistic SIQR model described in section 2.3. The rate parameters δ (detection rate), γ and π (recovery and death rates in infected but non detected individuals) cannot be estimated using only the available data sequence $\{Y_t, G_t, M_t\}$ (daily new positives, recoveries and deaths) without additional assumptions on their relationships with the rate parameters for detected cases (α_t and ϵ_t). We obtained from the literature $\delta = 0.009$ [30] and $\gamma + \pi = 1/10$ [14,30] and assumed that the ratio of the daily recovery probability to the daily death probability in non detected infectives is equal to this ratio in the detected individuals at outbreak, *i.e.* before the implementation of treatments, if any. From $\gamma/\pi = \alpha_0/\epsilon_0 \approx 5.1495$, we obtained $\gamma = 1/11.9419$ and $\pi = 1/61.4953$.

Two demographic parameters are required in the SIQR model: the daily recruitment rate of susceptibles (through births and immigration) η (individuals/day) and the per capita natural mortality rate μ (day^{-1}). Using the birth rate ρ_b (total births and net immigrations in a period of length L divided by the average population size $\overline{N}$ during this period), the recruitment rate η is estimated by

$$\eta = \frac{r_b \overline{N}}{L}. \tag{60}$$

Under "natural" (*i.e.* disease–free) conditions where $N_t = S_t$, the variation ΔN of the population size N_t over a period of length L satisfies

$$\Delta N = \left(\frac{\eta}{\mu} - N_i\right)\left(1 - e^{-\mu L}\right) \tag{61}$$

where N_i is the population at the begining of the period. The equation (61) follows by (24) with $I_0 = 0$. The variation ΔN of the population size is given by $\Delta N = r_b \overline{N} - r_d \overline{N}$, where $r_b \overline{N}$ represents the total recruitment during the period and $r_d \overline{N}$ represents the total number of deaths with ρ_d the mortality rate (individuals/day). Consequently, μ can be obtained by solving (61) for μ using $\Delta N = (r_b - r_d)\overline{N}$.

We considered $L = 365.25$ days, $\overline{N} = 401,861,254$, $N_i = 397,429,929$ [27]. Using the annual birth (32.816/1000) and death rates (7.952/1000) [31] and the net annual immigrations (-177,000 individuals) in West Africa [27], we obtained the rates $r_b = (32.816/1000) - (177,000/\overline{N}) = 32.371/1000$ and $r_d = 7.952/1000$. By (60) and (61), we then found and used for our analyses on West Africa, $\eta = 35615.35$ individuals/day and $\mu = 2.1745 \times 10^{-5}$ day^{-1}. We plotted the daily number of new infections, infectives and recovered individuals, as well as the reproduction number in the West African population.

3.2.3. Standard Error and Confidence Interval

Standard errors were obtained for quantities calculated using estimated model parameters by the delta method [32]. For a positive definite parameter or calculated quantity ϕ in general, we first found the estimate $\hat{\phi}$ and its logarithmic scale-standard error $\hat{\sigma}_\phi$ by the delta method and computed its logarithmic scale-mean given by $\hat{\mu}_\phi = \log \hat{\phi} - 0.5\hat{\sigma}_\phi^2$. We then obtained the bounds of its shortest confidence interval as described by [33].

4. Results

4.1. Growth Curve for New Positives and Logistic Regressions for Removals

The maximum likelihood estimates of growth and logistic regression model parameters are presented in Table 4. The growth model involving the generic growth curve was retained. Indeed, the combination of an early exponential growth and the generic growth models was found to be the best growth model for the new positive cases in West Africa, as compared to the combinations of the exponential growth with the Bertalanffy-Richards (likelihood ratio $LR = 60.06$ on $df = 1$ degrees of freedom, p-value <0.001), the hyper-

logistic ($LR = 240.33$, $df = 1$, p-value <0.001) and the hyper-Gompertz ($LR = 512.91$, $df = 1$, p-value <0.001) growth models.

Table 4: **Estimate, standard error (*SE*), Wald test statistic (z-value), p-value (P(> |z|)) and 95% confidence interval ($CI_{95\%}$) for the parameters of the combination of an early exponential growth curve with a generic growth curve (fitted to daily PCR-confirmed positives) and logistic regression parameters (fitted to daily numbers of recoveries and deaths) using West African COVID-19 data from February 28th to August 31st, 2020.**

