scientific reports

OPEN

GrowthPredict: A toolbox and tutorial-based primer for fitting and forecasting growth trajectories using phenomenological growth models

Gerardo Chowell[1✉], Amanda Bleichrodt[1], Sushma Dahal[1], Amna Tariq[2], Kimberlyn Roosa[3], James M. Hyman[4] & Ruiyan Luo[1]

Simple dynamic modeling tools can help generate real-time short-term forecasts with quantified uncertainty of the trajectory of diverse growth processes unfolding in nature and society, including disease outbreaks. An easy-to-use and flexible toolbox for this purpose is lacking. This tutorial-based primer introduces and illustrates *GrowthPredict*, a user-friendly MATLAB toolbox for fitting and forecasting time-series trajectories using phenomenological dynamic growth models based on ordinary differential equations. This toolbox is accessible to a broad audience, including students training in mathematical biology, applied statistics, and infectious disease modeling, as well as researchers and policymakers who need to conduct short-term forecasts in real-time. The models included in the toolbox capture exponential and sub-exponential growth patterns that typically follow a rising pattern followed by a decline phase, a common feature of contagion processes. Models include the 1-parameter exponential growth model and the 2-parameter generalized-growth model, which have proven useful in characterizing and forecasting the ascending phase of epidemic outbreaks. It also includes the 2-parameter Gompertz model, the 3-parameter generalized logistic-growth model, and the 3-parameter Richards model, which have demonstrated competitive performance in forecasting single peak outbreaks. We provide detailed guidance on forecasting time-series trajectories and available software (https://github.com/gchowell/forecasting_growthmodels), including the full uncertainty distribution derived through parametric bootstrapping, which is needed to construct prediction intervals and evaluate their accuracy. Functions are available to assess forecasting performance across different models, estimation methods, error structures in the data, and forecasting horizons. The toolbox also includes functions to quantify forecasting performance using metrics that evaluate point and distributional forecasts, including the weighted interval score. This tutorial and toolbox can be broadly applied to characterizing and forecasting time-series data using simple phenomenological growth models. As a contagion process takes off, the tools presented in this tutorial can help create forecasts to guide policy regarding implementing control strategies and assess the impact of interventions. The toolbox functionality is demonstrated through various examples, including a tutorial video, and the examples use publicly available data on the monkeypox (mpox) epidemic in the USA.

Abbreviations

AIC	Akaike information criterion
ARIMA	Auto regressive integrated moving average
CDC	Centers for disease control and prevention
CI	Confidence interval
GGM	Generalized growth model

[1]Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA. [2]School of Medicine, Stanford University, Stanford, CA, USA. [3]National Institute for Mathematical and Biological Synthesis (NIMBioS), University of Tennessee, Knoxville, TN, USA. [4]Department of Mathematics, Center for Computational Science, Tulane University, New Orleans, LA, USA. ✉email: gchowell@gsu.edu

GLM	Generalized logistic growth model
IS	Interval score
MAE	Mean absolute error
MLE	Maximum likelihood
MSE	Mean squared error
ODE	Ordinary differential equations
PI	Prediction interval
RIC	Richards model
GOM	Gompertz model
SARS	Severe acute respiratory syndrome
USA	United States of America
WIS	Weighted interval score

Reliable short-term forecasts of time series describing an evolving transmission or growth process are essential for decision-making in all aspects of life, including predicting the weather, commercial product demand, the number of cases of an emerging infectious disease, and the growth or decline of the economy. Simple statistical models such as autoregressive integrated moving average (ARIMA) models have been popular for forecasting time series[1–5]. In contrast, dynamical models based on rates of change equations (i.e., differential equations) are less frequently applied by non-specialists in specific scientific fields. However, forecasts based on dynamic models can provide more information about the process of interest by characterizing specific mechanisms and parameters involved in their dynamics[6–9]. For instance, the simple phenomenological growth models discussed in this tutorial can help characterize growth rates, scaling of growth, doubling times, reproduction numbers, and turning points, as well as predicting the size of a growing population (i.e., carrying capacity) or an epidemic outbreak (i.e., epidemic size) at different time horizons with quantified uncertainty[10–17]. Hence, there is a need for an easy-to-use and flexible toolbox to generate short-term forecasts from simple phenomenological growth models with quantified uncertainty of the trajectory of diverse growth processes observed in nature and society, such as infectious disease outbreaks[6].

This tutorial paper introduces a user-friendly MATLAB toolbox to fit and forecast time-series trajectories of infectious diseases using phenomenological dynamic growth models based on ordinary differential equations (ODEs) that will find broad applications in the natural and social sciences. This toolbox is written for various audiences, including students training in time-series forecasting, dynamic growth modeling, parameter estimation, parameter uncertainty and identifiability, model comparison, performance metrics, and forecast evaluation. The toolbox is also helpful for researchers and policymakers who need to conduct short-term forecasts by relying on historical and real-time trajectory data of the process of interest, such as an unfolding epidemic.

This forecasting toolbox utilizes a variety of phenomenological growth models based on ordinary differential equations (ODEs), such as the generalized-logistic growth model (GLM) and the Richards model, which have shown competitive performance in modeling epidemic outbreaks in prior studies[10,18,19]. These models capture exponential and sub-exponential growth patterns that typically follow a rising pattern followed by a decline phase, a common feature of contagion processes[10,14,15,20].

Specifically, models in the toolbox include the 2-parameter generalized-growth model (GGM)[21,22], which has proved useful in characterizing and forecasting the ascending phase of epidemic outbreaks[18], the 2-parameter Gompertz model, the 3-parameter GLM, and the Richards model, which have each demonstrated competitive performance in forecasting single peak epidemics[10,18,19]. The toolbox includes nonlinear least squares estimation (LSQ) and maximum likelihood estimation (MLE) methods with different assumptions about the error structure of the observed data, including normal, Poisson, and negative binomial distributions, as well as uncertainty quantification based on a parametric bootstrapping approach[6,19]. The toolbox also includes functions to quantify forecasting performance using metrics that evaluate point and distributional forecasts, including the weighted interval score (WIS). The toolbox's functions are illustrated using weekly publicly available case data from the monkeypox (mpox) epidemic in the USA.

This tutorial-based primer is organized as follows. After providing an overview of the toolbox functions for users, we introduce the parameter estimation methods included in the toolbox and then describe the underlying methodology and user parameters and functions to calibrate, evaluate, and display model fits to data. Finally, we use specific examples in the context of the monkeypox epidemic in the USA to demonstrate the functions that generate, display, and quantify the performance of model-based forecasts, as well as estimate doubling times and the effective reproduction number. A tutorial video that demonstrates the toolbox functionality is available at: https://www.youtube.com/channel/UC6IzIu-pPcMLlLYAho43loQ.

Implementation
Installing the toolbox

- Download the MATLAB code located in folder '**forecasting_growthmodels code**' from the GitHub repository: https://github.com/gchowell/forecasting_growthmodels
- Create 'input' folder in your working directory where your input data will be stored.
- Create 'output' folder in your working directory where the output files will be stored.
- Open a MATLAB session.

Overview of the toolbox functions
Table 1 lists the names of user functions associated with the toolbox, along with a brief description of their role. The internal functions associated with the toolbox are given in supplementary file 1 (Table 1S). The user needs

to specify the parameters related to model fitting and forecasting in the `options_fit.m` and `options_forecast.m` files.

Overview of the tutorial

The workflow described in this tutorial, summarized in Fig. 1, is composed of 6 main sections: (1) plotting preliminary model simulations, (2) fitting the models to data through statistical inference, (3) plotting the resulting model fits, (4) generating short-term forecasts with quantified uncertainty, (5) plotting the resulting short-term forecasts and the associated performance metrics, and (6) calculating the effective reproduction number in the context of epidemic time series.

Parameter estimation method

Let $y_{t_1}, y_{t_2}, \ldots, y_{t_{n_d}}$ denote the time series of the epidemic trajectory used in calibrating the model. Here, t_j, $j = 1, 2, \ldots, n_d$, are the time points for the time series data, and n_d is the number of observations. Let $f(t, \Theta)$ denote the expected curve of the epidemic trajectory. We can estimate the set of model parameters, denoted by Θ, by fitting the model solution to the observed data via nonlinear least squares (LSQ)[23]; within the MATLAB toolbox, this is realized by setting the parameter `<method1>` to 0 in the `options_fit.m` or `options_forecast.m` files. Least squares estimation is achieved by searching for the set of parameters, Θ, that minimizes the sum of squared differences between the observed data $y_{t_1}, y_{t_2}, \ldots, y_{t_{n_d}}$ and the best fit of the model (model mean) which corresponds to $f(t, \Theta)$. That is, Θ is estimated by $\widehat{\Theta} = \mathrm{argmin} \sum_{j=1}^{n_d} (f(t_j, \Theta) - y_{t_j})^2$. In the following section, we describe different phenomenological growth models included in this toolbox for the expected epidemic trajectory curve $f(t, \Theta)$.

This parameter estimation method (LSQ) weights each data point equally and does not require a specific distributional assumption for y_t, except for the first moment $E[y_t] = f(t; \Theta)$. That is, the mean of the observed data at time t equals the expected count denoted by $f(t, \Theta)$ at time t[24]. Under mild regularity conditions, this method yields asymptotically unbiased point estimates regardless of any misspecification of the variance–covariance error structure. Hence, the estimated model $f(t, \widehat{\Theta})$ best fits the mean of the observed data in terms of the L2 norm. We can use the *fmincon* function in MATLAB to set the optimization problem. Finally, we also employ MATLAB's **MultiStart** feature to specify the number of random initial guesses of the model parameters using the parameter `<numstartpoints>` in the `options_fit.m` or `options_forecast.m` files to thoroughly search for the best-fit parameter estimates.

In addition to nonlinear least squares fitting, we can also estimate model parameters via maximum likelihood estimation (MLE)[25] with specific assumptions about the error structure in the data (e.g., Poisson, Negative binomial) via parameter `<method1>`. The log-likelihood expressions derived for Poisson and negative binomial error structures are given below.

(a) Poisson

For a Poisson error structure, the full log-likelihood of Poisson (`<method1>=1`) is given by:

Function	Role
`options_fit.m`	Specifies the parameters related to model fitting, including the characteristics of the time series data, the model, the parameter estimation method, the error structure, and the calibration period. The structure of the `options_fit.m` file is described in Text 1S (supplementary file 1)
`options_forecast.m`	Specifies the parameters related to model forecasting, including the characteristics of the time series data, the model, the parameter estimation method, the error structure, the calibration period, and forecasting period. The structure of the `options_forecast.m` file is described in Text 2S (supplementary file 1)
`options_Rt.m`	Specifies the parameters related to the estimation of the effective reproduction number, namely the type of generation interval distribution and the mean and variance of the generation interval for the infectious disease of interest
`plotGrowthModel.m`	Plots model solutions where the user provides the type of growth model, parameter values, and the initial condition
`Run_Fit_GrowthModels.m`	Fits a model to data with quantified uncertainty
`Run_Forecasting_GrowthModels.m`	Fits a model to data with quantified uncertainty and generates a model-based forecast with quantified uncertainty
`plotFit_GrowthModels`	Displays the model fit and the empirical distribution of the estimated parameters. It also saves output `.csv` files in the output folder with the model fit, the parameter estimates with 95% confidence intervals (CIs), the calibration performance metrics, and the doubling times estimated from the incidence trajectory
`plotForecast_GrowthModels`	Displays the model-based forecast and the associated performance metrics. Moreover, the data associated with the forecasts, the parameter estimates, the doubling times estimated from the incidence trajectory, and the calibration and forecasting performance metrics, are saved as `.csv` files in the output folder
`plotFit_ReproductionNumber`	Displays the model fit and the effective reproduction number. It also saves `.csv` files in the output folder with the model fit, the parameter estimates with 95% CIs, the calibration performance metrics, and the doubling times estimated from the incidence trajectory
`plotForecast_ReproductionNumber`	Displays the model-based forecast and the corresponding effective reproduction number. Moreover, the data associated with the forecasts, the parameter estimates, the doubling times estimated from the incidence trajectory, and the calibration and forecasting performance metrics, are saved as `.csv` files in the output folder

