

# Prediction of the COVID-19 outbreak based on a realistic stochastic model

Yuan Zhang<sup>\*,1,2</sup>, Chong You<sup>\*\*3</sup>, Zhenghao Cai<sup>1</sup>, Jiarui Sun<sup>1</sup>,  
Wenjie Hu<sup>1</sup>, Xiao-Hua Zhou<sup>\*\*\*,3,4,2</sup>

1. School of Mathematical Sciences,  
Peking University, Beijing, China, 100871

2. Center for Statistical Sciences,  
Peking University, Beijing, China, 100871

3. Beijing International Center for Mathematical Research,  
Peking University, Beijing, China, 100871

4. Department of Biostatistics, School of Public Health  
Peking University, Beijing, China, 100871

\* First author \*\* Joint first author

\*\*\* Corresponding author: [azhou@math.pku.edu.cn](mailto:azhou@math.pku.edu.cn)

10 March 2020

## Abstract

The current outbreak of coronavirus disease 2019 (COVID-19) has become a global crisis due to its quick and wide spread over the world. A good understanding of the dynamic of the disease would greatly enhance the control and prevention of COVID-19. However, to the best of our knowledge, the unique features of the outbreak have limited the applications of all existing models. In this paper, a novel stochastic model is proposed which aims to account for the unique transmission dynamics of COVID-19 and capture the effects of intervention measures implemented in Mainland China. We find that, (1) instead of aberration, there is a remarkable amount of asymptomatic individuals, (2) an individual with symptoms is approximately twice more likely to pass the disease to others than that of an asymptomatic patient, (3) the transmission rate has reduced significantly since the implementation of control

measures in Mainland China, (4) it is expected that the epidemic outbreak would be contained by early March in the the selected provinces and cities.

*Keywords:* Compartment Stochastic Model, Continuous Time Markov Process, COVID-19, reproduction number

## 1 Introduction

The current outbreak of coronavirus disease 2019 (COVID-19) has become a global crisis due to its quick and wide spread over the world. According to [Official Report](#) by the [National Health Commission of the People's Republic of China \(2020b\)](#), as of March 4, 2020, the outbreak of COVID-19 has caused 80 409 confirmed cases and 3012 fatalities in Mainland China. For the purpose of control and prevention, various containment measures have been implemented by the Chinese authorities, including traffic restrictions, medical tracking, entry or exit screening, isolation, quarantine and awareness campaigns since January 19, 2020. Especially on January 23, 2020, a [strict travel restriction](#) was introduced in Hubei province, cities in Hubei have been locked down since then ([Government of the People's Republic of China \(2020\)](#)).

A good understanding of the epidemic dynamic would greatly enhance the control and prevention of COVID-19 as well as other infectious diseases, while dynamic model is probably one of the oldest and best-known mathematical tools to study the law of epidemic development. To the best of our knowledge, the SIR model proposed by [Kermack and McKendrick \(1927\)](#) almost a century ago is the first (and one of the most important) attempt to describe the development of an epidemic with a deterministic ordinary differential equation (ODE) system. Over the years, numerous modifications and generalizations have been made on the original SIR model to accommodate different epidemic characteristics, among which, the SIS model and SEIR model are most widely accepted ones. It is not surprising various deterministic dynamic models have been again designed and employed to the current outbreak of COVID-19. We hereby present a brief review on some of the representative works as follows.

Recently [Tang et al. \(2020\)](#) proposed a deterministic compartmental model by taking the clinical progression, epidemiological status, and the intervention measures into account. However, it implicitly assumed that the exposed population were not infectious, which is not the case in COVID-19. In addition, it was assumed that quarantine was implemented as soon as the infection occurred, which fails to reflect the inevitable latency brought by medical tracking. In the study of [Wu et al. \(2020\)](#), it proposed an extended SEIR model by considering transmissions among cities. However, it did not take the control measures

such as quarantine and medical tracking into account. Furthermore, it was also assumed that exposed population are not infectious. For more similar or simpler deterministic ODE models to COVID-19, we refer to [Liu et al. \(2020b\)](#) for an overview.

Besides the ODE models, an even older approach is to discretize the time and consider a difference equation (DE) model, which can be seen as a self-propelled sequence of numbers (or vectors). In such models, the increment or decrement of different compartment at each step is a function of the configuration in the previous step(s). Earliest examples of a DE include the well-known geometric and Fibonacci sequence, which can both be used to describe the growth of a biological population. Recently, a discrete time DE model was introduced in [Yang et al. \(2020\)](#) to predict the epidemic trend of COVID-19. The proposed model correctly took the infectious incubation into account. However, this model did not consider the time lag between symptoms onset and diagnosis or the medical tracking. Besides, the rationale behind the assumption of the equal transmission probability between symptomatic and asymptomatic patients is questionable.

In contrast to the deterministic models (ODE or DE) summarized above, the transmission of a real world disease is inevitably random in nature. As a result, numerous stochastic dynamics models (see Section 3 for references) have been developed since the pioneering randomization of SIR model in [Kendall \(1956\)](#). In fact, a deterministic ODE model can often be seen as the mean-field equation of the corresponding stochastic counterpart. Under certain conditions, the mean-field equation may represent the evolution of the expectation of the corresponding stochastic model. In some more generalized cases, the mean-field equation is a large scale approximation of the corresponding stochastic model, which can be seen as a process version of *Law of Large Numbers*. However, when the size of outbreak is not comparable to that of the total population, the approximation aforementioned may not be interpreted that the stochastic and deterministic models are close to each other by themselves, since they have both been rescaled to approximately a constant in these scenarios. See Section 3 for more details. Thus compared with deterministic models, the stochastic model may be a better choice to account for the non-negligible random nature of the COVID-19 outbreak. Furthermore, the stochastic dynamic model is also known for its expandability to incorporate individual variations [Athreya and Ney \(1972\)](#), or even spatial structures [Durrett \(1988\)](#), which may not be fully captured by its mean-field equations. To our knowledge, the stochastic dynamic modeling for COVID-19 is yet relatively rare comparing to its deterministic counterparts, though preliminary approaches such as statistic exponential growth models was considered in recent studies of [Liu et al. \(2020a\)](#); [Zhao et al. \(2020\)](#). Very recently in [Chinazzi et al. \(2020\)](#), an existing discrete time, stochastic model ([Balcan et al. \(2009\)](#)) was employed to estimate the “effect of travel restrictions on the spread” of COVID-19. However, unique features of SARS-CoV-2, such as the infectious incubation and asymptomatic carriers, as well as control measures such as medical

tracking, are still yet to be captured in [Chinazzi et al. \(2020\)](#).

To remedy the aforementioned issues in the existing studies, and to depict a more realistic transmission mechanism, we propose a novel stochastic compartmental model which captures the unique transmission dynamics of COVID-19 and the effects of intervention measures implemented in Mainland China. Our proposed stochastic model aims to study the COVID-19 outbreak in the following aspects:

- Estimation of key epidemiology parameters:

Parameters in the model have specific epidemiology meanings. For example,  $\rho$  stands for the proportion of symptomatic virus carriers,  $\theta_E$  represents the proportion of infectivity of asymptomatic carriers and symptomatic carriers and  $q$  can reflect the strength of quarantine in each region. These parameters have very limited information in medical reports for direct access and can be estimated through our model.

- Prediction of epidemic development:

With a set of estimated parameters and initial values of the stages, we are able to simulate the development of epidemic over time  $t$ . This is helpful for capturing the spread tendency of the epidemic.

- Estimation of un-observable carriers and epidemic containment date:

Un-observable virus carriers in the crowd are nonnegligible for assess of epidemic control. Instead of epidemic control assessment only through existing infections, our model is able to take the trend of un-observable carriers into account. Based on the simulated trend of un-observable carriers, epidemic containment date can also be calculated.

- Assessment of control measures

The controlled reproduction number,  $R_c$ , which reflects the transmission ability of the epidemic, is essential for the assessment of control measures and speed of disease transmission.  $R_c$  can be estimated from the ratio between the in-and-out flows of the active virus carriers in a given time period in our model. Moreover, we can also conduct hypothetical “controlled ” test and evaluate the effectiveness of medical tracking.

Our main findings are as follows:

1. There exist a non-negligible portion of asymptomatic virus carriers.
2. Patients with symptoms have approximately twice the potential of transmission comparing with that of asymptomatic carriers.

3. The epidemic containment times in the selected provinces and cities (see Section 5 for a precise definition of the epidemic containment times) are estimated to be around late February to early March.
4. The time dependent controlled reproduction numbers  $R_c$  are between 2 and 3 in the selected provinces and cities at the beginning of the epidemic, and have been significantly reduced since the implement of control measures.
5. Medical tracking endeavors contribute significantly in containing the epidemic in the selected provinces and cities.

The rest of this paper is structured as follows. In Section 2 we describe the data used in this study. Section 3 introduces the proposed stochastic dynamic model and compares its features with some representative models in the recent literature. Parameter estimations are described in Section 4. In Section 5 we present our findings. We discuss our results, advantages and limitations in Section 6.

## 2 Method

### 2.1 Data Sources

Data used in this study include the numbers of confirmed diagnosis, recoveries and fatalities in the following provinces and cities of China: Beijing, Shanghai, Chongqing, Guangdong, Zhejiang and Hunan. These publicly available data were retrieved from [Beijing Municipal Health Commission \(2020\)](#); [Shanghai Municipal Health Commission \(2020\)](#); [Chongqing Municipal Health Commission \(2020\)](#); [Health Commission of Guangdong Province \(2020\)](#); [Health Commission of Zhejiang Province \(2020\)](#) and [Health Commission of Hunan Province \(2020\)](#) based on a daily update (see Table 5 and Table 6 in Appendix A). The corresponding population of residents in each region is collected from [China National Bureau of Statistics \(2018\)](#) (see Table 7 in Appendix A).

Note that we exclude Hubei province which is the epicenter of the current outbreak in this study due to the following reasons: (1) the medical resources in Hubei province were overburdened at the beginning of the epidemic, not all individual with confirmed diagnosis could get immediate hospitalization; in fact according to the [Official Press Briefing](#) by the [Information Office of Hubei Provincial People's Government \(2020\)](#), as of February 8, 2020, there were 1499 patients remained with confirmed diagnosis and serve symptom who had yet been hospitalized; (2) the diagnostic criteria were [changed](#) overtime in Hubei which resulted in a [massive surge](#) of confirmed cases in mid February ([National Health Commission](#)

of the People’s Republic of China (2020a)); (3) the fatality rate in Hubei province was much higher than other regions in China. These features distinct the dynamic model in Hubei from the model in other regions of China, which will be considered in our future studies, see Section 6 for details.

## 2.2 Model Description

The evolution of an epidemic is usually described by either a deterministic dynamic model [Kermack and McKendrick \(1927\)](#) or a stochastic one [Kendall \(1956\)](#); [Bailey \(1975\)](#), among which, the stochastic model is generally considered more realistic than the deterministic counterpart. Stochastic models often use a continuous time, compartment Markov Process to describe the evolution of the epidemic. See for instances, [Kendall \(1956\)](#); [Kurtz \(1970\)](#); [Barbour \(1972\)](#); [Bailey \(1975\)](#) for pioneering works, and [Clémençon et al. \(2008\)](#); [Zhang and Wang \(2013\)](#); [Ji and Jiang \(2014\)](#); [Berrhazi et al. \(2017\)](#) for recent developments. See also [Wilkinson et al. \(2017\)](#) and the references therein for a brief review.

