Infectious Disease Modelling 5 (2020) 839–847

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Early dynamics of transmission and projections of COVID-19 in some West African countries

Jules-Clement Assob-Nguedia [a], David Dongo [b], Pierre Evariste Nguimkeu [c, *]

[a] Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Cameroon
[b] Department of Mathematics & Computer Sciences, Faculty of Science, University of Dschang, Cameroon
[c] Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA

ARTICLE INFO

Article history:
Received 12 May 2020
Received in revised form 15 September 2020
Accepted 14 October 2020
Handling Editor: Dr. J Wu

ABSTRACT

The initial cases of novel coronavirus (COVID-19) were identified in most West African countries between late February and early March of 2020. But it is only after March 15, 2020 that the number of cases started rising significantly in these countries. This study analyzes the transmission dynamics of the outbreak in West Africa nearly 5 months after the effective onset. We focus on Cameroon, Ghana, Guinea and Nigeria, which are the four West African countries with the highest numbers of infected cases. We combine models of mathematical epidemiology and publicly available data to estimate the main disease transmission characteristics. In particular, we estimate the initial doubling time, the peak time, the peak rate, the final size and the short-term transmission forecasts of the COVID-19 epidemic for these countries. Policy implications for the effectiveness of control measures and for assessing the potential impact on public health in West Africa are discussed.

1. Introduction

The current novel coronavirus disease 2019 (COVID-19) that burst out in December 2019 in China was declared an epidemic of Public Health Emergency of International Concern as of January 30, 2020 by WHO (Minesh et al., 2020), and to a pandemic on March 11, 2020. As of February 11, 2020, the epidemic registered 42,708 cases in China and spread to 25 countries that reported a total of 395 cases. All continents have now reported confirmed cases of COVID-19. Africa confirmed its first case in Egypt on February 14, 2020. Because China is one of Africa's leading commercial partner; it is believed that the large travel volumes through which severe acute respiratory syndrome coronavirus could reach the continent would have promoted the propagation of the virus in the continent. Yet, at the same date only 03 African countries were infected (Gilbert et al., 2020).

Many studies have been undertaken to analyze the early dynamics of the COVID-19 transmission in almost all continents (e.g., Li et al., 2020; Pongkitivanichkul et al., 2020; Shim et al., 2020). However, little attention has been given to the situation of African countries. This is perhaps due to the slow transmission rate of the epidemic observed in Africa, especially sub-Saharan Africa to date. Indeed, the initial cases of COVID-19 were identified in most West African countries (particularly,

* Corresponding author.
E-mail addresses: jcassob@hotmail.com (J.-C. Assob-Nguedia), dongodavid@yahoo.fr (D. Dongo), nnguimkeu@gsu.edu (P.E. Nguimkeu).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2020.10.006

Cameroon, Guinea, Ghana, and Nigeria) between late February and early March of 2020. But it is only after March 15, 2020 that the number of cases started rising significantly to reach nearly 165,000 cases in these four countries as of September 13, 2020. Nonetheless, nearly 5 months after the effective onset of the pandemic in West Africa, it seems useful to analyze and understand the characteristics of this epidemic in the context of this region in order to envisage appropriate steps.

The purpose of this paper is therefore to fill this gap by providing a data-driven analysis of the COVID-19 pandemic in West Africa. We combine models of mathematical epidemiology and statistical modeling to estimate the main disease transmission characteristics in four major West African countries using publicly available data. This is, to our knowledge, the first study to provide statistical estimates of the dynamics of this epidemic in this region, with the aim of providing insights to the discussion related to mitigation efforts and containment measures. While several approaches such as the Susceptible-Infected-Recovered (SIR) model and its extension (SEIR, SIS, etc.) are often used to model the transmission of epidemics (Murray 2003; Tsanou et al. 2016, 2018; Tsanou et al., 2016), the logistic growth models are very popular and useful to predict the final size of the epidemic (Batista, 2020a,b, Pell et al., 2018, Viboud et al., 2016). The latter has the interesting property that its "*S*-shaped" (Sigmoid curve) feature is suitable to describe processes that consist of a slow early transmission stage, followed by a phase of rapid transmission which then tails off as the susceptible population becomes saturated. Logistic functions are often used in demography, medicine, telecommunication, physics, linguistics, and agriculture. We use it in the present context to estimate the initial doubling time, the peak time, the peak rate, the final size and the short-term transmission forecasts of the COVID-19 pandemic for these countries. Policy implications for the effectiveness of control measures and for assessing the potential impact on public health are discussed in order to accompany these governments and others in their efforts to curb down the disease progression.

While there might be other channels of inception of this type of epidemic in a population, the current infection came to Africa through an exogenous route, that is, through imported cases from abroad (Gilbert et al., 2020). This further justifies the use of the abovementioned mathematical models to assess its early dynamics. These models rely on the assumption that the initial number of infected cases is exogenously given and generates all the subsequent cases. The rest of the paper is organized as follows. Section 2 formulates the models that we use to assess the disease transmission, and the main quantities of interest are derived analytically. Section 3 provides the empirical analysis where the data and background are described, the estimation approach is discussed, and results are presented. Concluding remarks and policy implications are discussed in Section 4.