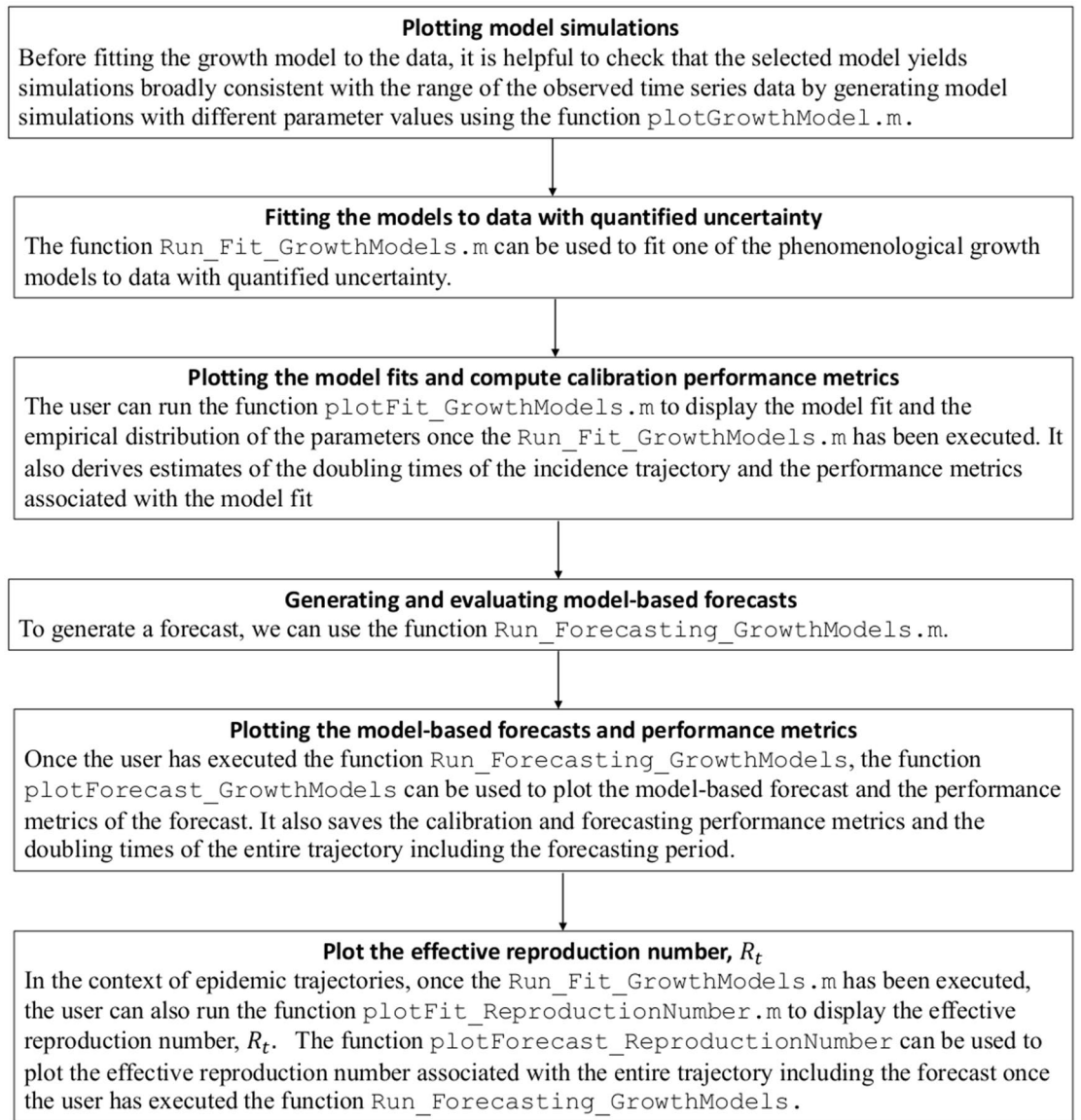**Table 1.** Description of the user functions associated with the toolbox.

Plotting model simulations

Before fitting the growth model to the data, it is helpful to check that the selected model yields simulations broadly consistent with the range of the observed time series data by generating model simulations with different parameter values using the function `plotGrowthModel.m`.

Fitting the models to data with quantified uncertainty

The function `Run_Fit_GrowthModels.m` can be used to fit one of the phenomenological growth models to data with quantified uncertainty.

Plotting the model fits and compute calibration performance metrics

The user can run the function `plotFit_GrowthModels.m` to display the model fit and the empirical distribution of the parameters once the `Run_Fit_GrowthModels.m` has been executed. It also derives estimates of the doubling times of the incidence trajectory and the performance metrics associated with the model fit

---

Generating and evaluating model-based forecasts

To generate a forecast, we can use the function `Run_Forecasting_GrowthModels.m`.

Plotting the model-based forecasts and performance metrics

Once the user has executed the function `Run_Forecasting_GrowthModels`, the function `plotForecast_GrowthModels` can be used to plot the model-based forecast and the performance metrics of the forecast. It also saves the calibration and forecasting performance metrics and the doubling times of the entire trajectory including the forecasting period.

Plot the effective reproduction number, R_t

In the context of epidemic trajectories, once the `Run_Fit_GrowthModels.m` has been executed, the user can also run the function `plotFit_ReproductionNumber.m` to display the effective reproduction number, R_t. The function `plotForecast_ReproductionNumber` can be used to plot the effective reproduction number associated with the entire trajectory including the forecast once the user has executed the function `Run_Forecasting_GrowthModels`.

Figure 1. Overview of the workflow for fitting and forecasting time series trajectories described in this tutorial.

$$\sum_{i=1}^{n} \left\{ y_i ln(\mu_i) - ln(y_i!) - \mu_i \right\}$$

where $\mu_i = f(t_i, \Theta)$ denotes the mean of y_i at time t_i and $f(t_i, \Theta)$ is the mean curve to be estimated from the differential equation.

(b)   Negative binomial

Let $r > 0$ denote the number of failures until the experiment is stopped and $p \in [0, 1]$ denote the success probability in each experiment. Then the number of successes, y, before the r-th failure occurs has a **negative binomial** distribution:

$$f(y|r,p) = \binom{r + y - 1}{y} p^y (1 - p)^r = \frac{1}{y!} \prod_{j=0}^{y-1} (j + r) . p^y (1 - p)^r$$

with mean $= \mu = \frac{rp}{(1-p)}$ and variance$=\sigma^2 = \frac{rp}{(1-p)^2} > \mu$. For n observations $y_1, \ldots, y_n$, the full log-likelihood is

$$l(r,p) = \sum_{i=1}^{n} \left\{ \left\{ \sum_{j=0}^{y_i-1} ln(j+r) \right\} + y_i ln(p) + r ln(1-p) - ln(y_i!) \right\}. \tag{1.1}$$

We can express the likelihood with μ and σ^2 by substituting $p = 1 - \frac{\mu}{\sigma^2}$ and $r = \frac{\mu^2}{\sigma^2 - \mu}$, where $\mu = f(t, \Theta)$ is the mean curve to be estimated from the differential equation.

There are different types of variances commonly used in a negative binomial distribution. If the variance scales linearly with the mean, then $\sigma^2 = \mu + \alpha\mu$,(i.e.,`<method1>=3` in `options_fit.m` or `options_forecast.m`), $p = \frac{\alpha}{1+\alpha}$ and $r = \mu/\alpha$. The full log-likelihood (1.1) can be expressed as follows:

$$l(\Theta, \alpha) = \sum_{i=1}^{n} \left\{ \left\{ \sum_{j=0}^{y_i - 1} ln(j + \alpha^{-1} f(t_i, \Theta)) \right\} + y_i ln(\alpha) - (y_i + \alpha^{-1} f(t_i, \Theta)) ln(1 + \alpha) - ln(y_i!) \right\}. \quad (1.2)$$

If the variance scales quadratically with the mean, then $\sigma^2 = \mu + \alpha\mu^2$ (ie., `<method1>=4` in `options_fit.m` or `options_forecast.m`), $p = \frac{\alpha\mu}{1+\alpha\mu}$ and $r = 1/\alpha$. The full log-likelihood (1.1) can be expressed as follows:

$$l(\Theta, \alpha) = \sum_{i=1}^{n} \left\{ \left\{ \sum_{j=0}^{y_i - 1} ln(j + \alpha^{-1}) \right\} + y_i ln(\alpha f(t_i, \Theta)) - (y_i + \alpha^{-1}) ln(1 + \alpha f(t_i, \Theta)) - ln(y_i!) \right\}. \quad (1.3)$$

The more general form of variance is $\sigma^2 = \mu + \alpha\mu^d$ (i.e.,`<method1>=5` in `options_fit.m` or `options_forecast.m`) with any $-\infty < d < \infty$. Then the full log-likelihood (1.1) can be expressed as follows:

$$l(\Theta, \alpha) = \sum_{i=1}^{n} \left[\left\{ \sum_{j=0}^{y_i - 1} ln(j + \alpha^{-1} \mu_i^{2-d}) \right\} + y_i ln(\alpha\mu_i^{d-1}) - (y_i + \alpha^{-1}\mu_i^{2-d}) ln(1 + \alpha\mu_i^{d-1}) - ln(y_i!) \right]$$
$$(1.4)$$

where $\mu_i = f(t_i, \Theta)$.

Finally, it is worth noting that the above full log-likelihood expressions allow the selection or comparison of models based on different error structures via their AIC_c (corrected Akaike Information Criterion) values. However, if the goal is to focus on different models with the same type of error structure (e.g., normal), we could use simplified likelihood expressions by removing the constants to speed up running time.

Parametric bootstrapping

To quantify parameter uncertainty, we follow a parametric bootstrapping approach, which allows the computation of standard errors and related statistics without closed-form formulas[26]. We generate B bootstrap samples from the best-fit model $f(t, \widehat{\Theta})$, with an assumed error structure specified using parameter `<dist1>` in the `options_fit.m` or `options_forecast.m` files, to quantify the uncertainty of the parameter estimates and construct confidence intervals. The bootstrapping algorithm is given in ref[6]. Typically, the error structure in the data is modeled using a probability model such as the normal, Poisson or negative binomial distribution. Using nonlinear least squares (`<method1>=0`), in addition to a normally distributed error structure (`<dist1>=0`), we can also assume a Poisson (`<dist1>=1`) or a negative binomial distribution (`<dist1>=2`), whereby the variance-to-mean ratio is empirically estimated from the time series. To estimate this constant ratio, we group a fixed number of observations (e.g., 7 observations for daily data into a bin across time), calculate the mean and variance for each bin, and then estimate a constant variance-to-mean ratio by calculating the average of the variance-to-mean ratios over these bins. For non-normal error structure, we estimate parameters using maximum likelihood estimation (MLE) assuming Poisson or negative binomial error structures in the data (`<method1>=1` & `<dist1>=1` for Poisson and `<method1>=3`, `<method1>=4` & `<dist1>=4`, and `<method1>=5` & `<dist1>=5` for the different negative binomial error structures described above).

Specifically, using the best-fit model $f(t, \widehat{\Theta})$, we generate B-times replicated simulated datasets of size n_d, where the observation at time t_j is sampled from the corresponding distribution specified by `<dist1>`. Next, we refit the model to each of the B simulated bootstrap datasets and re-estimate the parameters using the same estimation method as for the original data. The new parameter estimates for each realization are denoted by $\widehat{\Theta}_b$, where $b = 1, 2, \ldots, B$. Using the B sets of re-estimated parameters $\left(\widehat{\Theta}_b \right)$, we can characterize the empirical distribution of each parameter estimate, calculate the variance, and construct confidence intervals for each parameter. The resulting uncertainty around the model fit can similarly be obtained from $f\left(t, \widehat{\Theta}_1 \right)$, $f\left(t, \widehat{\Theta}_2 \right), \ldots, f(t, \widehat{\Theta}_B)$. For the purposes of this tutorial, we characterize the uncertainty using 300 bootstrap realizations (i.e., parameter `<B>=300` in the `options_fit.m` or `options_forecast.m` files).

Model-based forecasts with quantified uncertainty

Based on the best-fit model $f\left(t, \widehat{\Theta} \right)$, we can forecast h days ahead based on the estimate $f(t + h, \widehat{\Theta})$. The uncertainty of the forecasted value can be obtained using the previously described parametric bootstrap method. Let

$$f\left(t + h, \widehat{\Theta}_1 \right), f\left(t + h, \widehat{\Theta}_2 \right), \ldots, f(t + h, \widehat{\Theta}_B)$$

denote the forecasted value of the current state of the system propagated by a horizon of h time units, where $\widehat{\Theta}_b$ denotes the estimation of parameter set Θ from the b_{th} bootstrap sample. We can use these values to calculate the bootstrap variance to measure the uncertainty of the forecasts and use the 2.5% and 97.5% percentiles to construct the 95% prediction intervals (95% PI). We can set the forecasting horizon (h) using the parameter `<forecastingperiod1>` in the `options_forecast.m` file.

For illustration, we fit the models through LSQ with a normal error structure (i.e., `<method1>=0` and `<dist1>=0`) for the monkeypox data. In the `options_fit.m` or `options_forecast.m` files, the values of the parameters related to the parameter estimation method and parametric bootstrapping follow:

```
% <=========================================================>
% <========== Parameter estimation and bootstrapping =======>
% <=========================================================>

method1=0; % Type of estimation method. See below:

% Nonlinear least squares (LSQ)=0,
% MLE Poisson=1,
% MLE (Neg Binomial)=3, with VAR=mean+alpha*mean;
% MLE (Neg Binomial)=4, with VAR=mean+alpha*mean^2;
% MLE (Neg Binomial)=5, with VAR=mean+alpha*mean^d;

dist1=0; % Define dist1 which is the type of error structure.
See below:

%dist1=0; % Normal distribution to model error structure
(method1=0)
%dist1=1; % Poisson error structure (method1=0 OR method1=1)
%dist1=2; % Neg. binomial error structure where var =
factor1*mean where
                % factor1 is empirically estimated from the
time series
                % data (method1=0)
%dist1=3; % MLE (Neg Binomial) with VAR=mean+alpha*mean
(method1=3)
%dist1=4; % MLE (Neg Binomial) with VAR=mean+alpha*mean^2
(method1=4)
%dist1=5; % MLE (Neg Binomial)with VAR=mean+alpha*mean^d
(method1=5)

switch method1
    case 1
        dist1=1;
    case 3
        dist1=3;
    case 4
        dist1=4;
    case 5
        dist1=5;
end

numstartpoints=10; % Number of initial guesses for parameter
estimation procedure using MultiStart

B=300; % number of bootstrap realizations to characterize
parameter uncertainty
```

Phenomenological growth models

Below we describe the growth models included in the toolbox. We use $C(t)$ to denote the cumulative case count at time t and $C'(t)$ denotes the expected epidemic trajectory curve $f(t, \Theta)$.