Parameter	Estimate	*SE*	z-value*	P(> |z|)	$CI_{95\%}$
t_e (*day*)	29.4781	1.2368	80.1417	<0.001	[26.9413, 31.7865]
Ω (*ind.*)	191290.8	6444.5420	360.9696	<0.001	[178756.4, 204008.2]
ω (*day^{-1}*)	0.0148	0.0007	-87.1715	<0.001	[0.0134, 0.0162]
ν	3.7640	0.5280	9.3782	<0.001	[2.7685, 4.8240]
ρ	0.1202	0.0169	-15.1710	<0.001	[0.0884, 0.1541]
τ (*day*)	171.3210	2.4252	70.6431	<0.001	[166.5678, 176.0742]
σ (log *ind.*)	0.3962	0.0201	-18.4774	<0.001	[0.3572, 0.4361]
κ_0	-4.0609	0.0122	-333.6829	<0.001	[-4.0848, -4.0370]
κ	0.0059	0.0001	68.5372	<0.001	[0.0058, 0.0061]
λ_0	-5.7136	0.0682	-83.7346	<0.001	[-5.8473, -5.5799]
λ	-0.0126	0.0006	-22.4195	<0.001	[-0.0137, -0.0115]
ω_o (*day^{-1}*)	0.1660	0.0011	-261.8024	<0.001	[0.1659, 0.1662]
τ_0 (*day*)	-7.2208	0.0226	-319.1971	<0.001	[-7.2651, -7.1764]
ξ (*ind.*)	200.3128	2.7771	382.2758	<0.001	[194.8864, 205.7716]

Table notes: *ind.* = *individuals*; * z-value was computed at logarithmic scale for positive definite parameters (t_e, Ω, ω, ν, ρ, σ and ω_0), so that a p-value < 0.05 indicates significant difference from 1 at 5% level.

The χ^2 test for overall goodness-of-fit indicated that there is lack-of-fit ($\chi^2_{175} = 45700.68$, p-value <0.001) and the overall adjusted-deviance reduction ratio is $r^2_{dev} = 11.60\%$. Looking for the sub-models, we noticed that the estimated growth curve is significantly different from the corresponding null model fit ($\chi^2_6 = 3601.57$, p-value <0.001) and does not lack fit ($\chi^2_{179} = 173.04$, p-value = 0.6115). Indeed, the adjusted-coefficient of determination is $r^2_a = 99.96\%$ and the adjusted-deviance reduction ratio is $r^2_{dev} = 95.26\%$. The overall lack of fit is due to the logistic regression fits for the daily recoveries ($\chi^2_{184} = 45028.51$, p-value <0.001, $r^2_{dev} = 9.25\%$) and deaths ($\chi^2_{184} = 499.13$, p-value <0.001, $r^2_{dev} = 49.08\%$). We nevertheless kept these fits because there are significantly different from the corresponding null model fits (for recoveries: $\chi^2_1 = 4861.55$, p-value <0.001; for deaths: $\chi^2_1 = 486.40$, p-value <0.001).

The Wald test results in Table 4 agree with the likelihood ratio tests considered to select the growth model for the sub-exponential growth phase. Indeed, the 95% confidence bounds for the parameters ν ($\hat{\nu} = 3.76$, $CI(\nu) = [2.77, 4.82]$) and ρ ($\hat{\rho} = 0.12$, $CI(\rho) = [0.09, 0.15]$) indicate that none of the Bertalanffy-Richards growth model ($\rho \to 0$), the hyper-logistic growth model ($\nu = 1$), the logistic growth model ($\rho \to 0$, $\nu = 1$), the hyper-Gompertz growth model ($\nu \to 0$, $\omega\nu^{1+\rho} \to \tilde{\omega}$) and the Gompertz growth model ($\rho \to 0$, $\nu \to 0$, $\omega\nu \to \tilde{\omega}$) are appropriate for this dataset.

The exponential growth phase lasted about one month ($\hat{t}_e = 29.48$, $CI(t_e) = [26.94, 31.79]$ days) after the outbreak (Table 4). The growth curve fitted to the cumulative positive cases is given by

$$C_t = \begin{cases} e^{0.1660 \times (t+7.2208)} & \text{if } 0 < t \leq 29.48 \\ 200.3128 + \dfrac{191290.8}{\left\{1+[1+0.0067 \times (t-171.3210)]^{8.3185}\right\}^{0.2656}} & \text{if } t > 29.48 \end{cases} \qquad (62)$$

