Rational evaluation of various epidemic models based on the COVID-19 data of China

Wuyue Yang [b,1], Dongyan Zhang [c,1], Liangrong Peng [d], Changjing Zhuge [c,*], Liu Hong [a,*]

[a] *School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, PR China*
[b] *Yau Mathematical Sciences Center, Tsinghua University, Beijing, 100084, PR China*
[c] *Beijing Institute for Scientific and Engineering Computing, Faculty of Science, Beijing University of Technology, Beijing 100124, PR China*
[d] *College of Mathematics and Data Science, Minjiang University, Fuzhou, 350108, PR China*

ARTICLE INFO

ABSTRACT

In this paper, based on the Akaike information criterion, root mean square error and robustness coefficient, a rational evaluation of various epidemic models/methods, including seven empirical functions, four statistical inference methods and five dynamical models, on their forecasting abilities is carried out. With respect to the outbreak data of COVID-19 epidemics in China, we find that before the inflection point, all models fail to make a reliable prediction. The Logistic function consistently underestimates the final epidemic size, while the Gompertz's function makes an overestimation in all cases. Towards statistical inference methods, the methods of sequential Bayesian and time-dependent reproduction number are more accurate at the late stage of an epidemic. And the transition-like behavior of exponential growth method from underestimation to overestimation with respect to the inflection point might be useful for constructing a more reliable forecast. Compared to ODE-based SIR, SEIR and SEIR-AHQ models, the SEIR-QD and SEIR-PO models generally show a better performance on studying the COVID-19 epidemics, whose success we believe could be attributed to a proper trade-off between model complexity and fitting accuracy. Our findings not only are crucial for the forecast of COVID-19 epidemics, but also may apply to other infectious diseases.

1. Background

During the study of epidemics, one of the most significant and challenging problems is to forecast the future trends, like how many individuals might be infected each day, when the epidemics stop spreading, what kinds of policies and actions have to be taken and how they will influence the epidemics, and so forth (Li, 2018; Lutz et al., 2019; Basu and Andrews, 2013; Gingras et al., 2016). The importance of epidemic forecast cannot be emphasized too much.

In the literature, various forecasting models/methods have been reported (Chowell et al., 2016; Walters et al., 2018; Funk et al., 2019; Stocks et al., 2018a; Roosa and Chowell, 2019). Among them, empirical functions, statistical inference methods and dynamical models (difference equations, differential equations, stochastic equations) are three major routines. Empirical functions, especially those with explicit forms, play an unreplaceable role in this field. They are simple, easily understandable, fast implemented and analyzable. The statistical inference methods are also highly welcomed, especially in the presence of a large amount of first-hand data. The basic goal of most statistical

methods in epidemics is to estimate the basic/effective reproduction number, which serves as a key to evaluate the severe condition of an infectious disease. In dynamical models, the basic/effective reproduction number is transformed into reaction rate coefficients. Based on compartment assumptions on populations involved in epidemics, classical SI, SIR, SEIR model and many other generalized models are built. They show a great ability to correctly reproduce the basic features of an infectious disease, to uncover the hidden dynamics, like the numbers of exposed cases and asymptomatic carriers which are hard to be learnt from usual epidemiological investigation, to forecast the future trends of epidemics, as well as to evaluate the influence of diverse control policies and actions in quantity.

However, in the face of so many possible choices, which method is the best? Especially for the purpose of a reliable estimation on the epidemic trend in the future? In this paper, based on the COVID-19 data of Shanghai and other six provinces/cities in China during the spring of 2020, we explore this critical issue systematically. The performance of seven widely used empirical functions, four statistical

---

inference methods and five dynamical models reported in the literature are compared in detail. The basic evaluation criteria, models/methods and data are summarized in Section 2. Detailed analyses, comparisons and evaluations among various epidemic models are carried out in Section 3. In Section 4, we forecast the first epidemic wave happening in Austria, Malaysia, Norway and Republic of Korea from March to June in 2020 as a further validation. The last section contains a conclusion and some brief discussions.

2. Methods

2.1. Criteria and quantities for model evaluation

It is far from a trivial problem to evaluate the forecast ability of various functions/models/methods in a rational way (Tabataba et al., 2017; Funk et al., 2019; Roosa and Chowell, 2019). Many competing requirements should be considered at the same time. Here we employ three basic criteria as a general guiding principle, which can be measured through explicitly calculable quantities (see next section), *i.e.*

- **Complexity v.s. Accuracy.** We seek for a well-balance between the model complexity and fitting accuracy. Neither too complicated models with numerous free parameters and unverified mechanisms, nor over-simplified models without sufficient capability to mimic the real situations is welcomed. This issue is also closely related to the over-fitting and under-fitting problems met in numerics.
- **Fitting v.s. Prediction.** It is a very one-sided pursuit of the least fitting errors (measured by the mean square error, root mean square error, correlation coefficients, *etc.*) for a predictive model, though it is often the case in most published works! In fact, there are tremendous evidences showing that the best fitting does not always lead to the best forecast (see Figs. S2 and S3 in SI for example). Just as an old Chinese proverb says, going too far is as bad as falling short. So we need to make a trade-off between the short-term best fittings and long-term promising predictions. A practical choice would be the statistical average of all possible results based on their weights (*e.g.* the Boltzmann factor).
- **Robustness v.s. Sensitivity.** On one hand, we hope our model is sensitive to parameter changes in order to model the influence of different situations and strategies, *etc.* On the other hand, the model is expected to be robust (insensitive in other words) against perturbations arising from various sources, such as numerical errors, data noise, incomplete knowledge about epidemic mechanisms, *etc.* Obviously, these two opposite pursuits cannot be satisfied at the same time. Therefore, we turn to the reproducibility of key dynamical features (like the inflection point, half time) and the asymptotic stability (basic/effective reproduction number) of the model instead.

The three criteria above reflect the competition and compromise between model complexity and simplicity, over-fitting and under-fitting, short-term and long-term goals, robustness and sensitivity, as well as energetic and entropic, deterministic and statistical, local and global views. The over-emphasis of one aspect would lead to unsatisfactory predictions. As a perfect reflection of the central spirit of Confucianism – Doctrine of the mean, which says that in all activities and thoughts one has to adhere to moderation, our three criteria provide a practical solution to overcome above difficulties both qualitatively and quantitatively. And thus they can be used for making a rational evaluation of different functions/methods/models for epidemic forecast and other related scientific problems.

To make model evaluation quantitatively, more concrete and easily measurable mathematical quantities are needed. Corresponding to each criterion discussed above, we consider the following factors:

(1) **The Akaike information criterion (AIC)** and its various modified versions, like AICc, AICu, QAIC, BIC, *etc.* AIC was introduced

by Japanese statistician Akaike in the early 1970s (Akaike, 1974). It is based on the concept of entropy, and incorporates the model complexity and its goodness of fit together.

$$AIC = 2K - 2\ln(L), \tag{1}$$

in which K is the total number of free parameters in a model, while L is the likelihood function. Models with less free parameters and higher fitting accuracy will have lower AIC values. In this work, we use $AICc = -2\ln(L)/N + (N + K)/(N - K - 2)$ proposed by McQuarrie and Tsai (1998) to remove the dependence on data size N. According to Sugiura (1978), when $K > N/40$, namely when the number of parameters is large in comparison to the number of time points, AICc should be adopted instead of AIC. The Akaike information criterion and its various modified versions have been widely used for model evaluation in the literature, e.g. see Refs. Martcheva (2015), Weston et al. (2020). Especially, based on the early COVID-19 epidemic data of Wuhan, Weston et al. made a preliminary comparison between the SIR and SEIR models by using the AIC value (Weston et al., 2020).