(a) Generalized-growth model (GGM)

Models commonly used to study the growth pattern of infectious disease outbreaks often assume exponential growth in the absence of control interventions (e.g., compartmental models); however, growth patterns are likely slower than exponential for some diseases, depending on the transmission mode and population structure. For example, Ebola spreads via close contact, therefore, sub-exponential growth patterns would be expected in a constrained population contact structure[27]. The generalized growth model (GGM)[21] includes a "deceleration of growth" parameter, p (range: [0, 1]), that relaxes the assumption of exponential growth. A value of $p = 0$ represents constant (linear) growth, while a value of $p = 1$ indicates exponential growth. If $0 < p < 1$, the growth pattern is characterized as sub-exponential or polynomial.

The GGM is as follows:

$$\frac{dC(t)}{dt} = C'(t) = rC(t)^p,$$

where the derivative $C'(t)$ describes the incidence curve over time t. The positive parameter r is the growth rate parameter ($r > 0$), and p is the scaling of growth parameter[21]. For this model, we estimate $\Theta = (r, p)$. This model can be selected by setting `<flag1>=0` in the `options_fit.m` and `options_forecast.m` files.

(b) Generalized logistic growth model (GLM)

The generalized logistic growth model (GLM) has three parameters and is given by:

$$\frac{dC(t)}{dt} = C'(t) = rC^p(t)\left(1 - \frac{C(t)}{K_0}\right)$$

The growth scaling parameter, p, is also used in the GLM to model a range of early epidemic growth profiles ranging from constant incidence $(p = 0)$, polynomial $0 < p < 1$, and exponential growth dynamics $(p = 1)$. When $p = 1$, this model reduces to the logistic growth model (`<flag1> = 3`). The remaining model parameters are as follows: r is the growth rate, and K_0 is the final cumulative epidemic size. For this model, $\Theta = (r, p, K_0)$. The GLM has been employed to generate short-term forecasts of Zika, Ebola, and COVID-19 epidemics[10,14,15,28]. This model can be selected by setting `<flag1>=1` in the `options_fit.m` and `options_forecast.m` files.

(c) Richards model (RIC)

The well-known Richards model extends the simple logistic growth model and relies on three parameters. It extends the simple logistic growth model by incorporating a scaling parameter, *a*, that measures the deviation from the symmetric simple logistic growth curve[6,29,30]. The Richards model is given by the differential equation:

$$\frac{dC(t)}{dt} = rC(t)\left[1 - \left(\frac{C(t)}{K_0}\right)^a\right]$$

where r is the growth rate, *a* is a scaling parameter and K_0 is the final epidemic size. The Richards model has been employed to generate short-term forecasts of SARS, Zika, Ebola, and COVID-19 epidemics[10,11,14,15,28]. For this model, we estimate $\Theta = (r, a, K_0)$. This model can be selected by setting `<flag1>=4` in the `options_fit.m` and `options_forecast.m` files. A 4-parameter extension of the Richards model is the generalized Richards model (`<flag1>=2`), which incorporates the growth scaling parameter p used in the GGM and GLM.

(d) Gompertz model (GOM)

The 2-parameter Gompertz model is given by:

$$\frac{dC(t)}{dt} = C'(t) = rC(t)e^{-bt}$$

where r is the growth rate and $b > 0$ describes the exponential decline of the growth rate. For this model, we estimate $\Theta = (r, b)$. The GOM model has been employed to generate short-term forecasts of Zika and COVID-19 epidemics[17,31,32]. This model can be selected by setting `<flag1>=5` in the `options_fit.m` and `options_forecast.m` files.

Initial condition

Besides the parameters of the dynamic growth models, it is also possible to estimate the initial number of cases in the time series, rather than fixing $C(0)$ according to the first data point in the time series by specifying the Boolean variable `<fxI0>` in the `options_fit.m` and `options_forecast.m` files. Specifically, `<fixI0>=1`

fixes the initial condition according to the first data point in the time series, whereas `<fixI0>=0` estimates the initial condition along the other model parameters.

Quality of model fit

To assess the quality of model fit, we can compare the AIC_c (corrected Akaike Information Criterion) values of the best-fit models. The AIC_c is given by[33,34]:

$$AIC_c = -2log(likelihood) + 2m + \frac{2m(m+1)}{n_d - m - 1}$$

where m is the number of model parameters and n_d is the number of data points. Specifically for normal distribution, the AIC_c is

$$AIC_c = n_d log(SSE) + 2m + \frac{2m(m+1)}{n_d - m - 1}$$

where $SSE = \sum_{j=1}^{n_d} \left(f\left(t_j, \widehat{\Theta}\right) - y_{t_j}\right)^2$. Thus, this metric accounts for model complexity regarding the number of model parameters and is used for model selection. In the `options_fit.m` and `options_forecast.m` files, the values of the parameters related to the selection of the growth model follow:

```
% <===============================================================>
% <============================ Growth model =====================>
% <===============================================================>


EXP=-1;  % -1 = EXP
GGM=0;   % 0 = GGM
GLM=1;   % 1 = GLM
GRM=2;   % 2 = GRM
LM=3;    % 3 = LM
RICH=4;  % 4 = Richards
GOM=5;   % 5 = Gompertz


flag1=GLM; % Growth model considered in the epidemic trajectory

model_name1='GLM';   % A string provided by the user for the name of the
model

fixI0=1; % 0=Estimate the initial number of cases; 1 = Fix the initial
number of cases according to the first data point in the time series
```

Plotting simulations of the growth model

Before fitting the growth model to the data, it is helpful to check that the selected model yields simulations broadly consistent with the range of the time series data by generating model simulations with different parameter values. For example, if data show systematic differences that contrast with the model solutions (e.g., multiple peaks or plateaus), it may suggest that the model is not the best choice for the data at hand.

The function `plotGrowthModel.m` can be used to plot model solutions where the user provides the type of growth model, parameter values, and initial condition as passing input parameters to the function in the following order: `<flag1>`, r, p, a, K, $C(0)$, and finally the duration of the simulation (i.e., the time span over which the model is numerically solved). For example, the following call plots a simulation of the GLM model (`<flag1>=1`) with the following parameter values: $r = 0.18, p = 0.9$, and $K = 1000$. The initial condition is indicated by $C(0) = 1$, and the total duration of the simulation is set at 200.

`>>plotGrowthModel(1,0.18,0.9,[],10000,1,200)`

Of note, in the above call, the value of parameter a is passed empty ([]) since the GLM model does not use this parameter. This function will generate a figure (Fig. 2A) that shows the corresponding model solution, $dC(t)/dt$. Additional representative simulations from other growth models supported in the toolbox are shown in Fig. 2

Model and code testing

Before fitting the model to the data, it is helpful to check that the code works properly with simulated data before applying the toolbox to actual data. If we can estimate the true parameter values used to generate simulated data from the model, we can confidently move forward that the code is working appropriately with simulated data. For this purpose, we used the following selected values $r = 0.16, p = 0.94$ and $K = 10000$, to generate
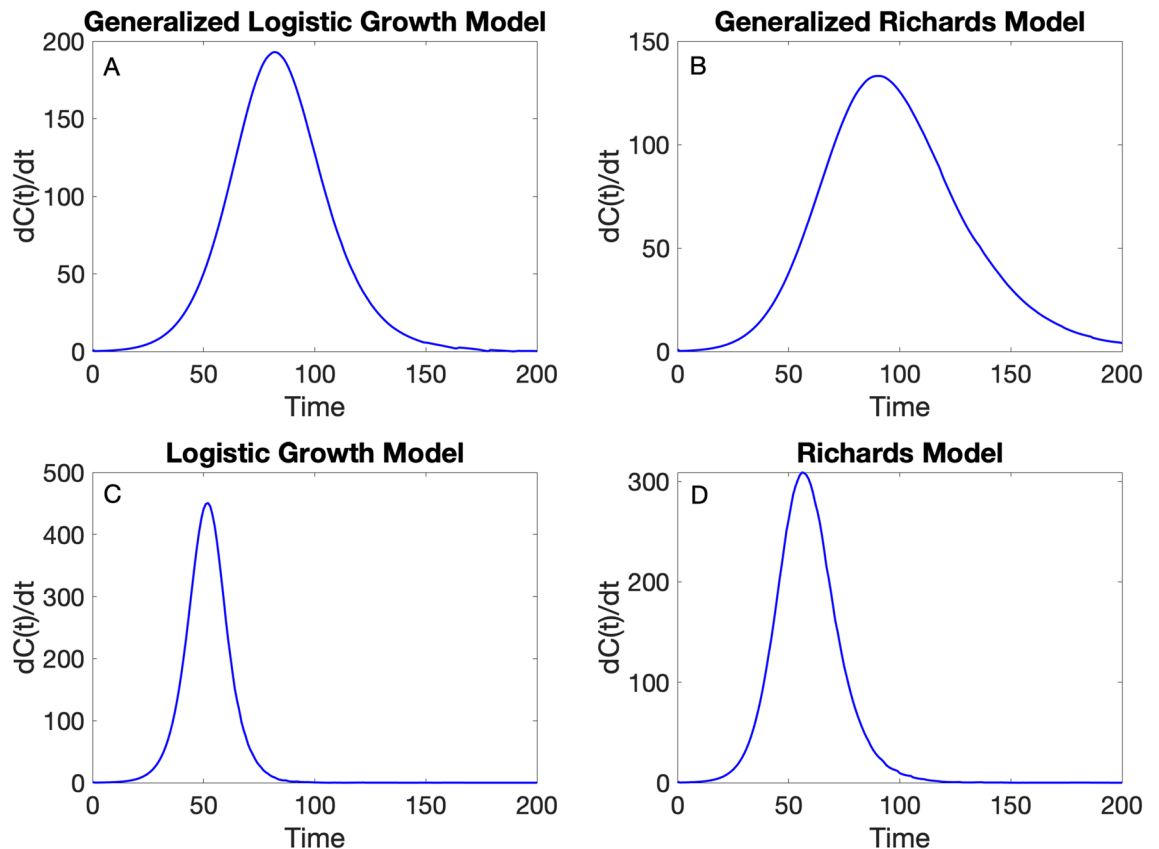
Figure 2. Representative simulations from various growth models, namely (**A**) `<flag1>=1`, (**B**) `<flag1>=2`, (**C**) `<flag1>=3`, and (**D**) `<flag1>=4`, supported in the toolbox using parameter values: $r = 0.18, p = 0.9, a = 0.6, K = 10000$. Here, r refers to the growth parameter, p is a growth scaling parameter (GLM and GGM models), and a is a scaling parameter employed with the Richards model. K refers to the size of the epidemic. The initial condition is $C(0) = 1$, and the total duration of the simulation is set at 200.

a dataset from the GLM with Poisson error structure, initial condition $C(0) = 1$, and duration of 120 days. We then proceeded to estimate the three model parameters for the simulated data using maximum likelihood (i.e., `<method1>=1,<dist1>=1`). The corresponding `options_fit.m` file is given in supplementary file 1 (Code File 1S). Figure 3 shows the GLM fit and the empirical distribution of the estimated parameters for the simulated data. These findings indicate that the parameter estimates align with the "true" parameter values used to simulate the data. Moreover, the model is well calibrated to the data with the 95% prediction interval coverage falling at 95%.

Performance metrics

To assess calibration and forecasting performance, we used four performance metrics: the mean absolute error (MAE), the mean squared error (MSE), the coverage of the 95% prediction intervals (95% PI), and the weighted interval score (WIS)[35]. While it is possible to generate h-time units ahead forecasts of an evolving process, those forecasts looking into the future can only be evaluated when sufficient data for the h-time units ahead has been collected. In the `options_forecast.m` file, the parameter `<getperformance>` is a Boolean variable (0/1) to indicate whether the user wishes to compute the performance metrics of the forecasts when sufficient data is available.