Especially, when the size of outbreak is not comparable to that of the total population, the randomness is more significant, hence a stochastic model is needed to quantify the uncertainty in estimates and predictions in such case. In our study, none of selected provinces and cities has more than 2000 accumulated confirmed by now (see table 5 in Appendix A). These number, though alerting, are not comparable to the total population in provinces or cities, which are of an order 10 million-100 million (see table 7 in Appendix A). Hence, a novel stochastic dynamic model is designed to capture the unique features of the COVID-19 outbreak, where the **unique features** here refer to

- Infectious incubation period: according to an [Official Press Briefing](#) by the [State Council Information Office of the People’s Republic of China \(2020\)](#), unlike the SARS-CoV, SARS-Cov-2 can induce an infectious incubation period.
- Large portion of asymptomatic virus carriers: according to an [Official Report](#) by the [Japanese Ministry of Health, Labour and Welfare \(2020\)](#), it has been found that the proportion of asymptomatic infected population is considerable.
- Unprecedented contact control and medical tracking measures: since January 19, 2020, various containment measures have been implemented by the Chinese authorities. Especially on January 23, 2020, [strict travel restriction](#) was introduced in Hubei province at a unprecedented scale, cities in Hubei have been locked down since then ([Government of the People’s Republic of China \(2020\)](#)). At the same time, great efforts have also been taken in the medical tracking and quarantine of close contacts. For instance, [remarkable efforts](#) have been made by the [The People’s Govern-](#)

ment of Zhejiang Province (2020), which till March 2, 2020 have successfully tracked more than forty thousand close contacts (People's Government of Zhejiang Province (2020)).

To our knowledge, these features have not yet been fully captured by the existing stochastic dynamic models for the epidemic.

Under mild assumptions that (1) motions of all individuals in the system are independent, (2) the total population in the system is a fixed number of  $N$ , we devise the stochastic model with states listed as follow,

- $S$ : Susceptible.
- $E$ : Exposed. It is divided into four sub-states:
  - $E_1$ : will become symptomatic in the future and is not traceable in medical tracking.
  - $E_2$ : will become symptomatic in the future and is traceable in medical tracking.
  - $A_1$ : won't become symptomatic in the future and is not traceable in medical tracking.
  - $A_2$ : won't become symptomatic in the future and is traceable in medical tracking.
- $Q$ : Quarantined. It is divided into two sub-states:
  - $E_q$ : will become symptomatic in the future.
  - $A_q$ : won't become symptomatic in the future.
- $IN$ : Infected, symptomatic, but not yet admitted to hospital. They are divided into two sub-states:
  - $IN_1$ : not traceable in medical tracking.
  - $IN_2$ : traceable in medical tracking.
- $IH$ : Infected, symptomatic and currently under hospitalization. They are divided into two sub-states:
  - $IH_L$ : with light symptom.
  - $IH_S$ : with severe symptom.
- $R$ : Recovered. They are divided into three sub-states:



- $R_A$ : recover from state  $A_1$  and  $A_2$ .
  - $R_N$ : recover from state  $IN_1$  and  $IN_2$ .
  - $R_H$ : recover from state  $IH_L$ .
- D: Dead.

Note that each individual can be classified into one of the above states at a specific time. For the purpose of conciseness, we denote  $S(t)$ ,  $E_1(t)$ ,  $E_2(t)$  and so on as the population sizes in the corresponding states at time  $t$ . The evolution of  $\xi(t) = \{S(t), E_1(t), E_2(t), E_q(t), A_1(t), A_2(t), A_q(t), IN_1(t), IN_2(t), IH_L(t), IH_S(t), R_A(t), R_N(t), R_H(t), D(t)\}$  over time  $t$  forms a continuous time Markov Process with state space  $\{0, 1, 2, \dots, N\}^{15}$ , with transitions described as follows:

- Infection: Every primary case (including those of states  $E_1$ ,  $E_2$ ,  $A_1$ ,  $A_2$ ,  $IN_1$  and  $IN_2$ ) passes a pathogen to its secondary case at Poisson rate  $\lambda_E = \lambda_{IN}\theta_E$ ,  $\lambda_A = \lambda_{IN}\theta_A$  and  $\lambda_{IN}$  respectively. To be specific, a primary case chooses an individual randomly from the total population, and the individual is infected if it is of state  $S$ . At each transmission event,
  - with probability  $\rho$ , the secondary case becomes  $E_1$  or  $E_2$ , meanwhile, this contact is traceable with probability  $q$ ;
  - with probability  $1 - \rho$ , the secondary case becomes  $A_1$  or  $A_2$ , meanwhile, this contact is traceable with probability  $q$ .
- Quarantine: Every individual of state  $E_2$ ,  $A_2$  and  $IN_2$  will be quarantined and consequently transfer into  $E_q$ ,  $A_q$  and  $IH$  respectively at a Poisson rate of  $r_q$ . Note we assume such individuals will keep quarantined or will be hospitalized until recovery or death.
- Symptom onset: Every individual of state  $E_1$ ,  $E_2$  or  $E_q$  will develop symptoms and transfer into  $IN_1$ ,  $IN_2$  and  $IH$  respectively at a rate of  $r_s$ .
- Hospitalization: Every individual of state  $IN$  will admitted to hospital at a Poisson rate of  $r_H$ . With probability  $p_1$ , it will transfer into  $IH_L$  and with probability  $1 - p_1$ , it will transfer into  $IH_S$ .
- Symptom relief: Every individual of state  $IH_S$  will transfer into  $IH_L$  at a Poisson rate of  $r_b$ .
- Recovery: Every individual of state  $A_1$ ,  $A_2$ ,  $IN$  and  $IH_L$  will recover at Poisson rates of  $\gamma_A$ ,  $\gamma_{IN}$  and  $\gamma_{IH}$  respectively.



- Death: Every individual of state  $IN$  and  $IH_S$  will die at Poisson rates of  $\delta_{IN}$  and  $\delta_{IH}$  respectively.

The process can be illustrated by Figure 1. The precise definition of the continuous time Markov process can be found in Appendix B.

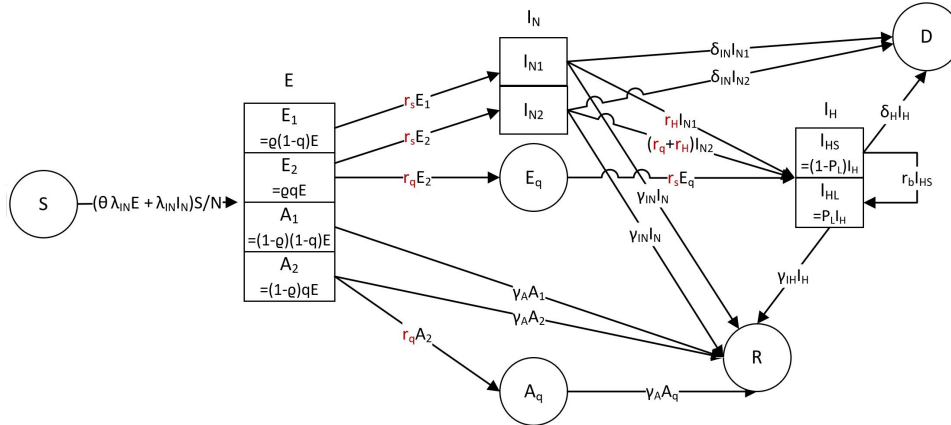


Figure 1: Process Illustration

From a more theoretical, functional analysis point of view, the stochastic dynamic model defined above can be equivalently described with its Markov Semigroup/ infinitesimal generator, see [Liggett \(1985\)](#); [Ethier and Kurtz \(1986\)](#) for details. From those operators, it is natural for us to consider the following **Mean-field Differential Equation System**, which itself serves as a deterministic counterpart of the stochastic model, that is,

$$\tilde{S}(t)' = -\frac{\tilde{S}(t)}{N} \left( \lambda_{IN}\theta_E\tilde{E}(t) + \lambda_{IN}\theta_A\tilde{A}(t) + \lambda_{IN}\tilde{I}\tilde{N}(t) \right), \quad (2.1)$$

$$\tilde{E}_1(t)' = \frac{\tilde{S}(t)}{N} \left( \lambda_{IN}\theta_E\tilde{E}(t) + \lambda_{IN}\theta_A\tilde{A}(t) + \lambda_{IN}\tilde{I}\tilde{N}(t) \right) \rho(1-q) - r_s\tilde{E}_1(t), \quad (2.2)$$

$$E_2(t)' = \frac{\tilde{S}(t)}{N} \left( \lambda_{IN}\theta_E\tilde{E}(t) + \lambda_{IN}\theta_A\tilde{A}(t) + \lambda_{IN}\tilde{I}\tilde{N}(t) \right) \rho q - (r_s + r_q)\tilde{E}_2(t), \quad (2.3)$$

$$\tilde{E}_q(t)' = r_q\tilde{E}_2(t) - r_s\tilde{E}_q(t), \quad (2.4)$$

$$\tilde{A}_1(t)' = \frac{\tilde{S}(t)}{N} \left( \lambda_{IN}\theta_E\tilde{E}(t) + \lambda_{IN}\theta_A\tilde{A}(t) + \lambda_{IN}\tilde{I}\tilde{N}(t) \right) (1-\rho)(1-q) - \gamma_A\tilde{A}_1(t), \quad (2.5)$$

$$\tilde{A}_2(t)' = \frac{\tilde{S}(t)}{N} \left( \lambda_{IN}\theta_E\tilde{E}(t) + \lambda_{IN}\theta_A\tilde{A}(t) + \lambda_{IN}\tilde{I}\tilde{N}(t) \right) (1-\rho)q - (\gamma_A + r_q)\tilde{A}_2(t), \quad (2.6)$$

$$\tilde{A}_q(t)' = r_q\tilde{A}_2(t) - \gamma_A\tilde{A}_q(t), \quad (2.7)$$

$$\tilde{I}\tilde{N}_1(t)' = r_s\tilde{E}_1(t) - (r_H + \gamma_{IN} + \delta_{IN})\tilde{I}\tilde{N}_1(t), \quad (2.8)$$

$$\tilde{I}\tilde{N}_2(t)' = r_s\tilde{E}_2(t) - (r_H + r_q + \gamma_{IN} + \delta_{IN})\tilde{I}\tilde{N}_2(t), \quad (2.9)$$

$$\tilde{I}\tilde{H}_L(t)' = p_1 \left[ r_H\tilde{I}\tilde{N}_1(t) + (r_q + r_H)\tilde{I}\tilde{N}_2(t) + r_s\tilde{E}_q(t) \right] + r_b\tilde{I}\tilde{H}_S(t) - \gamma_H\tilde{I}\tilde{H}_L(t), \quad (2.10)$$

$$\tilde{I}\tilde{H}_S(t)' = (1-p_1) \left[ r_H\tilde{I}\tilde{N}_1(t) + (r_q + r_H)\tilde{I}\tilde{N}_2(t) + r_s\tilde{E}_q(t) \right] - r_b\tilde{I}\tilde{H}_S(t) - \delta_H\tilde{I}\tilde{H}_S(t), \quad (2.11)$$

$$\tilde{R}_A(t)' = \gamma_A\tilde{A}(t), \quad (2.12)$$

$$\tilde{R}_N(t)' = \gamma_{IN}\tilde{I}\tilde{N}(t), \quad (2.13)$$

$$\tilde{R}_H(t)' = \gamma_H\tilde{I}\tilde{H}_L(t), \quad (2.14)$$

$$\tilde{D}(t)' = \delta_{IN}\tilde{I}\tilde{N}(t) + \delta_H\tilde{I}\tilde{H}_S(t). \quad (2.15)$$