(2) **The root mean square error (RMSE).** The RMSE is extensively used to quantify the accuracy of regression models. It is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2}, \tag{2}$$

in which N denotes the data size, x_i and y_i are the true values and predicted ones separately. In this study, since we are dealing with the epidemic of a large province/city which includes at least millions of populations and hundreds of infected cases, the data (*e.g.* the number of confirmed infected cases) follows a Gaussian distribution according to the central limit theorem, whose variance is generally proportional to the root square of the epidemic size. Therefore, RMSE is a natural quantity to characterize the prediction accuracy of various epidemic models.

(3) **The robustness coefficient (RC).** There are plenty of ways to quantify model robustness. Here we adopt a simple definition based on the confidence interval. By randomly sampling the free parameter space, the best-fit values to the epidemic data and the 95% confidence intervals are determined through Markov Chain Monte Carlo (MCMC) algorithms (Chen et al., 2000). Then the robustness coefficient is defined as the ratio between the smallest and the largest final epidemic size within 95% confidence interval,

$$RC = \frac{the\ smallest\ final\ epidemic\ size\ among\ all\ predictions}{the\ largest\ final\ epidemic\ size\ among\ all\ predictions}. \tag{3}$$

A RC value close to one indicates that the model predictions are consistent and reliable.

The robustness coefficient is closely related to the so-called "nonidentifiability" in the literature (Raue et al., 2009; Lintusaari et al., 2016), which means that a group of model parameters, giving the same good fit to the data but leading to completely different model predictions, cannot be uniquely determined during model calibration. The appearance of nonidentifiability may largely influence the reliability of model predictions and result in a relatively low robustness coefficient.

It should be noted in the current study AICc is calculated based on the training data, while the RMSE is calculated based on the test data, so that they are not synchronous. Furthermore, the robustness coefficient basically depends on the mathematical structure of the model, showing no direct correlation with AICc and RMSE.

2.2. Model specification

Far from complete, in the current study we collect seven empirical functions with explicit forms — linear, exponential, logistic, Hill's, Gompertz's, Richards', and generalized logistic functions (Zhao et al., 2019); four statistical inference methods — exponential growth, maximum likelihood, sequential Bayesian and time-dependent reproduction

number (Obadia et al., 2012; Li et al., 2020); as well as five dynamical models based on ordinary differential equations (ODEs) – SIR, SEIR, SEIR-QD (Peng et al., 2020), SEIR-AHQ (Tang et al., 2020) and SEIR-PO (http://swarma.blog.caixin.com/archives/220791) (see Fig. 1). Most of them have been frequently used in the literature to study the spreading of infectious diseases.

2.2.1. Empirical functions

To describe the growth of cumulative number of infected cases due to an infectious disease, like COVID-19, empirical functions in explicit forms are widely used (Zhao et al., 2019). Here, the linear, logistic, exponential, Hill's, Gompertz's, Richards', and generalized logistic functions are summarized in the upper left corner of Fig. 1.

2.2.2. Statistical methods

Assuming a population to be totally susceptible, the basic reproduction number R_0 is defined as the average number of secondary infectious cases produced by one infectious case during a disease outbreak. The basic reproduction number R_0 plays a key role in studying the epidemics of infectious diseases. And many different statistical methods are designed for estimating R_0 (Li et al., 2020), some of which have been implemented with the "R_0 package" in R (Obadia et al., 2012).

(1) **Exponential growth estimation**.

Exponential growth estimation method assumes that the number of infected cases increases exponentially, which is more suitable in the early stage of an epidemic. In this case, the basic reproduction number (Wallinga and Lipsitch, 2007) is given by

$$R_0 = 1/M(-\gamma) = 1/\int_0^\infty e^{-\gamma\tau}\omega(\tau)d\tau,$$

where γ is the growth rate and M is the moment generating function of the generation time distribution $\omega(\tau)$. The latter is generally assumed to follow the Gamma distribution.

(2) **Maximum likelihood estimation**.

This method assumes the number of infected cases generated from the first case follows the Poisson distribution, whose mean is directly proportional to the basic reproduction number and can be estimated by using the maximum likelihood method (Forsberg White and Pagano, 2008),

$$ll(R_0) = \sum_{i=1}^T \log\left(\frac{e^{-\mu_i}\mu_i^{dI_i}}{dI_i!}\right), \ \mu_i = R_0\sum_{k=1}^i dI_{i-k}\omega_k.$$

Here $ll(R_0)$ is the likelihood depending on R_0. μ_i and $dI_i = I_i - I_{i-1}$ are the number of daily new infected cases and incident cases at discrete time point i, w_i is the generation time distribution. This method also requires the period during which the exponential growth is happening to be identified from the data by statistical tools.

(3) **Sequential Bayesian method**.

The sequential Bayesian method, or real-time Bayesian, starts with a non-informative prior and tries to predict the posterior distribution of the basic reproduction number R_0 by referring to the Bayesian formula (Bettencourt and Ribeiro, 2008),

$$P(R_0|dI_0, \ldots, dI_{i+1}) = \frac{P(dI_{i+1}|R_0, dI_0, \ldots, dI_i) P(R_0|dI_0, \ldots, dI_i)}{P(dI_0, \ldots, dI_i)},$$

where $P(dI_{i+1}|R_0, dI_0, \ldots, dI_i)$ is the likelihood of observing incident cases at time $i+1$ given the value of R_0 and past observations of incident cases from time 0 to i, $P(R_0|dI_0, \ldots, dI_i)$ is a prior distribution of the basic reproduction number, and $P(dI_0, \ldots, dI_i)$ is the joint probability of observing the incident cases. The number of daily new infected cases is also assumed to be Poisson distributed with the mean $\mu_i = dI_{i-1}e^{\gamma(R_0-1)}$.

(4) **Estimation of time dependent reproduction numbers**.

This method computes the basic reproduction numbers by averaging over all transmission networks compatible with observations (Wallinga

and Teunis, 2004). The relative likelihood p_{ij}, that a case onset at time i was infected by a case onset at time j, is given by $p_{ij} = \omega_{i-j}/\sum_{k=0}^{i-1}\omega_{i-k}$. Consequently, the time-dependent effective reproduction number for case j is defined as $R_j = \sum_i p_{ij}$, and the basic reproduction number is the average of all R_j, i.e. $R_0 = \frac{1}{T}\sum_{j=1}^T R_j$.

2.2.3. ODE-based dynamical equations

Without considering time delay and spatial heterogeneity, ordinary differential equations are the most widely used models for describing the spreading process of epidemics. Here we summarize five different dynamical models reported in the literature for studying COVID-19.

(1) **SIR model**

The classical SIR model divides populations into three compartments, that is *susceptible*, *infectious* (with infectious capacity and not yet recovered) and *recovered* cases (recovered and not be either infectious or infected once again) denoted by $S(t)$, $I(t)$ and $R(t)$ separately. As shown in Fig. 1, the coefficients β and δ represent the infection rate and recovery rate separately.

(2) **SEIR model**

To account for the infected cases which are still in a latent period and not yet being infectious, a new *exposed* population $E(t)$ is introduced in SEIR model (Wang et al., 2020), in which a new coefficient γ denotes the transition rate from exposed individuals to the infected.

(3) **SEIR-QD model**

To take the effects of quarantine and self-protection into consideration, Peng et al. (2020) proposed to generalize the classical SEIR model by introducing a new *quarantined* state between *infectious* and *recovery*. The numbers of *death* and *unsusceptible* are denoted as $D(t)$ and $S^A(t)$ separately. In SEIR-QD model, the coefficients $\alpha, \lambda, \delta, \kappa$ denote the protection rate of susceptible individuals, the transition rate from infectious individuals to the quarantined infected class, the recovery rate and death rate, respectively.

(4) **SEIR-AHQ model**

To incorporate appropriate compartments relevant to interventions such as quarantine, isolation and treatment, Tang et al. (2020) generalized the SEIR model. They stratified the populations as *susceptible* (S), *exposed* (E), infectious but not yet symptomatic (*pre-symptomatic*) (A), *infectious* with symptoms (I), *hospitalized* (H) and *recovered* (R) compartments, and further included *quarantined susceptible* (S_q) and *isolated exposed* (E_q) compartments.