The *mean absolute error* (MAE) is given by:

$$\text{MAE} = \frac{1}{N} \sum_{h=1}^{N} \left| f\left(t_h, \widehat{\Theta}\right) - y_{t_h} \right|,$$

where t_h are the time points of the time series data[36], and N is the calibration or forecasting period length. Similarly, the *mean squared error* (MSE) is given by:

$$\text{MSE} = \frac{1}{N} \sum_{h=1}^{N} \left(f(t_h, \widehat{\Theta}) - y_{t_h} \right)^2.$$
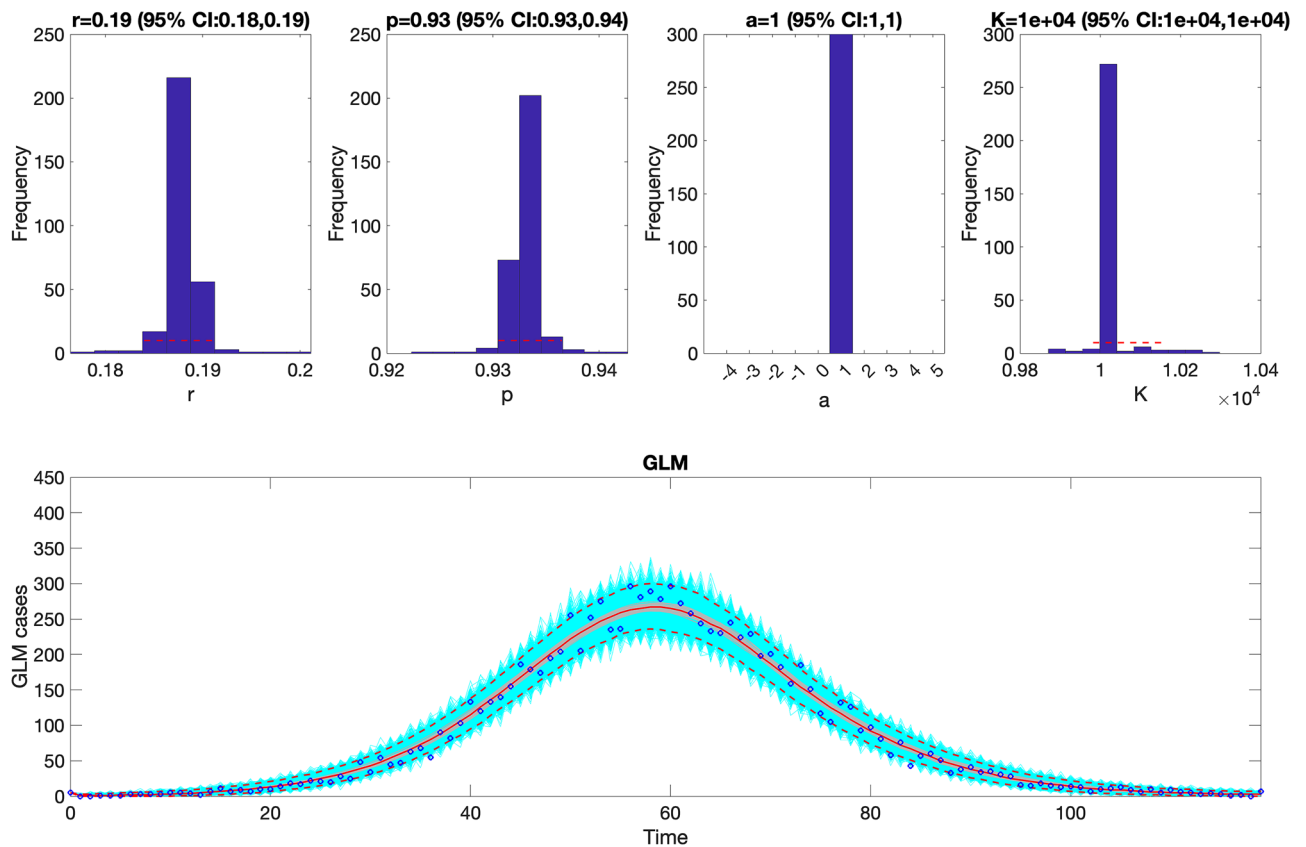
Figure 3. The generalized logistic-growth model (GLM) fit, and the empirical distribution of the estimated parameters for the simulated data. These findings indicate that the parameter estimates of r (growth rate), p (growth scaler) and K (epidemic size) are in line with the "true" parameter values used to simulate the data. The scaling parameter, a, is not estimated when employing the GLM. In the bottom panel, the solid red line is the median model fit and the dashed lines correspond to the 95% prediction intervals. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit.

The coverage of the *95% prediction interval (95% PI)* corresponds to the fraction of data points that fall within the 95% PI, and is calculated as

$$95\% \text{ PI coverage} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}\{Y_t > L_t \cap Y_t < U_t\},$$

where L_t and U_t are the lower and upper bounds of the 95% PIs, respectively, Y_t are the data and $\mathbf{1}$ is an indicator variable that equals 1 if Y_t is in the specified interval and 0 otherwise.

The *weighted interval score* (WIS)[35,37], which is a proper score recently embraced for quantifying model forecasting performance in epidemic forecasting studies[38–42], provides quantiles of predictive forecast distribution by combining a set of Interval Score (IS) for probabilistic forecasts. An IS is a simple proper score that requires only a central $(1 - \alpha) * 100\%$ PI[35] and is described as

$$IS_\alpha (F, y) = (u - l) + \frac{2}{\alpha} \times (l - y) \times 1(y < l) + \frac{2}{\alpha} \times (y - u) \times 1(y > u).$$

In this Eq. 1 refers to the indicator function, meaning that $1(y < l) = 1$ if $y < l$ and 0 otherwise. The terms l and u represent the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the forecast F. The IS consists of three distinct quantities:

1. The sharpness of F, given by the width $u - l$ of the central $(1 - \alpha) \times 100\%$ PI.
2. A penalty term $\frac{2}{\alpha} \times (l - y) \times 1(y < l)$ for the observations that fall below the lower end point l of the $(1 - \alpha) \times 100\%$ PI. This penalty term is directly proportional to the distance between y and the lower end l of the PI. The strength of the penalty depends on the level α.
3. An analogous penalty term $\frac{2}{\alpha} \times (y - u) \times 1(y > u)$ for the observations falling above the upper limit u of the PI.

To provide more detailed and accurate information on the entire predictive distribution, we report several central PIs at different levels $(1 - \alpha_1) < (1 - \alpha_2) < \cdots < (1 - \alpha_K)$ along with the predictive median, $\widetilde{y}$, which can be seen as a central prediction interval at level $1 - \alpha_0 \rightarrow 0$. This is referred to as the WIS, and it can be evaluated as follows:

$$WIS_{\alpha_{0:K}}(F, y) = \frac{1}{K + \frac{1}{2}} \cdot (w_0 \cdot |y - \widetilde{y}| + \sum_{k=1}^{K} w_k \cdot IS_{\alpha_k}(F, y))$$

where, $w_k = \frac{\alpha_k}{2}$ for $k = 1, 2, \ldots . K$ and $w_0 = \frac{1}{2}$. Hence, WIS can be interpreted as a measure of how close the entire distribution is to the observation in units on the scale of the observed data[39,43].

In the `options_forecast.m` file, we can specify the parameters related to the epidemic forecasts, including the forecasting horizon and an indicator variable to specify whether the forecasting performance metrics should be computed:

```
% <===========================================================================>
% <===================== Forecasting parameters =====================>
% <===========================================================================>

getperformance=1; % flag or indicator variable (1/0) to calculate
forecasting performance or not

forecastingperiod=4; % forecast horizon (number of time units ahead)
```

Doubling times

Doubling times characterize the sequence of times at which the cumulative incidence doubles. Denote the times at which cumulative incidence doubles by t_{d_j}, such that $2C(t_{d_j}) = C(t_{d_{j+1}})$ where $t_{d_0} = 0$, $C(t_{d_0}) = C_0$, $j = 1, 2, 3, \ldots, n_g$ and n_g is the total number of times cumulative incidence doubles[44,45]. The actual sequence of "doubling times" is defined as follows:

$$d_j = \Delta t_{d_j} = t_{d_j} - t_{d_{j-1}} \text{where} j = 1, 2, 3, \ldots, n_g.$$

For exponential growth, doubling times remain invariant and are given by $(\ln 2)/r$, whereas the doubling times increase when the growth pattern follows sub-exponential growth[44]. We can characterize the doubling times and their uncertainty from the best-fit model $f\left(t, \widehat{\Theta}\right)$[46]. We can evaluate the uncertainty of the sequence of doubling times and the overall doubling time using the model parameter estimates derived from bootstrapping $\left(\widehat{\Theta}_b\right)$, where $b = 1, 2, 3, \ldots, B$. That is, $d_j(\widehat{\Theta}_b)$ provides a sequence of doubling times for a set of bootstrap parameter estimates, $\widehat{\Theta}_b$, where $b = 1, 2, 3, \ldots, B$. We can use these curves to derive 95% CIs for the sequence of doubling times and quantify the probability of observing a given number of doublings.

The effective reproduction number, R_t

While the basic reproduction number, commonly denoted by R_0, gauges the transmission potential at the onset of an epidemic[47], the effective reproduction number, R_t, captures changes in transmission potential throughout the epidemic[48,49]. We can characterize the effective reproduction number and its uncertainty from the best-fit model $f\left(t, \widehat{\Theta}\right)$[46]. We can derive the 95% CI of R_t from the uncertainty associated with the parameter estimates derived from bootstrapping $(\widehat{\Theta}_b)$ where $b = 1, 2, 3, \ldots, B$. That is, $R_t(\widehat{\Theta}_b)$ provides a curve of the effective reproduction number for a set of parameter values $\widehat{\Theta}_b$ where $b = 1, 2, 3, \ldots, B$. Denote the incidence at calendar time t_j by $I(t_j, \widehat{\Theta}_b)$, and the discretized probability distribution of the generation interval by ρ_{t_j}. The effective reproduction number $R_t(\widehat{\Theta}_b)$ can be estimated using the renewal equation[48,49]:

$$R_{t_j}\left(\widehat{\Theta}_b\right) = \frac{I\left(t_j, \widehat{\Theta}_b\right)}{\sum_{k=0}^{j} I\left(t_{j-k}, \widehat{\Theta}_b\right) \rho_{t_k}}$$

where the denominator represents the total number of cases that contribute (as primary cases) to generate the number of new cases I_{t_j} (as secondary cases) at calendar time t_j[49].

In the `options_Rt.m` file, we can specify the parameters related to the estimation of the effective reproduction number, namely the type of generation interval distribution (`type_GId1`), the mean of the generation interval (`mean_GI1`) and the variance of the generation interval (`var_GI1`) of the infectious disease of interest. For monkeypox, we employ a gamma distributed generation interval with a mean of 1.78 weeks and variance of 0.65 weeks[250] as shown below:

11

```
function [type_GId1,mean_GI1,var_GI1]=options_Rt


%<=======================================================================>
% <============= Reproduction number number parameters ================>
% <=======================================================================>

type_GId1=1; % type of Generation interval distribution (1=Gamma,
2=Exponential, 3=Delta)

mean_GI1=1.78;  % mean of the generation interval distribution

var_GI1=0.65; % variance of the generation interval distribution
```

Rolling window analysis

A rolling window analysis can be useful to assess the stability of the model parameters and forecasts over time and requires the specification of three parameters in the options_fit.m or options_forecast.m files: the start time (<tstart1>) of the first rolling window, the window size (<windowsize1>), and the end time (<tend1>) which corresponds to the start time of the last rolling window. Hence, the first rolling window contains observations for period <tstart1> to <tstart1> + <windowsize1> − 1 , the second rolling window contains observations for period <tstart1> + 1 through <windowsize1>, and so on. Therefore, <windowsize1> corresponds to the length of the calibration period for each model fit. The outputs obtained from the rolling window analysis correspond to the parameter estimates and their uncertainty for each rolling window subsample. A plot of the parameter estimates over the rolling windows can help examine how the estimates change with time. The parameters can be specified in the options_fit.m and options_forecast.m files as shown below:

```
% <==============================================================>
% <============= Parameters of the rolling window analysis ========>
% <==============================================================>

windowsize1=32;  % moving window size

tstart1=1; % time point for the start of rolling window analysis

tend1=1;  %time point for the end of the rolling window analysis
```

However, they can also be passed as input parameters to the fitting and forecasting functions described in the following section.

Results and discussion
The input dataset

For this toolbox, the time series data will be stored in the 'input' folder and needs to be a text file with the extension *.txt. The first column should correspond to the time index: 0,1,2,3, …, and the second column corresponds to the temporal incidence data. If the time series file contains cumulative incidence count data, the name of the time series data file must start with "cumulative".

To illustrate the methodology presented in this tutorial paper, we employ the weekly incidence curve of monkeypox cases reported in the USA from the publicly available data published by the Centers for Disease Control and Prevention (CDC) from the week of 12 May 2022 through the week of 15 December 2022[51]. The data file is pre-loaded in the input folder within the toolbox's working directory (data file path: ./input/ Most_ Recent_Timeseries_US-CDC.txt). A snapshot in Excel of the contents of the file is shown in Fig. 4.

