

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

Wuyue Yang<sup>1a)</sup>, Dongyan Zhang<sup>2a)</sup>, Liangrong Peng<sup>3</sup>, Changjing Zhuge<sup>2b)</sup>, Liu Hong<sup>1b)</sup>

<sup>1</sup>Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing, 100084, P.R.C.

<sup>2</sup>Beijing Institute for Scientific and Engineering Computing, College of Applied Sciences, Beijing University of Technology, Beijing, 100124, P.R.C.

<sup>3</sup>College of Mathematics and Data Science, Minjiang University, Fuzhou, 350108, P.R.C.

---

<sup>a)</sup> Those authors contribute equally to this work.

<sup>b)</sup> Author to whom correspondence should be addressed. Electronic mail: zcamhl@tsinghua.edu.cn(L.Hong), zhuge@bjut.edu.cn(C.Zhuge)

During the study of epidemics, one of the most significant and also challenging problems is to forecast the future trends, on which all follow-up actions of individuals and governments heavily rely. However, to pick out a reliable predictable model/method is far from simple, a rational evaluation of various possible choices is eagerly needed, especially under the severe threat of COVID-19 epidemics which is spreading worldwide right now.

In this paper, based on the public COVID-19 data of seven provinces/cities in China reported during the spring of 2020, we make a systematical investigation on the forecast ability of eight widely used empirical functions, four statistical inference methods and five dynamical models widely used in the literature. We highlight the significance of a well balance between model complexity and accuracy, over-fitting and under-fitting, as well as model robustness and sensitivity. We further introduce the Akaike information criterion, root mean square errors and robustness index to quantify these three golden means and to evaluate various epidemic models/methods.

Through extensive simulations, we find that the inflection point plays a crucial role in the choice of the size of dataset in forecasting. Before the inflection point, no model considered here could make a reliable prediction. We further notice the Logistic function steadily underestimate the final epidemic size, while the Gomertz's function makes an overestimation in all cases. Since the methods of sequential Bayesian and time-dependent reproduction number take the non-constant nature of the effective reproduction number with the progression of epidemics into consideration, we suggest to employ them especially in the late stage of an epidemic. 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. Towards the dynamical models based on ODEs, it is observed that the SEIR-QD and SEIR-PO models generally show a better performance than SIR, SEIR and SEIR-AHQ models on the COVID-19 epidemics, whose success could be attributed to the inclusion of self-protection and quarantine, and a proper trade-off between model complexity and fitting accuracy.

Keywords: COVID-19, Model evaluation, Epidemic size, Forecast, Akaike information criterion

## I. INTRODUCTION

During the study of epidemics, one of the most significant and challenging problem 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 taken and how they influence the epidemics, and so forth<sup>1-4</sup>. The importance of epidemic forecast cannot be emphasized too much. After the outbreak of an epidemic, all actions of individuals and government heavily depend on our understandings on its future trend. The SARS in 2003<sup>5</sup>, H1N1 flu in 2009<sup>6,7</sup> and recent COVID-19 are several well-known examples.

Since the accurate epidemic forecast is so critical, there are diverse methods reported in the literature to try to achieve this goal<sup>8-12</sup>. Among them, empirical functions, methods based on statistical inference and dynamical models (difference equations, ODEs and PDEs) are three major routines (see Fig. 1). Empirical functions, especially those with explicit forms, are most popular ones. They are simple, easily understandable, fast implemented and analyzable. So many merits lay down the unreplaceable role of empirical functions in this field. 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 quantity to evaluate the future trends of an infectious disease. In dynamical models, the basic/effective reproduction number is transformed into reaction coefficients. Based on compartment assumptions on populations involved in epidemics, such as susceptible, exposed, infected, quarantined, hospitalized, recovered and death, classical SI, SIR, SEIR and other generalized models show a great ability to correctly reproduce the basic features of the spreading process of an infectious disease, to reveal the hidden dynamics, like the numbers of exposed cases and asymptomatic carriers which are hard to be learnt from usual epidemiology investigation, to forecast the future trends of epidemics, as well to evaluate the influence of diverse control policies and actions against the spreading of infectious diseases in quantity.

However, in the face of so many possible choices, which one would be our best choice, especially for the purpose of a reliable estimation on the future epidemic trend? In this paper, based on the COVID-19 data of Shanghai and other six provinces/cities in China in the spring of 2020, we explore this critical issue systematically by comparing the performance

of eight widely used empirical functions, four statistical inference methods and five dynamical models reported in the literature. The basic guiding principles and evaluation criteria are summarized in Section II. Detailed analyses, comparisons and evaluations among various epidemic models are carried out in Section III. In Section IV, we forecast the epidemic trends in South Korea, Italy and Iran in the near future. The last section contains a conclusion and some brief discussions.