2. Model formulation

While many of the approaches used to model disease transmissions are based on the Susceptible-Infected-Recovered (SIR) models and its extensions (Murray 2003; Tsanou et al. 2016, 2018), the Logistic growth model (sometimes referred to as the Verhulst model) is one of the most popular approach (Batista, 2020a,b; Chowell et al., 2014; Pell et al., 2018; Viboud et al., 2016). While both models may give similar results for countries at an advanced stage of the epidemic such as China, Italy, France, USA (e.g. Lega & Brown, 2017), SIR models have different goals and would require a relatively large number of data points that are not available for the countries studied here (Becker & Grenfell, 2017; Lega & Brown, 2017). The logistic model seems more appropriate for countries at their early epidemic stage, and is given by explicit formulas which make statistical analysis much simpler.[1]

When one uses a phenomenological approach, the epidemic dynamics can be described following the logistic growth model. The underlying assumption of this model is that the rate of change in the number of new cases per capita linearly decreases with the number of cases. Hence, if C_t is the accumulated number of cases observed at time t, then the model is given by

$$\frac{dC_t}{dt} = r\left(1 - \frac{C_t}{K}\right)C_t \tag{1}$$

where $r$is the infection rate, and $K$is the final epidemic size. The initial number of cases is $C_0 > 0$and is given. Solving for this equation gives the logistic response function defined by

$$C_t = K[1 + A\exp(-rt)]^{-1} \tag{2}$$

where A is the integration constant, given by $A = \frac{K}{C_0} - 1$.

The number of new cases, $\frac{dC_t}{dt}$, reaches its maximum when its slope changes sign, implying $\frac{d^2C_t}{dt^2} = 0$.

By solving this condition, we obtain that the peak in the number of new cases occurs at time $t_p = \frac{\ln A}{r}$,

The corresponding number of cases is $C_p = \frac{K}{2}$, and the peak rate is given by $f_p = \frac{dC_t}{dt}_{t=t_p} = \frac{rK}{4}$.

Denote by Δ_t the doubling time of the epidemic, that is, the time it takes to double the number of cases. Then $\Delta_t$is obtained by solving the equation $C_{t+\Delta_t} = 2C_t$, which yields:

[1] In the SIR model, one must solve a system of ordinary differential equations on each optimization step, making it less numerically appealing. It is however useful to compute $\mathscr{R}_0$, the basic reproduction number, an important statistic in epidemiology.

Table 1
Summary statistics of the epidemic and background of the countries[a].

Country	Cases	Recov.	Deaths	Start date	Popul. (millions)	GDP capita ($)	Med-age (years)	Popul. 65+ (%)	Density (km^2)
Cameroon	20,009	18,837	415	Feb 24	26.393	1533.7	18.7	3	56
Ghana	45,434	44,342	286	Mar 13	30.922	2202.3	21.5	3	134
Guinea	10,020	9251	63	Mar 14	1.9604	851.00	18.8	3	70
Nigeria	56,177	44,088	1078	Feb 28	204.92	2028.2	18.1	3	226

[a] As of December 09, 2020.

$$\Delta_t = -\frac{1}{r}\ln\left(\frac{1}{2} - \frac{e^{rt}}{2A}\right),\ t < \frac{\ln A}{r} = t_p.$$

Note that $\Delta_t$is positive and increasing with t, and we have $\Delta_t \to \infty$ and $C_t \to \frac{K}{2} = C_p$,when $t \to \frac{\ln A}{r}$. This means the doubling time can only be defined on the left hand side of the peak time, and lost its meaning after that.

Given $\widehat{A}$ and $\widehat{r}$, the estimated values for A and r whose estimation procedure is described in the next section, the average doubling time, Δ, can then be estimated at the early stages of the epidemic (i.e. before the epidemic reaches its peak when $T < \frac{\ln A}{r}$), by

$$\widehat{\Delta} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta}_t = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widehat{r}}\ln\left(\frac{1}{2} - \frac{e^{\widehat{rt}}}{2\widehat{A}}\right).$$

One of the drawbacks of the logistics growth model is that it tends to underestimate the final epidemic size such that the actual number of cases may be slightly larger than that predicted by the logistics model (Batista, 2020b). If one observes that the actual number of cases is starting to systematically go beyond the predicted final state, then the model will no longer be applicable as a second phase of the epidemic is likely to arise.

3. Empirical analysis

This section estimates the model and quantities discussed above using the data. We describe the data and provide a brief statistical background of the sample countries. We then explain the estimation approach and present the results as well as the short-term forecasts that are derived.

3.1. Data and summary statistics

The daily confirmed cases of COVID-19 in Cameroon, Ghana, Guinea, and Nigeria were extracted from publicly available data including Our World in Data website,[2] Worldometer,[3] and Wikipedia. Although the initial cases occurred from late February to early March in these countries, it is only around March 15, 2020 that the number of new cases started rising and the phenomenon became increasingly intense. Hence, in our analysis, we assume that the epidemic started on the latter date with the initial cases corresponding to the reported cases, and the entire populations from these countries were assumed initially susceptible.