In SEIR-AHQ model, the parameters $\{c, \beta, q, \sigma, \lambda, \rho, \delta_I, \delta_q, \gamma_I, \gamma_A, \gamma_H, \alpha\}$ represent the contact rate, probability of transmission per contact, quarantined rate of exposed individuals, transition rate from exposed individuals to the infected, release rate of uninfected contacts from quarantine, probability of having symptoms among infected individuals, transition rate of symptomatic infected individuals to the quarantined infected class, transition rate from quarantined exposed individuals to the quarantined infected class, recovery rates of symptomatic infected individuals, asymptomatic infected individuals and quarantined infected individuals, as well as disease-induced death rate.

(5) **SEIR-PO model**

By incorporating the public opinion on COVID-19, Zhang et al. (http://swarma.blog.caixin.com/archives/220791) further classified the populations of *susceptible* and *exposed* in SEIR model into *unconscious* (S^U, E^U) and *conscious* (S^A, E^A) based on their different knowledge on epidemics and self-protection.

In SEIR-PO model, the parameters $\{\gamma, \delta, \beta, \eta, \eta_1, \eta_2, \eta_3, \alpha\}$ denote the transition rate from exposed individuals to the infected, recovery rate of infected individuals, infection rate of unconscious susceptible population, reduced infection ratio of conscious susceptible individuals, effective infection factors of infectious individuals, unconscious and conscious exposed individuals, as well as the spreading rate of knowledge about COVID-19 among individuals.
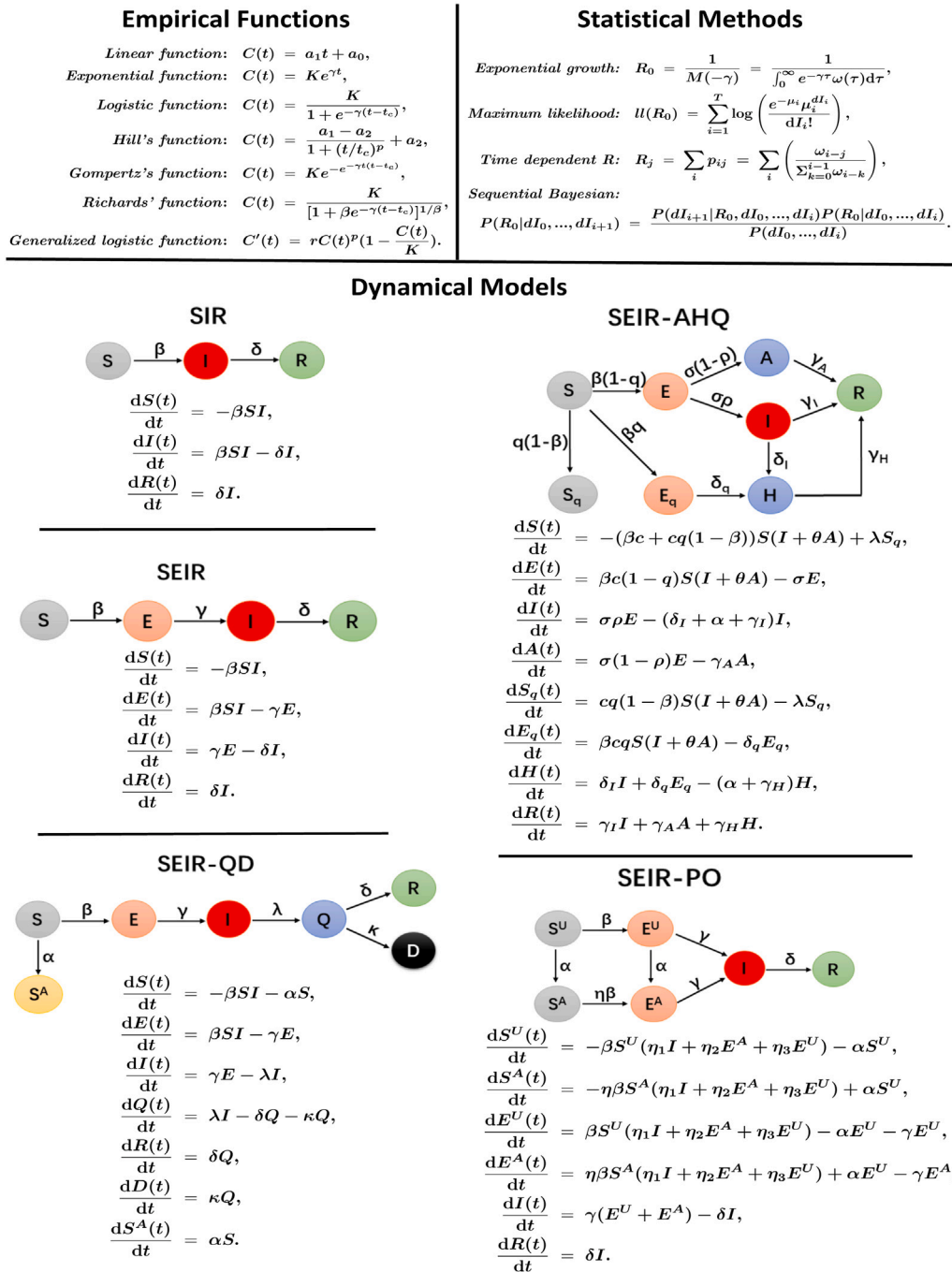
Empirical Functions

$$\text{Linear function:} \quad C(t) = a_1 t + a_0,$$

$$\text{Exponential function:} \quad C(t) = K e^{\gamma t},$$

$$\text{Logistic function:} \quad C(t) = \frac{K}{1 + e^{-\gamma(t - t_c)}},$$

$$\text{Hill's function:} \quad C(t) = \frac{a_1 - a_2}{1 + (t/t_c)^p} + a_2,$$

$$\text{Gompertz's function:} \quad C(t) = K e^{-e^{-\gamma t(t - t_c)}},$$

$$\text{Richards' function:} \quad C(t) = \frac{K}{[1 + \beta e^{-\gamma(t - t_c)}]^{1/\beta}},$$

$$\text{Generalized logistic function:} \quad C'(t) = r C(t)^p \left(1 - \frac{C(t)}{K}\right).$$

Statistical Methods

$$\text{Exponential growth:} \quad R_0 = \frac{1}{M(-\gamma)} = \frac{1}{\int_0^\infty e^{-\gamma \tau} \omega(\tau) d\tau},$$

$$\text{Maximum likelihood:} \quad ll(R_0) = \sum_{i=1}^{T} \log\left(\frac{e^{-\mu_i} \mu_i^{dI_i}}{dI_i!}\right),$$

$$\text{Time dependent } R: \quad R_j = \sum_i p_{ij} = \sum_i \left(\frac{\omega_{i-j}}{\sum_{k=0}^{i-1} \omega_{i-k}}\right),$$

$$\text{Sequential Bayesian:}$$

$$P(R_0 | dI_0, ..., dI_{i+1}) = \frac{P(dI_{i+1} | R_0, dI_0, ..., dI_i) P(R_0 | dI_0, ..., dI_i)}{P(dI_0, ..., dI_i)}.$$