## II. THREE GOLDEN MEANS FOR MODEL EVALUATION

It is far from a trivial problem to evaluate the forecast ability of various functions/models/method in a rational way<sup>10,12,13</sup>. Many aspects should be considered at the same time. Here, as a very general guiding principle, we propose three golden means which can be measured through explicitly calculable quantities, *i.e.*

- **Complexity v.s. Accuracy.** We seek for a well-balance between the model complexity and fitting accuracy. Neither too complicated models with too many free parameters and unverified mechanisms nor over simplified models without sufficient capability to mimic the real complicated situations and all available data is welcomed. This issue is also closely related to the over-fitting and under-fitting problem met in numerics.
- **Fitting v.s. Prediction.** It is a very one-sided pursuit of the lowest fitting errors (measured by MSE, RMSE, correlation coefficients, *etc.*) for a predictive model, though it is always the case in most published works! In fact, there are tremendous evidences to show that the best fitting does not always lead to the best forecast (see Fig. 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 more reasonable choice in practice would be the statistical average of all possible results according to their weights (*e.g.* the Boltzmann factor).
- **Robustness v.s. Sensitivity.** On one hand, we hope our model be sensitive to parameter changes in order to model the influence of different situations and strategies,

to pick out the best fitting parameters, *etc.* On the other hand, the model is expected to be robust (insensitive in other words) to against perturbations coming from various sources, such as numerical errors, data noise, incomplete knowledge about epidemic mechanisms, *etc.* Obviously, these are two opposite pursuits just like two sides of the same coin and can not be optimized simultaneously. So that we need to relax our requirement on model robustness and sensitivity, and focus on the reproducibility of key dynamical features (like the inflection point, half time, *etc.*), the asymptotic stability (basic/effective reproduction number  $R_0$ ), the bifurcation point, *etc.*

Above three golden means reflect the struggle and compromise between 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. These issues are met here and there, now and then. The overemphasis of one aspect by omitting the other either deliberately or unintentionally would lead to unsatisfactory predictions. As a perfect reflection of the central spirit of Confucianism – Zhongyong, our three golden means provide a practical solution to above difficulties both qualitatively and quantitatively, and can be used for making a rational evaluation of different fuctions/methods/models for epidemic forecast and other related scientific problems.

To quantify these three golden means, here we propose three easily calculable factors for model evaluation.

(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<sup>14</sup>. It is based on the concept of entropy, and incorporate the model complexity and its goodness of fit together in accordance with our first golden mean.

$$AIC = 2k - 2 \ln(L), \quad (1)$$

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 high 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<sup>15</sup> to remove the size dependence on data set  $N$ .

(2) **The robustness index (RB).** There are plenty of ways to quantity model robust-

ness. A common formula is

$$\text{model robustness} = \frac{\text{variation of prediction}}{\text{strength of perturbation}}. \quad (2)$$

By either introducing noises to training data or make perturbations on model parameters and initial values, the variations of model predictions on the basic/effective reproduction, inflection point, half time, the bifurcation point, *etc.* could be measured in quantity. For models involving multiple free parameters, a more practical way to quantify its robustness is to perform Monte Carlo simulations and look at the variance of predictions. Here we simply define RB as the ratio between the 5% quantile and 95% quantile of all simulations.

(3) **The root mean square error (RMSE).** It is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2}, \quad (3)$$

in which  $N$  denotes the data size,  $x_i$  and  $y_i$  are the true values and predicted ones separately. The RMSE is widely used to quantify the accuracy of regression models and needs no more explanation.

### III. MODEL EVALUATION BASED ON DATA FOR COVID-19 EPIDEMICS IN CHINA

In this part, we apply our three golden means to evaluate most widely used models/methods in the field of epidemics. Far from complete, we collect eight empirical function with explicit formulas – linear, quadratic, cubic, exponential, Hill’s, Logistic, Gompertz’s and Richards’ functions<sup>16</sup>, four statistical inference methods – exponential growth, maximum likelihood, sequential Bayesian and time-dependent reproduction number<sup>17,18</sup>, as well as five ODE models – SIR, SEIR, SEIR-QD<sup>19</sup>, SEIR-AHQ<sup>20</sup> and SEIR-PO (see Fig. 1). The unknown parameters involved in different models are estimated respectively by fitting the models to the epidemic data through the standard non-linear least squares approach.