In the options_fit.m and options_forecast.m files, the user specifies the parameters related to model fitting and forecasting, respectively, as shown below. For instance, the parameter < cadfilename1> is a string used to indicate the name of the data file, parameter <caddisease> is a string used to indicate the name of the disease related to the time series data, and < datatype> is a string parameter indicating the nature of the data (e.g., cases, deaths, hospitalizations).
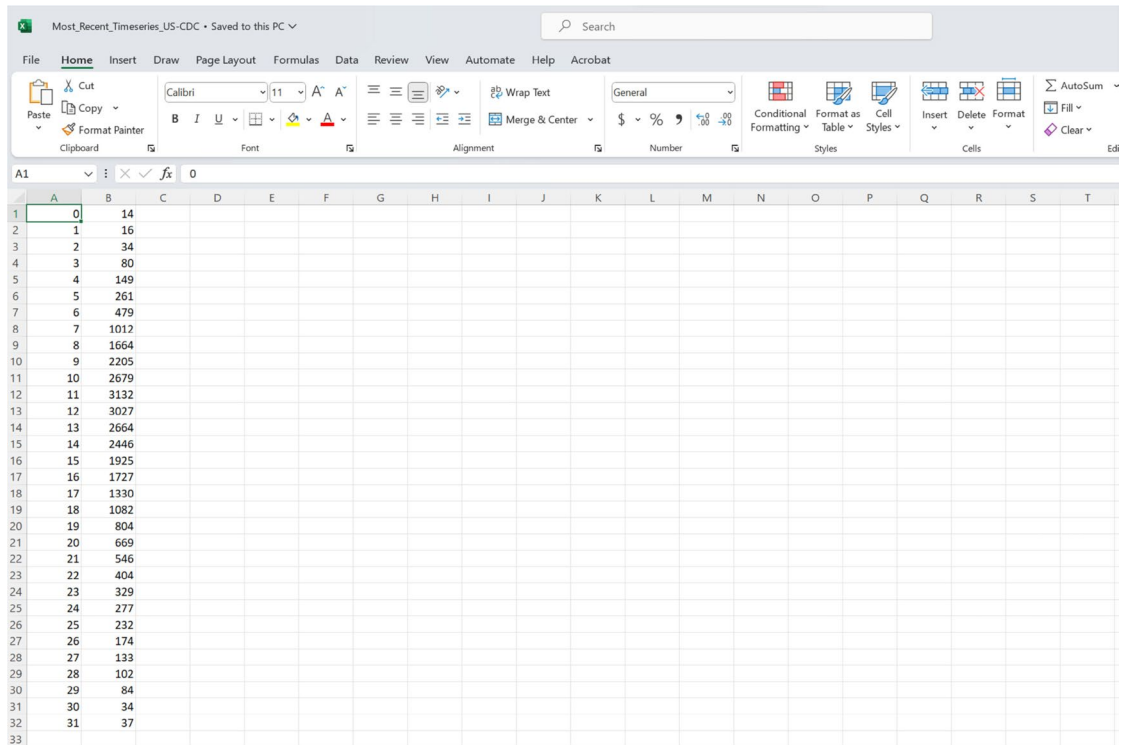
Figure 4. A screenshot of the weekly incidence curve of monkeypox (mpox) cases reported in the USA published by the Centers for Disease Control and Prevention (CDC) for the week of 12 May 2022 through the week of 15 December 2022[51]. The file is in the input folder within the working directory (data file path:./input/ Most_Recent_Timeseries_US-CDC.txt). The first column of the file corresponds to the time index, and the second column corresponds to the weekly incidence curve of monkeypox cases.

```
% <================================================================>
% <====================== Datasets properties ====================>
% <================================================================>
% Located in the input folder, the time series data file is a text file
with extension *.txt.
% The time series data file contains the incidence curve of the epidemic
of interest.
% The first column corresponds to time index: 0,1,2, ... and the second
% column corresponds to the temporal incidence data. If the time series
file contains cumulative incidence count data,
% the name of the time series data file must start with "cumulative".

cadfilename1='Most_Recent_Timeseries_US-CDC'; % Name of the data file
containing the incidence curve

caddisease='monkeypox'; % string indicating the name of the disease
related to the time series data

datatype='cases'; % string indicating the nature of the data (cases,
deaths, hospitalizations, etc)
```

Fitting the models to data with quantified uncertainty

The function `Run_Fit_GrowthModels.m` can be used to fit one of the phenomenological growth models to data with quantified uncertainty. The function uses the input parameters specified by the user in the
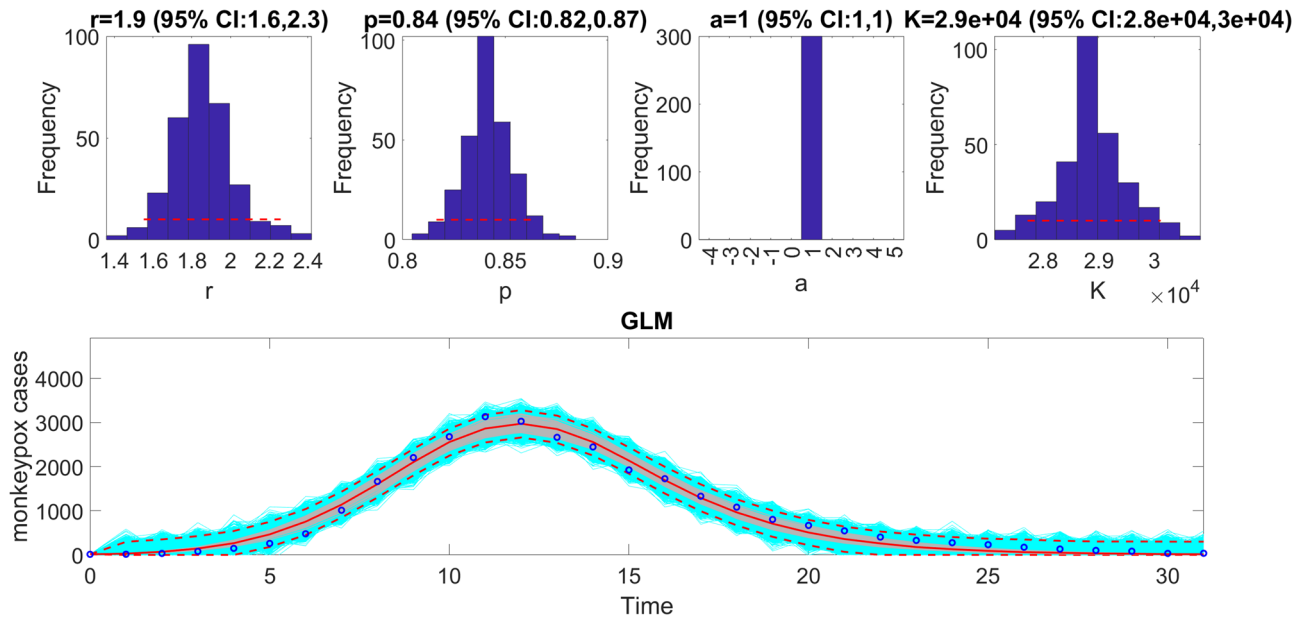
Figure 5. Fitting the generalized logistic-growth model (GLM) to the entire incidence curve of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 15 December 2022. The model provides a good fit to the entire incidence curve. The model supports sub-exponential growth dynamics (i.e., $p \sim 0.8$–0.9), and has a growth rate (r) estimated between 1.6 and 2.3. The epidemic size parameter, K, was estimated at $\sim 28{,}000$–30,000 cases, and the scaling parameter, a, is not estimated when employing the GLM. The horizontal dashed lines in the top panels show the range of the 95% CIs of the parameter estimates. In the bottom panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit.

options_fit.m file. However, the function may receive the parameters related to the rolling window analysis (`<tstart1>`, `<tend1>`, and `<windowsize1>`) as passing input parameters with the remaining inputs accessed from the `options_fit.m` file.

For example, we can fit the generalized logistic growth model (`<flag1>=1` in `options_fit.m` file) to the weekly incidence curve of monkeypox in the USA pre-loaded in the input folder (data file path: `./input/Most_Recent_Timeseries_US-CDC.txt`). We assume a normal error structure (i.e., `<dist1>=0` in the `options_fit.m` file) and examine the fit of the model to the data. Since the monkeypox epidemic curve comprises 32 weeks of data, we can pass the rolling window parameters to the function call in MATLAB as follows:

>>**Run_Fit_GrowthModels**(1,1,32).

In the above call to the function, `<tstart1>=1`, `<tend1>=1`, and `<windowsize>=32`. Hence, this function will generate a single model fit and store several output MATLAB files related to the model fit, parameter estimates, and the quality of model fit in the output folder. For each model fit, it will also generate a figure with the model fit and the corresponding empirical distributions of the estimated parameters (Fig. 5).

Plot the mean model fits and compute calibration performance metrics

Once the `Run_Fit_GrowthModels.m` has been executed, the user can run the function `plotFit_GrowthModels.m` to display the model fit and the empirical distribution of the parameters. It also saves `.csv` files in the output folder with the model fit, the parameter estimates with 95% CIs, the Monte Carlo standard errors (MCSE) of the parameter estimates, the AIC_c values, the calibration performance metrics, and the estimated doubling times of the incidence trajectory. The call to the function from MATLAB's command line follows:

>>**plotFit_GrowthModels**(1,1,32)

The function uses the same rolling window parameters employed in the `Run_Fit_GrowthModels.m` function call. This function will store the following `.csv` files in the output folder:

(1) The model fit to the data:
`Fit-flag1-1-i-1-fixI0-1-method-0-dist-0-calibrationperiod-32-horizon-0-monkeypox-cases.csv`
(2) Model parameter estimates:
`parameters-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-1-calibrationperiod-32-horizon-0-monkeypox-cases.csv`
(3) Monte Carlo standard errors:

Model	MAE	MSE	Coverage 95% PI	WIS
Generalized logistic growth model	110.34	17,277.54	100.00	63.39
Richards model	69.59	7595.55	100.00	43.17

Table 2. Calibration performance metrics for a 32-week calibration period quantifying how well the fits of the generalized logistic growth model and the Richards model captured the epidemic curve of monkeypox in the USA. The metrics indicate that the Richards model yields a better fit to the data in terms of the MAE, MSE, and WIS while both models achieved a 100% coverage of the 95% prediction interval.

```
MCSES-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-1-calibra-
    tionperiod-32-horizon-0-monkeypox-cases.csv
```
(4)   AIC$_c$ values:
```
AICcs-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-1-calibra-
    tionperiod-32-horizon-0-monkeypox-cases.csv
```
(5)   Calibration performance metrics:
```
parameters-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-
    1-calibrationperiod-32-horizon-0-monkeypox-cases.csv
```
(6)   Doubling times:
```
doublingtimes-flag1-1-tstart-1-fixI0-1-method-0-dist-0-calibrationperiod-
    32-horizon-0-monkeypox-cases.csv
```

For this example, the resulting calibration performance metrics indicate that the 95% prediction intervals include all the data points comprising the epidemic curve of monkeypox (mpox) in the USA (i.e., coverage of the 95% PI is 100%) (Table 2). Moreover, the sequence of doubling times increased from 1.04 (95% CI: 0.94, 1.16) for the 1st doubling, 1.12 (95% CI: 1.05, 1.26) for the 2nd doubling to 3.68 (95% CI: 2.59, 7.12) for the 8th doubling time. The output also shows the probability of observing the ith doubling time, which decreases from about 1.0 during the first eight doublings to 0.023 for the 9th doubling. Hence, it is improbable to observe more than 8 doublings.

This function also plots the temporal sequence of parameter estimates and their uncertainty obtained from the rolling-window analysis whenever the value of the parameter <tend1> is greater than parameter <tstart1>. For instance, after running the function Run_Fit_GrowthModels(1,3,30) to generate a rolling window analysis of model fits to capture the outbreak's trajectory, we employed a window size of 30 with start times at 1 (end time: 30), 2 (end time: 31), and 3 (end time: 32), we can run the function plotFit_GrowthMod-els(1,3,30) from MATLAB's command line to generate the rolling window analysis plot (Fig. 6).

We can also assess the fit of the Richards model to the monkeypox incidence curve (<flag1>=4 in options_fit.m file) and compare the quality of the model fit to that obtained using the generalized logistic growth model using the performance metrics. Figure 7 shows the fit of the Richards model to the epidemic curve and the empirical distribution of the parameters.

The calibration performance metrics of the generalized logistic growth model and the Richards model (Table 2) indicate that the Richards model yields a better fit to the data in terms of the MAE, MSE, and WIS while both models achieved a 100% coverage of the 95% PI.

Plotting and assessing model-based forecasts

To generate a forecast, we can use the function Run_Forecasting_GrowthModels.m. This function uses the input parameters provided by the user in the options_forecast.m file. However, the function can also receive <tstart1>, <tend1>, <windowsize1>, and <forecastingperiod> as passing input parameters with the remaining input parameters accessed from the options_forecast.m file.

For example, we can fit the generalized logistic growth model (<flag1>=1 in options_forecast.m file) to the first 10 weeks of the monkeypox epidemic in the USA assuming a normal error structure (i.e., <dist1>=0 in options_forecast.m file) and generate a 4-week ahead prediction by running the function from MATLAB's command line as follows:

>>**Run_Forecasting_GrowthModels(**1,1,10,4**)**

This function will generate a single model fit and store several output MATLAB files related to the model fit and forecast, parameter estimates, and the calibration and forecasting performance metrics. It will also generate a figure with the model fit and 4-week ahead forecast and the corresponding empirical distributions of the parameters (Fig. 8). Overall, the 4-week ahead forecast shown in Fig. 8 underpredicted the trajectory of the epidemic. Figure 1S (supplementary file 1) illustrates the results of a timing study related to the employed modeling methodology utilized in Fig. 8. For comparison, Fig. 9 also shows the model fit and 4-week ahead forecast when the model is calibrated using the first 12 weeks of the epidemic curve instead of the first 10 weeks. The 4-week ahead forecast calibrated on the first 12 weeks of data forecasted better the epidemic curve than the one calibrated on the first 10 weeks of data.