where t is the time (day) from the outbreak. Figure 2 (A) shows the daily confirmed positives cases and the fitted growth curve based on a log-normal error structure. The observed peak of new positives happened 148 days after the outbreak (July 24th, 2020) and amounted 2626 positive cases. However, the number of positive cases showed a high variability arround this date (16–29/07/2020), with most daily records roughly ranging between 1600 and 2000 new positive cases (Figure 2 (A)) around an average of 1803 cases (with standard error $SE = 86.48$). The estimated peak time for the new positives cases is around July 15th, 2020, *i.e.* about 139 days after the outbreak ($\hat{t}_p = 138.87$, $SE = 2.26$, and $CI(t_p) = [134.45, 143.31]$ days), and the estimate of the peak size is about 1805 new positive cases ($\hat{\dot{C}}_p = 1804.90$, $SE = 83.40$ and $CI(\dot{C}_p) = [1643.19, 1969.86]$ new positive cases). Assuming a log-normal distribution, the 95% prediction interval for the peak size is $PI(\dot{C}_p) = [1368.93, 2669.55]$ new positive cases, which includes the observed value (2626 new positive cases). The 95% prediction interval for the peak time is $PI(t_p) = [126.59, 151.65]$ days, which also includes the observed peak time (143 days).
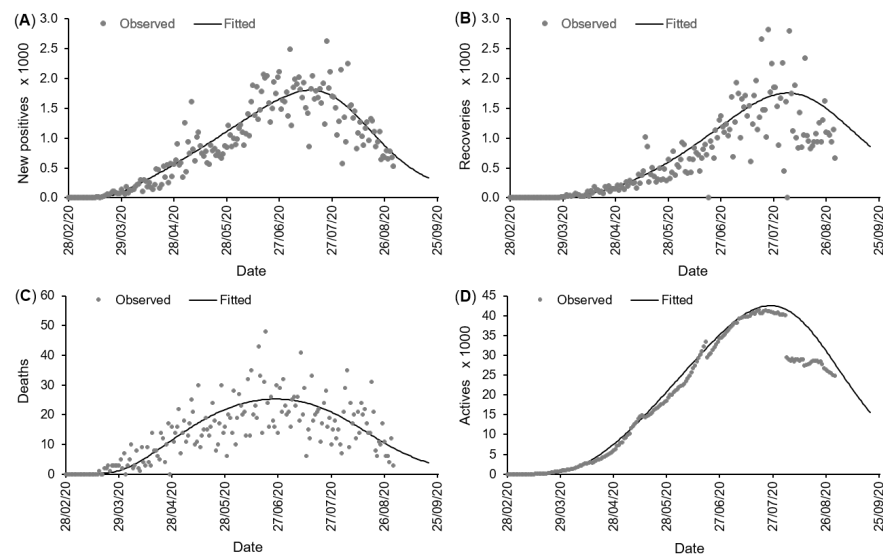
Figure 2. Records of new positive cases $\dot{C}_t$ (A), daily recoveries $\alpha_t Q_t$, (B), daily deaths $\epsilon_t Q_t$ (C) and known actives cases (quarantined at home/hospital) Q_t (D) in COVID-19 daily case reporting data from West Africa (February 28th to August 31st, 2020). The fitted curves are based on a combination of an early exponential growth model and a generic growth model with log-normal error structure for the daily new positive cases $\dot{C}_t$, two logistic regression models for the probabilities of recovery (α_t) and death (ϵ_t), and the combination of $\dot{C}_t$, α_t and ϵ_t (using (14)) for actives Q_t. Two outlying data points (6006 recoveries on 20/06/2020 and 11468 recoveries on 04/08 2020) were removed from the graph (**B**) for a better visualization.

Based on the logistic regression parameters shown in Table 4, the probabilities of removals from the actives (quanrantined) are at time t given by $\hat{\alpha}_t = e^{-4.0609 + 0.0059t}/(1 + e^{-4.0609 + 0.0059t})$ for recovery and $\hat{\epsilon}_t = e^{-5.7136 + -0.0126t}/(1 + e^{-5.7136 + -0.0126t})$ for death. The probabilities of recovery and death were thus $\hat{\alpha}_0 = 0.0169$ and $\hat{\epsilon}_0 = 0.0033$ respectively at outbreak ($t = 0$). The recovery probability then improved, with an odd ratio (recover/not recover) increasing by $\hat{\kappa} = 0.59\%$ ($CI(\kappa) = [0.58, 0.61]\%$) each day. The death probability on contrary decreased, with an odd ratio (die/not die) decreasing by $\hat{\lambda} = -1.26\%$ ($CI(\lambda) = [-1.37, -1.15]\%$) each day. Figure 2 (B-C) show the removals (daily recovery and death) and the fitted values based on the logistic regression models for removal probabilities. We noticed that the lack-of-fit (indicated by the residual deviance test) is due to the very large variability of the observed daily proportions of recoveries and deaths. However, despite the lack-of-fit in the logistic regression fits, the use of the related recovery and death probabilities (α_t and ϵ_t) along with the fitted growth

curve ($\dot{C}_t$), resulted in fitted active cases (Q_t) agreeing to a large extent with the observed daily actives (Figure 2 (D)), with an adjusted-coefficent of determination of 97.08%. The peak of known active cases (Q_t) was observed 143 days after the outbreak (July 19th, 2020) and amounted 41435 actives. The fitted peak is $\widehat{Q}_{max} = 42507.01$ ($SE = 1449.81$, $CI(Q_{max}) = [39687.48, 45368.24]$ actives) about 150 days after the outbreak, *i.e.* around July 26th, 2020 ($t_{Q_{max}} = 149.67$, $SE = 1.78$, $CI(t_{Q_{max}}) = [146.18, 153.17]$ days). The 95% prediction interval is $PI(Q_{max}) = [34807.25, 50893.54]$ actives for the maximum of active cases and $PI(t_{Q_{max}}) = [139.92, 159.71]$ days for the peak time $t_{Q_{max}}$ (July 16th to August 5th, 2020).