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 collect the data of daily reported confirmed infected cases  $C(t)$  from the China CDC (<http://www.chinacdc.cn/>). The forty-day data from Jan. 20th, 2020 to Feb. 28th, 2020 of Shanghai, which is located in eastern China and is considered as one of the most successful

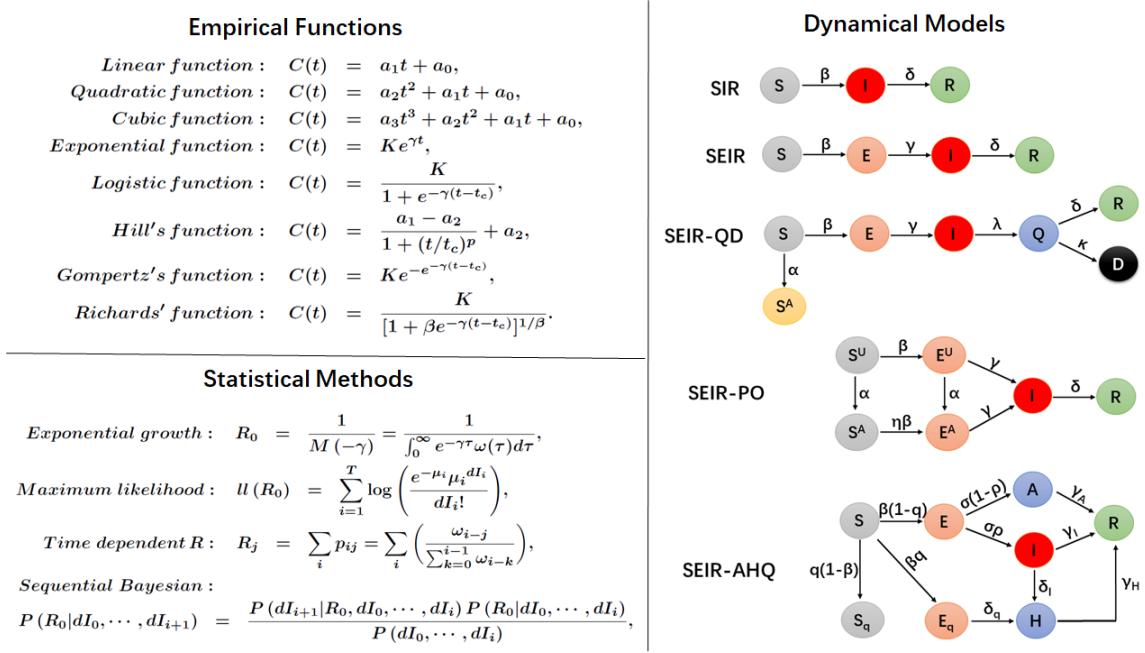


FIG. 1. A summary on the empirical functions, statistical inference methods and dynamical models for epidemics evaluated in the current study. See SI for details.

controlled big cities in China during the battle against COVID-19, are equally divided into four sequential series every ten days. By separately fitting various models to different training data set, their forecasting abilities are compared in quantity through RMSE, AICc and RB values based on the test data set.

Then similar approaches are carried out for the COVID-19 epidemics during almost the same time period in other six different regions in China selected mainly according to their geographic location (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). However, their data set are divided into two instead of four and only one comparison is made for each case for simplicity.

Our main findings are summarized as follows:

(1) ***The best model for epidemic forecast can be picked out based on AIC and RB.*** The best model (judged by RMSE of test data) can be easily picked out from each group based on the values of AICc and RB (depending only on the model and training data!). The lower the AICc value is, the better trade-off between model complexity and