Once the user has executed the function Run_Forecasting_GrowthModels, the function plotForecast_GrowthModels can be used to plot the model-based forecast and the performance metrics of the forecast (MSE, MAE, 95% PI, WIS) based on the inputs indicated in the options_forecast.m file. However, this function can also receive <tstart1>, <tend1>, <windowsize1>, and <forecastingperiod>
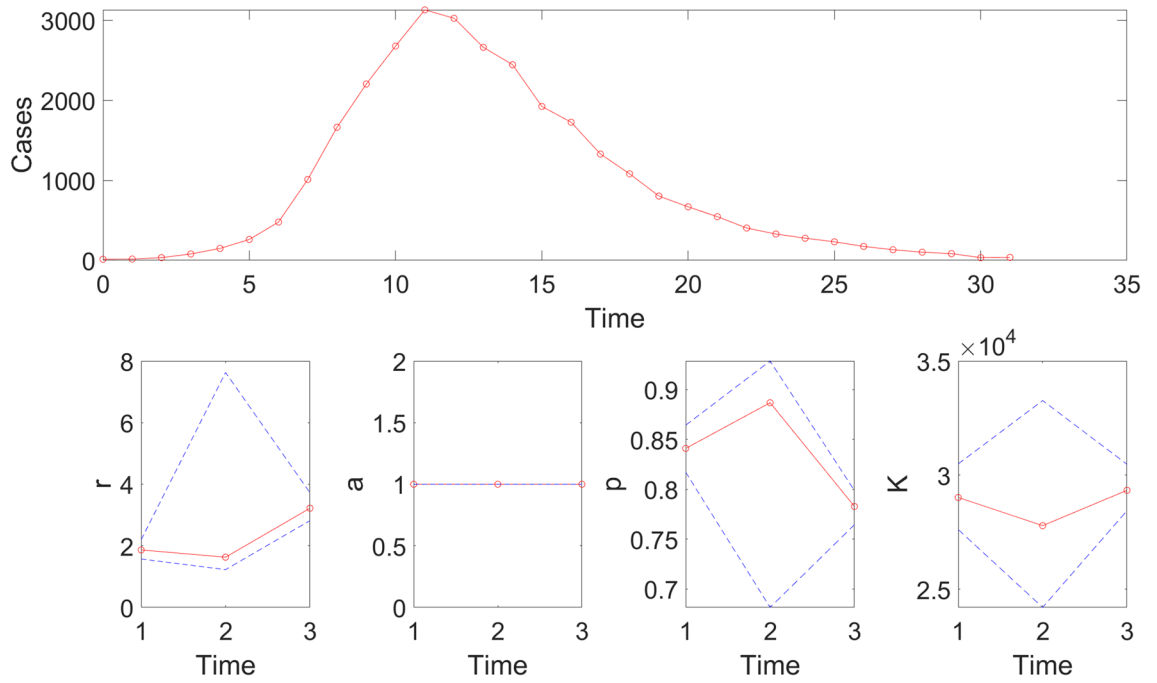
Figure 6. Results of the rolling window analysis after running the function `plotFit_GrowthModels(1,3,30)`. The top panel shows the monkeypox epidemic curve in the USA for reference. The bottom panels show the temporal sequence of parameter estimates (-o-, red line), and their 95% CIs (blue dashed lines) for three different moving time windows (1–30, 2–31, and 3–32).
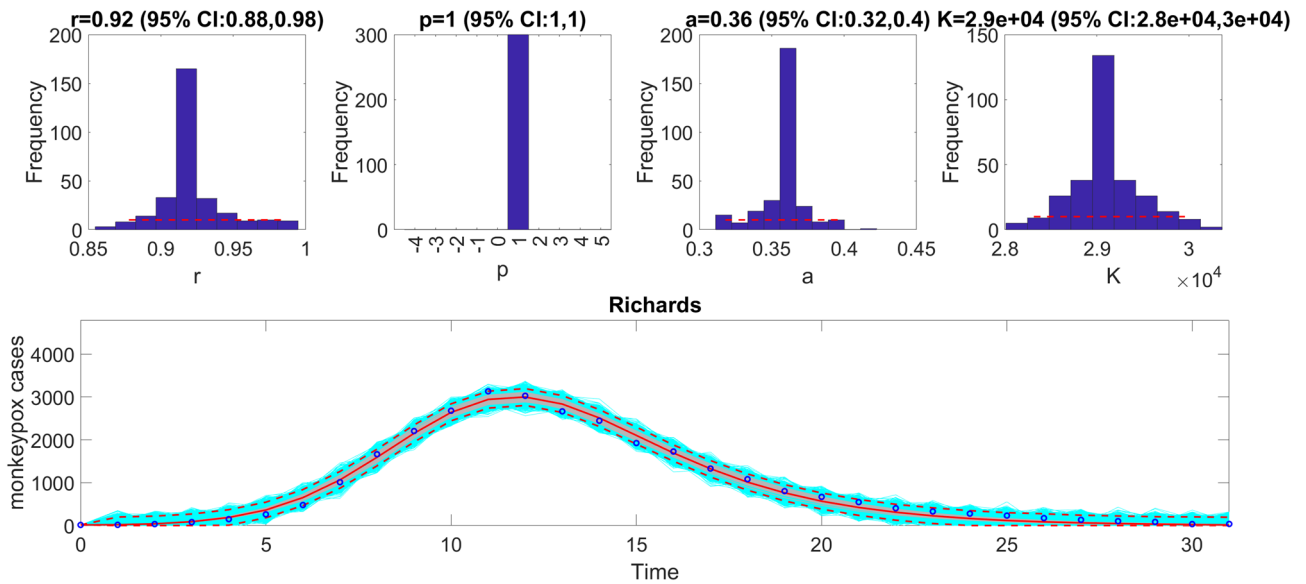

Figure 7. The fit of the Richards model to the entire incidence curve of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 15 December 2022. The model provides a good fit to the entire incidence curve. The epidemic size parameter, K, was estimated at ∼28,000–30,000 cases, similar to that estimated using the generalized logistic-growth model. The growth rate, r, was estimated to be between 0.88 and 0.98, and the scaling parameter, a, between 0.32 and 0.4. The growth scaling parameter, p, is not estimated for the Richards model. The horizontal dashed lines in the top panels show the range of the 95% CIs of the parameter estimates. In the bottom panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit.
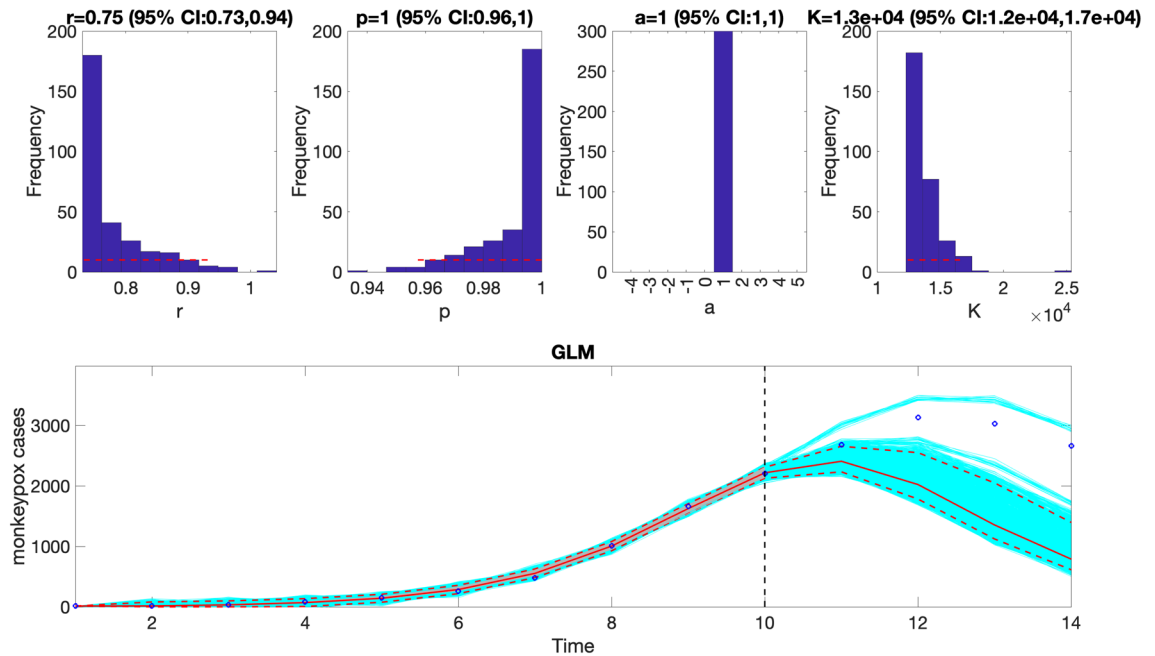
Figure 8. The generalized logistic-growth model (GLM) fit, and 4-week ahead forecast based on the first 10 weeks of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 14 July 2022. The model fit is consistent with early exponential growth dynamics (i.e., $p \sim 1$), and has an estimated growth rate (r) between 0.73 and 0.94. The epidemic size parameter, K, was estimated to be between 120,000 and 170,000 cases, and the scaling parameter, a, is not estimated when employing a GLM. The horizontal dashed lines in the top panels show the range of the 95% CIs of the parameter estimates. In the bottom panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit. The vertical dashed line separates the 10-week calibration (left) and the 4-week ahead forecasting period (right). Overall, the 4-week ahead forecast underpredicted the trajectory of the epidemic.

as passing input parameters while the remaining inputs are retrieved from the `options_forecast.m` file. Moreover, the data associated with the forecasts, the parameter estimates, the Monte Carlo standard errors (MCSE) of the parameter estimates, the AIC_c values, the calibration and forecasting performance metrics, and the doubling times of the entire trajectory, including the forecasting period, are saved as `.csv` files in the output folder. For example, the following line illustrates the execution of the function from MATLAB's command window:

>>**plotForecast_GrowthModels**(1,1,10,4)

This function plots the model fit based on the first 10-weeks of data as the calibration period and produces a 4-week ahead forecast and the empirical distribution of the estimated parameters (Fig. 8). It also displays the associated forecasting performance metrics (Fig. 10).

Similarly, we can generate the forecast using the Richards model by specifying <flag1>=4 in the `options_forecast.m` file and compare the forecasting performance of this model with that obtained employing the generalized logistic growth model using the performance metrics. Figure 11 shows the corresponding forecast of the Richards model based on the first 10 weeks of the epidemic curve and the resulting empirical distribution of the parameters.

The forecasting performance metrics of the generalized logistic growth model and the Richards model based on the first 10 weeks of the epidemic curve (Table 3) indicate that the generalized logistic growth model outperformed the Richards model in terms of forecasting performance.

Plotting the effective reproduction number, R_t

Once the `Run_Fit_GrowthModels.m` has been executed, the user can also run the function `plotFit_ReproductionNumber.m` to display the effective reproduction number, R_t based on the inputs indicated in the `options_fit.m` and `options_Rt.m` files (Fig. 12). It also saves .csv files in the output folder with the effective reproduction number, the model fit, the parameter estimates, including 95% CIs, the Monte Carlo standard errors (MCSE) of the parameter estimates, the AIC_c values, the calibration performance metrics, and the estimated doubling times of the incidence trajectory. The call to the function from MATLAB's command line follows:

>>**plotFit_ReproductionNumber(**1,1,32**)**

The function uses the same rolling window parameters employed in the Run_Fit_GrowthModels.m function call. This function will store the following .csv files in the output folder:
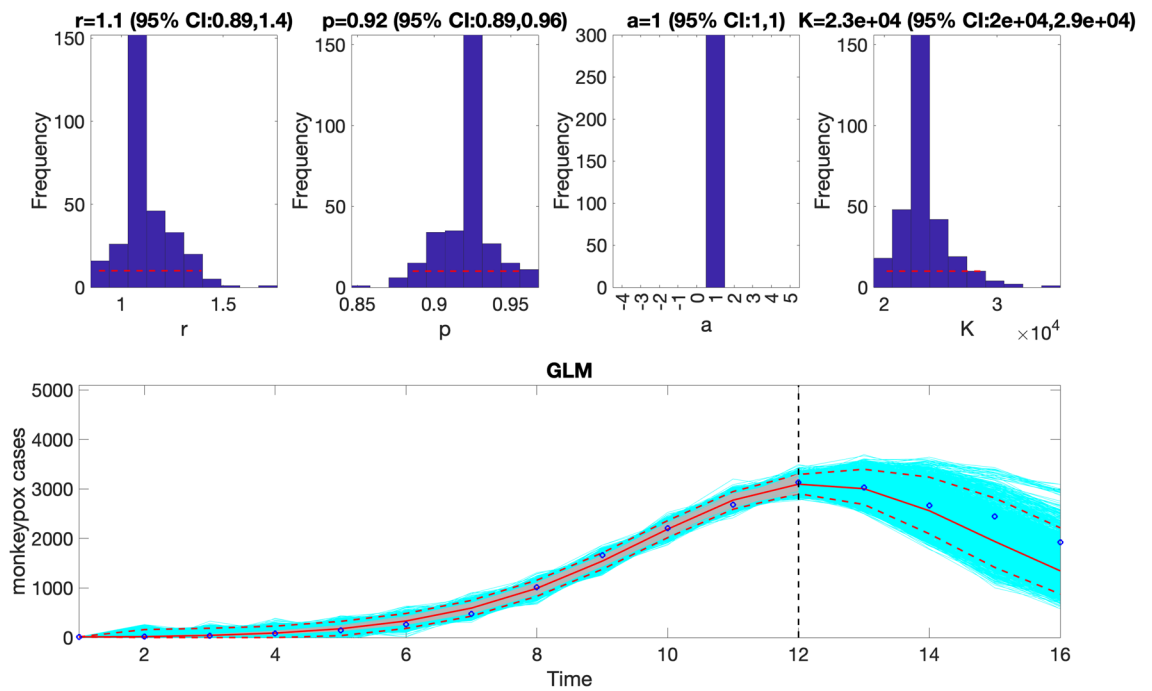
Figure 9. The generalized logistic-growth model (GLM) fit, and 4-week ahead forecast based on the first 12 weeks of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 28 July 2022. The model fit is consistent with early sub-exponential growth dynamics (i.e., $p \sim 0.92$), with a growth rate (r) estimated between 0.89 and 1.4. The epidemic size parameter, K, was estimated to be between 20,000 and 29,000 cases, and the scaling parameter, a, is not estimated when employing a GLM. The horizontal dashed lines in the top panels show the range of the 95% Cis of the parameter estimates. In the bottom panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% prediction intervals. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit. The vertical dashed line separates the 12-week calibration (left) and the 4-week ahead forecasting period (right). Overall, the 4-week ahead forecast using the GLM with a 12-week calibration period tracked the trajectory of the epidemic well.