Dynamical Models

SIR

$$\frac{dS(t)}{dt} = -\beta S I,$$

$$\frac{dI(t)}{dt} = \beta S I - \delta I,$$

$$\frac{dR(t)}{dt} = \delta I.$$

SEIR

$$\frac{dS(t)}{dt} = -\beta S I,$$

$$\frac{dE(t)}{dt} = \beta S I - \gamma E,$$

$$\frac{dI(t)}{dt} = \gamma E - \delta I,$$

$$\frac{dR(t)}{dt} = \delta I.$$

SEIR-QD

$$\frac{dS(t)}{dt} = -\beta S I - \alpha S,$$

$$\frac{dE(t)}{dt} = \beta S I - \gamma E,$$

$$\frac{dI(t)}{dt} = \gamma E - \lambda I,$$

$$\frac{dQ(t)}{dt} = \lambda I - \delta Q - \kappa Q,$$

$$\frac{dR(t)}{dt} = \delta Q,$$

$$\frac{dD(t)}{dt} = \kappa Q,$$

$$\frac{dS^A(t)}{dt} = \alpha S.$$

SEIR-AHQ

$$\frac{dS(t)}{dt} = -(\beta c + c q(1 - \beta)) S(I + \theta A) + \lambda S_q,$$

$$\frac{dE(t)}{dt} = \beta c(1 - q) S(I + \theta A) - \sigma E,$$

$$\frac{dI(t)}{dt} = \sigma \rho E - (\delta_I + \alpha + \gamma_I) I,$$

$$\frac{dA(t)}{dt} = \sigma(1 - \rho) E - \gamma_A A,$$

$$\frac{dS_q(t)}{dt} = c q(1 - \beta) S(I + \theta A) - \lambda S_q,$$

$$\frac{dE_q(t)}{dt} = \beta c q S(I + \theta A) - \delta_q E_q,$$

$$\frac{dH(t)}{dt} = \delta_I I + \delta_q E_q - (\alpha + \gamma_H) H,$$

$$\frac{dR(t)}{dt} = \gamma_I I + \gamma_A A + \gamma_H H.$$

SEIR-PO

$$\frac{dS^U(t)}{dt} = -\beta S^U(\eta_1 I + \eta_2 E^A + \eta_3 E^U) - \alpha S^U,$$

$$\frac{dS^A(t)}{dt} = -\eta \beta S^A(\eta_1 I + \eta_2 E^A + \eta_3 E^U) + \alpha S^U,$$

$$\frac{dE^U(t)}{dt} = \beta S^U(\eta_1 I + \eta_2 E^A + \eta_3 E^U) - \alpha E^U - \gamma E^U,$$

$$\frac{dE^A(t)}{dt} = \eta \beta S^A(\eta_1 I + \eta_2 E^A + \eta_3 E^U) + \alpha E^U - \gamma E^A,$$

$$\frac{dI(t)}{dt} = \gamma(E^U + E^A) - \delta I,$$

$$\frac{dR(t)}{dt} = \delta I.$$

Fig. 1. A summary on the empirical functions, statistical inference methods and dynamical models for epidemics evaluated in the current study. See Methods for details.

2.3. Data

To make a quantitative comparison, here we focus on the outbreaks of COVID-19 caused by the novel coronavirus — SARS-CoV-2, which currently spreads severely worldwide. We download the data of daily reported confirmed infected cases $C(t)$ from the China CDC (http://www.chinacdc.cn/). As a first example, the public epidemic data of Shanghai is studied. Shanghai is in the east of China and is considered as one of the best controlled cities in China during the battle against COVID-19. A forty-day period from Jan. 20th, 2020 to Feb. 28th, 2020 is equally divided into four sequential series by every ten days.

Similarly, a larger data set is collected during almost the same time period in six different regions in China selected mainly according to their geographic locations (see Fig. 2), including Heilongjiang province (northeast China), Tianjin (northern), Guangdong province (southern), Chongqing (southwest), Hunan province (central) and Xiaogan city in Hubei province (central, the city with the second largest reported infected populations). Each data set is divided into two instead of four for simplicity.

2.4. Parameter estimation

The fitting procedure of all empirical functions is done through the nonlinear fitting mode in Origin 2019.

As to four statistical methods, since we aim at making predictions on the progression of epidemics, we need to combine them with further assumptions on the dynamics. A widely adopted one is the exponential growth, which assumes the number of infected populations
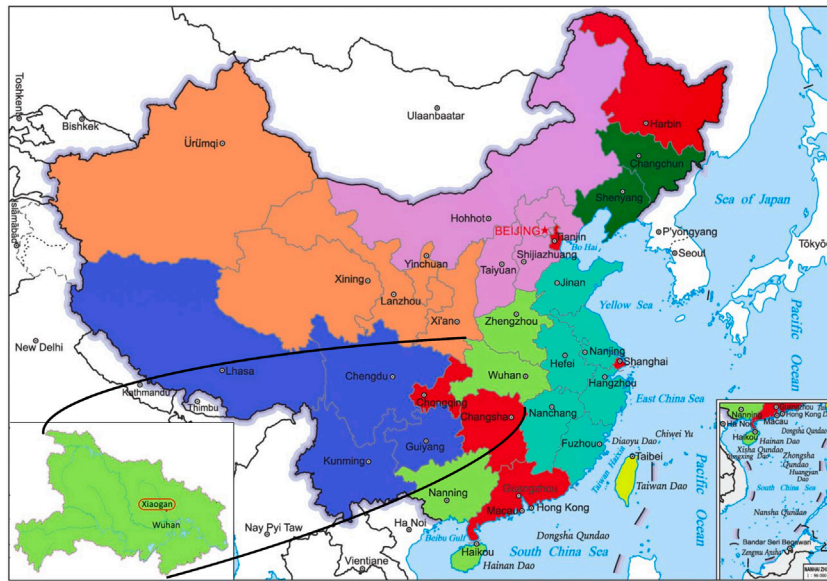
Fig. 2. The geographic locations of seven provinces/cities studied in the current research (colored in red). The maps are depicted based on the standard maps (GS(2019)1647,GS(2019)3333) from Standard Map Service (bzdt.ch.mnr.gov.cn) by Ministry of Natural Resources of the People's Republic of China. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

grows exponentially with the time and the exponent γ is correlated with the basic reproduction number as $R_0 = 1/\int_0^\infty e^{-\gamma\tau}\omega(\tau)d\tau$. In the current study, we assume the generating time distribution $\omega(t)$ obeys the Gamma distribution $\Gamma(k,\theta)$ (see Fig. S1 in SI), whose moment generating function is explicitly known as $M(t) = (1 - t\theta)^{-k}, \forall t < 1/\theta$. From it, we immediately see $\gamma = (R_0^{1/k} - 1)/\theta$. Then inserting γ into either the recurrence formula $\mu_i = dI_{i-1}e^{\gamma(R_0-1)}$ (no free parameter) or the Logistic function (two more free parameters), the progression of epidemics is fitted and predicted.

The unknown parameters involved in dynamical models are estimated respectively by fitting the models to the epidemic data through either the standard nonlinear least-squares approach (Peng et al., 2020) or the Markov-Chain Monte-Carlo (MCMC) algorithms. The MCMC algorithms are widely used in this field to sample the parameter space and to fit the model to the data (Chen et al., 2000; Tang et al., 2020). The MCMC is performed through an adaptive Metropolis–Hastings algorithm, which is implemented in the R package POMP (King et al., 2016). 80,000 iterations with a burn-in of the first 50,000 iterations are carried out, where non-informative uniform distributions are chosen as the prior distributions. From the posterior distributions, we obtain the best-fit values and their 95% confidence intervals.

3. Results

In this part, we apply our three basic criteria to evaluate several widely used models/methods in the field of epidemics. By fitting models to the training data set of Shanghai with varied time spans, their forecasting abilities are compared and quantified through AICc, RMSE and RC values (see Fig. 3 and Table 1). Analogously, the results for other six cities/provinces are summarized in Fig. 4 and Table 2.

Based on extensive numerical explorations of models for COVID-19 with epidemic data of China, our main findings are summarized as follows:

(1) *The model with the least RMSE can be picked out based on AIC and RC.* An astonishing finding of our current study is that the model with the least RMSE to the test data can be easily picked out by examining the values of AICc and RC, while the latter two depend on the model and training data only! As we learned in the previous section, the lower the AICc value is, the better trade-off between model complexity and fitting accuracy is achieved. Meanwhile, a medium

RC value (0.5–0.9 in general) could take the requirements on model robustness and sensitivity into consideration at the same time. Among 27 groups of epidemic data (9 cases times 3 groups of models) under comparison, the AICc value helps to find out 18 models with the least RMSE to the test data, and the rest 9 models all have the second lowest AICc values (see Tables 1 and 2 for details). This finding is consistent with previous reports based on Ebola epidemics that reactive models are of better performance for short-term weekly incidence if they have few parameters. (Viboud et al., 2018)

(2) *Sigmoid functions are more suitable for epidemic forecast.* Linear and exponential functions are not suitable for describing epidemic data in general, while Hill's, Logistic, generalized Logistic, Gompertz's and Richards' functions can well capture the typical S-shaped curve for the cumulative infected cases.