4.2. Overall Epidemic Dynamics

The estimate of the duration of the epidemic latency period (delay between the arrival of the first infectious individual and outbreak) is about 25 days ($\hat{t}_o = 24.78$, $SE = 2.55$, $CI(t_o) = [19.91, 29.87]$ days). Accordingly, the first imported COVID-19 case(s) in West Africa likely entered the region during the last week of January and the first week of February (January 28th–February 7th), 2020. The estimate of the basic reproduction number is $\widehat{\mathcal{R}}_o = 2.66$ ($SE = 0.11$, $CI(\mathcal{R}_o) = [2.60, 2.69]$). At outbreak, the number of infectives in the region is estimated at about 61 ($\widehat{I}_0 = 61.17$, $SE = 6.94$, $CI(I_0) = [47.98, 75.05]$) infectives. The estimate of the control reproduction number during the exponential growth phase after the outbreak is $\widehat{\mathcal{R}}_0 = 2.52$ ($SE = 0.12$, $CI(\mathcal{R}_0) = [2.29, 2.76]$).

Figure 3 shows the curves of the daily number of new infections ($\dot{T}_t$), the daily number of infectives (I_t) and the immune fraction of the population ($R_t = K_t + U_t$). As expected, the peak in new infections occurred before the peak in detected infected individuals (observed 143 days after the outbreak). Indeed, the number of new infections peaked about 131 days after the outbreak ($\hat{t}_{new} = 131.12$, $SE = 2.53$, $CI(t_{new}) = [126.18, 136.11]$ days), *i.e.* around July 7th (July 2nd-12th, 2020) to about 22353 ($\widehat{\dot{T}}_{max} = 22352.97$, $SE = 1067.46$, $CI(\dot{T}_{max}) = [20284.04, 24464.98]$) new infections. As of August 31st 2020, the number of known recoveries in the West African region was 140,249. The number of both known and unknown recovered people at this date is estimated at about 1,754,699 individuals ($\widehat{R}_{186} = 1754698.50$, $SE = 40665.66$, $CI(R_{186}) = [1675407.60, 1834783.00]$), *i.e.* about 0.44% of the population the region.
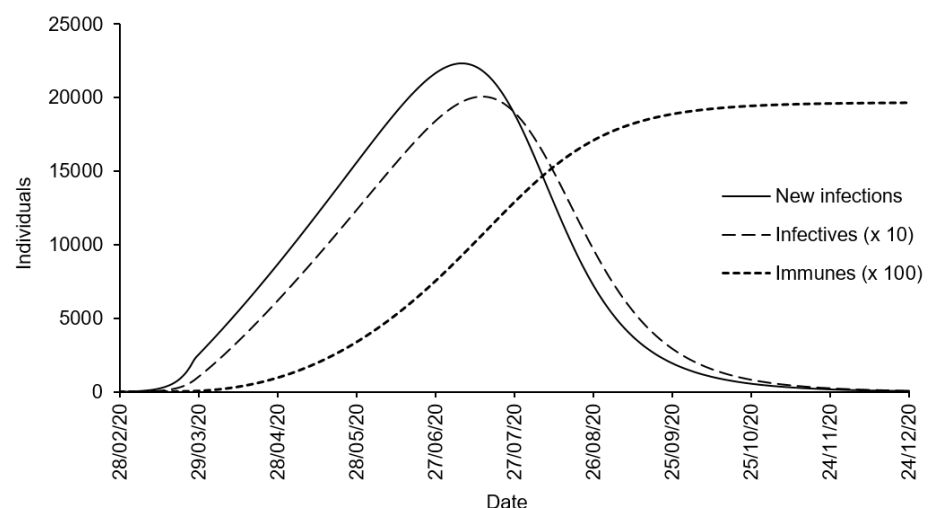
Figure 3. Estimates of the daily number of new infections, infectives and recovered individuals using the COVID-19 daily case reporting data from West Africa (February 28th to August 31st, 2020). The estimates are based on a SIQR model (see (17–20)) with rate parameters $\delta = 0.009$ day^{-1} (detection rate), $\gamma = 1/11.9419$ day^{-1} (recovery rate for non detected), $\pi = 1/61.4953$ day^{-1} (death rate for non detected), $\eta = 35615.35$ individuals/day (recruitment rate) and $\mu = 2.1745 \times 10^{-5}$ day^{-1} (natural mortality rate).

