

**Supplement Information:**  
**Epidemic analysis of COVID-19 in China by dynamical modeling**

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

---

<sup>a</sup> These 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)

**CONTENTS**

1. Definitions of rates used in Figure 2	3
2. The influence of time delay	3
3. Difference formula of the generalized SEIR model	3
4. Global optimization method for parameter estimation	3
5. Shooting method for inverse inference	5
References	5

## 1. DEFINITIONS OF RATES USED IN FIGURE 2

In Table S1, we summarize the definitions and formulas of quantities calculated in Figure 2 in the main text.

## 2. THE INFLUENCE OF TIME DELAY

Since patients could be treated for several days until they become recovered, it is reasonable to hypothesize that the number of daily new recovered cases depends on the number of confirmed cases several days before, in other words the new recovered cases may be a function of a delay of daily confirmed cases. A similar argument may also hold for the death number as shown Eq. (1).

$$\frac{dR}{dt} = \lambda Q(t - \tau_1), \frac{dD}{dt} = \kappa Q(t - \tau_2). \quad (1)$$

To check the above hypothesis, we plot the ratio of the numbers of daily new recovered cases and deaths to the daily confirmed cases with different delays for 0 to 7 days (Fig. S1), which can be formulated into the following equations (2).

$$\bar{\lambda}_\tau(t) := \frac{h(t+1) - h(t)}{r(t-\tau)}, \bar{\kappa}_\tau(t) := \frac{d(t+1) - d(t)}{r(t-\tau)}, \quad (2)$$

where  $h(t)$ ,  $d(t)$  and  $r(t)$  stand for the real data of total recovered cases, total deaths and daily confirmed cases at time  $t$  respectively.

According to Fig. S1, the inclusion of different time delays could not apparently suppress the variations of recovery rate and death rate with time. So we employ a time-dependent  $\lambda$  and  $\kappa$  rather than time delay in the current study.

## 3. DIFFERENCE FORMULA OF THE GENERALIZED SEIR MODEL

Since the data are published by NHC of China daily, it is natural to adopt difference formula of the generalized SEIR model in numerical simulations, i.e.

$$\begin{cases} S(i+1) = S(i) - \beta S(i)I(i)/N - \alpha S(i) \\ E(i+1) = E(i) + \beta S(i)I(i)/N - \gamma E(i) \\ I(i+1) = I(i) + \gamma E(i) - \delta I(i) \\ R(i+1) = R(i) + \delta I(i) - \lambda(i)R(i) - \kappa(i)R(i) \\ H(i+1) = H(i) + \lambda(i)R(i) \\ D(i+1) = D(i) + \kappa(i)R(i) \\ P(i+1) = P(i) + \alpha S(i) \end{cases} \quad (3)$$

where  $i = 1, 2, \dots, n$  is the number of days.  $N$  is the total population of the region concerned.  $X(i)$  represents the number of population  $X$  on the date  $i$ , where  $X = S, P, E, I, R, H, D$  stands for the population in *susceptible*, *insusceptible*, *exposed* (infected by not infectious), *infectious* (not quarantined), *quarantined*, *recovered* and *death* cases respectively.

## 4. GLOBAL OPTIMIZATION METHOD FOR PARAMETER ESTIMATION

Here we search for the unknown parameters and initial conditions for the generalized SEIR model by global optimization. This is done automatically by implementing the simulated annealing algorithm and the nonlinear