(3) *At the early stage of an epidemic, no model can make long-term reliable forecast.* With respect to very limited data in the early stage of an epidemic, there is no way to tell which model is superior than the rest. They may either overestimate or underestimate the epidemic size in an unpredictable way. Since the model with fewer parameters is more robust, we suggest adopting either the exponential function or even the linear function, though their valid regions are quite narrow. A more elegant way is to combine the knowledge of the basic reproduction number derived from statistical methods and the forecast ability of exponential or logistic functions. However, it should be noted that during the early stage the variance of the derived basic/effective reproduction number is generally very large (see Fig. S1 in SI), making long-term reliable forecast almost impossible.

(4)*The inflection point is crucial for forecast.* The inflection point plays an essential role in forecast. It was suggested by Zhao et al. (2019) in Zika research that, when the epidemic passes the inflection point, predictions on the final epidemic size by the sigmoid empirical functions, such as Logistic, Gompertz's and Richards' functions, will converge to the true values. Here we basically reproduce their results. As shown in the last row of Fig. 3, the RMSE of predictions on the COVID-19 epidemic data of Shanghai decays in an exponential way with respect to the size of training data, meanwhile the RMSE of fitting keeps nearly unchanged. Interestingly, before Jan. 31st which is also the inflection point of Shanghai, all functions fail to make a reliable prediction (and some functions fail even earlier) and their RC values drop to zero rapidly.
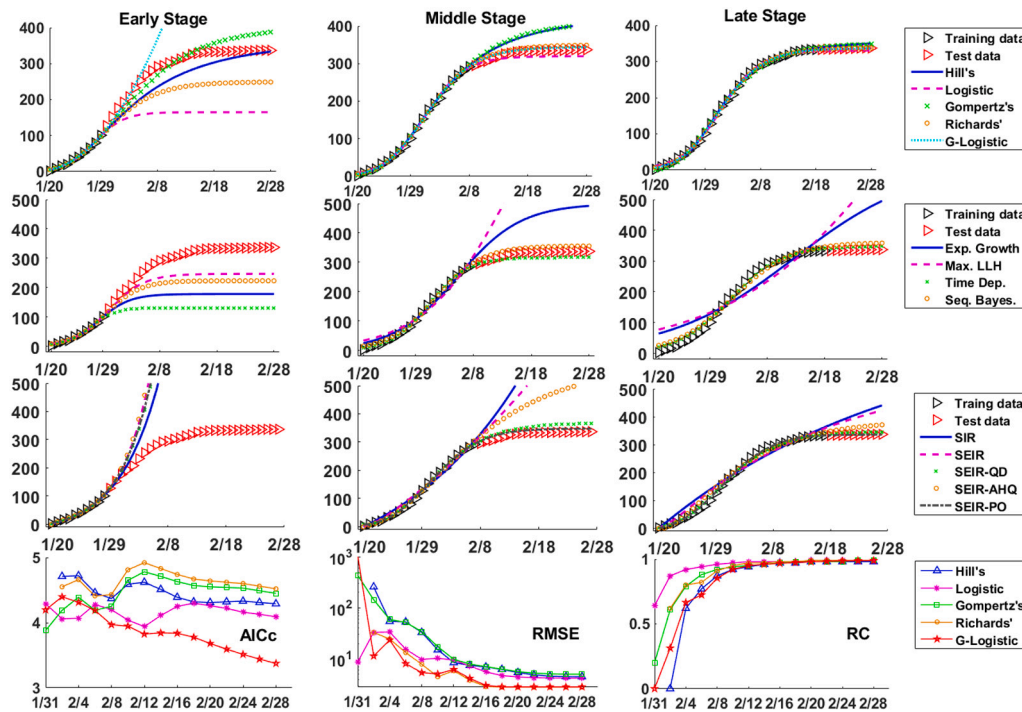
Fig. 3. Forecast of the COVID-19 epidemic in Shanghai from 01/20/2020 to 02/28/2020 based on the data of first 10 (early), 20 (middle) and 30 (late) days respectively. The first three panels give the results of (upper) five explicit functions, (middle) four different statistical inference methods combined with the Logistic function (the exponent γ derived from R_0), (lower) and five ODE models. The one with the smallest RMSE to the training data is drawn. The last row shows the variations of AICc (for training data), RMSE (for test data) and RC for four explicit functions with respect to different sizes of training data set (from Jan. 20th to the date as marked).

Table 1

Summary of AICc (for training data), RMSE (for test data) and RC values for different models calculated based on the epidemic data of Shanghai. Note the negative AICc values result from the fact that data points are fewer than the free model parameters.

Shanghai	Early stage			Middle stage			Late stage		
Model	AICc	RMSE	RC	AICc	RMSE	RC	AICc	RMSE	RC
Hill's	5.2	31	0.47	4.4	33	0.85	4.3	6.2	0.97
Logistic	4.4	120	0.68	4.2	10	0.96	4.3	4.7	0.99
Gompertz's	3.9	25	0.37	4.2	34	0.92	4.6	6.2	0.98
Richards'	4.9	65	0.73	4.0	7.8	0.92	3.7	2.8	0.99
G-Logistic	4.5	448	0.01	4.0	5.4	0.85	3.8	2.8	0.97
Exp.Growth	3.8	112	0.53	6.7	85	0.84	8.6	68	0.88
Max.LLH	4.1	61	0.21	7.5	268	0.11	9.0	101	1.0e−3
Seq.Bayes.	4.0	78	1.4e−4	5.1	13.3	0.39	6.6	16	0.71
Time Dep.	4.2	148	0.48	4.1	11.9	0.72	6.0	11	0.81
SIR	3.5	3.2e3	0.17	6.4	281	0.02	7.6	43	0.04
SEIR	3.5	1.1e4	0.76	6.2	184	0.11	7.2	35	0.60
SEIR-QD	8.9	5.1e3	1.2e−4	4.7	16	0.44	5.1	6.0	0.69
SEIR-AHQ	−3.6	1.0e4	1.7e−5	10	84	2.8e−3	7.9	17	0.15
SEIR-PO	−17.8	7.2e3	6.9e−5	5.8	8.2	0.14	4.9	3.6	0.83

(5) ***The Logistic function underestimates the epidemic size while Gompertz's function overestimates it.*** In all nine cases (including three cases for Shanghai), we notice the Logistic functions always underestimate the total number of infected cases, while the Gompertz's function makes an overestimate (see Fig. 4A). This finding needs further exploring and would be useful for estimating the lower and upper bounds for the real total infected populations, though it still requires further validation. The results of the other three functions are not so consistent and their goodness-of-fit varies from case to case.