Table 1 gives summary statistics of the epidemic in the countries in the sample, including the total number of cases, deaths, recovery cases, and the initial dates where first cases where detected. To understand the background of these countries, we also provided basic country statistics such as population size, GDP (gross domestic product), median age, proportion of population aged 65+, and density of the population. These countries have a relatively young population with median age ranging between 18 and 22 and the proportion of elderly (65+) is 3 percent of total population in these countries. The fact that individuals aged 65 and over are at higher risk than younger ones who might be less likely to become infected (Li et al., 2020) may partly explain the slow growth in the number of infected cases in these countries.[4] Nigeria has the largest population, representing more than 6 times the population of any other country in the sample, and, as a consequence, also has the highest density of population per squared kilometer.

Fig. 1 compares the evolution of infected cases in the sample countries. Overall, the number of infected cases has been consistently higher in Nigeria compared to the other countries. This might be due to the fact that the population of Nigeria is much larger than any other country in the sample and so the number of susceptible cases in Nigeria is likely to be larger as well (at least given the country's relatively high population density, see Table 1). The number of cases in Ghana has also been

[2] See https://ourworldindata.org/coronavirus-source-data.
[3] See https://www.worldometers.info/world-population.
[4] More rigorous studies are however needed to fully assess this hypothesis.

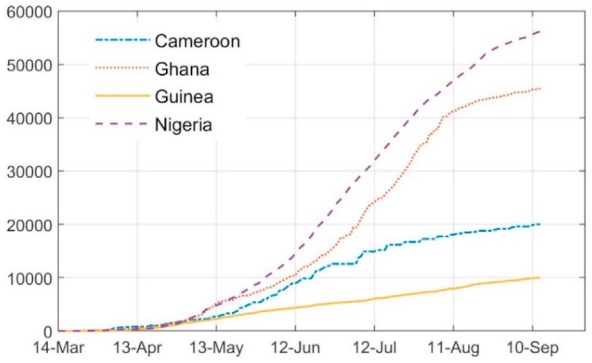


Fig. 1. Trends in total infected cases in select West African countries.

rising drastically in recent months compared to Cameroon, although the death toll remains relatively lower in Ghana as shown in Table 1.

In general, Fig. 1 shows features and patterns that suggest that the number of cases have been growing drastically in these countries and is likely to continue to grow. This therefore calls for the need to harness the pandemic more efficiently. An important step is to understand the transmission characteristics of the epidemic and get a sense of what one should expect in the near future. This is where statistical estimation and short-term forecasting may come in handy.

3.2. Estimation

Clearly, all the relationships derived above are made in a deterministic way for simplicity. In reality, there is a substantial amount of uncertainty that surrounds the intertemporal relationship among stochastic processes. Our empirical analysis accounts for this possibility by specifying an empirical model for the Logistic equation (2) defined by:

$$C_t = K[1 + A\exp(-rt)]^{-1} + u_t,\ t = 1, 2, \ldots, T \tag{3}$$

for a sample of observed time series data of accumulated number of cases $\{C_t,\ t = 1,..,T\}$We assume that the disturbance terms, u_t, are stationary with constant mean and variance.

To fit this infectious disease model to the data, we use numerical optimization that requires reasonable start values for the parameters. A good way to choose those values should use insightful information about the structure of the data. For these countries, we fitted the logistic model which has three unknown parameters, K, r, and Athat need to be estimated. Because the model is nonlinear, the initial guesses need to be made carefully. We proceed by grid search, assuming that by the end of the epidemic, the final size of the epidemic would be a fraction κ of the population thus setting the starting value of K as κN. Then, using the same reasoning as in Nguimkeu and Rekkas (2011), we can derive the starting value for Aas $\frac{\kappa N}{C_0} - 1$ and that of r as $-\ln\left(\frac{(\kappa N - C_1)C_0}{(\kappa N - C_0)C_1}\right)$ By varying the values of κ, (e.g. $\kappa = 0.001,\ 0.0005,\ 0.0001$.) and evaluating the corresponding objective function (i.e. sum of squared residuals or likelihood function), we select the appropriate starting values for K, r, and A. Another excellent method can be found in Batista (2020a,b). We estimated the parameters of this model to fit the data of the sample countries using the maximum likelihood method assuming normality.

Preliminary analysis suggested that the disturbance terms in Equation (3) are serially correlated with an AR (1) structure (autoregressive of order 1) such that

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{4}$$

where $|\rho| < 1$, and the error terms ε_t can be assumed to be independently and normally distributed, i.e. $\varepsilon_t \tilde{} N(0,\ \sigma^2)$.The graphical inspection of u_t is given in Fig. 2 and confirms the suggested error structure with positive autocorrelation. Denoting $\theta = (K,\ r,\ A)$ and $g_t(\theta) = K[1 + A\exp(-rt)]^{-1}$,the estimation can then proceed by finding the values of the parameter vector $(\theta,\ \rho, \sigma^2)$ that maximizes the log-likelihood function defined by[5]

$$L(\theta, \rho) = -\frac{n}{2}\ln 2\,\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}[C - g(\theta)]'\Omega(\rho)[C - g(\theta)]$$

where $C = [C_1, \ldots, C_T]'$; $g(\theta) = [g_1(\theta), \ldots,\ g_T(\theta)]'$;and the weighting matrix $\Omega(\rho)$ is defined by

[5] The derivation of this the log-likelihood function assuming AR(1) errors can be found Nguimkeu and Rekkas (2011). Given the large sample inference that we use here, the normality assumption is innocuous, and all the parameter estimators will be same if we use nonlinear least squares with AR(1) errors. The difference is only in the asymptotic variance.