Intuitively, the mean-field ODE system above serves as a degenerated case of our stochastic model, where all randomness has been averaged out. When the differential equations are linear, which is not true in our case, the ODE also describes the evolution of the expectation of the stochastic system (Dynkin's Formula, see [Da Prato and Zabczyk \(2014\)](#) Section 4.4 for example). Besides, according to [Kurtz \(1970\)](#) and [Kurtz \(1971\)](#), we may see that this deterministic model can also serve, under a much weaker condition, as a scaling approximation of the stochastic one after rescaled by the renormalizing factor  $N$ . To be more specific, the renormalized random path  $\xi_N(t)/N$  as a stochastic process is “almost” deterministic, and fluctuates closely around the deterministic trajectory  $\tilde{\xi}_N(t)/N$  if the total population  $N$  is large and the renormalized initial values  $\xi_N(0)/N$  and  $\tilde{\xi}_N(0)/N$  are close to each other, where  $\tilde{\xi}(t)$  denote the collections of  $\left\{ \tilde{S}(t), \tilde{E}_1(t), \tilde{E}_2(t), \tilde{E}_q(t), \tilde{A}_1(t), \tilde{A}_2(t), \tilde{A}_q(t), \tilde{I}\tilde{N}_1(t), \tilde{I}\tilde{N}_2(t), \tilde{I}\tilde{H}_L(t), \tilde{I}\tilde{H}_S(t), \tilde{R}_A(t), \tilde{R}_N(t), \tilde{R}_H(t), \tilde{D}(t) \right\}$ , and  $\tilde{\xi}_N(t)$  and  $\xi_N(t)$  be copies of the deterministic and stochastic model with total population  $N$ . Using a mathematical lan-

guage, with probability one there is the pathwise convergence

$$\lim_{N \rightarrow \infty} \max_{t \leq t_0} \frac{\|\xi_N(t) - \tilde{\xi}_N(t)\|}{N} = 0 \quad (2.16)$$

for all  $t_0 \in [0, \infty)$ . However, we would like to specifically bring to the reader's attention that, NO MATTER the size of the total population  $N$ , as long as the size of epidemic outbreak is NOT comparable to  $N$ , the convergence above in (2.16) DOES NOT imply that the un-renormalized  $\xi(t)$  is non-random by itself or that the epidemic involved populations. For example,  $E_1(t)$  and  $\tilde{E}_1(t)$  are close to each other in any sense without the rescaling factor. Actually the fact that the size of epidemic outbreak is not comparable to  $N$  itself already implies that

$$\lim_{N \rightarrow \infty} \frac{\xi_N(t)}{N} = \lim_{N \rightarrow \infty} \frac{\tilde{\xi}_N(t)}{N} = (1, 0, 0, \dots, 0), \quad (2.17)$$

which perfectly satisfies (2.16) while at the same time provides no information on whether the actual values of  $E_1(t)$  and  $\tilde{E}_1(t)$  are close to each other. To summarize, when the size of epidemic outbreak is NOT comparable to  $N$ , the stochastic model possesses intrinsic randomness, and may not be well approximated by a deterministic ODE in any non-degenerate sense.

## 2.3 Choices of initial values and estimation of model parameters

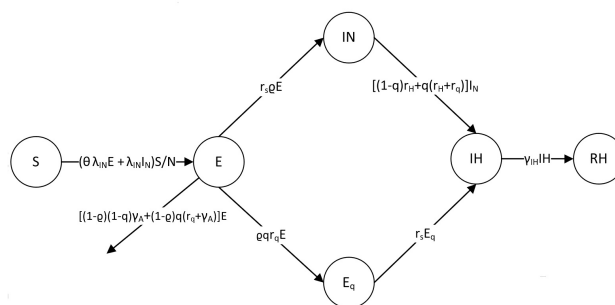


Figure 2: A Simplified Version of process

The parameters and initial values of the proposed stochastic model can be estimated based on a simplified model in Figure 2. Here the stage  $D$  is removed as fatality rate is extremely low in all the selected regions. The populations of  $IH(t)$ ,  $R_H(t)$  can be observed directly from the collected data, that is, the number of existing confirmed cases and reported recoveries at time  $t$  respectively, while, the remaining states are latent, namely not observable. Among the latent states, the initial value  $S(0)$  can be approximated by

the population of permanent residents in the city or province,  $E_q(0)$  is set to be zero as there was no quarantine implemented before January 23, 2020 and  $R_N(0)$  can be set to any number as it would not affect the estimation and prediction of the model. However, the choices for initial values,  $IN(0)$  and  $E(0)$ , is also non-observable and could be a challenge to determine a priori Peng et al. (2020). In this study,  $IN(0)$  and  $E(0)$  are treated as unknown parameters and to be estimated together with other model parameters as described below.

There is a total of 9 model parameters in the proposed model for each selected region. They are  $\lambda_{IN}$ ,  $\theta_E$ ,  $\rho$ ,  $q$ ,  $\gamma_{IH}$ ,  $\gamma_A$ ,  $r_s$ ,  $r_q$  and  $r_H$ , among which,  $r_s$ ,  $r_q$  and  $r_H$  are related to the clinical characteristics of the disease and can be prefixed through existing studies. To be more specific,  $r_H$  is the inverse of the average time from symptoms onset to diagnosis,  $r_s$  is the inverse of the mean incubation period, while  $r_q$  is the inverse of mean difference between infectious period and serial interval. Based on preliminary trials, we find there is very limited information of  $\gamma_A$  which can be obtained from the data, and the estimate is highly influenced by the choices of prior. A possible explanation towards it is that  $\gamma_A$  is less related to the observations. Hence, instead of estimating  $\gamma_A$  with large uncertainty, we prefix  $\gamma_A = 1/10$ . Sensitivity analysis is conducted on the different choices of  $\gamma_A$ .

The rest of parameters would be estimated from the model. Among which, the parameters,  $\rho$  and  $\theta_E$ , are directly related to the nature of the disease, and hence are considered as constants in China, while,  $\lambda_{IN}$ ,  $q$  and  $\gamma_{IH}$  may vary in different regions depending the local medical resources, population densities and containment measures. Furthermore, it is more realistic to consider  $\lambda_{IN}$  and  $\gamma_{IH}$  as time varying parameters to reflect the effect of intervention measures and improvement of the medical treatment. In this study, a simple setting for the time varying function is used, that is,  $\lambda_{IN}(t) = 1_{\{t < T_1\}}\lambda_{IN_0} + 1_{\{t > T_1\}}a\lambda_{IN_0}$  and  $\gamma_{IH}(t) = 1_{\{t < T_2\}}b\gamma_{IH_0} + 1_{\{t > T_2\}}\gamma_{IH_0}$ . The time  $T_1$  is set to be January 29 as there is an obvious change of rates occurred on January 29 illustrated in Figure 3 (You et al., 2020). We use observed  $\frac{R_H(t+1)-R_H(t)}{IH(t)}$  to approximate  $\gamma_{IH}$  on day  $t$ . The time  $T_2$  is selected to be the time when  $\gamma_{IH}$  has a significant change for each province/city.

The likelihood function is obtained by assuming that daily confirmed cases and recoveries are independent Poisson random variables (Wu et al., 2020), that is,

$$L(\lambda_{IN}, \theta_E, \rho, q, \gamma_{IH}) = \prod_{i=1}^k \frac{e^{-\lambda_i} \lambda_i^{\Delta H_i}}{\Delta H_i!} \prod_{j=1}^k \frac{e^{-\gamma_j} \gamma_j^{\Delta R_j}}{\Delta R_j!}$$

where  $\Delta H_t$  and  $\Delta R_t$  are the newly confirmed and recovered cases on day  $t$ ,  $\lambda_t$  and  $\gamma_t$  are the functions of model parameters  $E(0)$ ,  $IN(0)$ ,  $\lambda_{IN}$ ,  $\theta_E$ ,  $\rho$ ,  $q$  and  $\gamma_{IH}$  based on the Mean-field Differential Equation System. Parameters are estimated by the posterior means through the Metropolis-Hastings algorithm implemented in the R package POMP (King et al., 2020, 2016) where non-informative uniform distributions are chosen as the prior distributions, see Table 1 and 2 for more details.

Table 1: Priors for parameter estimation and values for prefixed parameters

Priors for estimation		
Parameter	Prior	Notes
$\rho$	Uniform(0,1)	$\rho$ is the same for all provinces/cities.
$\theta_E$	Uniform(0,1)	$\theta$ is the same for all provinces/cities.
$\lambda_{IN}$	Uniform(0,0.7)	Combining with the prefixed parameter $r_H$ , the upper bound of $\lambda_{IN}$ here means $R_0$ is at least $0.7*5$ , about 3.5, which is a large enough upper bound based on existing estimates for $R_0$ . ( <a href="#">Wu et al., 2020</a> )
q	Uniform(0,1)	
$\gamma_{IH}$	Uniform(1/30,1)	a patient should have recovered after a month in hospital, which means $\gamma_{IH}$ is larger than 1/30.
a	Uniform(0,0.2)	Existing data show newly infections have begun to decrease. This means the speed of infection has dropped to a relatively low value, so a should be quite small.
b	Uniform(0,1)	
Prefixed parameters		
Parameter	value	notes
$r_H$	t<7: 1/5.6	$r_H, r_s, r_q$ can be prefixed through the average time from symptoms onset to diagnosis, the mean incubation period and mean difference between infectious period and serial interval shown in <a href="#">You et al. (2020)</a> .
	t≥7: 0.5	
$r_s$	1/5.3	
$r_q$	t<7: 1/6.6	We assume that $r_H$ and $r_q$ increase to 0.5 due to improvement in testing means and control measures after these measures are taken.
	t≥7: 0.5	

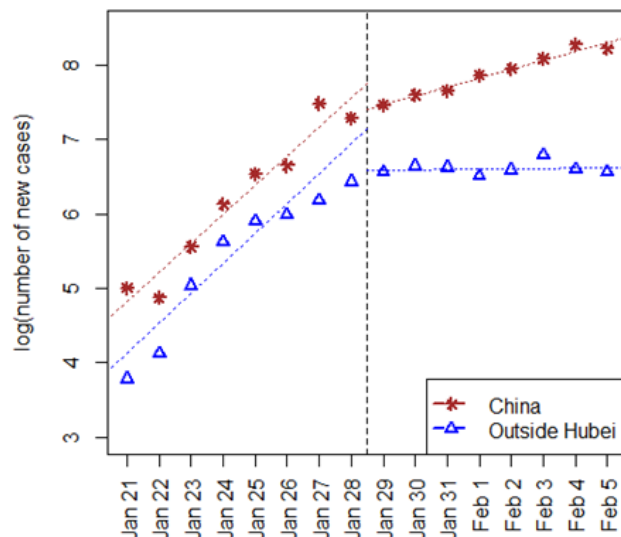


Figure 3: Visualization of daily numbers of newly confirmed case along with date, an obvious change of rate occurred on January 29, 2020.