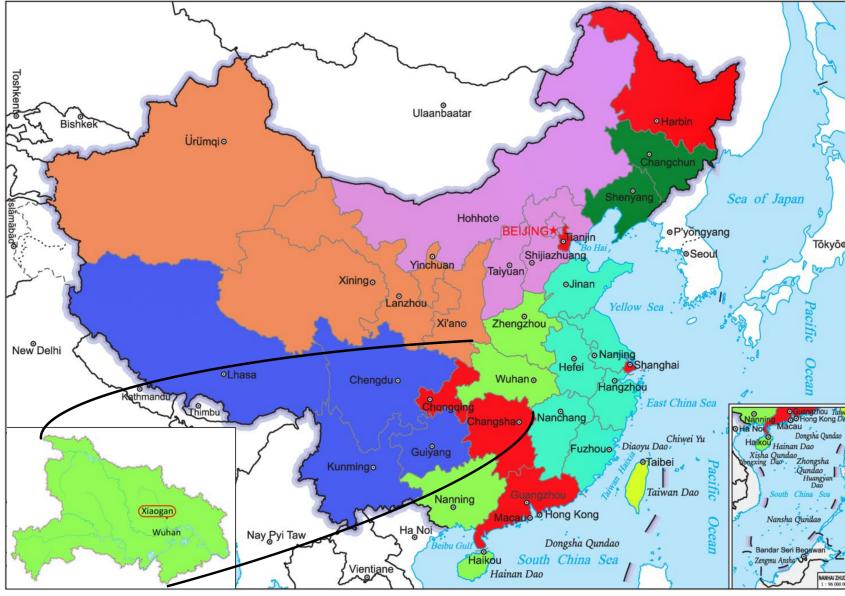


FIG. 2. The COVID-19 epidemics in seven provinces/cities considered in the current study (colored in red) and their geographic locations in China.

fitting accuracy is achieved. Meanwhile, a medium RB 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 (9 cases times 3 groups of models) under comparison, the AICc value helps to find out 18 best models (by RMSE), and the rest 9 best models all have the second lowest AICc values (see Table. I and Table. II for details).

(2) ***Sigmoid functions are more suitable for epidemic forecast.*** Linear, quadratic, cubic and exponential functions are not suitable for describing epidemic data in general (see SI), while Hill's, 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 is reliable.*** In the case of very limited data during the early stage of an epidemic, there is no way to tell which model is superior than the other. They may either overestimate or underestimate the real situation of a epidemic in an unpredictable way. Since the model with fewer parameters are more robust, we suggest to adopt either exponential function or even linear function, though their valid region is quite narrow. A more delicate way is to combine the knowledge of the reproduction number by statistics 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 SI), making a long-term reliable forecast almost impossible.

(4) ***The inflection is crucial for forecast.*** The inflection point plays a role in forecast. It was suggested by Zhao *et al.*<sup>16</sup> in Zika research that in the period when enough data are collected, typically 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 grows 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 fitting (though some functions fail even earlier) and their RB values rapidly drop to zero.

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

TABLE I. Summary of AICc, RB and RMSE values for three stages in Shanghai based on different models. Note the negative AICc values are caused by fewer data points than free model parameters.

(5) ***The observation that Logistic function underestimates the epidemic size***