(6) ***Methods of sequential Bayesian and time-dependent reproduction number are more accurate at the late stage of an epidemic.*** For statistical methods, since sequential Bayesian and time-dependent reproduction number methods take the non-constant nature of the effective reproduction number with the progression of epidemics into consideration (see Fig. S1 in SI), their predictions appear to be more accurate than the exponential growth and maximum likelihood methods in the late stage (see Fig. 4B). In addition, the sequential Bayesian

method seems to be less robust than the time-dependent reproduction number method. The latter inherits the merit of Logistic function by slightly underestimating the true epidemic size. The nice performance of Bayesian method has been observed for Ebola too, in which ensemble Bayesian method outperformed other 8 methods including Logistic function and SEIR model (Viboud et al., 2018). It is further observed that the basic reproduction number R_0 estimated by the exponential growth method exhibits a transition from overestimation to underestimation with respect to the inflection point, which is in accordance with the S-shaped curve for the total infected populations. As a consequence, with respect to the early stage data of Shanghai COVID-19 epidemic, the exponential growth method combining with the Logistic function makes an underestimation on the final epidemic size, and a contrary overestimation based on accumulated data in the late stage. Finally, we find the maximum likelihood method overestimates the epidemic size to a large extent in all seven cases, indicating this method may not be suitable for studying COVID-19 epidemics.
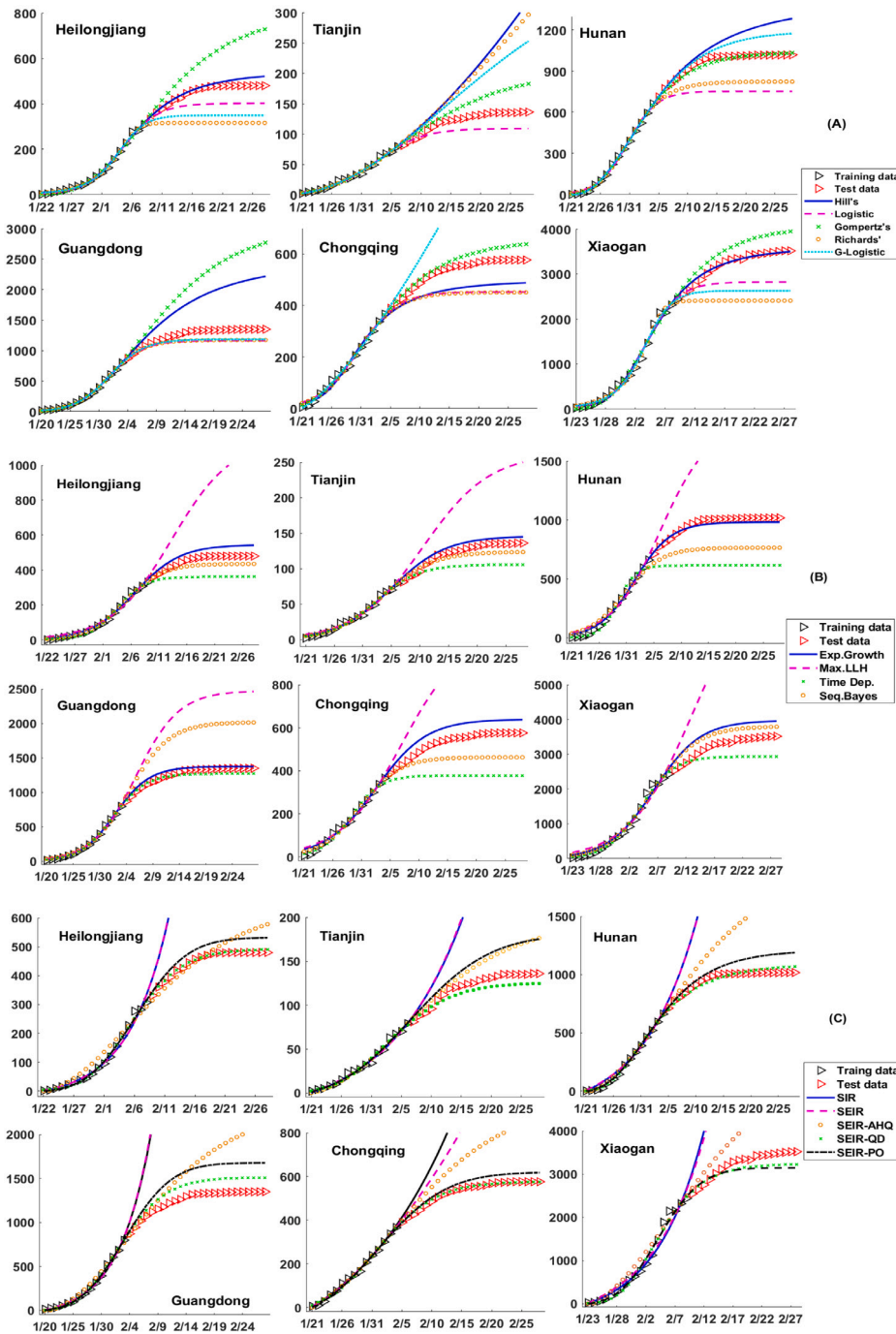
Fig. 4. Forecast of COVID-19 epidemics in Heilongjiang province (data for first 18 days of 38 in total are used for modeling fitting (training set), while the rest 20 data points are used for validation (test set)), Tianjin (17/39), Hunan province (15/39), Guangdong province (15/40), Chongqing (15/39) and Xiaogan city (18/37) in Hubei province (central, city with the second largest reported infected populations). The upper panel (A) shows the results of five empirical functions, the middle one (B) for four different statistical inference methods combined with the Logistic function (the exponent γ derived from R_0), while the lower panel (C) gives the results of five ODE models. The one with the smallest RMSE to the training data is drawn.

(7) ***The SEIR-QD and SEIR-PO models are suitable for modeling COVID-19 epidemics.*** The dynamical models generally require more training data to achieve a reliable forecast than empirical functions, since the former usually involves more free parameters and more complicated mathematical structure. Based on their performance, the dynamical models can be classified into three groups. As shown in Fig. 3 and Fig. 4C, The classical SIR model and SEIR model seem to be inadequate to describe the outbreak of COVID-19, especially the final equilibration phase. Contrarily, the SEIR-AHQ model involves too many free parameters as reflected through the large AICc value. As a

consequence, its robustness is also the poorest among all five models. The SEIR-QD and SEIR-PO models are two suitable ones for modeling COVID-19 by appropriately incorporating the effects of quarantine and self-protection.

It is noted that all of our above statements have to be remained specific to COVID-19 in the countries where data were fitted. Application to other scenarios should be done with great care and requires further exploration.
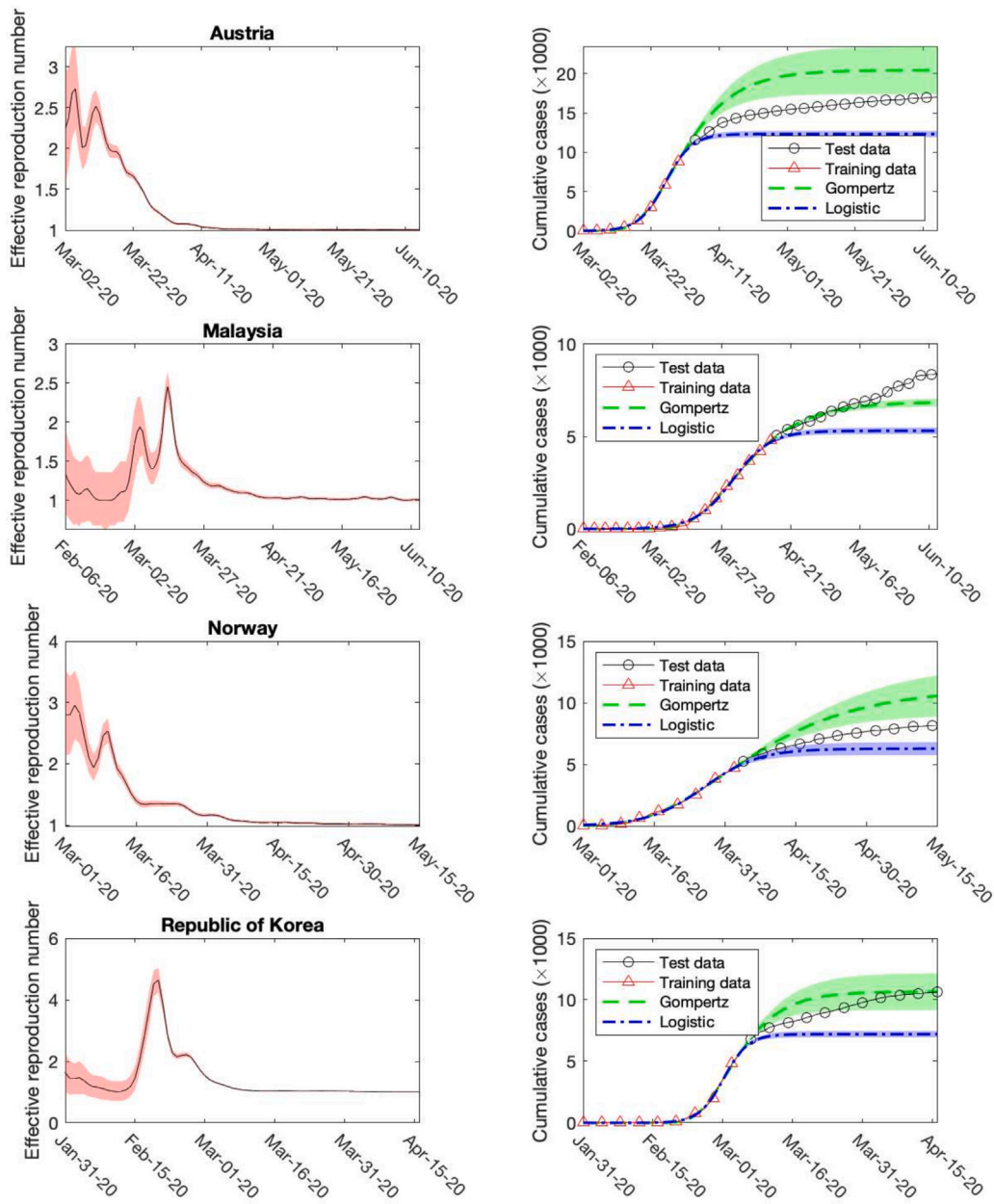
Fig. 5. Forecast of the COVID-19 epidemics in Austria, Malaysia, Norway and Republic of Korea in 2020. Left column shows the effective reproduction number derived from the method of time-dependent reproduction number, while the right column gives the reported confirmed cases and predicted ones by using Gompertz's and Logistic functions.