The time-varying effective reproduction number is shown on Figure 4. It appears that the effective reproduction number first decreased during the sub-exponential growth phase (from 2.52 on February 27th, 2020), reaching 1 on July 15th and 0.66 on August 31st, 2020. The effective reproduction number attained a minimum value of 0.61 on 29 September 2020 and then increased with a dynamics indicating $\mathcal{R}_\infty = 1$.
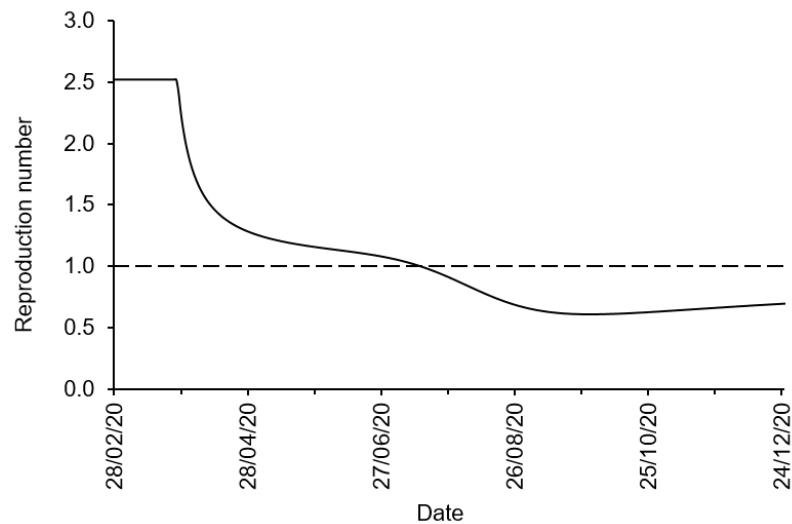
Figure 4. **Time varying effective reproduction number of the 2020 COVID-19 epidemic in West Africa using daily case reporting data (February 28th to August 31st, 2020).** The estimate is based on a SIQR model (see (17–20)) with rate parameters $\delta = 0.009$ day^{-1} (detection rate), $\gamma = 1/11.9419$ day^{-1} (recovery rate for non detected), $\pi = 1/61.4953$ day^{-1} (death rate for non detected), $\eta = 35615.35$ individuals/day (recruitment rate) and $\mu = 2.1745 \times 10^{-5}$ day^{-1} (natural mortality rate).

5. Discussion

The importance of mathematical models in understanding and predicting the course of an epidemic outbreak and in assessing the impacts of public health control measures has been well documented in the current context of the COVID-19 pandemic [17,34–36]. Mathematical models for epidemic dynamics are usually divided into two groups. These include phenomenological models which use an empirical approach ignoring the physical mechanisms underlying the observed patterns in the data, and mechanistic models which structure the population under study in different epidemiological states and incorporate key physical processes in order to explain patterns in the observed data [5].

This study proposes a hybrid modelling framework which combines phenomenological and mechanistic modelling approaches to assess the dynamics of epidemic outbreaks. The proposal uses a combination of the exponential growth model for the initial dynamics of the epidemic and a generic growth curve [8] to capture the observed patterns in the number of detected positive individuals. The observed dynamics of these individuals (recovery and death) is also modelled using logistic regression models. Then, these empirical models are integrated into a deterministic SIQR model [15], to provide an overall view of the dynamics of the target epidemic. This mechanistic model explicitly acknowledges the isolation of the detected positive individuals. It does not however include an exposed-but-noninfectious (E) state as in the SEIQR model [37]. We have illustrated our description of the different epidemiological aspects that the hybrid modelling framework deals with using COVID-19 data from West Africa (February 28th to August 31st, 2020).

The considered generic growth model allows to recover simpler models such as the logistic model, the generalized logistic model [9,10], the hyperlogistic model [8,11], the hyper-Gompertz [8] and the Gompertz curves [12,13] when this is supported by the observed data. Inference in the phenomenological models is conducted using the maximum

likelihood approach. This allows to select the effective parsimonious model fitting the observed data based on likelihood ratio tests [24,25] or information criteria such as the Akaike's Information Criterion [26]. The effectiveness of this approach to phenomenological modelling has been demonstrated on COVID-19 data [7]. Our application on COVID-19 data from West Africa has nevertheless shown that the logistic regression of recoveries and deaths in the identified positive individuals against time can lack fit, as measured by an asymptotic χ^2 test on the residual deviance statistic. Nevertheless, these fits can be improved by adding explanatories (different from time, but related to available health facilities) in the logistic regression models.