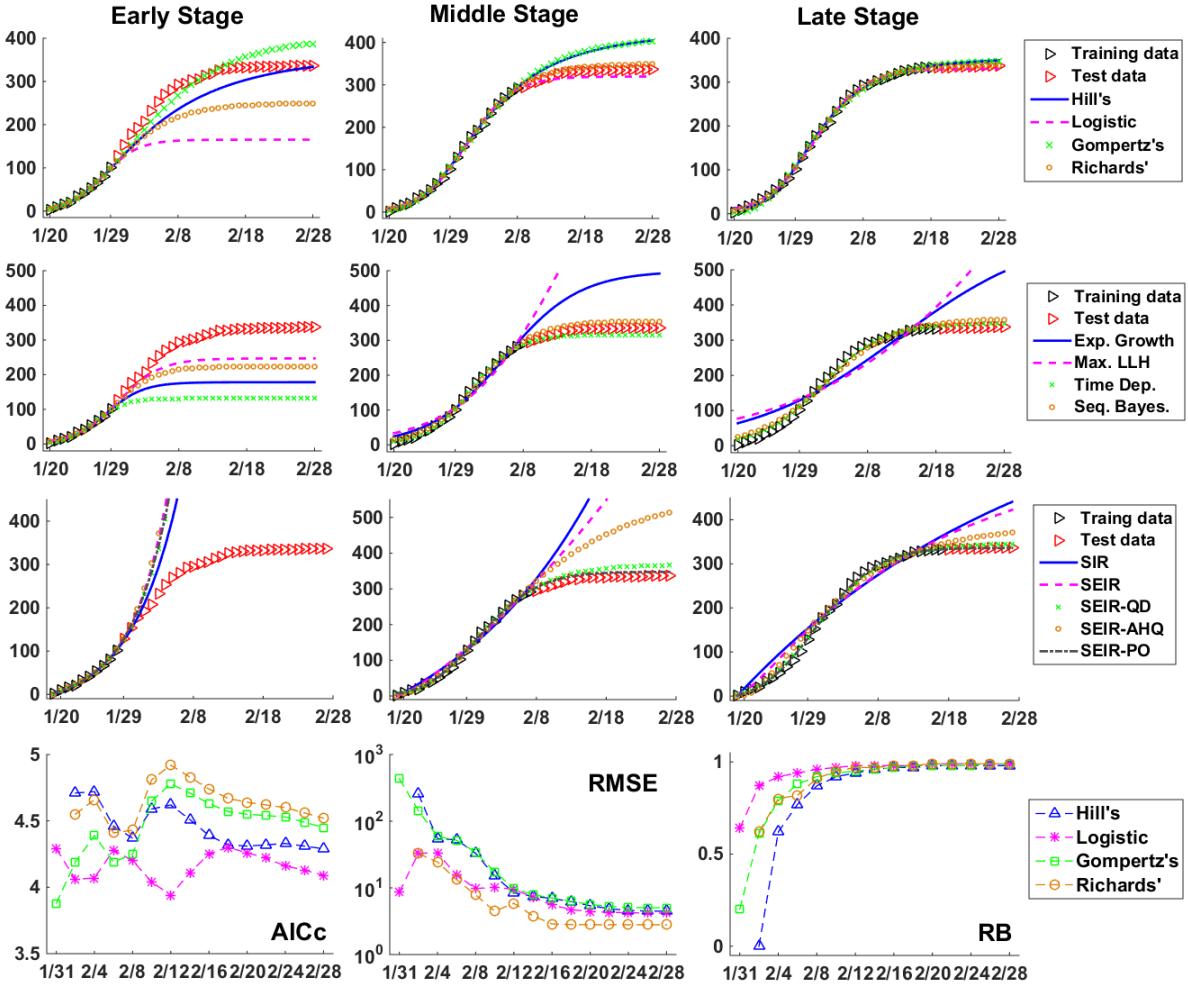


FIG. 3. Forecast of the COVID-19 epidemic in Shanghai from 01/20/2020 to 02/28/2020 based on data of first 10 (early), 20 (middle) and 30 (late) days respectively. The first three panels give the results of (upper) four explicit functions, (middle) four different statistical inference methods combined with the Logistic function (the exponent  $\gamma$  derived from  $R_0$ ), (lower) and five ODE dynamical models. The last row shows the variations of AICc, RMSE and RB in forecast of four explicit functions with respect to different sizes of training data set (from Jan. 20th to the date as marked).

**while Gompertz's function overestimates it requires further validation.** In all nine cases (including three cases for Shanghai), we notice the Logistic functions always underestimate the total number of infected cases, while the Gomertz's function makes an overestimate (see Fig. 4). If this is not a coincidence, it would be extremely useful for estimating the lower and upper bounds for the real total infected populations, though it still

requires further more evidences. The predictions of the other two functions are not so stable and their goodness-of-fit varies from case to case, which reflects the lacking of theoretical support of empirical functions during application.

(6) ***Methods fo 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, they appear to be more accurate than the exponential growth and maximum likelihood methods in the late stage. 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 underestimate the true epidemic size. It is further observed that the estimated basic reproduction number  $R_0$  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, in the early stage of Shanghai COVID-19 epidemic, the exponential growth method after 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.

(7) ***The SEIR-QD and SEIR-PO models are suitable for modeling COVID-10 epidemics.*** The dynamical models generally requires more reliable data to achieve a reliable forecast than empirical functions, since the former usually involves more free parameters and more complicated mathematical structure than the latter. Based on their performance, the dynamical models can be classified into three groups. 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.