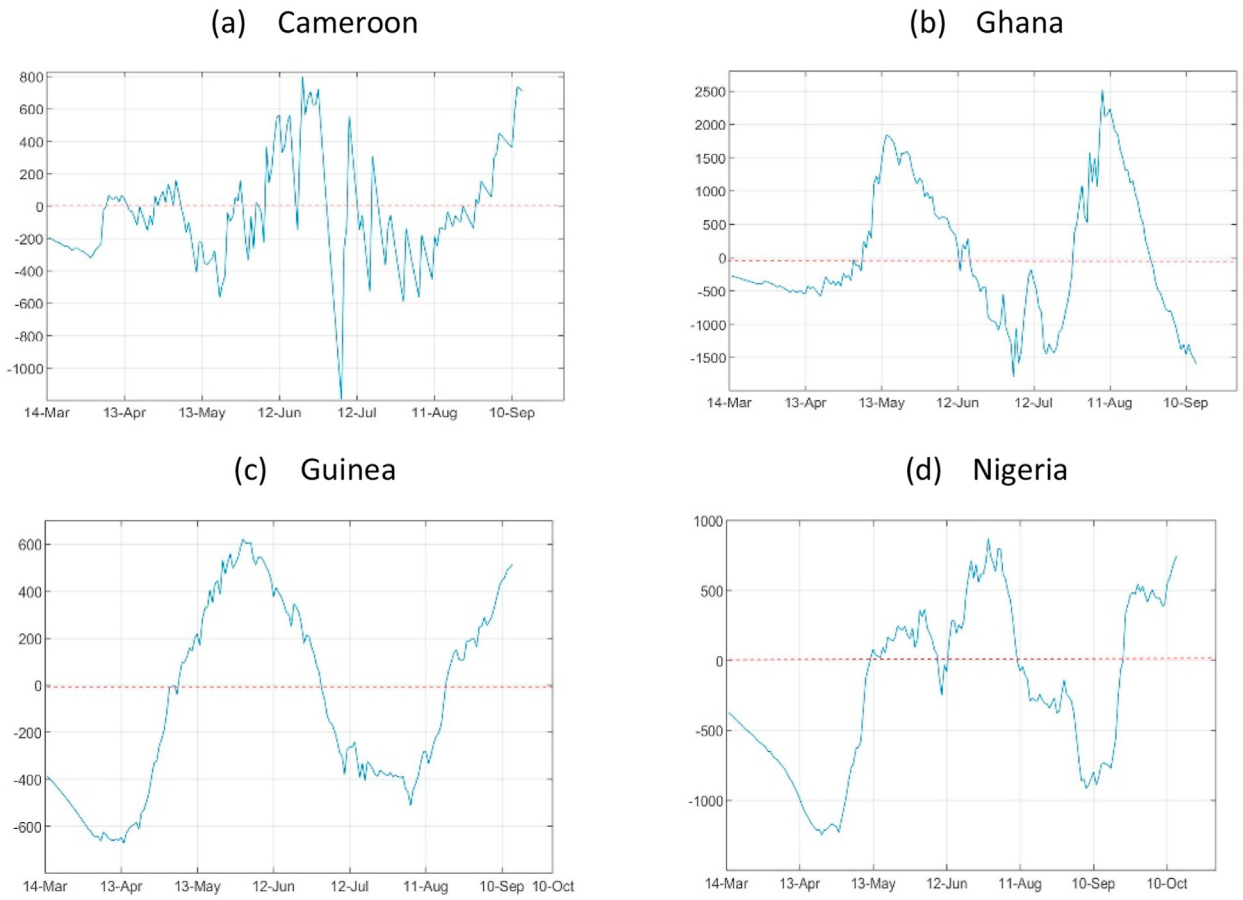


Fig. 2. Residual Plots of the disturbances.u_t

$$\Omega(\rho) = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}$$

All data and code required to reproduce the analysis are available from the authors. From the parameter estimates of the model, $\widehat{K}$, $\widehat{r}$, $\widehat{A}$, $\widehat{\rho}$, and $\widehat{\sigma}^2$, one can easily derive estimates for t_p, C_p, Δ and f_p by plugging-in these parameter values in the corresponding formulas obtained above. Their standard errors can then be computed using the delta method, and the standard errors of the predicted values of the model can be computed as well.

3.3. Results

Table 2 reports the maximum likelihood estimation of the model (3) with errors defined as in Equation (4), including parameter estimates, standard errors and significance. Given the substantial sample sizes, we use asymptotic variances to compute the p-values for evaluating the significance of the parameters.

The results show that the final sizes of the epidemic are estimated to be relatively larger in Nigeria, compared to all other countries with estimated 58,335 cases by November 09, 2020. For the remaining countries, Cameroon, Ghana and Guinea, the epidemic is estimated to reach its final size by October 10, 2020, with final sizes equal to 19,573, 50,254 and 10,399, respectively. The initial doubling time in these countries is estimated to range between 26.65 and 39.18 days on average. The peak time is estimated at 95 and 106 days from the onset for Cameroon and Guinea, respectively, whereas for Ghana and Nigeria, it is relatively larger and estimated at 120 days and 115 days, respectively. This means that the epidemic is progressing relatively slowly in the latter countries compared to the former. These countries are expected to reach their peak rates at very different numbers of cases per day. While Nigeria and Ghana have the highest expected peak rates to reach 637 and 541 cases

Table 2
Model estimation results.