Among interest quantities provided by the hybryd modelling framework, we have the epidemic latency period t_o (the time from the appearance of the first infectious case in the population to the outbreak). For the West African region, we obtained $\hat{t}_o = 24.78$ ($CI(t_o) = [19.91, 29.87]$) days. To the best of our knowledge, this is the first estimate of this duration in the region. The result indicates that the first imported COVID-19 case(s) in West Africa likely entered the region during the last week of January and the first week of February (about January 28th–February 7th), 2020. This epidemic latency period is much lower than the 40 days estimated for Italy [14]. This is in line with the relatively late arrival of the virus in the region, compared to the Asian and European continents, and the prevention and detection measures anticipated by many West African governments [29]. We obtained a basic reproduction number of $\widehat{\mathcal{R}}_o = 2.66$ ($CI(\mathcal{R}_o) = [2.60, 2.69]$), which is higher than the estimate $\widehat{\mathcal{R}}_o = 1.85$ ($CI(\mathcal{R}_o) = [1.84, 1.87]$) obtained by [17]. Our estimate is, however, closer to country-specific estimates such as $\widehat{\mathcal{R}}_o = 2.42$ ($CI(\mathcal{R}_o) = [2.37, 2.47]$) in Nigeria [38] and $\widehat{\mathcal{R}}_o = 2.68$ ($CI(\mathcal{R}_o) = [1.99, 3.37]$) in Ghana [39].

During the early phase of the epidemic after the outbreak in West Africa, the detection and isolation of a fraction of infected individuals reduced the reproduction number from $\mathcal{R}_o$ to a control reproduction number of $\widehat{\mathcal{R}}_0 = 2.52$ ($CI(\mathcal{R}_0) = [2.29, 2.76]$), *i.e.* about 5.26% decrease. We estimated the duration of this phase characterized by an exponential growth at $\hat{t}_e = 29.48$ ($CI(t_e) = [26.94, 31.79]$) days after the outbreak. This implies that the control measures implemented by West African governments to limit the transmission of the disease were not effective on average before April 2020. Indeed, apart from measures taken to limit the importation of new positive individuals (travel bans), many actions to limit the local propagation of the disease were first implemented in late March 2020 [29] (*e.g.* curfew set up on 21 March in Burkina-Faso, on 23 March in Ivory Coast, Mauritius and Senegal and on 26 March in Mali; city lockdown on 22 March in Ghana, and on 29 March in Nigeria, isolation of the capital from the rest of the country in Ivory Cost on 25 March 2020, and *cordon sanitaire* set up to isolate the South from the rest of the country on 30 March 2020 in Benin). Our results indicated that these measures started to impact the transmission dynamics of the disease from early April 2020.

After the exponential growth phase, the sub-exponential growth pattern allowed the epidemic to peak. The estimated peak time for the detected positive cases was $\hat{t}_p = 138.87$ ($CI(t_p) = [134.45, 143.31]$) days, *i.e.* around 15 July 2020, and close to the observed peak time (24 July 2020). This estimated date has a delay of about 8 days with respect to the estimated peak time of new infections $\hat{t}_{new} = 131.12$ ($CI(t_{new}) = [126.18, 136.11]$) days. This estimate is higher than the value $\hat{t}_{new} = 111$ ($CI(t_{new}) = [108, 112]$) days obtained by [30]. These contrasting results may be related to the more realistic SIQR model considered in this work as compared to the simpler SIR model used by [30] who ignored the quarantine-adjustment of the disease incidence [15]. On the contrary, the estimated maximum number of new infections $\widehat{\dot{T}}_{max} = 22352.97$ ($CI(\dot{T}_{max}) = [20284.04, 24464.98]$) agrees with the value $\widehat{\dot{T}}_{max} = 25267$ ($CI(\dot{T}_{max}) = [24239, 26294]$) new infections obtained by [30].

Our results showed that the time-varying effective reproduction number has decayed over April–August 2020, reaching 1 on about 15 July 2020, and 0.66 at the end of the considered period (31 August 2020). Based on the modelled dynamics, the effective reproduction number likely reached its minimum value 0.61 around 29 September 2020. However, the reproduction number likely increased again to approach $\mathcal{R}_\infty = 1$ in the

long run. Overall, the various measures decided and enforced by different West African governments, against the first COVID-19 epidemic wave in the region, were able to contain the propagation of the disease (importation of new cases and local transmission) in five months.

However, the COVID-19 pandemic will remain an important issue for a long time, and local region's endemic to the pathogen will likely appear in the long run. This is so because of the following factors: the re-opening of borders and airports in the region to limit the related economic feedback [40,41]; the relaxation of measures such as the ban of sport, political, cultural and religious gatherings [29,42]; and the natural evolution of the SARS-Cov-2 virus [43–46]. The limited resources and capacity of Sub-Sahara Africa countries in general [47–49] to immunize their population through vaccination will compound this threat in the region.