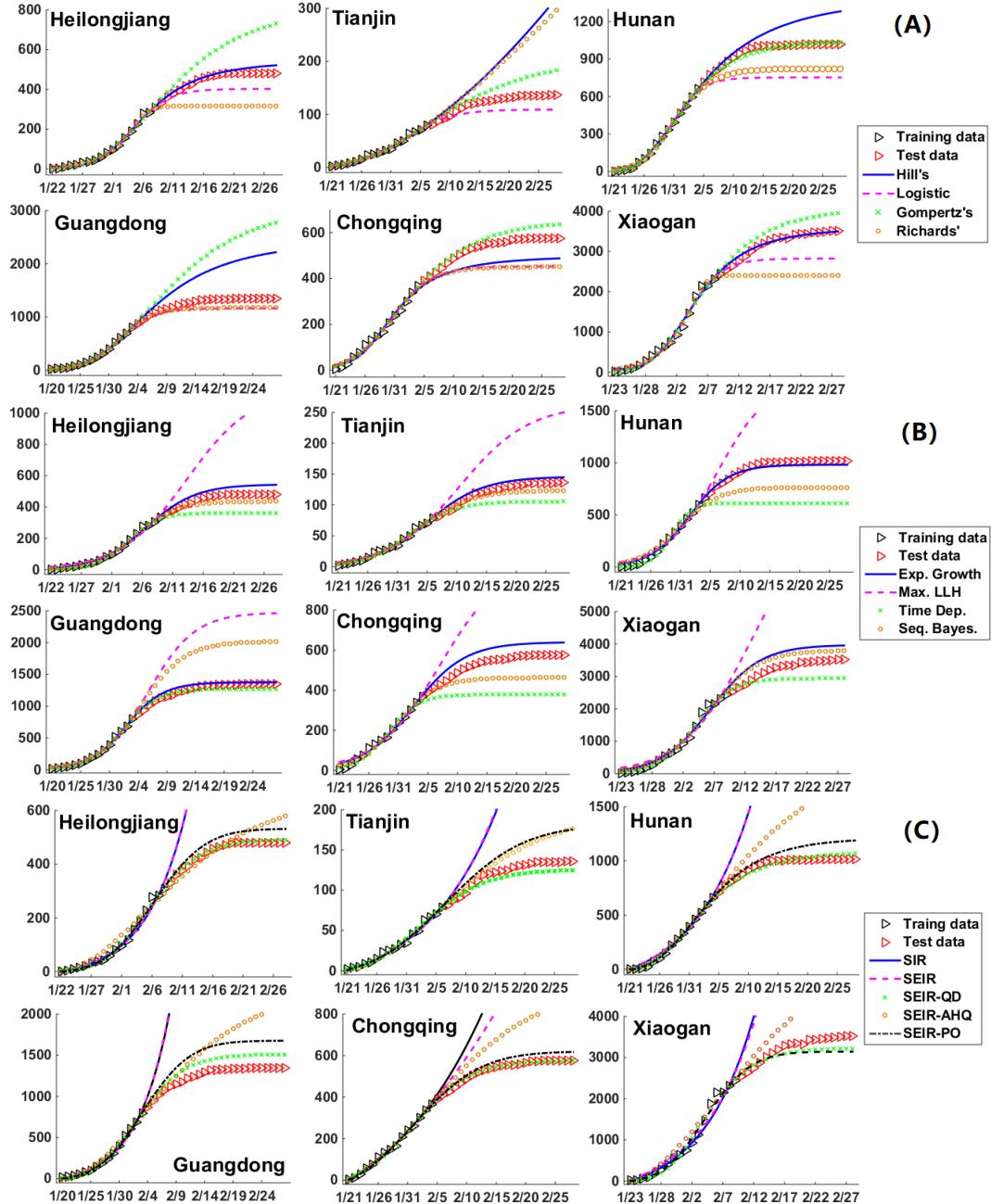


FIG. 4. Forecast of COVID-19 epidemics in Heilongjiang province (data for first 18 days of 38 in total are used for modeling fitting (training set), while the rest 20 data points are used for validation (test set)), Tianjin (17/39), Hunan province (15/39), Guangdong province (15/40), Chongqing (15/39) 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 (upper) four explicit functions, the middle one (B) for four different statistical inference methods combined with the Logistic function (the exponent  $\gamma$  derived from  $R_0$ ), while the lower panel (C) gives the results of five ODE dynamical models.

## IV. EPIDEMIC TRENDS IN SEVERAL HOT COUNTRIES

Based on our previous evaluations on different epidemic models and methods, we try to forecast the epidemic trends in several hot countries, which is currently facing severe threat of the outbreak of COVID-19 epidemics. However, after a very careful examination, we find that except China most countries with more than 100 confirmed COVID-19 cases are still within the early stage of epidemics (the inflection point has not been reached yet). It is really terrible! As a consequence, we refer to the statistical inference method, the time-dependent reproduction number to be exact, to try to derive the basic/effective reproduction number. Then we adopt the Gompertz's function (usually overestimate the final epidemic size as we claimed above), Logistic function, as well as Logistic function combined with  $R_0$  derived from the exponential growth model (usually underestimate the final epidemic size before the inflection point and make an overestimation after that) to try to provide more reasonable upper and lower bounds for the total confirmed infected case at least in the near future.