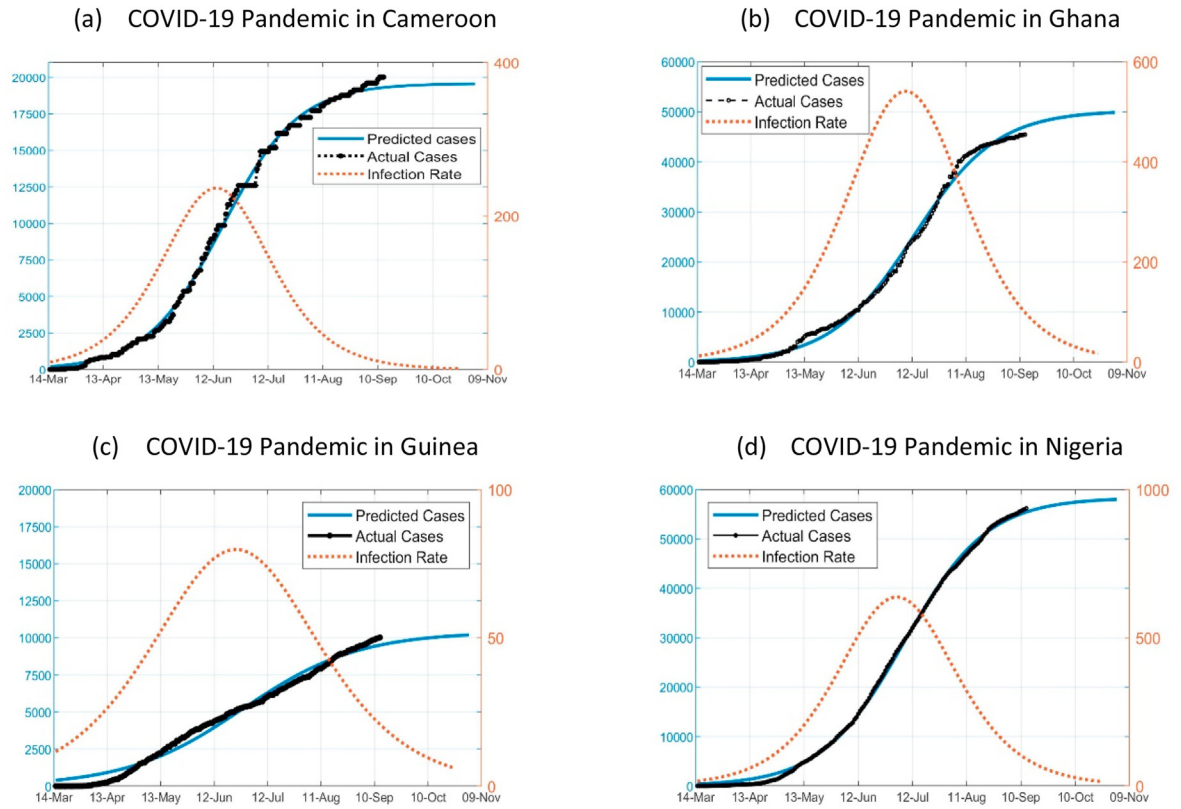
Parameters		Cameroon	Ghana	Guinea	Nigeria
Epidemic size:	K	19,573*** (1402.05)	50,254*** (3201.92)	10,399*** (599.51)	58,335*** (3841.49)
Growth rate:	r	0.0483*** (0.0122)	0.0431*** (0.0103)	0.0307** (0.0192)	0.0437 *** (0.0128)
Displacement:	A	99.00** (21.501)	181.01** (36.121)	260.00** (31.690)	155.00** (28.602)
Autocorrelation:	ρ	0.866***	0.979***	0.996***	0.993***
		(0.023)	(0.021)	(0.105)	(0.132)
Regression variance:	σ^2	171.673	229.98	42.152	86.872
Initial doubling time (day):	Δ	33.012	26.449	39.176	27.649
Peak date:	t_p	95.0561	120.749	105.890	115.490
Peak rate (cases/day):	f_p	236.344	541.487	79.812	637.309
Pseudo R^2		0.981	0.992	0.961	0.994

Note: Standard errors are in parenthesis.

per day, Cameroon and Guinea rates are 236 and 80, respectively. However, these countries growth rates are quite similar and are estimated to range between 0.031 and 0.048. All these estimates are significant at 5% significance level or less.

Fig. 3 depicts the estimation results, particularly the number of cases over time where the graphs in the figure represents the fit of the model, the number of new cases, and their predicted values. The left vertical axis reports the number of cases (actual and predicted), whereas the right vertical axis reports the infection rate (number of new cases).

On this figure, the peak dates are shown to have occurred around June 12th and 23rd for Cameroon and Guinea, respectively, and around July 4th and 10th for Nigeria and Ghana, respectively. These peak dates also correspond to the inflexion point in the disease trends. The post-estimation residual diagnostics confirm the independence assumption of the error terms ε_t, and a graphical inspection of these residuals is given in Fig. 4. There are many other sigmoid models that are often used to analyze disease spreads and diffusion processes in the literature (see, e.g., Meade and Islam 2002 for a review). However, as explained by Martino (2003), the alternative growth curve most commonly used by forecasters is the Gompertz



Note: Left vertical axis is for the number of cases (actual and predicted); Right vertical axis is for the infection rate

Fig. 3. Transmission of COVID-19 epidemic in Cameroon, Ghana, Guinea and Nigeria.