Table 2: Initial Values

Initial values			
City	Parameter	Prior	Notes
Beijing	$E(0)$	Uniform(0,179)	Upper bound for $IN(0)$ is obtained from the newly confirmed cases in the first 5 days. Upper bound for $E(0)$ is obtained from a coarse estimate based on a conservative $R_0$ shown in <a href="#">Wu et al. (2020)</a> .
	$IN(0)$	Uniform(0,66)	
Shanghai	$E(0)$	Uniform(0,136)	
	$IN(0)$	Uniform(0,50)	
Guangdong	$E(0)$	Uniform(0,424)	
	$IN(0)$	Uniform(0,156)	
Zhejiang	$E(0)$	Uniform(0,443)	
	$IN(0)$	Uniform(0,163)	
Chongqing	$E(0)$	Uniform(0,275)	
	$IN(0)$	Uniform(0,101)	
Hunan	$E(0)$	Uniform(0,364)	
	$IN(0)$	Uniform(0,134)	

## 3 Results

### 3.1 Interpretation of parameters

A summary of the estimated model parameters is given in Table 3, from which we find that

- The estimate of  $\rho$  is not sensitive to the choice of  $\gamma_A$ , about 30% infected individuals are asymptomatic.
- The estimate of  $\theta_E$  is reasonably close to each other with different choices of  $\gamma_A$ , patients with symptoms are about twice as likely to pass a pathogen to others as asymptomatic patients.
- The estimate of  $q$  is not sensitive to the choice of  $\gamma_A$  except in Shanghai. Zhejiang has the highest  $q$  in the selected regions, which is consistent with the remarkable efforts made by the The People's Government of Zhejiang Province (2020), which till March 2, 2020 has successfully tracked more than forty thousand close contacts (People's Government of Zhejiang Province (2020)).

The sensitive analysis of  $\gamma_A$  shows that, (1) the estimates of  $\gamma_{IH}$  and  $b$  are consistent regardless the choice of  $\gamma_A$  as the states  $RH$  and  $IH$  are both observable; (2) the estimated initial populations for states  $E$  and  $IN$  in each regions vary over different choices of  $\gamma_A$ , but are still in the same order of magnitude; (3) the estimate of  $\lambda_{IN}$  and  $a$  vary with the choices of  $\gamma_A$  notably in Guangdong, Zhejiang and Shanghai, cautions need to be taken when making inferences on  $\lambda_{IN}$  and  $a$ .

	$\gamma_A = 1/7$	$\gamma_A = 1/10$	$\gamma_A = 1/14$
$\rho$	0.6866	0.7101	0.7077
$\theta$	0.4958	0.4788	0.5459
$\lambda_{IN}^{gd}$	0.6120	0.5954	0.5249
$\lambda_{IN}^{zj}$	0.4082	0.4018	0.3508
$\lambda_{IN}^{hn}$	0.5384	0.5148	0.4412
$\lambda_{IN}^{bj}$	0.4882	0.4860	0.4244
$\lambda_{IN}^{sh}$	0.4700	0.4890	0.4035
$\lambda_{IN}^{cq}$	0.4061	0.4168	0.3724
$q^{gd}$	0.4328	0.4747	0.4549
$q^{zj}$	0.5311	0.527	0.5586
$q^{hn}$	0.4020	0.4136	0.4087



$q^{bj}$	0.37365	0.3811	0.3886
$q^{sh}$	0.4852	0.5411	0.4754
$q^{cq}$	0.3542	0.3658	0.3827
$a^{gd}$	0.0520	0.0822	0.0597
$a^{zj}$	0.0924	0.0717	0.1096
$a^{hn}$	0.0276	0.0313	0.0249
$a^{bj}$	0.1113	0.1027	0.1163
$a^{sh}$	0.0879	0.1169	0.0642
$a^{cq}$	0.1366	0.1310	0.1496
$E_0^{gd}$	152	147	177
$E_0^{zj}$	402	368	398
$E_0^{hn}$	217	222	248
$E_0^{bj}$	43	37	46
$E_0^{sh}$	75	61	75
$E_0^{cq}$	100	69	82
$IN_0^{gd}$	88	83	77
$IN_0^{zj}$	75	79	68
$IN_0^{hn}$	41	39	32
$IN_0^{bj}$	49	49	48
$IN_0^{sh}$	22	23	23
$IN_0^{cq}$	80	86	81
$\gamma_{IH}^{gd}$	0.04738	0.04748	0.04735
$\gamma_{IH}^{zj}$	0.04856	0.04863	0.04826
$\gamma_{IH}^{hn}$	0.07132	0.07071	0.07139
$\gamma_{IH}^{bj}$	0.03702	0.03674	0.03707
$\gamma_{IH}^{sh}$	0.11650	0.11480	0.11710
$\gamma_{IH}^{cq}$	0.06682	0.06838	0.06750
$b^{gd}$	0.1732	0.1721	0.1769
$b^{zj}$	0.3006	0.3078	0.3104
$b^{hn}$	0.4556	0.4594	0.4543
$b^{bj}$	0.5292	0.5347	0.5195
$b^{sh}$	0.1558	0.1563	0.1564
$b^{cq}$	0.2472	0.2419	0.2469

---

Table 3: parameter estimation

## 3.2 Prediction of the epidemic trend

Based on the estimated parameters, future trajectories of the epidemic are simulated using the proposed stochastic dynamic model. For each region, 1000 simulations are conducted to produce the 95% confidence interval for the epidemic evolution of some key populations. In this section, the predictions of populations of states, the containment time of the outbreak, the controlled reproduction number  $R_c$ , and a test on the effectiveness of the current medical tracking policy are reported.

Figure 4 plots the 95% confidence interval of:

- accumulated confirmed cases, namely, the sum of  $IH$ ,  $RH$  and  $D$ .
- the population of state  $IH$ , representing the people in hospitals.
- population of active virus carriers, consisting of the states  $E$  and  $IN$ .

Note that the first two populations can be directly observed, but population of  $E$  and  $IN$  is not observable. We can see that the observed accumulated confirmed cases and the population in  $IH$  lie in the calculated confidence interval well. Sensitivity analysis shows that the predicted confidence interval is stable against the uncertainty of  $\gamma_A$ , which gives us more confidence in that the proposed model can capture the spread tendency of the epidemic.

The containment time of the outbreak is defined as the moment when the number of active virus carriers is first less than a threshold  $T_c$ . Noting that the spread of the disease is due to the existing active virus carriers in the proposed model, thus when the number of the active virus carriers decreases to a certain level, the new infections are limited, which means the epidemic is under control. Figure 5 shows the 95% confidence interval of the containment time of the outbreak for each region when  $T_c = 10$ . Among the six regions in this study, Shanghai is predicted to have the earliest containment time of February 21, while the containment time in Guangdong is predicted to be the latest, around March 7. From the daily report of NHC of China, new infections in Shanghai from February 17 to February 21 are 0,0,0,1,0; and new infections in Guangdong in the same period are 6,3,1,1,6, which potentially supports our prediction. Notice that in sensitivity analysis on  $\gamma_A$ , we find that when the recover rate of asymptomatic patients  $\gamma_A$  gets smaller, the expected containment time is longer. This is because smaller  $\gamma_A$  means that infected asymptomatic patients need

more time to recover, which making the number of active virus carriers stays large for a longer time.

The controlled reproduction number,  $R_c$ , reflects the transmission ability of the epidemic, which is one of the most important quantities in epidemiology. In this study, the evolution of  $R_c$  is approximated by the ratio between the in-and-out flows of the active virus carriers in a given time period, that is, for a time interval of length  $\Delta t$ , say  $[t, t + \Delta t]$ , we keep tracking the transitions that leads to the increase/decrease of the active virus carrier population  $E$  and  $IN$ , with their accumulative numbers recorded. Precisely speaking, we define

$$RC\_IN([t, t + \Delta t]) = \text{card}(\{t \in [t, t + \Delta t], E(t) + IN(t) = E(t-) + IN(t-) + 1\})$$

and

$$RC\_OUT([t, t + \Delta t]) = \text{card}(\{t \in [t, t + \Delta t], E(t) + IN(t) = E(t-) + IN(t-) - 1\}).$$

Recalling that our stochastic epidemic model evolves as a non-explosive continuous time Markov process,  $RC\_IN(\cdot)$  and  $RC\_OUT(\cdot)$  are both non-negative finite integers. And when  $RC\_OUT([t, t + \Delta t]) \neq 0$ , we approximate the controlled reproduction number over  $[t, t + \Delta t]$  by its average value

$$\overline{R_c}([t, t + \Delta t]) = \frac{RC\_IN([t, t + \Delta t])}{RC\_OUT([t, t + \Delta t])}.$$

Simulation results for the approximated  $R_c$  in each region is shown in Figure 6. In most provinces and cities we find  $R_c$  is between 2 and 3 before control measures, and drops rapidly to about 0.2 during January 29 and February 1 in all selected regions.

Finally, we evaluate the effectiveness of the current medical tracking policy with a hypothetical controlled test as follows: while keeping the other parameters we estimated unchanged, we set the traceable ratio  $q = 0$  to represent the scenario where no medical tracking were implemented, and let the system evolve. Under this setting, the epidemic would still be contained, see Figure 7 for details, due to the contact control measures that reduce the values of  $\lambda$ 's and the reduction of diagnosis waiting time. However, for selected the provinces and cities in this study, there are significant delays in the dates of containment when  $q = 0$ , which indicates the current medical tracking policy contribute significantly in containing the epidemic.