Table 2
Summary of AICc (for training data), RMSE (for test data) and RC values for different models calculated based on the epidemic data of six provinces/cities in China.

Model	Heilongjiang			Tianjin			Hunan			Guangdong			Chongqing			Xiaogan		
	AICc	RMSE	RC	AICc	RMSE	RC	AICc	RMSE	RC	AICc	RMSE	RC	AICc	RMSE	RC	AICc	RMSE	RC
Hill's	5.9	16	0.62	3.58	69	0.79	6.09	122	0.67	6.76	447	0.42	7.42	61	0.83	10.71	73	0.66
Logistic	5.26	46	0.89	3.48	15	0.78	7.43	178	0.89	6.09	113	0.90	6.74	75	0.82	10.17	363	0.90
Gompertz's	5.87	112	0.53	3.36	20	0.45	6.16	18	0.86	6.55	719	0.58	6.15	29	0.71	10.57	240	0.69
Richards'	6.08	101	0.92	3.61	62	0.78	6.47	132	0.93	6.86	105	0.71	6.47	77	0.88	10.79	623	0.93
G-Logistic	5.23	79	0.78	3.58	47	0.01	6.00	76	0.39	6.40	98	0.57	5.70	314	0.01	10.28	481	0.79
Exp.Growth	6.21	34	0.85	3.55	6.6	0.60	7.92	28	0.89	6.43	41	0.86	6.96	46	0.80	10.96	322	0.95
Max.LLH	7.21	283	0.31	4.04	57	0.11	8.80	475	0.56	7.82	698	0.54	7.43	260	0.40	11.73	3.1e3	0.72
Seq.Bayes.	5.30	27	0.08	3.36	7	0.65	8.87	176	0.57	7.53	434	0.02	6.50	68	0.44	10.86	242	0.58
TimeDep.	5.50	71	0.72	3.29	17	0.05	8.49	279	0.88	6.08	43	0.62	7.00	126	0.82	10.04	296	0.88
SIR	6.75	2.1e3	0.94	3.45	161	0.99	8.13	3.2e3	0.92	7.78	1.9e4	0.93	5.25	666	0.98	11.56	8.3e3	0.99
SEIR	6.98	2.1e3	0.78	3.66	169	0.68	8.55	3.1e3	0.44	8.11	2.1e4	0.30	5.58	390	0.47	11.69	6.8e3	0.65
SEIR-QD	3.13	6.8	0.90	3.15	7.0	0.86	4.27	25	0.77	5.52	107	0.76	3.15	7.0	0.88	5.86	156	0.76
SEIR-AHQ	9.38	38	0.81	7.84	17	0.37	13.80	336	0.66	13.84	351	0.57	12.03	138	0.42	15.07	601	0.38
SEIR-PO	4.56	29	0.78	4.20	19	0.63	5.50	85	0.71	6.32	217	0.60	4.42	24	0.73	6.18	185	0.84

4. Epidemic trends in other countries

Based on our previous evaluations on different epidemic models and methods, we attempt to predict the epidemic trends in several other countries other than China, which also serves as a further validation of our statements.

In Fig. 5, we look into four examples — the first epidemic wave of Austria, Malaysia, Norway and Republic of Korea, which have been randomly selected from their representative categories (four groups of countries clustered by the k-means algorithm according to their diverse control policies, unpublished data). In each case, the effective reproduction number R_t is derived from the method of time-dependent reproduction number. It is well-known that the fact $R_t > 1$ indicates a rapid spreading of the coronavirus. In particular, an astonishing peak of R_t is observed for Republic of Korea on Feb. 19th, which might be attributed to the appearance of super-spreaders. With respect to the total confirmed infected cases reported in these four countries, data after the inflection point are predicted by the Gompertz's and Logistic functions separately, since according to our previous findings the Gompertz's function usually overestimates the final epidemic size while the Logistic function underestimates it. Here only one exception is observed (the test data of Malaysia is larger than the prediction by the Gompertz's function), which is likely caused by the early occurrence of the second epidemic wave. As a conclusion, we can still believe the Gompertz's function and the Logistic function provide reasonable upper and lower bounds for the total confirmed infected case at least for the near future.

5. Conclusion and discussion

In this paper, based on the COVID-19 data of seven provinces/cities in China during the spring of 2020, we make a systematical investigation on the forecast ability of seven widely used empirical functions, four statistical inference methods and five dynamical models reported in the literature. We highlight the significance of a good balance between model complexity and accuracy, over-fitting and under-fitting, as well as model robustness and sensitivity for model performance. Quantitative analyses are made with respect to the Akaike information criterion, root mean square error and robustness coefficient.

Through extensive simulations and detailed comparisons, we find that the inflection point plays a crucial role for making reliable forecasts, in agreement with previous reports (Zhao et al., 2019). The RMSE of model prediction decays exponentially with respect to the size of training data set, while the model robustness characterized through the variance of final epidemic size also approaches to unity rapidly after the inflection point. Furthermore, the forecast abilities of several epidemic models are also closely related to the inflection point. For example, the estimated basic reproduction number R_0 by the exponential growth method exhibits a transition from overestimation to underestimation with the increase of the training data set, and the inflection point acts as the demarcation.

We notice the Logistic functions always underestimate the total number of infected cases, while the Gompertz's function makes an overestimation in all cases we studied. Generalized Logistic, Hill's and Richards' functions do not have such a consistency. Since the sequential Bayesian and time-dependent reproduction number methods take the non-constant nature of the effective reproduction number with the progression of epidemics into consideration, we think they are more accurate than the exponential growth and maximum likelihood methods especially in the late stage of an epidemic. The transition of exponential growth method from underestimation to overestimation with respect to the inflection point could be quite useful for constructing a more reliable forecast. Towards the dynamic models based on ODEs, it is observed that the SEIR-QD and SEIR-PO models generally show a better performance than the other three, highlighting the significance of a trade-off between model complexity and fitting accuracy. The success

of the former two models could also be attributed to the inclusion of self-protection and quarantine during the progression of COVID-19 epidemics.