(1) The model fit to the data:

```
Fit-flag1-1-i-1-fixI0-1-method-0-dist-0-calibrationperiod-32-horizon-0-mon-
    keypox-cases.csv
```

(2) Model parameter estimates:

```
parameters-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-
    1-calibrationperiod-32-horizon-0-monkeypox-cases.csv
```

(3) Monte Carlo standard errors:

```
MCSES-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-1-calibra-
    tionperiod-32-horizon-0-monkeypox-cases.csv
```

(4) AIC$_c$ values:

```
AICcs-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-1-calibra-
    tionperiod-32-horizon-0-monkeypox-cases.csv
```

(5) Calibration performance metrics:

```
parameters-rollingwindow-flag1-1-fixI0-1-method-0-dist-0-tstart-1-tend-
    1-calibrationperiod-32-horizon-0-monkeypox-cases.csv
```

(6) Doubling times:

```
doublingtimes-flag1-1-tstart-1-fixI0-1-method-0-dist-0-calibrationperiod-
    32-horizon-0-monkeypox-cases.csv
```

(7) Effective reproduction number:

```
Rt-flag1-1-tstart-1-fixI0-1-method-0-dist-0-calibrationperiod-32-horizon-
    0-monkeypox-cases.csv
```

Once the user has executed the function `Run_Forecasting_GrowthModels`, the function `plot-Forecast_ReproductionNumber` can be used to plot the effective reproduction number associated with the entire trajectory including the forecast based on the inputs indicated in the `options_forecast.m` and `options_Rt.m` files. Moreover, the data associated with the forecasts, the parameter estimates, the doubling
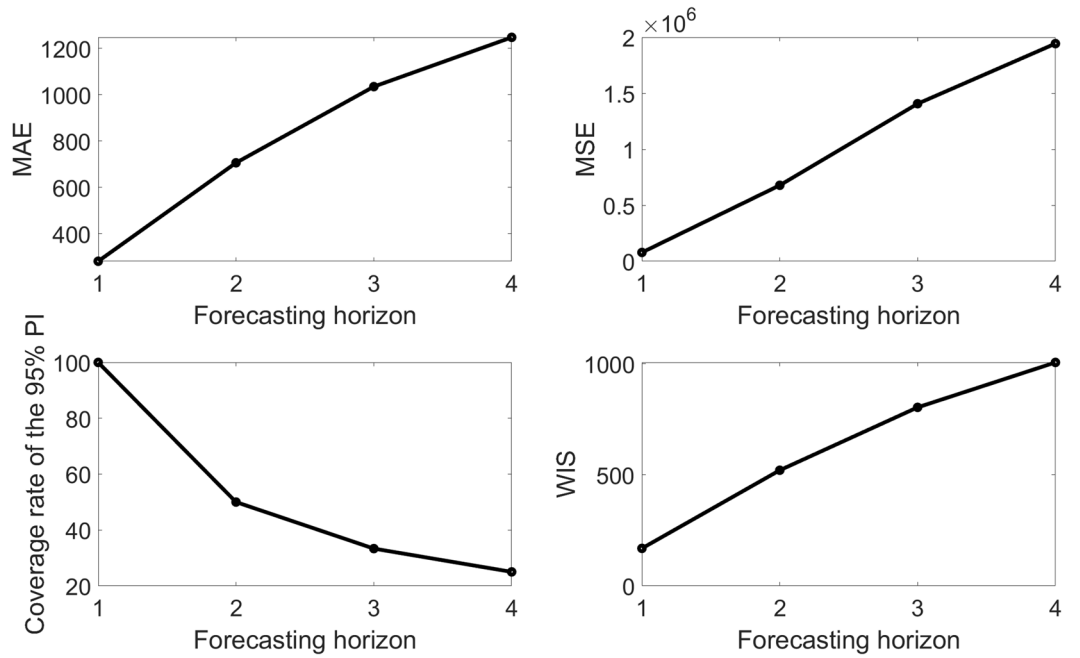
Figure 10. Forecasting performance metrics (MAE, MSE, 95% PI coverage, WIS) associated with the 4-week ahead forecasts obtained from fitting the generalized logistic growth model to the first 10 weeks of the monkeypox epidemic in the USA.
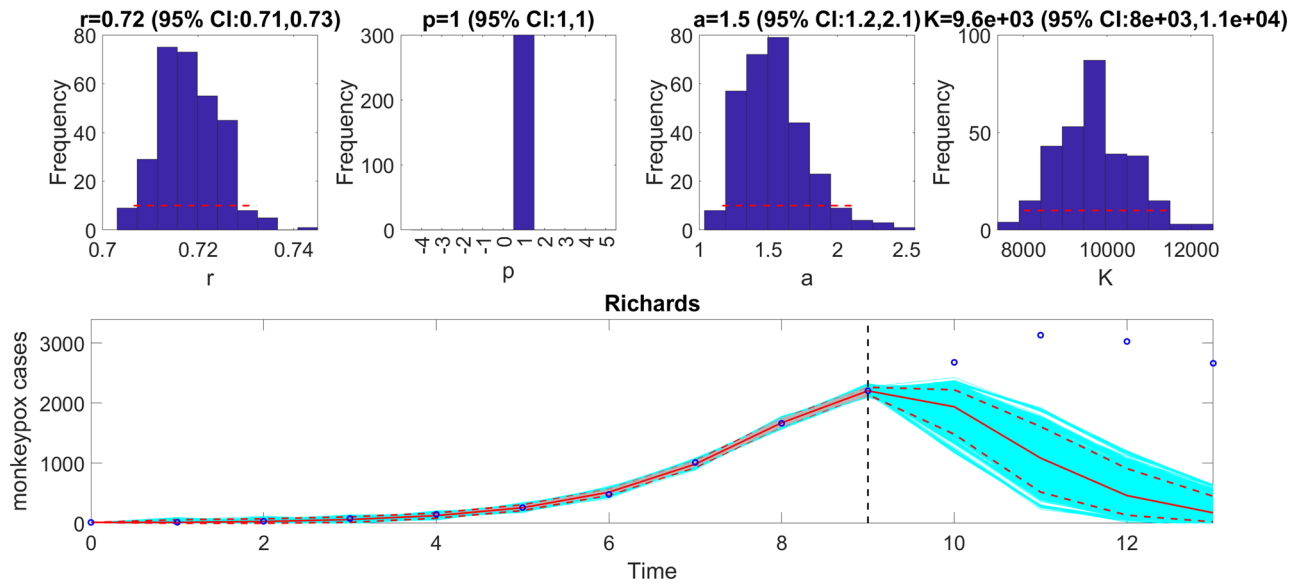

Figure 11. The Richards model fit, and 4-week ahead forecast based on the first 10 weeks of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 14 July 2022. The model fit estimates the epidemic size at $K \sim 8300$–11,000 cases, and the growth rate, r, was estimated to be between 0.71 and 0.73. The scaling parameter, a, fell between 1.2 and 2.1. The growth scaling parameter, p, is not estimated for the Richards model. The horizontal dashed lines in the top panels show the range of the 95% CIs of the parameter estimates. In the bottom panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines (closely aligned with the median model fit) correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit. The vertical dashed line separates the 10-week calibration period (left) and the 4-week ahead forecast period (right). Overall, the 4-week ahead forecast underpredicted the trajectory of the epidemic.

Model	MAE	MSE	Coverage 95% PI	WIS
Calibration performance				
Generalized logistic growth	18.53	813.60	90.00	13.30
Richards	11.97	300.82	100.00	8.65
Forecast performance				
Generalized logistic growth	1240.20	1,925,007.00	25.00	976.22
Richards	1960.40	4,381,564.00	0.00	1812.55

Table 3. Calibration and forecasting performance metrics obtained from the fits of the generalized logistic growth model and the Richards model based on the first 10 weeks of the monkeypox epidemic in the USA forecasting out 4-weeks. The metrics indicate that the Richards model yields a better fit to the data (better calibration performance) than the generalized logistic growth model in terms of the MAE, MSE, and WIS. However, the generalized logistic growth model outperformed the Richards model in terms of forecasting performance.
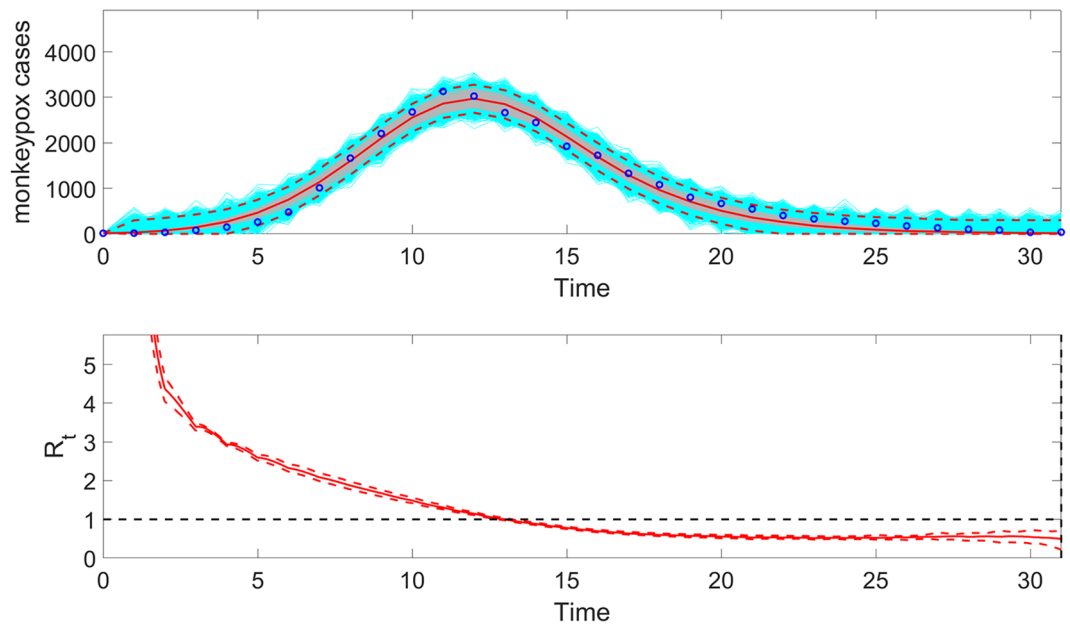
Figure 12. The top panel displays the fit of the generalized logistic-growth model to the entire incidence curve of the monkeypox epidemic in the USA for the week of 12 May 2022 through the week of 15 December 2022. In the top panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit. The bottom panel shows the corresponding effective reproduction number, R_t, assuming a gamma distribution for the generation interval with a gamma distributed generation interval with a mean of 1.78 weeks and variance of 0.65 weeks[250]. The solid red line corresponds to the mean effective reproduction number, while the dashed lines correspond to the 95% confidence bounds around the mean.

times, and the effective reproduction number are saved as `.csv` files in the output folder. For example, the following line illustrates the execution of the function from MATLAB's command window:

```
>>plotForecast_ReproductionNumber(1,1,10,4)
```

This function plots the model fit based on a 10-week calibration period, a 4-week ahead forecast, and the corresponding effective reproduction number (Fig. 13).