## 4 Discussion

In this article, we propose a novel stochastic dynamic model in order to depict the transmission mechanism of COVID-19. In comparison with some existing models on COVID-19,

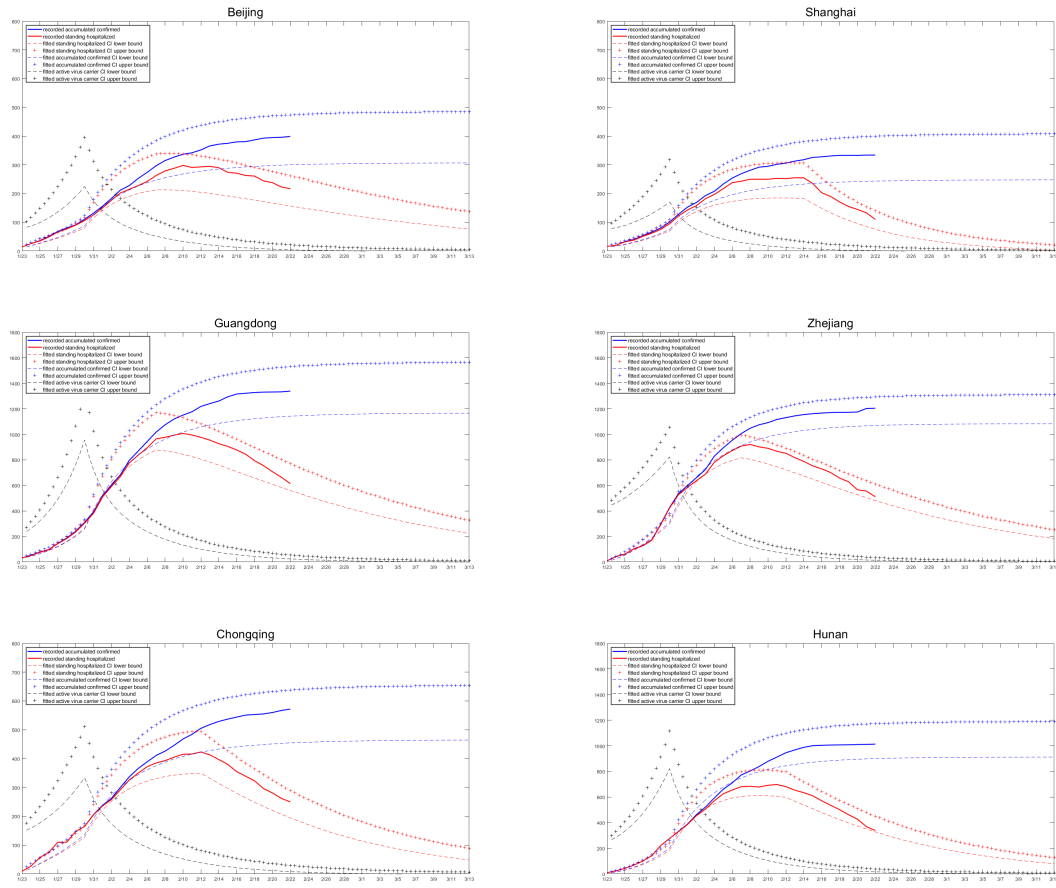


Figure 4: Predicted confidence interval for key populations

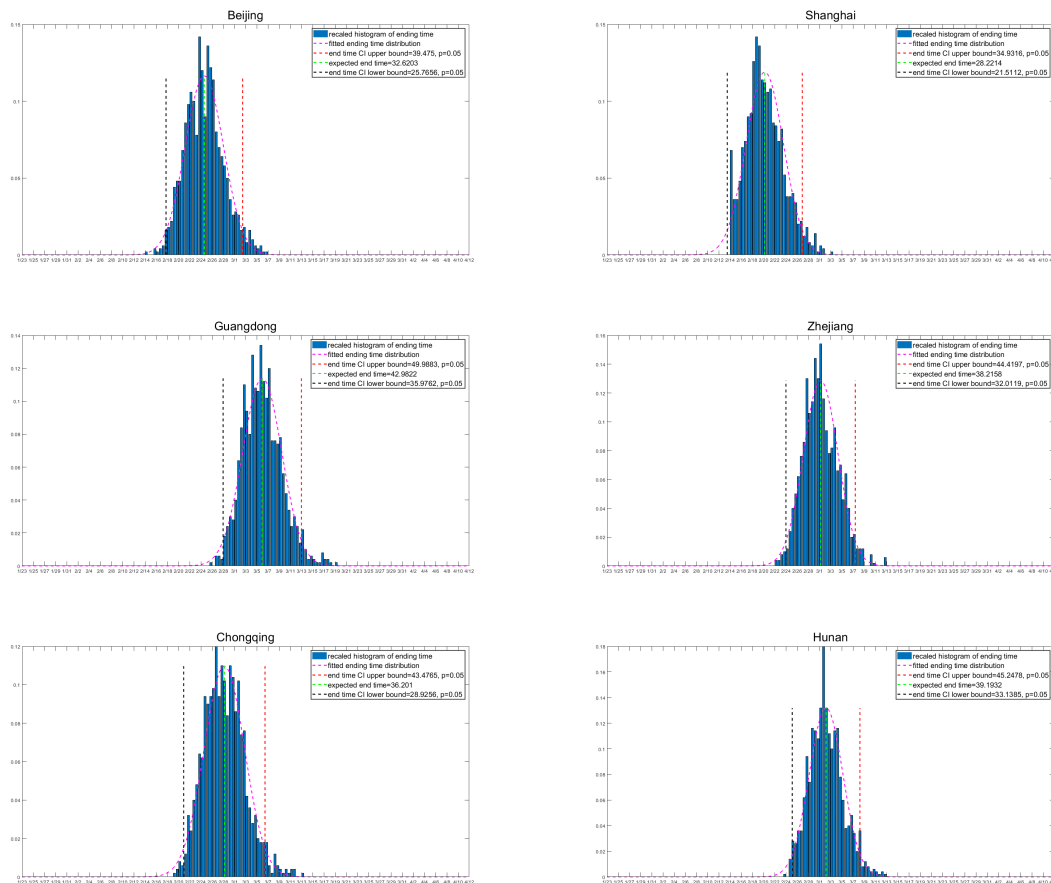


Figure 5: Prediction of the containment time of the outbreak. The containment time of 1000 simulations is plotted as a histogram and is fitted with normal distribution for each region. The  $y$  axis represents the density of containment time.

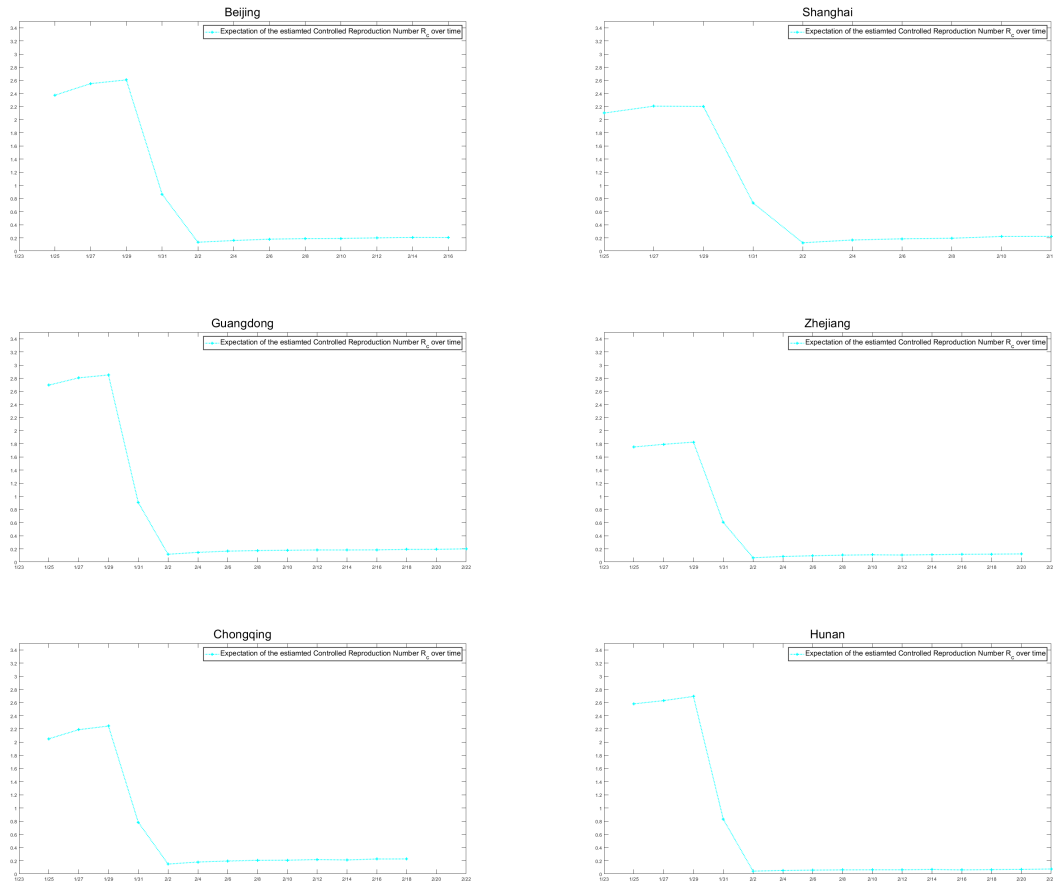


Figure 6: Predicted time-varying  $R_c$  curve.

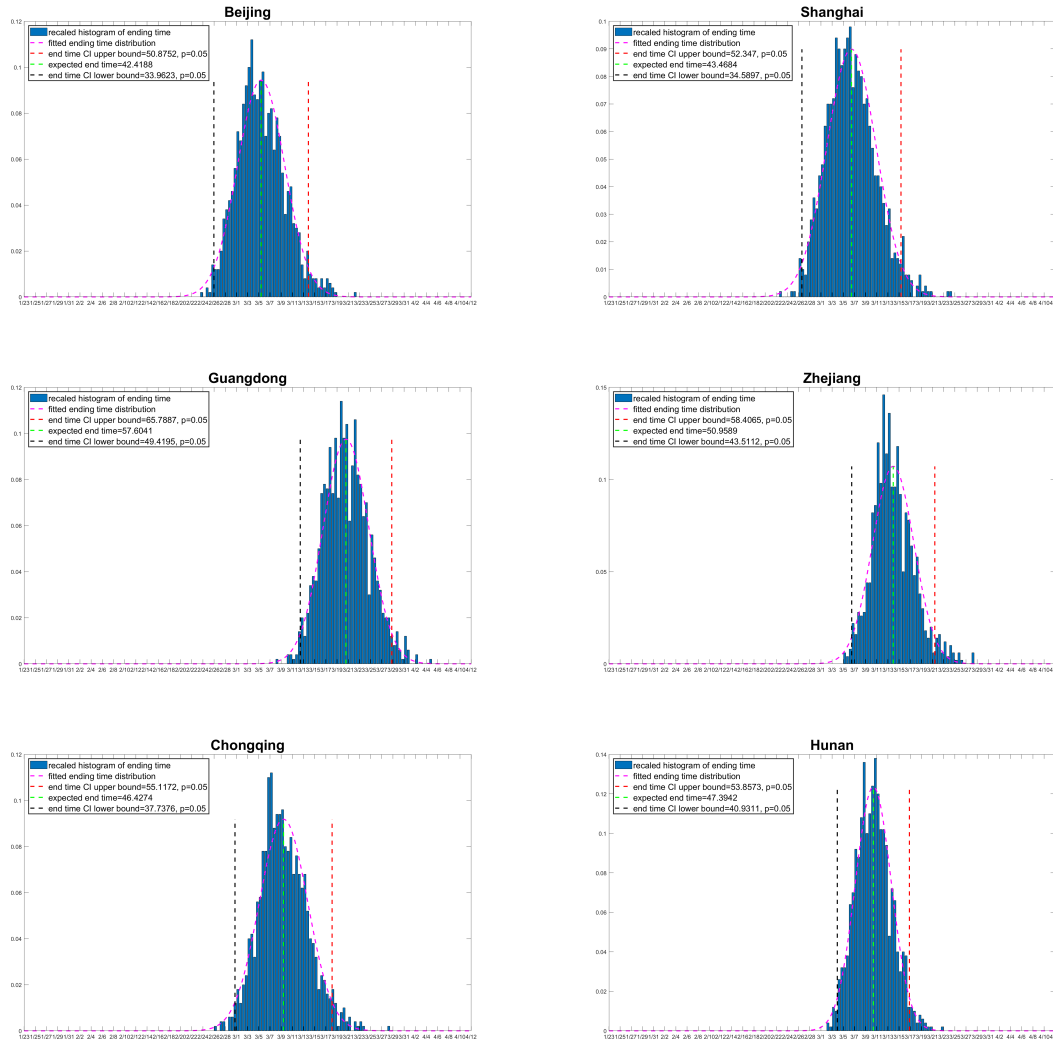


Figure 7: Prediction of the containment time of the outbreak with  $q = 0$ . The containment time of 1000 simulations is plotted as a histogram and is fitted with normal distribution for each region. The  $y$  axis represents the density of containment time.



our model features the employment of a stochastic dynamic as well as a comprehensive account for the infectious incubation period, the asymptomatic virus carriers, and the medical tracking measure with time latency, which make it probably closer to reality in these aspects. Moreover our proposed model also sets the basis for further studies with individual/network based models, which may not have an exact mean-field counterpart.