There are many factors, like changing the reporting rate, increasing the testing capacity, improving the social awareness and self-protection, promoting vaccine injection and herd immunity, may affect the epidemic to a great degree. Generally, these factors are highly time- and policy-dependent, varied from region to region, may or may not be fully considered in various models. In the current study, the influence of these factors has not been thoroughly examined, and we call the readers' attention to this point. Furthermore, besides ODE models, partial differential equations (Martcheva, 2015), stochastic equations (Ma and Jia, 2009) and time-delayed equations (Yue et al., 2020) have been applied to this field too. For example, it has been claimed that "stochastic models should be preferred to deterministic models in most circumstances because they afford improved accounting for real variability and increased opportunity for quantifying uncertainty" (King et al., 2014). How to generalize our current research to these models would be of great value. Interested readers may refer to Konishi and Kitagawa (2008), Stocks et al. (2018b), Gibson et al. (2018) for further details.

CRediT authorship contribution statement

Wuyue Yang: Collected the data, Analyzed the data, Writing – original draft, Writing – review & editing. **Dongyan Zhang:** Collected the data, Analyzed the data, Writing – review & editing. **Liangrong Peng:** Analyzed the data, Writing – original draft, Writing – review & editing. **Changjing Zhuge:** Designed the project, Analyzed the data, Writing – original draft, Writing – review & editing. **Liu Hong:** Designed the project, Analyzed the data, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2021.100501.

References

Akaike, Hirotugu, 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19 (6), 716–723.

Basu, Sanjay, Andrews, Jason, 2013. Complexity in mathematical models of public health policies: A guide for consumers of models. PLoS Med. 10 (10), 1–6.

Bettencourt, Luis M.A., Ribeiro, Ruy M., 2008. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS One 3 (5).

Chen, Ming-Hui, Shao, Qi-Man, Ibrahim, Joseph G., 2000. Monte Carlo Methods in Bayesian Computation. Springer-Verlag.

Chowell, Gerardo, Sattenspiel, Lisa, Bansal, Shweta, Viboud, Cécile, 2016. Mathematical models to characterize early epidemic growth: A review. Phys. Life Rev. 18, 66–97.

Forsberg White, Laura, Pagano, Marcello, 2008. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat. Med. 27 (16), 2999–3016.

Funk, Sebastian, Camacho, Anton, Kucharski, Adam J., Lowe, Rachel, Eggo, Rosalind M., Edmunds, W. John, 2019. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. PLoS Comput. Biol. 15 (2), 1–17.

Gibson, Gavin J., Streftaris, George, Thong, David, 2018. Comparison and assessment of epidemic models. Statist. Sci. 33 (1), 19–33.

Gingras, Guillaume, Guertin, Marie-Hélène, Laprise, Jean-François, Drolet, Mélanie, Brisson, Marc, 2016. Mathematical modeling of the transmission dynamics of clostridium difficile infection and colonization in healthcare settings: A systematic review. PLoS One 11 (9), 1–19.

King, Aaron A., Celles, Matthieu Domenech De, Magpantay, Felicia M.G., Rohani, Pejman, 2014. Avoidable errors in the modeling of outbreaks of emerging pathogens, with special reference to Ebola. Proc. R. Soc. Lond. [Biol.] 282 (1806).

King, A.A., Nguyen, D., Ionides, E.L., 2016. Statistical inference for partially observed Markov processes via the r package pomp. J. Stat. Softw. 069.

Konishi, Sadanori, Kitagawa, Genshiro, 2008. Information Criteria and Statistical Modeling. Springer.

Li, Michael Y., 2018. Important concepts in mathematical modeling of infectious diseases. In: An Introduction To Mathematical Modeling of Infectious Diseases. Springer International Publishing, Cham, pp. 1–33.

Li, Jinghua, Wang, Yijing, Gilmour, Stuart, Wang, Mengying, Yoneoka, Daisuke, Wang, Ying, You, Xinyi, Gu, Jing, Hao, Chun, Peng, Liping, Du, Zhicheng, Xu, Dong Roman, Hao, Yuantao, 2020. Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study. MedRxiv.

Lintusaari, J., Gutmann, M.U., Kaski, S., Corander, J., 2016. On the identifiability of transmission dynamic models for infectious diseases. Genetics 202, 911–918.

Lutz, Chelsea S., Huynh, Mimi P., Schroeder, Monica, Anyatonwu, Sophia, Dahlgren, F. Scott, Danyluk, Gregory, Fernandez, Danielle, Greene, Sharon K., Kipshidze, Nodar, Liu, Leann, Mgbere, Osaro, McHugh, Lisa A., Myers, Jennifer F., Siniscalchi, Alan, Sullivan, Amy D., West, Nicole, Johansson, Michael A., Biggerstaff, Matthew, 2019. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. BMC Publ. Health 19 (1), 1659.

Ma, Zhien, Jia, Li, 2009. Dynamical Modeling and Analysis of Epidemics. World Scientific.

Martcheva, 2015. An Introduction to Mathematical Epidemiology. Springer.

McQuarrie, Allan D.R., Tsai, Chih-Ling, 1998. Regression and Time Series Model Selection. World Scientific.

Obadia, Thomas, Haneef, Romana, Boëlle, Pierre-Yves, 2012. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med. Inf. Decis. Making 12 (1), 147.

Peng, Liangrong, Yang, Wuyue, Zhang, Dongyan, Zhuge, Changjing, Hong, Liu, 2020. Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv preprint arXiv:2002.06563.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25 (15), 1923–1929.

Roosa, Kimberlyn, Chowell, Gerardo, 2019. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theoret. Biol. Med. Model. 16 (1), 1.

Stocks, Theresa, Britton, Tom, Höhle, Michael, 2018a. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. kxy057,

Stocks, Theresa, Britton, Tom, Hohle, Michael, 2018b. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. Biostatistics.

Sugiura, Nariaki, 1978. Further analysis of the data by anaike's information criterion and the finite corrections. Commun. Stat. 7 (1), 13–26.

Tabataba, Farzaneh Sadat, Chakraborty, Prithwish, Ramakrishnan, Naren, Venkatramanan, Srinivasan, Chen, Jiangzhuo, Lewis, Bryan, Marathe, Madhav, 2017. A framework for evaluating epidemic forecasts. BMC Infect. Dis. 17 (1), 345.

Tang, Biao, Wang, Xia, Li, Qian, Bragazzi, Nicola Luigi, Tang, Sanyi, Xiao, Yanni, Wu, Jianhong, 2020. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. J. Clin. Med. 9 (2).

Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., 2018. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics 22, 13–21.

Wallinga, Jacco, Lipsitch, Marc, 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. Lond. [Biol.] 274 (1609), 599–604.

Wallinga, Jacco, Teunis, Peter, 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. 160 (6), 509–516.

Walters, Caroline E., Meslé, Margaux M.I., Hall, Ian M., 2018. Modelling the global spread of diseases: A review of current practice and capability. Epidemics 25, 1–8.

Wang, Huwen, Wang, Zezhou, Dong, Yinqiao, Chang, Ruijie, Xu, Chen, Yu, Xiaoyue, Zhang, Shuxian, Tsamlag, Lhakpa, Shang, Meili, Huang, Jinyan, et al., 2020. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. Cell Discov. 6 (1), 1–8.

Weston, C. Roda, Marie, B. Varughese, Donglin, Han, Michael, Y. Li, 2020. Why is it difficult to accurately predict the COVID-19 epidemics? infectious disease modelling. Infect. Disease Model. 5, 271–281.

Yue, Yan, Yu, Chen, Keji, Liu, Xinyue, Luo, Boxi, Xu, Yu, Jiang, Jin, Cheng, 2020. Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system. Sci. Sin. Math..

Zhao, Shi, Musa, Salihu S., Fu, Hao, He, Daihai, Qin, Jing, 2019. 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), 344.