Here we look into three examples – South Korea, Italy and Iran, which have the second to fourth largest confirmed infected populations by COVID-19 till Mar. 10th, 2020. Based on the effective reproduction number  $R_t$  derived from the method of time-dependent reproduction number, it is observed that  $R_t$  for South Korea becomes lower than 1, meaning the COVID-19 epidemic is under control. The number of total confirmed cases is estimated to be 8,000 – 10,800. In contrast,  $R_t$  for Italy and Iran are both larger than 1, suggesting a rapid spreading of COVID-19 in these countries. A conservative estimate of final confirmed cases is around 18,000 – 23,000 for Italy and 9,500 – 13,500 for Iran. Similar severe epidemic situations are observed in Germany, France and Spain (data not shown), meaning these European countries (and also some others) are in great danger too.

## V. 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 eight widely used empirical functions, four statistical inference methods and five dynamical models reported in the literature. We highlight the significance of a well balance between model complexity and accuracy, over-fitting and under-fitting, as well as model robustness and

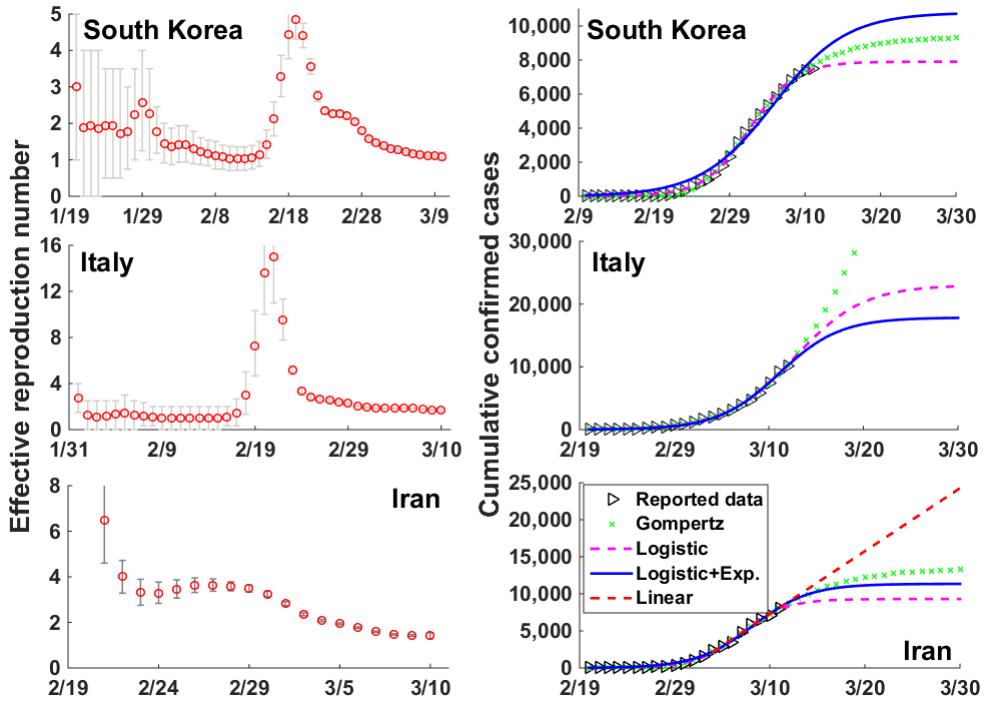


FIG. 5. Forecast of the COVID-19 epidemics in South Korea, Italy and Iran. Left column shows the effective reproduction number derived from the method of time-dependent reproduction number, while the right column gives the reported confirmed cases and predicted ones by using Gompertz's function, Logistic function, and Logistic function combined with the exponential growth model.

sensitivity. We further introduce the Akaike information criterion, root mean square errors and robustness index to quantify these three golden means and to evaluate various epidemic models/methods.

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<sup>16</sup>. 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 increasing in the size of 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 Gomertz's function makes an overestimate in all cases we studies. 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 essential 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.

This work has some limitations. The analyses are highly relied on the quality of the epidemic data. The initial conditions and parameter regions may also affect the results significantly, though our qualitative conclusions will not be changed. Besides five simple ODE models, many other epidemic models inclduing PDEs, stochastic equations and time-delayed equations are not considered. In addition, further validations than seven cases are needed to explore the applicable regions of all the methods/models.

## ACKNOWLEDGMENT

The authors acknowledged the financial supports from the National Natural Science Foundation of China (Grants No. 21877070, 11801020), Startup Research Funding of Minjiang University (mjy19033) and Special Pre-research Project of Beijing University of Technology for Fighting the Outbreak of Epidemics. Zhuge would like to thank Dr. Yi Wei for his support on data collection.