In the selected provinces and cities in our study, none of them has more than 2000 accumulated confirmed by now (See Table 5 in Appendix A). The numbers of confirmed cases, though alerting, are still not comparable to the total populations of these provinces, which are of the order 10 million-100 million (see Section 2 for exact population data). Moreover, for a system of size 1000 or so, it is not safe to argue the randomness is negligible due to the Law of Large Numbers, which makes it already sufficiently close to its hydrodynamic limit. This justifies the use of stochastic dynamic in our model.

Since the outbreak of COVID-19, the government has conducted large-scale medical tracking and quarantine measures. These non-negligible measures have a significant impact on the spread of COVID-19. The sub-states of  $E$ ,  $A$  and  $IN$  in our model are able to reflect such tracking and quarantine measures. Moreover, we also take the latency between medical tracking and quarantine into account. What's more, it has been found that the proportion of asymptomatic infected population is considerable. This justifies the introduction of state  $A$ . Finally, for an exposed and symptomatic individual, this model describes the complete process from infection to recovery(or death).

In comparison, the model of [Tang et al. \(2020\)](#)(hereafter abbreviated as T.model) failed to consider the infectivity of the incubative virus carriers( $E$ ). What's more, T.model assumed quarantine measures will complete at the moment of contact. It is not completely physical since there should always be a time lag between the transition and the medical tracking.

In [Wu et al. \(2020\)](#), they didn't explicitly consider the impact of medical tracking and quarantine measures in dynamic. As we mentioned above, there are large-scale medical tracking and quarantine measures after the outbreak of COVID-19, and these measures have a significant impact on the spread of the virus. The questionable assumption that the exposed population is non-infectious has also been made in these two models.

The recent work of [Yang et al. \(2020\)](#) proposed a discrete time difference equation(DE) which correctly accounted the infectious incubation period. However, they assumed a symptomatic patient is hospitalized ("quarantine" in their term, see Page 4 of [Yang et al. \(2020\)](#)) immediately at symptom onset, and did not consider the time lag between symptom onset diagnosis/hospitalization. Moreover, the probability of transmission  $b$  are questionably assumed to be the same for both the symptomatic and asymptomatic population. See Page 3-4 of [Yang et al. \(2020\)](#); and no medical tracking mechanism has been considered. Finally,

there seems to be multiple imperfections in the interpretation of dynamic and estimation of parameters.

The differences among models mentioned above are hereby summarized in Table 4:

	Our Model	Tang et al. (2020)	Wu et al. (2020)	Yang et al. (2020)
Model Type	Stochastic	ODE	ODE (with migration)	DE
Infectious Incubation	YES	NO	NO	YES
Asymptomatic Carrier	YES	YES	NO	NO
Medical Tracking	YES	YES	NO	NO
Latency MT*	YES	NO	NA	NA
Latency H**	YES	YES	NA	NO

Table 4: Comparison of Models

\* Latency MT= Latency in medical tracking

\*\* Latency H= Latency in hospitalization

From the estimated parameters, we find that about 30% of the infected are asymptomatic, patients with symptoms are about twice as likely to pass a pathogen to others as asymptomatic patients, and current containment measures are effective to reduce the contact and transmission rate. The containment time of the outbreak for different regions are predicted. Time-varying  $R_c$  is calculated and a rapid drop of  $R_c$  is observed due to containment measures that governments take. We have also conducted controlled test and find that in addition of the control measures reducing the contacts of people, the current medical tracking policy also contributes significantly in containing the epidemic.

However, we acknowledge there are limitations in the propose model,

1. Given the limited data available, certainly parameters, especially those “far away” from observation in the proposed model may have a potential risk of identification issues.
2. The current estimation method converges slowly;
3. The proposed model does not apply if significant changes apply to the current epidemic control/treatment measure in the future;
4. The proposed model need further modification if a non-negligible portion of the asymptomatic patients remain infectious after the end of quarantine.
5. Parameter estimates may lose precision if the stochastic model differs excessively from its simplification described in Section 4.

In our future study, we propose to complement/generalize our current stochastic model from the following aspects:

- **Improved medical tracking dynamic:** In this ongoing work, we are to introduce a more realistic dynamic for medical tracking. Our current model has already considered the latency between transmission and quarantine. While in this modification, medical tracking is triggered in a more physical manner by the affirmative diagnosis of the transmission source. In this new model the quarantine process of each exposed agent depends on his/her individual contact history. Thus an individual based, rather than compartment dynamic is needed. Then the parameters we have estimated can be used as a basis for the individual-based model.
- **Introduction of medical service capacity:** In this modification, the maximal capacity of the medical service system is to be considered, which might be overloaded when facing a massive outbreak. This idea was first inspired by an [Amateur Demonstration](#) made by [Ele laboratory \(2020\)](#) on a site named “bilibili”, and seems to be among the key features of the initial phase of the COVID-19 outbreak in Wuhan. We may also allow such capacity to be time/configuration dependent to model the contribution of building the “emergency hospitals” in Wuhan.
- **Population flows over cities:** We will model the migration of people over different regions, which could have played an essential role in the spread of the epidemic before the Chinese New Year of 2020. We will also consider the transmissions on board and even allow the population flow to react on the information they have about the epidemic situation.

## Acknowledgment

This research is supported by National Natural Science Foundation of China grant 8204100362 and Zhejiang University special scientific research fund for COVID-19 prevention and control.

## References

- Athreya, K. B. and Ney, P. E. (1972). Branching processes. Springer-Verlag, New York-Heidelberg. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- Bailey, N. T. J. (1975). The mathematical theory of infectious diseases and its applications. Hafner Press [Macmillan Publishing Co., Inc.] New York, second edition.

- Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences, 106(51):21484–21489.
- Barbour, A. D. (1972). The principle of the diffusion of arbitrary constants. J. Appl. Probability, 9:519–541.
- Beijing Municipal Health Commission (2020). Situation report (in chinese). Available at [http://wjw.beijing.gov.cn/xwzx\\_20031/xwfb/202003/t20200305\\_1679143.html](http://wjw.beijing.gov.cn/xwzx_20031/xwfb/202003/t20200305_1679143.html).
- Berrhazi, B.-e., El Fatini, M., Laaribi, A., Pettersson, R., and Taki, R. (2017). A stochastic SIRS epidemic model incorporating media coverage and driven by Lévy noise. Chaos Solitons Fractals, 105:60–68.
- China National Bureau of Statistics (2018). Annual data by province (in chinese). Available at <http://data.stats.gov.cn/easyquery.htm?cn=E0103&zb=A0301&reg=440000&sj=2018>.
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., and Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science.
- Chongqing Municipal Health Commission (2020). Situation report (in chinese). Available at <http://wsjkw.cq.gov.cn/yqxxyqtb/20200221/255637.html>.
- Cléménçon, S., Tran, V. C., and de Arazoza, H. (2008). A stochastic SIR model with contact-tracing: large population limits and statistical inference. J. Biol. Dyn., 2(4):392–414.
- Da Prato, G. and Zabczyk, J. (2014). Stochastic equations in infinite dimensions, volume 152 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, second edition.
- Durrett, R. (1988). Lecture notes on particle systems and percolation. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- Ele laboratory (2020). Computer simulation programs tell you why it’s not time to go out now. Available at [https://www.bilibili.com/video/av86478875/?spm\\_id\\_from=333.788.b\\_7265636f5f6c697374.2](https://www.bilibili.com/video/av86478875/?spm_id_from=333.788.b_7265636f5f6c697374.2).

Ethier, S. N. and Kurtz, T. G. (1986). Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. Characterization and convergence.

Government of the People's Republic of China (2020). Notice of the pneumonia outbreak prevention and control command of new coronary virus infection in wuhan (in chinese). Available at [http://www.gov.cn/xinwen/2020-01/23/content\\_5471751.htm](http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm).

Health Commission of Guangdong Province (2020). Situation report on new coronavirus pneumonia outbreak in guangdong province (in chinese). Available at [http://wsjkw.gd.gov.cn/zwyw\\_yqxx/content/post\\_2903465.html](http://wsjkw.gd.gov.cn/zwyw_yqxx/content/post_2903465.html).

Health Commission of Hunan Province (2020). Situation report on new coronavirus pneumonia outbreak in hunan province (in chinese). Available at [http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw\\_1/202002/t20200221\\_11187516.html](http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw_1/202002/t20200221_11187516.html).

Health Commission of Zhejiang Province (2020). Situation report on new coronavirus pneumonia outbreak in guangdong province (in chinese). Available at [http://www.zjwjw.gov.cn/art/2020/2/21/art\\_1202101\\_41958074.html](http://www.zjwjw.gov.cn/art/2020/2/21/art_1202101_41958074.html).

Information Office of Hubei Provincial People's Government (2020). "prevention and control of pneumonia outbreak of new coronary virus infection" press conference (in chinese). Available at [https://www.hubei.gov.cn/hbfb/xwfbh/202002/t20200210\\_2023490.shtml](https://www.hubei.gov.cn/hbfb/xwfbh/202002/t20200210_2023490.shtml).

Japanese Ministry of Health, Labour and Welfare (2020). About the new-style coronavirus infectious disease which was checked in the cruise ship which is being quarantined at yokohama port (in japanese). Available at [https://www.mhlw.go.jp/stf/newpage\\_09668.html](https://www.mhlw.go.jp/stf/newpage_09668.html).

Ji, C. and Jiang, D. (2014). Threshold behaviour of a stochastic SIR model. Appl. Math. Model., 38(21-22):5067–5079.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. IV, pages 149–165. University of California Press, Berkeley and Los Angeles.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772):700–721.