6. Conclusion

There are two common approaches to epidemiological modelling: phenomenological models and mechanistic models. This study proposes a hybrid framework which combines the two approaches, starting from fitting curves to observed data (confirmed positive cases, recoveries and deaths) and then providing an overall view of the epidemic dynamics by integrating the fitted curves into a compartmental model. The proposal allows to estimate the delay between the appearance of the first infectious case in the population and the outbreak ("epidemic latency period"), the duration of period during which the epidemic growths exponentially, the basic and control reproduction numbers, the peaks (time and size) in positive cases, active cases and new infections. An application to COVID-19 data from West Africa indicated that the hybrid modelling framework can be used to match effective control measures dictated by health policies with changes in the transmission dynamics of the studied disease.

Abbreviations

The following abbreviations are used in this manuscript:

SIQR	Susceptible, Infective, Quarantined, Recovered model
Gen	Generic growth model
BR	Bertalanffy-Richards growth model
HG	Hyper-Gompertz growth model
Gom	Gompertz growth model
pdf	probability density function
pmf	probability mass function
LR	Likelihood Ratio
AIC	Akaike's Information Criterion
ind.	Individuals
SE	Standard Error
CI	Confidence Interval
PI	Prediction Interval

References

1. Chowell, G.; Tariq, A.; Hyman, J.M. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. *BMC medicine* **2019**, *17*, 1–18.
2. Roda, W.C.; Varughese, M.B.; Han, D.; Li, M.Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* **2020**, *5*, 271–281.
3. Jahedi, S.; Yorke, J.A. When the best pandemic models are the simplest. *Biology* **2020**, *9*, 353.
4. Golinski, A.; Spencer, P.D. Modeling the Covid-19 Epidemic using Time Series Econometrics. *medRxiv* **2020**. doi:https://doi.org/10.1101/20
5. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling* **2017**, *2*, 379–398.
6. Chowell, G.; Nishiura, H.; Bettencourt, L.M. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface* **2007**, *4*, 155–166.
7. Tovissodé, C.F.; Lokonon, B.E.; Glèlè Kakaï, R. On the use of growth models to understand epidemic outbreaks with application to COVID-19 data. *Plos one* **2020**, *15*, e0240578.
8. Turner, M.E.; Bradley, E.L.; Kirk, K.A.; Pruitt, K.M. A theory of growth. *Mathematical Biosciences* **1976**, *29*, 367 – 373. doi:https://doi.org/10.1016/0025-5564(76)90112-7.
9. Von Bertalanffy, L. Quantitative laws in metabolism and growth. *The quarterly review of biology* **1957**, *32*, 217–231.
10. Richards, F. A flexible growth function for empirical use. *Journal of experimental Botany* **1959**, *10*, 290–301.
11. Turner Jr, M.E.; Blumenstein, B.A.; Sebaugh, J.L. 265 Note: A generalization of the logistic law of growth. *Biometrics* **1969**, pp. 577–580.
12. Gompertz, B. XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical transactions of the Royal Society of London* **1825**, pp. 513–583.
13. Winsor, C.P. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences of the United States of America* **1932**, *18*, 1.
14. Pedersen, M.G.; Meneghini, M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy. *ResearchGate Preprint (online 21 March 2020)* **2020**, *10*.
15. Hethcote, H.; Zhien, M.; Shengbing, L. Effects of quarantine in six endemic models for infectious diseases. *Mathematical biosciences* **2002**, *180*, 141–160.
16. Hethcote, H.W. The mathematics of infectious diseases. *SIAM review* **2000**, *42*, 599–653.
17. Taboe, H.B.; Salako, K.V.; Tison, J.M.; Ngonghala, C.N.; Kakaï, R.G. Predicting COVID-19 spread in the face of control measures in West Africa. *Mathematical biosciences* **2020**, *328*, 108431.
18. Weiss, H.H. The SIR model and the foundations of public health. *Materials matematics* **2013**, pp. 0001–17.
19. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
20. MATLAB. *version 9.0.0 (R2016a)*; The MathWorks Inc.: Natick, Massachusetts, 2016.
21. Nelder, J.A.; Wedderburn, R.W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* **1972**, *135*, 370–384.
22. Williams, D. Generalized linear model diagnostics using the deviance and single case deletions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1987**, *36*, 181–191.
23. Zhang, D. A coefficient of determination for generalized linear models. *The American Statistician* **2017**, *71*, 310–316.
24. Neyman, J.; Pearson, E.S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. *Biometrika* **1928**, pp. 263–294.
25. Wilks, S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics* **1938**, *9*, 60–62.

26. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* **1986**, *81*, 26853.
27. Worldometers. *Population data*, 2020. Accessed on 20 December, 2020.
28. Roser, M.; Ortiz-Ospina, E. Global Education. *Our World in Data* **2020**. Accessed on 15 October, 2020.
29. Bonnet, E.; Le Marcis, F.; Faye, A.; Sambieni, E.; Fournet, F.; Boyer, F.; Coulibaly, A.; Kadio, K.; Diongue, F.B.; Ridde, V.; others. The COVID-19 Pandemic in Francophone West Africa: From the First Cases to Responses in Seven Countries **2020**.
30. Honfo, S.H.; Taboe, B.H.; Kakai, R.G. Modeling COVID-19 dynamics in the sixteen West African countries. *medRxiv* **2020**.
31. Macrotrends. *Macrotrends*, 2020. Accessed on 20 December, 2020.
32. Cox, C. Delta method. *Encyclopedia of biostatistics* **2005**, *2*.
33. Dahiya, R.C.; Guttman, I. Shortest confidence and prediction intervals for the log-normal. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* **1982**, pp. 277–291.
34. Thomas, D.M.; Sturdivant, R.; Dhurandhar, N.V.; Debroy, S.; Clark, N. A primer on COVID-19 Mathematical Models. *Obesity* **2020**, *28*, 1375–1377.
35. Baba, I.A.; Yusuf, A.; Nisar, K.S.; Abdel-Aty, A.H.; Nofal, T.A. Mathematical model to assess the imposition of lockdown during COVID-19 pandemic. *Results in Physics* **2021**, *20*, 103716.
36. Osayomi, T.; Adeleke, R.; Taiwo, O.J.; Gbadegesin, A.S.; Fatayo, O.C.; Akpoterai, L.E.; Ayanda, J.T.; Moyin-Jesu, J.; Isioye, A. Cross-national variations in COVID-19 outbreak in West Africa: Where does Nigeria stand in the pandemic? *Spatial Information Research* **2020**, pp. 1–9.
37. Jumpen, W.; Wiwatanapataphee, B.; Wu, Y.; Tang, I. A SEIQR model for pandemic influenza and its parameter identification. *International Journal of Pure and Applied Mathematics* **2009**, *52*, 247–265.
38. Adekunle, A.I.; Adegboye, O.; Gayawan, E.; McBryde, E. Is Nigeria really on top of COVID-19? Message from effective reproduction number. *Epidemiology & Infection* **2020**, *148*.
39. Asamoah, J.K.K.; Owusu, M.A.; Jin, Z.; Oduro, F.; Abidemi, A.; Gyasi, E.O. Global stability and cost-effectiveness analysis of COVID-19 considering the impact of the environment: using data from Ghana. *Chaos, Solitons & Fractals* **2020**, *140*, 110103.
40. Amewu, S.; Asante, S.; Pauw, K.; Thurlow, J. The economic costs of COVID-19 in sub-Saharan Africa: insights from a simulation exercise for Ghana. *The European Journal of Development Research* **2020**, *32*, 1353–1378.
41. Renzaho, A. The need for the right socio-economic and cultural fit in the COVID-19 response in Sub-Saharan Africa: examining demographic, economic political, health, and socio-cultural differentials in COVID-19 morbidity and mortality. *International journal of environmental research and public health* **2020**, *17*, 3445.
42. Gilbert, M.; Pullano, G.; Pinotti, F.; Valdano, E.; Poletto, C.; Boëlle, P.Y.; d'Ortenzio, E.; Yazdanpanah, Y.; Eholie, S.P.; Altmann, M.; others. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet* **2020**, *395*, 871–877.
43. Koyama, T.; Weeraratne, D.; Snowdon, J.L.; Parida, L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens* **2020**, *9*, 324.
44. Van Der Made, C.I.; Simons, A.; Schuurs-Hoeijmakers, J.; Van Den Heuvel, G.; Mantere, T.; Kersten, S.; Van Deuren, R.C.; Steehouwer, M.; Van Reijmersdal, S.V.; Jaeger, M.; others. Presence of genetic variants among young men with severe COVID-19. *Jama* **2020**, *324*, 663–673.
45. Ghosh, D.; Bernstein, J.A.; Mersha, T.B. COVID-19 pandemic: The African paradox. *Journal of Global Health* **2020**, *10*.
46. Lone, S.A.; Ahmad, A. COVID-19 pandemic–an African perspective. *Emerging microbes & infections* **2020**, *9*, 1300–1308.
47. Dzinamarira, T.; Dzobo, M.; Chitungo, I. COVID-19: A perspective on Africa's capacity and response. *Journal of medical virology* **2020**, *92*, 2465–2472.
48. Ihekweazu, C.; Agogo, E. Africa's response to COVID-19. *BMC medicine* **2020**, *18*, 1–3.
49. Gaye, B.; Khoury, S.; Cene, C.W.; Kingue, S.; N'Guetta, R.; Lassale, C.; Baldé, D.; Diop, I.B.; Dowd, J.B.; Mills, M.C.; others. Socio-demographic and epidemiological consideration of Africa's COVID-19 response: what is the possible pandemic course? *Nature Medicine* **2020**, *26*, 996–999.