Conclusion

In this tutorial-based primer, we have introduced the first toolbox, which will be broadly applicable to fit and forecast time-series trajectories using phenomenological dynamic growth models. In particular, the models included in the toolbox have been frequently applied to characterize and forecast epidemic trajectories in near real-time[10,11,14,15,18,21,28]. The toolbox can be used as part of the curriculum of student training in mathematical biology, applied statistics, infectious disease modeling, and specialty courses in epidemic modeling and
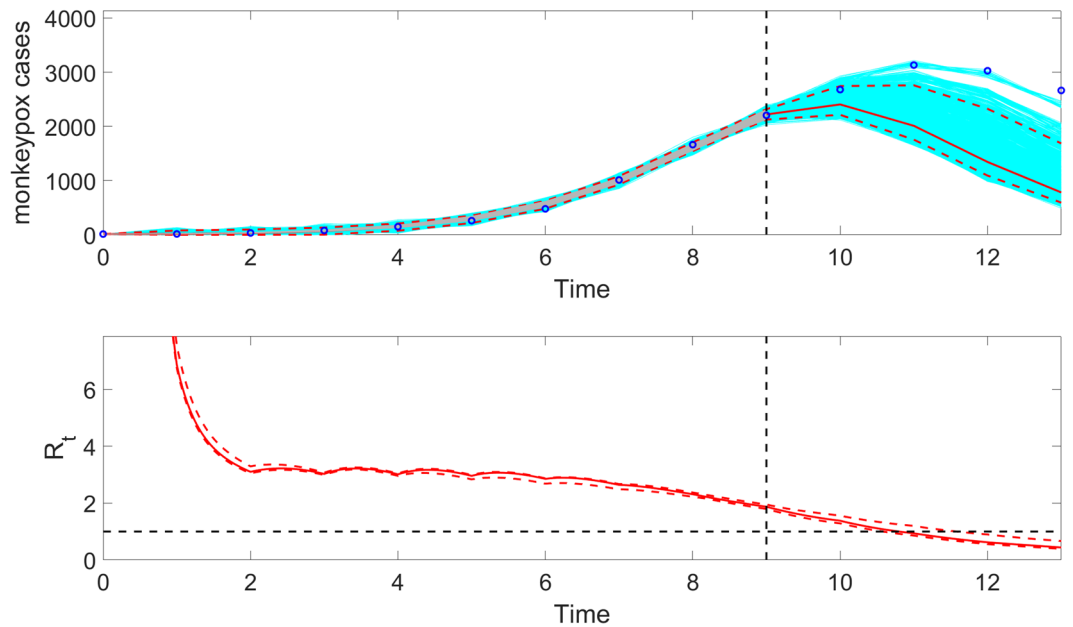
Figure 13. The top panel displays the generalized logistic-growth model fit, and 4-week ahead forecast based on the first 10 weeks of the monkeypox (mpox) epidemic in the USA, the week of 12 May 2022 through the week of 14 July 2022. In the top panel, the solid red line is the median model fit, and the dashed lines correspond to the 95% PIs. The blue dots indicate the observed data points. The gray lines correspond to the mean of the model fits obtained from the parametric bootstrapping with 300 bootstrap realizations, and the cyan lines indicate the predictive uncertainty around the model fit. The vertical dashed line separates the 10-week calibration period (left) and the 4-week ahead forecast period (right). The bottom panel shows the corresponding effective reproduction number, R_t, assuming a gamma distribution for the generation interval with a mean of 1.78 weeks and variance of 0.65 weeks[250]. The solid red line corresponds to the mean effective reproduction number, while the dashed lines correspond to the 95% confidence bounds around the mean.

time-series forecasting. It is also a helpful resource for researchers and policymakers who need to conduct short-term forecasts by relying on historical and real-time trajectory data of a process of interest.

We note some limitations and areas for future work. First, the toolbox is currently intended for users with minimal programming skills. We plan to develop a web interface with intuitive navigation to enhance the tool-box's usability and accessibility, making it more widely accessible to users with varying technical expertise. Similarly, we plan to exploit parallel computing techniques in future software versions to speed up the running time.

Data availability

Project name: Forecasting time series with phenomenological growth models. Project home page: https://github.com/gchowell/forecasting_growthmodels. Operating system(s): Platform independent. Programming language: MATLAB. Other requirements: NA. License: This program is free software: it can be redistributed or modified under the GNU Public License as published by the Free Software Foundation, version 3 of the License. Any restrictions to use by non-academics: None.

References
1. Dimri, T., Ahmad, S. & Sharif, M. Time series analysis of climate variables using seasonal ARIMA approach. *J. Earth Syst. Sci.* **129**, 149 (2020).
2. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* (OTexts, 2018).
3. Mondal, P., Shit, L. & Goswami, S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *Int. J. Sci. Eng. Appl.* **4**, 13 (2014).
4. Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A. & Vahdat, S. F. In *International Conference on Environment and Industrial Innovation, IPCBEE.*
5. Tektaş, M. Weather forecasting using ANFIS and ARIMA models. *Environ. Res. Eng. Manag.* **51**, 5–10 (2010).
6. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398. https://doi.org/10.1016/j.idm.2017.08.001 (2017).
7. Chowell, G. *et al.* Model parameters and outbreak control for SARS. *Emerg. Infect. Dis.* **10**, 1258 (2004).
8. Keeling, M. J. *et al.* Predictions of COVID-19 dynamics in the UK: Short-term forecasting and analysis of potential exit strategies. *PLoS Comput. Biol.* **17**, e1008619 (2021).
9. Viboud, C. *et al.* The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics.* **22**, 13–21 (2018).

10. Chowell, G. *et al.* Using Phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. *PLoS Curr.* https://doi.org/10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583 (2016).
11. Hsieh, Y. H. & Cheng, Y. S. Real-time forecast of multiphase outbreak. *Emerg. Infect. Dis.* **12**, 122–127. https://doi.org/10.3201/eid1201.050396 (2006).
12. Lasky, J. R., Hooten, M. B. & Adler, P. B. What processes must we understand to forecast regional-scale population dynamics?. *Proc. R. Soc. B.* **287**, 20202219 (2020).
13. Lega, J. & Brown, H. E. Data-driven outbreak forecasting with a simple nonlinear growth model. *Epidemics* **17**, 19–26 (2016).
14. Pell, B., Kuang, Y., Viboud, C. & Chowell, G. Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics* **22**, 62–70. https://doi.org/10.1016/j.epidem.2016.11.002 (2018).
15. Roosa, K. *et al.* Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect. Dis. Model.* **5**, 256–263. https://doi.org/10.1016/j.idm.2020.02.002 (2020).
16. Wei, Z.-L., Wang, D.-F., Sun, H.-Y. & Yan, X. Comparison of a physical model and phenomenological model to forecast groundwater levels in a rainfall-induced deep-seated landslide. *J. Hydrol.* **586**, 124894 (2020).
17. Zhao, S., Musa, S. S., Fu, H., He, D. & Qin, J. Simple framework for real-time forecast in a data-limited situation: The Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. *Parasites Vectors.* **12**, 1–13 (2019).
18. Shanafelt, D. W., Jones, G., Lima, M., Perrings, C. & Chowell, G. Forecasting the 2001 foot-and-mouth disease epidemic in the UK. *EcoHealth.* **15**, 338–347 (2018).
19. Chowell, G. & Luo, R. Ensemble bootstrap methodology for forecasting dynamic growth processes using differential equations: Application to epidemic outbreaks. *BMC Med. Res. Methodol.* **21**, 1–18 (2021).
20. Chowell, G., Viboud, C., Simonsen, L., Merler, S. & Vespignani, A. Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: Lessons and the way forward. *BMC Med.* **15**, 1–8 (2017).
21. Viboud, C., Simonsen, L. & Chowell, G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics.* **15**, 27–37. https://doi.org/10.1016/j.epidem.2016.01.002 (2016).
22. Chowell, G., Viboud, C., Simonsen, L. & Moghadas, S. M. Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics. *J. R. Soc. Interface.* **13**, 20160659 (2016).
23. Banks, H. T., Hu, S. & Thompson, W. C. *Modeling and Inverse Problems in the Presence of Uncertainty* (CRC Press, 2014).
24. Myung, I. J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **47**, 90–100 (2003).
25. Roosa, K., Luo, R. & Chowell, G. Comparative assessment of parameter estimation methods in the presence of overdispersion: A simulation study. *Math. Biosci. Eng.* **16**, 4299–4313. https://doi.org/10.3934/mbe.2019214 (2019).
26. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
27. Chowell, G., Viboud, C., Hyman, J. M. & Simonsen, L. The Western Africa Ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. *PLoS Curr.* https://doi.org/10.1371/currents.outbreaks.8b55f4bad99ac5c5db3663e916803261 (2014).
28. Roosa, K. *et al.* Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang China: February 13–23, 2020. *J. Clin. Med.* https://doi.org/10.3390/jcm9020596 (2020).
29. Wang, X. S., Wu, J. & Yang, Y. Richards model revisited: Validation by and application to infection dynamics. *J. Theor. Biol.* **313**, 12–19 (2012).
30. Richards, F. J. A flexible growth function for empirical use. *J. Exp. Bot.* **10**, 290–301. https://doi.org/10.1093/jxb/10.2.290 (1959).
31. Bürger, R., Chowell, G. & Lara-Díiaz, L. Y. Comparative analysis of phenomenological growth models applied to epidemic outbreaks. *Math. Biosci. Eng.* **16**, 4250–4273 (2019).
32. Harvey, A. & Kattuman, P. Time series models based on growth curves with applications to forecasting coronavirus. *Harv. Data Sci. Rev.* (2020).
33. Sugiura, N. Further analysts of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. Theory Methods.* **7**, 13–26 (1978).
34. Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika.* **76**, 297–307 (1989).
35. Gneiting, T. & Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**, 359–378 (2007).
36. Kuhn, M. & Johnson, K. *Applied Predictive Modeling* Vol. 26 (Springer, 2013).
37. M4Competition. *Competitor's Guide: Prizes and Rules.* <https://www.m4.unic.ac.cy/wp-content/uploads/2018/03/M4-Competitors-Guide.pdf> (2018).
38. Tariq, A. *et al.* An investigation of spatial-temporal patterns and predictions of the coronavirus 2019 pandemic in Colombia, 2020–2021. *PLoS Negl. Trop. Dis.* **16**, e0010228. https://doi.org/10.1371/journal.pntd.0010228 (2022).
39. Bracher, J., Ray, E. L., Gneiting, T. & Reich, N. G. Evaluating epidemic forecasts in an interval format. *PLoS Comput. Biol.* **17**, e1008618. https://doi.org/10.1371/journal.pcbi.1008618 (2021).
40. Hwang, E. Prediction intervals of the COVID-19 cases by HAR models with growth rates and vaccination rates in top eight affected countries: Bootstrap improvement. *Chaos Solitons Fractals.* **155**, 111789–111789. https://doi.org/10.1016/j.chaos.2021.111789 (2022).
41. Roosa, K., Tariq, A., Yan, P., Hyman, J. M. & Chowell, G. Multi-model forecasts of the ongoing Ebola epidemic in the Democratic Republic of Congo, March 2013-October 2019. *J. R. Soc. Interface.* **17**, 20200447. https://doi.org/10.1098/rsif.2020.0447 (2020).
42. Dahal, S., Luo, R., Subedi, R. K., Dhimal, M. & Chowell, G. Transmission dynamics and short-term forecasts of COVID-19: Nepal 2020/2021. *Epidemiologia.* **2**, 639–659 (2021).
43. Cramer, E. Y. *et al.* Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2113561119. https://doi.org/10.1073/pnas.2113561119 (2022).
44. Smirnova, A., DeCamp, L. & Chowell, G. Mathematical and statistical analysis of doubling times to investigate the early spread of epidemics: Application to the COVID-19 pandemic. *Mathematics.* **9**, 625 (2021).
45. Muniz-Rodriguez, K. *et al.* Doubling time of the COVID-19 epidemic by province, China. *Emerg. Infect. Dis.* **26**, 1912 (2020).
46. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).
47. Anderson, R. M. & May, R. M. Directly transmitted infections diseases: Control by vaccination. *Science.* **215**, 1053–1060 (1982).
48. Chowell, G., Viboud, C., Simonsen, L. & Moghadas, S. M. Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics. *J. R. Soc. Interface* **13**(123), pii: 20160659 (2016).
49. Nishiura, H. & Chowell, G. In *Mathematical and Statistical Estimation Approaches in Epidemiology* (eds G. Chowell, J. M. Hyman, L.M.A. Bettencourt, & C. Castillo-Chavez) 103–121. (Springer, 2009).
50. Guzzetta, G. *et al.* Early estimates of monkeypox incubation period, generation time, and reproduction number, Italy, May–June 2022. *Emerg. Infectious Dis.* **28**(10), 2078 (2022).
51. Centers for Disease Control and Prevention. U.S. Monkeypox Case Trends Reported to CDC. https://www.cdc.gov/poxvirus/mpox/response/2022/mpx-trends.html (2022).

Acknowledgements

Author contributions

G.C. conceived and developed the first version of the toolbox and wrote the first draft of the tutorial; G.C., A.B, S.D., A.T., K.R., J.M., R.L contributed to analysis and writing subsequent drafts of the tutorial. A.B. produced the tutorial video.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-51852-8.

Correspondence and requests for materials should be addressed to G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.