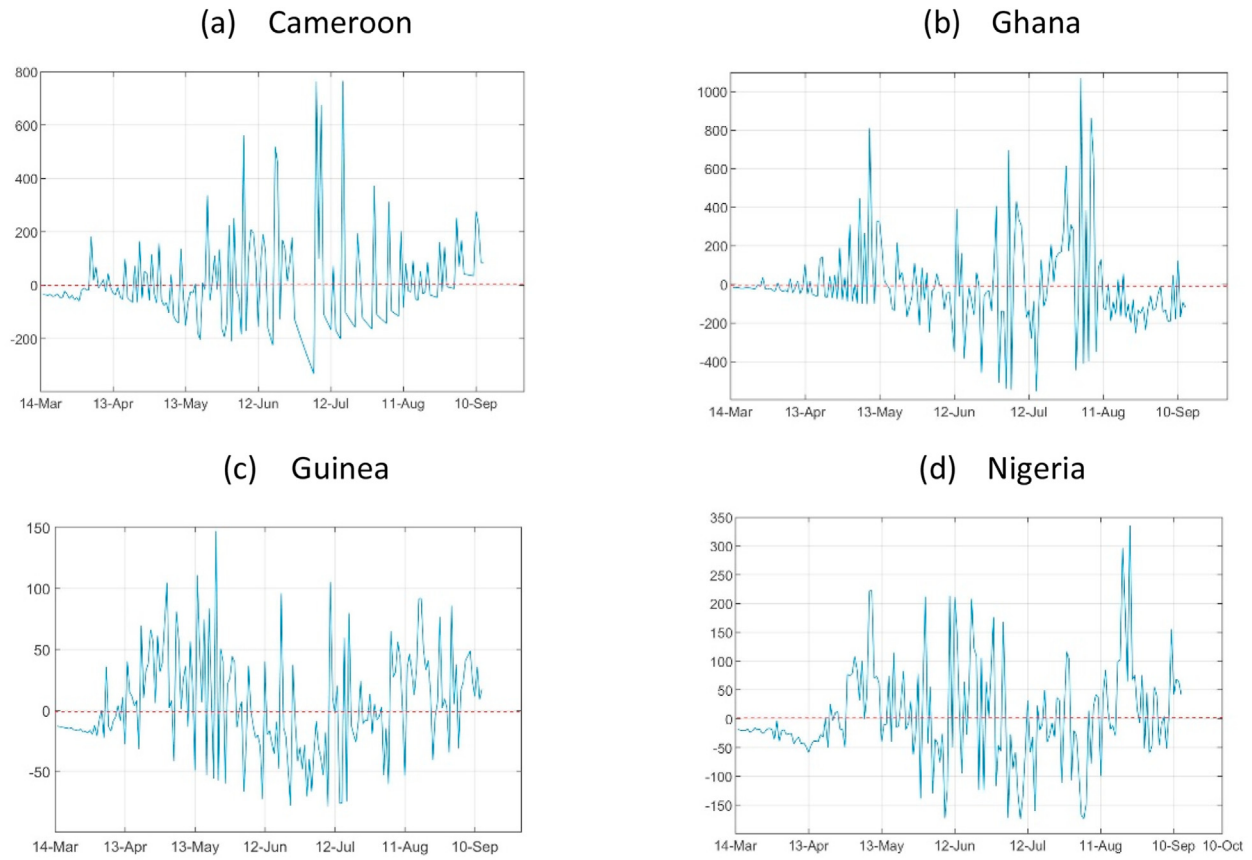


Fig. 4. Residual diagnostic of the disturbances.ε_t

model. Using a selection test developed by Nguimkeu (2014), we found that the Logistic curve better fits our data than the Gompertz.

3.4. Short-term forecasting

The parameter estimates obtained for the models are data-driven. Hence they are as reliable as the data are. These estimates should therefore be understood and interpreted with caution as the quality of the available data is uncertain and likely misreported. The uncertainties in these data are especially more serious in African countries where testing and recording procedures are limited. Nonetheless, if we assume that these data have some informative content, the derived estimates provide useful insights about what we are to expect in the future about the dynamics of the disease spread. Fig. 3 also shows the evolution of the projected number of cases and infection rate over time for the next 50–60 days. In Table 3, we present the estimates that compare the actual values and the predicted values obtained from the model for a period of 20 days, covering 10 days in-sample periods (the before dates) and 10 days out-of-sample periods (the after dates). Standard errors of the projected values are also reported to assess the precision of these projections.

Overall, the in-sample predictions are pretty accurate, with low standard errors and all prediction errors (i.e. percent difference between the actual and the predicted values) falling below 10% of the actual numbers. This suggests that the out-of-sample predictions should also be satisfactory. In particular, the results show that by September 23, (that is, ten days from the estimation date), Cameroon, Ghana, Guinea and Nigeria will have on average 19403, 48110, 9727 and 56424 cases, respectively, with standard errors 1420, 3344, 618, and 3975 respectively, if no further actions are taken in the fight against the pandemic in these countries.