- King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S. P., Ferrari, M. J., Kendall, B. E., Lavine, M., Nguyen, D., Reuman, D. C., Wearing, H., and Wood, S. N. (2020). pomp: Statistical Inference for Partially Observed Markov Processes. R package, version 2.7.
- King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. Journal of Statistical Software, 69(12):1–43.
- Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. Journal of applied Probability, 7(1):49–58.
- Kurtz, T. G. (1971). Limit theorems for sequences of jump markov processes approximating ordinary differential processes. Journal of Applied Probability, 8(2):344–356.
- Liggett, T. M. (1985). Interacting particle systems, volume 276 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York.
- Liu, T., Hu, J., Xiao, J., He, G., Kang, M., Rong, Z., Lin, L., Zhong, H., Huang, Q., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Gong, D., Wan, D., Chen, S., Guo, L., Li, Y., Sun, L., Liang, W., Song, T., He, J., and Ma, W. (2020a). Time-varying transmission dynamics of novel coronavirus pneumonia in china. bioRxiv.
- Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020b). The reproductive number of covid-19 is higher compared to sars coronavirus. Journal of travel medicine.
- National Health Commission of the People’s Republic of China (2020a). Update on the outbreak of new coronavirus pneumonia as of 24 hours on 12 february (in chinese). Available at <http://www.nhc.gov.cn/xcs/yqtb/202002/26fb16805f024382bff1de80c918368f.shtml>.
- National Health Commission of the People’s Republic of China (2020b). Update on the outbreak of new coronavirus pneumonia as of 24 hours on 4 march (in chinese). Available at <http://www.nhc.gov.cn/xcs/yqtb/202003/6e02433f7786402d8ea162fb5b80b5a0.shtml>.
- Peng, L., Yang, W., Zhang, D., Zhuge, C., and Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv Preprint: 2002.06563.
- People’s Government of Zhejiang Province (2020). Notice of new coronavirus pneumonia in zhejiang province on march 3, 2020 (in chinese). Available at [http://www.zj.gov.cn/art/2020/3/3/art\\_1228996608\\_42060522.html](http://www.zj.gov.cn/art/2020/3/3/art_1228996608_42060522.html).

Shanghai Municipal Health Commission (2020). Situation report (in chinese). Available at <http://wsjkw.sh.gov.cn/xwfb/20200222/0a10b6df11c845368af2d627d9551ed1.html>.

State Council Information Office of the People's Republic of China (2020). State council information office holds press conference on joint prevention and control of pneumonia outbreak with new coronavirus infection(in chinese). Available at <http://www.scio.gov.cn/xwfbh/xwfbh/wqfbh/42311/42478/index.htm>.

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., and Wu, J. (2020). Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. Journal of Clinical Medicine, 9(2):462.

The People's Government of Zhejiang Province (2020). The office of the leading group for the prevention and control of new coronavirus infection in zhejiang province issued a notice (provisional) on the prevention and control of new coronavirus infection in rural areas of zhejiang province (in chinese). Available at [http://www.zj.gov.cn/art/2020/2/10/art\\_1228996604\\_41898059.html](http://www.zj.gov.cn/art/2020/2/10/art_1228996604_41898059.html).

Wilkinson, R. R., Ball, F. G., and Sharkey, K. J. (2017). The relationships between message passing, pairwise, Kermack-McKendrick and stochastic SIR epidemic models. J. Math. Biol., 75(6-7):1563–1590.

Wu, J. T., Leung, K., and Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. Lancet, doi:10.1016/S0140-6736(20)30260-9.

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., and He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease, doi: 10.21037/jtd.2020.02.64.

You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou, F., Pang, C. H., Zhang, Y., Chen, Z., and Zhou, X.-H. (2020). Estimation of the time-varying reproduction number of covid-19 outbreak in China. medRxiv: 10.1101/2020.02.08.20021253.

Zhang, X. and Wang, K. (2013). Stochastic SIR model with jumps. Appl. Math. Lett., 26(8):867–874.

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., and Wang, M. H. (2020). Preliminary estimation of the basic reproduction



number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. [bioRxiv](#).

## Appendix A: Data

date	Beijing	Shanghai	Chongqing	Guangdong	Zhejiang	Hunan
2020.01.22	14	16	9	32	10	9
2020.01.23	26	20	27	53	43	24
2020.01.24	36	33	57	78	62	43
2020.01.25	51	40	75	98	104	69
2020.01.26	68	53	110	146	128	100
2020.01.27	80	66	110	188	173	143
2020.01.28	91	80	147	241	296	221
2020.01.29	111	101	165	311	428	277
2020.01.30	132	128	206	393	537	332
2020.01.31	156	153	238	520	599	389
2020.02.01	183	169	262	604	661	463
2020.02.02	212	193	300	683	724	521
2020.02.03	228	208	337	797	829	593
2020.02.04	253	233	366	870	895	661
2020.02.05	274	254	389	944	954	711
2020.02.06	297	269	411	1018	1006	772
2020.02.07	315	281	426	977	1048	803
2020.02.08	327	292	446	1120	1075	838
2020.02.09	337	295	468	1151	1092	879
2020.02.10	342	302	484	1177	1117	912
2020.02.11	352	306	505	1219	1131	946
2020.02.12	366	313	518	1241	1145	968
2020.02.13	372	318	529	1261	1155	988
2020.02.14	375	326	537	1294	1162	1001
2020.02.15	380	328	544	1316	1167	1004
2020.02.16	381	331	551	1322	1171	1006
2020.02.17	387	333	553	1328	1172	1007

2020.02.18	393	333	555	1331	1173	1008
2020.02.19	395	333	560	1332	1175	1010
2020.02.20	396	334	567	1333	1203	1011
2020.02.21	399	334	572	1339	1205	1013

Table 5: Cumulative confirmed cases. Data for the six regions is available at [Beijing Municipal Health Commission \(2020\)](#); [Shanghai Municipal Health Commission \(2020\)](#); [Chongqing Municipal Health Commission \(2020\)](#); [Health Commission of Guangdong Province \(2020\)](#); [Health Commission of Zhejiang Province \(2020\)](#); [Health Commission of Hunan Province \(2020\)](#).

date	Beijing	Shanghai	Chongqing	Guangdong	Zhejiang	Hunan
2020.01.22	0	0	0	0	1	0
2020.01.23	0	0	0	2	1	0
2020.01.24	1	1	0	2	1	0
2020.01.25	2	1	0	2	1	0
2020.01.26	2	1	0	2	1	0
2020.01.27	2	3	0	4	3	0
2020.01.28	2	4	0	5	3	0
2020.01.29	4	5	0	6	4	0
2020.01.30	5	5	1	11	9	2
2020.01.31	5	9	1	12	15	3
2020.02.01	5	10	3	12	23	8
2020.02.02	9	10	7	14	36	16
2020.02.03	12	10	9	20	48	22
2020.02.04	24	15	14	32	63	35
2020.02.05	31	15	15	49	81	56
2020.02.06	33	25	24	52	98	91
2020.02.07	34	30	31	97	127	119
2020.02.08	37	41	39	125	173	159
2020.02.09	37	44	51	143	201	186
2020.02.10	48	48	66	181	250	213
2020.02.11	56	53	79	241	279	263
2020.02.12	68	57	102	284	327	312

2020.02.13	79	62	128	332	367	352
2020.02.14	97	90	152	386	409	389
2020.02.15	105	124	184	436	437	430
2020.02.16	114	140	207	473	470	467
2020.02.17	122	161	225	530	514	501
2020.02.18	145	177	254	571	544	542
2020.02.19	153	186	274	619	609	578
2020.02.20	169	199	299	664	647	638
2020.02.21	178	221	316	720	694	670

Table 6: Cumulative recoveries. The data for the six regions are available at [Beijing Municipal Health Commission \(2020\)](#); [Shanghai Municipal Health Commission \(2020\)](#); [Chongqing Municipal Health Commission \(2020\)](#); [Health Commission of Guangdong Province \(2020\)](#); [Health Commission of Zhejiang Province \(2020\)](#); [Health Commission of Hunan Province \(2020\)](#).

City/Province	Permanent population
Beijing	21,540,000
Shanghai	24,240,000
Guangdong	113,460,000
Chongqing	31,020,000
Hunan	68,990,000
Zhejiang	57,370,000

Table 7: Permanent population. The data for the six regions are available at [China National Bureau of Statistics \(2018\)](#).

## Appendix B: Formal Definition of the Markov Process

In this appendix, we present the transition rates of the system  $\xi(t) = \{S(t), E_1(t), E_2(t), E_q(t), A_1(t), A_2(t), A_q(t), IN_1(t), IN_2(t), IH_L(t), IH_S(t), R_A(t), R_N(t), R_H(t), D(t)\}$ , which uniquely determines the continuous time Markov Process in our proposed model.

1. infection:

- $[S, E_1, \dots, D] \rightarrow [S - 1, E_1 + 1, \dots, D]$  at rate

$$[\lambda_{IN}\theta_E (E_1 + E_2) + \lambda_{IN}\theta_A (A_1 + A_2) + \lambda_{IN}IN] * \frac{S}{N} * \rho * (1 - q).$$

- $[S, E_1, E_2, \dots, R, D] \rightarrow [S - 1, E_1, E_2 + 1, \dots, R, D]$  at rate  $[\lambda_{IN}\theta_E (E_1 + E_2) + \lambda_{IN}\theta_A (A_1 + A_2) + \lambda_{IN}IN] * \frac{S}{N} * \rho * q.$
- $[S, \dots, A_1, \dots, R, D] \rightarrow [S - 1, \dots, A_1 + 1, \dots, R, D]$  at rate  $[\lambda_{IN}\theta_E (E_1 + E_2) + \lambda_{IN}\theta_A (A_1 + A_2) + \lambda_{IN}IN] * \frac{S}{N} * (1 - \rho) * (1 - q).$
- $[S, \dots, A_2, \dots, R, D] \rightarrow [S - 1, \dots, A_2 + 1, \dots, R, D]$  at rate  $[\lambda_{IN}\theta_E (E_1 + E_2) + \lambda_{IN}\theta_A (A_1 + A_2) + \lambda_{IN}IN] * \frac{S}{N} * (1 - \rho) * q.$

2. quarantine:

- $[S, E_1, E_2, E_q, \dots, D] \rightarrow [S, E_1, E_2 - 1, E_q + 1, \dots, D]$  at rate  $E_2 * r_q.$
- $[S, \dots, A_1, A_2, A_q, \dots, D] \rightarrow [S, \dots, A_1, A_2 - 1, A_q + 1, \dots, D]$  at rate  $A_2 * r_q.$
- $[S, \dots, IN_2, \dots, IH_L, \dots, D] \rightarrow [S, \dots, IN_2 - 1, \dots, IH_L + 1, \dots, D]$  at rate  $IN_2 * r_q * p_1.$
- $[S, \dots, IN_2, \dots, IH_S, \dots, D] \rightarrow [S, \dots, IN_2 - 1, \dots, IH_S + 1, \dots, D]$  at rate  $IN_2 * r_q * (1 - p_1).$

3. producing symptoms:

- $[S, E_1, \dots, IN_1, \dots, D] \rightarrow [S, E_1 - 1, \dots, IN_1 + 1, \dots, D]$  at rate  $E_1 * r_s.$
- $[S, E_1, E_2, \dots, IN_2, \dots, D] \rightarrow [S, E_1, E_2 - 1, \dots, IN_2 + 1, \dots, D]$  at rate  $E_2 * r_s.$
- $[S, E_1, E_2, E_q, \dots, IH_L, \dots, D] \rightarrow [S, E_1, E_2, E_q - 1, \dots, IH_L + 1, \dots, D]$  at rate  $E_q * r_s * p_1.$
- $[S, E_1, E_2, E_q, \dots, IH_S, \dots, D] \rightarrow [S, E_1, E_2, E_q - 1, \dots, IH_S + 1, \dots, D]$  at rate  $E_q * r_s * (1 - p_1).$

4. hospitalization:

- $[S, \dots, IN_1, \dots, IH_L, \dots, D] \rightarrow [S, \dots, IN_1 - 1, \dots, IH_L + 1, \dots, D]$  at rate  $IN_1 * r_H * q_1.$
- $[S, \dots, IN_1, \dots, IH_S, \dots, D] \rightarrow [S, \dots, IN_1 - 1, \dots, IH_S + 1, \dots, D]$  at rate  $IN_1 * r_H * (1 - q_1).$
- $[S, \dots, IN_2, \dots, IH_L, \dots, D] \rightarrow [S, \dots, IN_2 - 1, \dots, IH_L + 1, \dots, D]$  at rate  $IN_2 * r_H * q_1.$
- $[S, \dots, IN_2, \dots, IH_S, \dots, D] \rightarrow [S, \dots, IN_2 - 1, \dots, IH_S + 1, \dots, D]$  at rate  $IN_2 * r_H * (1 - q_1).$

5. symptom relief:

- $[S, \dots, IH_L, IH_S, \dots, D] \rightarrow [S, \dots, IH_L + 1, IH_S - 1, \dots, D]$  at rate  $IH_S * r_b.$

6. recovery:

- $[S, \dots, A_1, \dots, R_A, R_N, R_H, D] \rightarrow [S, \dots, A_1 - 1, \dots, R_A + 1, R_N, R_H, D]$  at rate  $A_1 * \gamma_A$ .
- $[S, \dots, A_2, \dots, R_A, R_N, R_H, D] \rightarrow [S, \dots, A_2 - 1, \dots, R_A + 1, R_N, R_H, D]$  at rate  $A_2 * \gamma_A$ .
- $[S, \dots, IN_1, \dots, R_A, R_N, R_H, D] \rightarrow [S, \dots, IN_1 - 1, \dots, R_A, R_N + 1, R_H, D]$  at rate  $IN_1 * \gamma_{IN}$ .
- $[S, \dots, IN_2, \dots, R_A, R_N, R_H, D] \rightarrow [S, \dots, IN_2 - 1, \dots, R_A, R_N + 1, R_H, D]$  at rate  $IN_2 * \gamma_{IN}$ .
- $[S, \dots, IH_L, \dots, R_A, R_N, R_H, D] \rightarrow [S, \dots, IH_L - 1, \dots, R_A, R_N, R_H + 1, D]$  at rate  $IH_L * \gamma_{IH}$ .

7. death:

- $[S, \dots, IN_1, \dots, D] \rightarrow [S, \dots, IN_1 - 1, \dots, D + 1]$  at rate  $IN_1 * \delta_{IN}$ .
- $[S, \dots, IN_2, \dots, D] \rightarrow [S, \dots, IN_2 - 1, \dots, D + 1]$  at rate  $IN_2 * \delta_{IN}$ .
- $[S, \dots, IH_S, \dots, D] \rightarrow [S, \dots, IH_S - 1, \dots, D + 1]$  at rate  $IH_S * \delta_{IH}$ .

## References

- Athreya, K. B. and Ney, P. E. (1972). Branching processes. Springer-Verlag, New York-Heidelberg. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- Bailey, N. T. J. (1975). The mathematical theory of infectious diseases and its applications. Hafner Press [Macmillan Publishing Co., Inc.] New York, second edition.
- Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences, 106(51):21484–21489.
- Barbour, A. D. (1972). The principle of the diffusion of arbitrary constants. J. Appl. Probability, 9:519–541.
- Beijing Municipal Health Commission (2020). Situation report (in chinese). Available at [http://wjw.beijing.gov.cn/xwzx\\_20031/xwfb/202003/t20200305\\_1679143.html](http://wjw.beijing.gov.cn/xwzx_20031/xwfb/202003/t20200305_1679143.html).
- Berrhazi, B.-e., El Fatini, M., Laaribi, A., Pettersson, R., and Taki, R. (2017). A stochastic SIRS epidemic model incorporating media coverage and driven by Lévy noise. Chaos Solitons Fractals, 105:60–68.

- China National Bureau of Statistics (2018). Annual data by province (in chinese). Available at <http://data.stats.gov.cn/easyquery.htm?cn=E0103&zb=A0301&reg=440000&sj=2018>.
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., and Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science.
- Chongqing Municipal Health Commission (2020). Situation report (in chinese). Available at <http://wsjkw.cq.gov.cn/yqxxxyqtb/20200221/255637.html>.
- Cléménçon, S., Tran, V. C., and de Arazoza, H. (2008). A stochastic SIR model with contact-tracing: large population limits and statistical inference. J. Biol. Dyn., 2(4):392–414.
- Da Prato, G. and Zabczyk, J. (2014). Stochastic equations in infinite dimensions, volume 152 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, second edition.
- Durrett, R. (1988). Lecture notes on particle systems and percolation. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- Ele laboratory (2020). Computer simulation programs tell you why it’s not time to go out now. Available at [https://www.bilibili.com/video/av86478875/?spm\\_id\\_from=333.788.b\\_7265636f5f6c697374.2](https://www.bilibili.com/video/av86478875/?spm_id_from=333.788.b_7265636f5f6c697374.2).
- Ethier, S. N. and Kurtz, T. G. (1986). Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. Characterization and convergence.
- Government of the People’s Republic of China (2020). Notice of the pneumonia outbreak prevention and control command of new coronary virus infection in wuhan (in chinese). Available at [http://www.gov.cn/xinwen/2020-01/23/content\\_5471751.htm](http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm).
- Health Commission of Guangdong Province (2020). Situation report on new coronavirus pneumonia outbreak in guangdong province (in chinese). Available at [http://wsjkw.gd.gov.cn/zwyw\\_yqxx/content/post\\_2903465.html](http://wsjkw.gd.gov.cn/zwyw_yqxx/content/post_2903465.html).
- Health Commission of Hunan Province (2020). Situation report on new coronavirus pneumonia outbreak in hunan province (in chinese). Available at [http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw\\_1/202002/t20200221\\_11187516.html](http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw_1/202002/t20200221_11187516.html).

- Health Commission of Zhejiang Province (2020). Situation report on new coronavirus pneumonia outbreak in guangdong province (in chinese). Available at [http://www.zjwjw.gov.cn/art/2020/2/21/art\\_1202101\\_41958074.html](http://www.zjwjw.gov.cn/art/2020/2/21/art_1202101_41958074.html).
- Information Office of Hubei Provincial People’s Government (2020). ”prevention and control of pneumonia outbreak of new coronary virus infection” press conference (in chinese). Available at [https://www.hubei.gov.cn/hbfb/xwfbh/202002/t20200210\\_2023490.shtml](https://www.hubei.gov.cn/hbfb/xwfbh/202002/t20200210_2023490.shtml).
- Japanese Ministry of Health, Labour and Welfare (2020). About the new-style coronavirus infectious disease which was checked in the cruise ship which is being quarantined at yokohama port (in japanese). Available at [https://www.mhlw.go.jp/stf/newpage\\_09668.html](https://www.mhlw.go.jp/stf/newpage_09668.html).
- Ji, C. and Jiang, D. (2014). Threshold behaviour of a stochastic SIR model. Appl. Math. Model., 38(21-22):5067–5079.
- Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. IV, pages 149–165. University of California Press, Berkeley and Los Angeles.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772):700–721.
- King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S. P., Ferrari, M. J., Kendall, B. E., Lavine, M., Nguyen, D., Reuman, D. C., Wearing, H., and Wood, S. N. (2020). pomp: Statistical Inference for Partially Observed Markov Processes. R package, version 2.7.
- King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. Journal of Statistical Software, 69(12):1–43.
- Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. Journal of applied Probability, 7(1):49–58.
- Kurtz, T. G. (1971). Limit theorems for sequences of jump markov processes approximating ordinary differential processes. Journal of Applied Probability, 8(2):344–356.
- Liggett, T. M. (1985). Interacting particle systems, volume 276 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York.



Liu, T., Hu, J., Xiao, J., He, G., Kang, M., Rong, Z., Lin, L., Zhong, H., Huang, Q., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Gong, D., Wan, D., Chen, S., Guo, L., Li, Y., Sun, L., Liang, W., Song, T., He, J., and Ma, W. (2020a). Time-varying transmission dynamics of novel coronavirus pneumonia in china. [bioRxiv](https://doi.org/10.1101/2020.03.10.20033803).

Liu, Y., Gayle, A. A., Wilder-Smith, A., and Rocklöv, J. (2020b). The reproductive number of covid-19 is higher compared to sars coronavirus. [Journal of travel medicine](https://doi.org/10.1016/j.jtm.2020.03.002).

National Health Commission of the People's Republic of China (2020a). Update on the outbreak of new coronavirus pneumonia as of 24 hours on 12 february (in chinese). Available at <http://www.nhc.gov.cn/xcs/yqtb/202002/26fb16805f024382bff1de80c918368f.shtml>.

National Health Commission of the People's Republic of China (2020b). Update on the outbreak of new coronavirus pneumonia as of 24 hours on 4 march (in chinese). Available at <http://www.nhc.gov.cn/xcs/yqtb/202003/6e02433f7786402d8ea162fb5b80b5a0.shtml>.

Peng, L., Yang, W., Zhang, D., Zhuge, C., and Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. [arXiv Preprint: 2002.06563](https://arxiv.org/abs/2002.06563).

People's Government of Zhejiang Province (2020). Notice of new coronavirus pneumonia in zhejiang province on march 3, 2020 (in chinese). Available at [http://www.zj.gov.cn/art/2020/3/3/art\\_1228996608\\_42060522.html](http://www.zj.gov.cn/art/2020/3/3/art_1228996608_42060522.html).

Shanghai Municipal Health Commission (2020). Situation report (in chinese). Available at <http://wsjkw.sh.gov.cn/xwfb/20200222/0a10b6df11c845368af2d627d9551ed1.html>.

State Council Information Office of the People's Republic of China (2020). State council information office holds press conference on joint prevention and control of pneumonia outbreak with new coronavirus infection(in chinese). Available at <http://www.scio.gov.cn/xwfbh/xwfbh/wqfbh/42311/42478/index.htm>.

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., and Wu, J. (2020). Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. [Journal of Clinical Medicine](https://doi.org/10.1016/j.jcm.2020.03.002), 9(2):462.

The People's Government of Zhejiang Province (2020). The office of the leading group for the prevention and control of new coronavirus infection in zhejiang province issued a notice (provisional) on the prevention and control of new coronavirus infection in rural areas of zhejiang province (in chinese). Available at [http://www.zj.gov.cn/art/2020/2/10/art\\_1228996604\\_41898059.html](http://www.zj.gov.cn/art/2020/2/10/art_1228996604_41898059.html).

- Wilkinson, R. R., Ball, F. G., and Sharkey, K. J. (2017). The relationships between message passing, pairwise, Kermack-McKendrick and stochastic SIR epidemic models. J. Math. Biol., 75(6-7):1563–1590.
- Wu, J. T., Leung, K., and Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. Lancet, doi:10.1016/S0140-6736(20)30260-9.
- Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., and He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease, doi: 10.21037/jtd.2020.02.64.
- You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou, F., Pang, C. H., Zhang, Y., Chen, Z., and Zhou, X.-H. (2020). Estimation of the time-varying reproduction number of covid-19 outbreak in China. medRxiv: 10.1101/2020.02.08.20021253.
- Zhang, X. and Wang, K. (2013). Stochastic SIR model with jumps. Appl. Math. Lett., 26(8):867–874.
- Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., and Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. bioRxiv.