## REFERENCES

<sup>1</sup>Michael Y. Li. *Important Concepts in Mathematical Modeling of Infectious Diseases*, pages 1–33. Springer International Publishing, Cham, 2018.

<sup>2</sup>Chelsea S. Lutz, Mimi P. Huynh, Monica Schroeder, Sophia Anyatonwu, F. Scott Dahlgren, Gregory Danyluk, Danielle Fernandez, Sharon K. Greene, Nodar Kipshidze, Leann Liu, Osaro Mgbere, Lisa A. McHugh, Jennifer F. Myers, Alan Siniscalchi, Amy D. Sullivan, Nicole West, Michael A. Johansson, and Matthew Biggerstaff. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health*, 19(1):1659, December 2019.

<sup>3</sup>Sanjay Basu and Jason Andrews. Complexity in mathematical models of public health policies: A guide for consumers of models. *PLOS Medicine*, 10(10):1–6, 10 2013.

<sup>4</sup>Guillaume Gingras, Marie-Hélène Guertin, Jean-François Laprise, Mélanie Drolet, and Marc Brisson. Mathematical modeling of the transmission dynamics of clostridium difficile infection and colonization in healthcare settings: A systematic review. *PLOS ONE*, 11(9):1–19, 09 2016.

<sup>5</sup>Xiao-Na Han, Sake J. De Vlas, Li-Qun Fang, Dan Feng, Wu-Chun Cao, and J. Dik F. Habbema. Mathematical modelling of sars and other infectious diseases in china: a review. *Tropical Medicine & International Health*, 14(s1):92–100, 2009.

<sup>6</sup>Catherine A. A. Beauchemin and Andreas Handel. A review of mathematical models of influenza a infections within a host or cell culture: lessons learned and challenges ahead. *BMC Public Health*, 11(1):S7, February 2011.

<sup>7</sup>Anirudh Mutalik. Models to predict h1n1 outbreaks: a literature review. *International Journal Of Community Medicine And Public Health*, 4(9):3068–3075, 2017.

<sup>8</sup>Gerardo Chowell, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66 – 97, 2016.

<sup>9</sup>Caroline E. Walters, Margaux M.I. Meslé, and Ian M. Hall. Modelling the global spread of diseases: A review of current practice and capability. *Epidemics*, 25:1 – 8, 2018.

<sup>10</sup>Sebastian Funk, Anton Camacho, Adam J. Kucharski, Rachel Lowe, Rosalind M. Eggo, and W. John Edmunds. Assessing the performance of real-time epidemic forecasts: A case study of ebola in the western area region of sierra leone, 2014-15. *PLOS Computational Biology*, 15(2):1–17, 02 2019.

<sup>11</sup>Theresa Stocks, Tom Britton, and Michael Höhle. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics*, 09 2018. kxy057.

- <sup>12</sup>Kimberlyn Roosa and Gerardo Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1):1, January 2019.
- <sup>13</sup>Farzaneh Sadat Tabataba, Prithwish Chakraborty, Naren Ramakrishnan, Srinivasan Venkatramanan, Jiangzhuo Chen, Bryan Lewis, and Madhav Marathe. A framework for evaluating epidemic forecasts. *BMC Infectious Diseases*, 17(1):345, May 2017.
- <sup>14</sup>Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- <sup>15</sup>Allan D. R. McQuarrie and Chih-Ling Tsai. *Regression and Time Series Model Selection*. World Scientific, 1998.
- <sup>16</sup>Shi Zhao, Salihu S Musa, Hao Fu, Daihai He, and Jing Qin. 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, 2019.
- <sup>17</sup>Thomas Obadia, Romana Haneef, and Pierre-Yves Boëlle. The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC medical informatics and decision making*, 12(1):147, 2012.
- <sup>18</sup>Jinghua Li, Yijing Wang, Stuart Gilmour, Mengying Wang, Daisuke Yoneoka, Ying Wang, Xinyi You, Jing Gu, Chun Hao, Liping Peng, Zhicheng Du, Dong Roman Xu, and Yuantao Hao. Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study. *medRxiv*, 2020.
- <sup>19</sup>Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong. Epidemic analysis of covid-19 in china by dynamical modeling. *arXiv preprint arXiv:2002.06563*, 2020.
- <sup>20</sup>Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2), 2020.

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

TABLE II. Summary of AICc, RB and RMSE values for six provinces/cities in China based on different models.