As one should expect, these predictions are less accurate as the prediction horizon increases. Assuming that from the current date, stronger measures are taken in these countries to fight the pandemic or there is a new wave of contaminations otherwise, then our predictions may not hold as much since they are based on a *statu quo* hypothesis. However, our predictions would still be useful as they could serve as a counterfactual (or control) basis to evaluate the impact of the policies implemented in these countries. In particular, by comparing the actual number of cases with those predicted in our study, one could assess the extent of the efforts conceded to fight the pandemic.

Table 3
Short term forecasting of infected cases in select West African countries.

	Cameroon	Ghana	Guinea	Nigeria								
Date	Actual	Pred.	Std Error	Actual	Pred.	Std Error	Actual	Pred.	Std Error	Actual	Pred.	Std Error
4-Sep	19604	19151.81	1390.9	44713	45643.6	3060.3	9526	9254.19	581.0	54587	54131.71	3698.1
5-Sep	19604	19171.29	1393.3	44777	45820.74	3078.8	9579	9285.07	583.4	54743	54298.90	3716.7
6-Sep	19604	19189.9	1395.7	44777	45991.71	3096.9	9649	9315.22	585.7	54905	54459.94	3734.8
7-Sep	19604	19207.66	1397.9	44777	46156.67	3114.5	9722	9344.65	587.9	55005	54614.98	3752.4
8-Sep	19604	19224.62	1400.0	45012	46315.8	3131.8	9798	9373.35	590.1	55160	54764.23	3769.5
9-Sep	19604	19240.8	1402.0	45012	46469.26	3148.6	9848	9401.37	592.3	55456	54907.87	3786.2
10-Sep	19848	19256.24	1403.9	45313	46617.2	3165.0	9885	9428.70	594.4	55632	55046.08	3802.4
11-Sep	20009	19270.98	1405.7	45313	46759.8	3181.0	9946	9455.35	596.5	55829	55179.03	3818.2
12-Sep	20009	19285.04	1407.5	45388	46897.21	3196.7	9979	9481.35	598.5	56017	55306.91	3833.5
13-Sep	20009	19298.46	1409.1	45434	47029.58	3211.9	10020	9506.69	600.5	56177	55429.88	3848.5
14-Sep	–	19311.07	1410.6	–	47157.07	3226.7	–	9531.39	602.4	–	55548.11	3863.0
15-Sep	–	19323.48	1412.0	–	47279.87	3241.2	–	9555.48	604.3	–	55661.75	3877.1
16-Sep	–	19335.13	1413.4	–	47398.07	3255.3	–	9578.96	606.2	–	55770.97	3890.7
17-Sep	–	19346.25	1414.7	–	47511.86	3269.0	–	9601.83	608.0	–	55875.93	3904.0
18-Sep	–	19356.85	1415.8	–	47621.36	3282.4	–	9624.11	609.8	–	55976.78	3916.9
19-Sep	–	19366.97	1416.9	–	47726.72	3295.4	–	9645.83	611.5	–	56073.65	3929.4
20-Sep	–	19376.61	1418.0	–	47828.08	3308.1	–	9666.98	613.1	–	56166.70	3941.6
21-Sep	–	19385.81	1418.9	–	47925.57	3320.4	–	9687.58	614.8	–	56256.07	3953.3
22-Sep	–	19394.59	1419.8	–	48019.33	3332.3	–	9707.64	616.4	–	56341.88	3964.7
23-Sep	–	19402.95	1420.6	–	48109.48	3344.0	–	9727.17	617.9	–	56424.27	3975.8

Note: Actual is the actual number of cases, Pred. Is the predicted value from the estimation, and Std Error is the standard error of the predicted value.

4. Conclusions and discussion

Using publicly available data, we estimated the early dynamics of transmissions and projections of the COVID-19 in four major West African countries (Cameroon, Ghana, Guinea and Nigeria). On the basis of this information and employing both models of mathematical epidemiology and statistical inference, we estimated the main disease transmission characteristics of the epidemic in these countries. We found evidence of early sustained transmission of COVID-19 in West Africa, with relatively low growth. We predicted the number of infected cases that are expected to arise in the near future, if no measures are taken or if no new and surprising information becomes available. In all countries, the final size of the epidemic is estimated to be at least 10,000 cases as in Guinea and could reach up to 60,000 cases in Nigeria by early November 2020.

In spite of this seemingly optimistic results in most of these countries compared to other parts of the world, this region remains one with a high potential risk, and considerable efforts to reduce transmission is therefore required to control and prevent the spread of the epidemic. This includes not only social distancing but also the application and execution of strict measures of detection, prevention, and control (Gilbert et al., 2020). These measures include heightened surveillance and rapid identification of suspected cases, followed by patient transfer and isolation, rapid diagnosis, tracing, and follow-up of potential contacts (Minesh et al., 2020). For instance, the incidence of the disease in Ghana and in Nigeria has been evolving in almost similar patterns and close sizes suggesting a better control strategy and preparedness in Nigeria at the early stage (given its 6 times larger population). However, projections suggest that the pandemic may become more severe for Nigeria over time while slowing down in Ghana.

The results found in this study may therefore appear as perhaps surprising overall, given the concerns that the WHO initially raised regarding the potential vulnerability of African countries, including those from West Africa such as Guinea, Ghana, and Nigeria (WHO, 2020). Whether it is a matter of faulty detection, climatic factors or simple fluke, the remarkably low rate of coronavirus infection in African countries, with their fragile health systems, continues to puzzle. This adds to the optimistic predictions and estimation of the epidemic characteristics that we find in this study. Many arguments in favor of the environmental factors in Africa, as well as their demographic and epidemiologic characteristics discussed earlier are advanced as an advantage that countervails the transmission of the COVID-19. However, this should not pose as a disincentive for governments to take adequate measures to halt the progression of the diseases. This is being emphasized as the outbreak is gradually gaining ground in these West African countries and their neighboring countries. Should better strategies to cope with the pandemic be implemented in these countries, the predicted number of cases obtained in this study as well as other disease characteristics (growth rate, final size of the epidemic, etc.) may not hold going forward. The comparison between these predictions and the observed characteristics may then serve to evaluate the impact of the strategies implemented to fight the pandemic.

Finally, the hypothesis that demographic, epidemiologic and geographic factors prevailing in sub-Saharan Africa may explain the slow COVID-19 transmission - in spite of the limited capacities to diagnose and handle such outbreaks - is one that merits more in-depth research and analysis. The authors plan to examine this question in a future research.

Declaration of Competing Interest statements

The Authors have no interests to declare.

Acknowledgement

We thank the editor and an anonymous referee for their comments and suggestions that greatly improved this paper. We are grateful to Valentin Azi, Alain Fofeh and Valery Jiongo for their helpful feedback.

References

Batista, M. (2020a). Estimation of the final size of the second phase of the coronavirus epidemic by the logistic model. *MedRxiv*. https://doi.org/10.1101/2020.03.11.20024901, 03.11.20024901.

Batista, M. (2020b). Estimation of the final size of the COVID-19 epidemic. *MedRxiv*. https://doi.org/10.1101/2020.02.16.20023606, 02.16.20023606.

Becker, A. D., & Grenfell, B. T. (2017). tsiR: An R package for time-series Susceptible-Infected-Recovered models of epidemics. *PLoS ONE, 12*(9), Article e0185528. https://doi.org/10.1371/journal.pone.0185528

Chowell, G., Simonsen, L., Viboud, C., Kuang, Y., & PLOS.. (2014). *Is West Africa approaching a catastrophic phase or is the ebola epidemic slowing down? Different models yield different answers for Liberia*.

Gilbert, M., Giulia, P., Francesco, P., Eugenio, V., Chiara, P., Boëlle, P.-Y., D'Ortenzio, E., Yazdanpanah, Y., Paul Eholie, S., Mathias, A., Bernardo, G., Moritz, U. G. K., & Vittoria, C. (2020). Preparedness and vulnerability of African countries against importations of COVID-19: A modelling study. *The Lancet, 395*, 871–877.

Lega, J., & Brown, H. E. (2017). Data-driven outbreak forecasting with a simple nonlinear growth model. *Epidemics, 17*, 19–26.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., & Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine, 382*(13), 1199–1207. https://doi.org/10.1056/NEJMoa2001316

Martino, J. (2003). A review of selected recent advances in technological forecasting. *Technological Forecasting and Social Change, 70*, 719–733.

Meade, N., & Islam, T. (1998). Technological forecasting, model selection, model stability, and combining models. *Management Science, 44*, 11–15.

Minesh, S., Christina, S., & Katie, W. (2020). *One health update*. CDC.

Nguimkeu, P. (2014). A simple selection test between the Gompertz and logistics growth models. *Technological Forecasting and Social Change, 88*, 98–105.

Nguimkeu, P. E., & Rekkas, M. (2011). Third-order inference for Autocorrelation in nonlinear Regression models. *Journal of Statistical Planning and Inference, 141*, 3413–3421.

Pell, B., Kuang, Viboud, C., & Chowell, G. (2018). Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics, 22*, 62–70.

Pongkitivanichkul, C., Samart, D., Tangphati, T., Koomhin, P., Pimton, P., Punsiri Dam, -O., & Channuie, P. (2020). Estimating the size of COVID-19 epidemic outbreak. *MedRxiv, 95*(8), 1–7. https://doi.org/10.1101/2020.03.28.20044339

Shim, E., Tariq, A., Choi, W., Lee, Y., & Chowell, G. (2020). Transmission potential and severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*. https://doi.org/10.1016/j.ijid.2020.03.031

Tsanou, B., Bowong, S., Lubuma, J., Mann, M., & Luther, M. (2018). Modeling ebola virus disease transmissions with reservoir in a complex virus life ecology. *Mathematical Biosciences and Engineering, 15*, 21–56.

Tsanou, B., Lubuma, J., Moremedi, M., Morris, N., Kaondera-Shava, & Roselyn. (2016). A simple mathematical model for ebola in Africa. *Journal of Biological Dynamics, 11*. https://doi.org/10.1080/17513758.2016.1229817

Viboud, C., Simonsen, L., & Chowell, G. (2016). A generalized-growth model to characterize the early ascendingphase of infectious disease outbreaks. *Epidemics, 15*, 27–37.

World Health Organization. (2020). *Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS)*. Available from: http://www.who.int/csr/sars/country/en/.