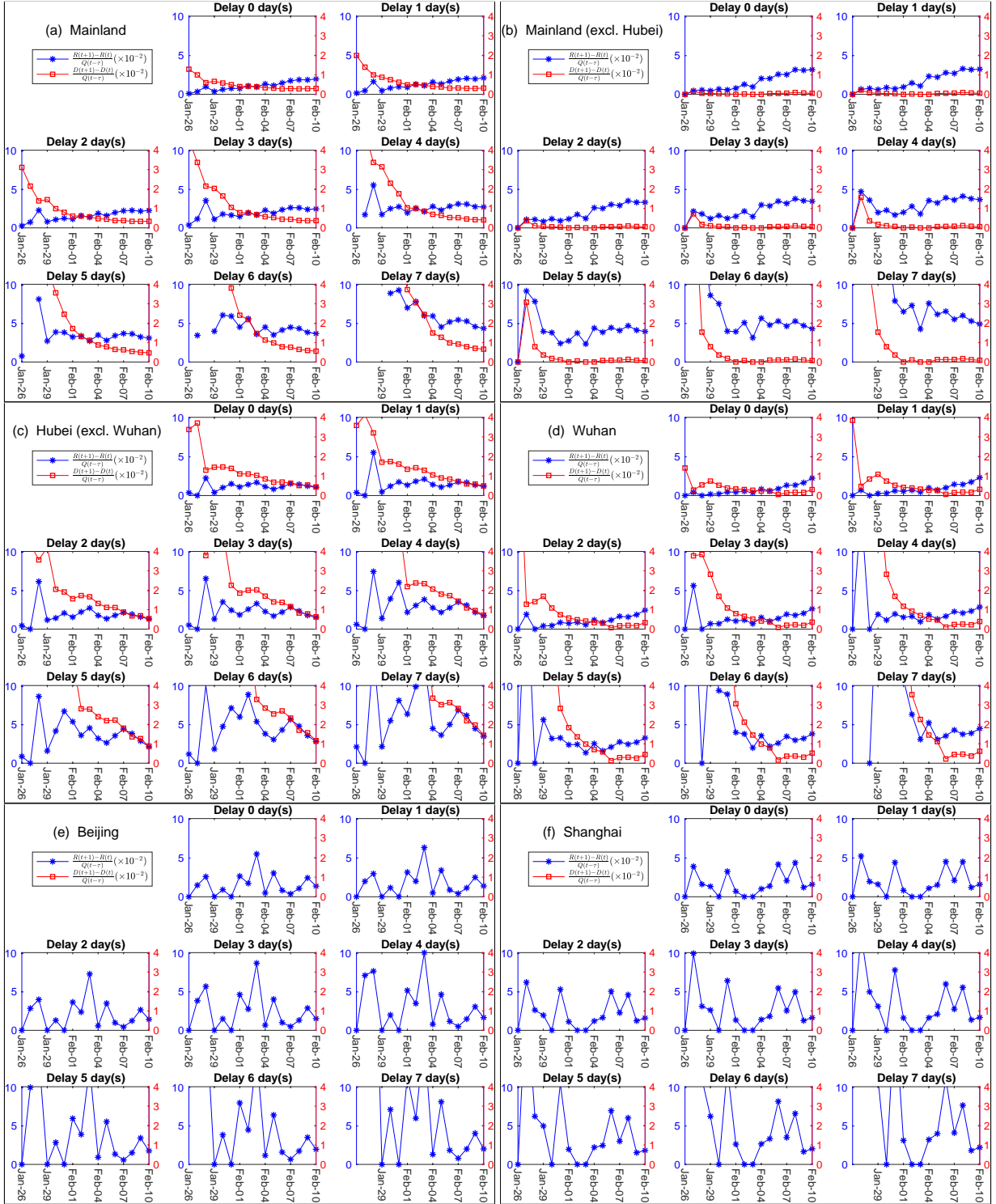


FIG. S1. Effect of time delay on the recovery rate and the mortality rate for various areas. The starred blue curves represent the ratio of new recovered cases to the confirmed cases under treatment from 0 to 7 days ago. The red squared curves represent the ratio of newly death cases to the confirmed cases. Data are all from official websites of National Health Commission of P. R. China and Health Commissions of provinces and cities involved in the figure. (a) Mainland, (b) Mainland (excl. Hubei), (c) Hubei Province (excl. Wuhan), (d) Wuhan, (e) Beijing and (f) Shanghai. The delays on the daily number of confirmed cases is shown by the title of each subplots. The mortality rates for Beijing and Shanghai are not shown due to the limited death cases reported in the two cities.

least square regression method. During successive iterations, the sum of the squared errors between the original data and the predicted values of the dynamical model is minimized, which finally gives the best parameters for data fitting. The code is available at <https://github.com/THU-ZCAM/SEIRHD-difference-model>.

## 5. SHOOTING METHOD FOR INVERSE INFERENCE

In order to determine the initial conditions  $\{E(0), I(0), R(0)\}$  and number of days  $T$  which correctly the desired pre-given values of  $\{E(T), I(T), R(T)\}$ , we adopt the shooting method. By generating random seeds (random values for  $\{E(0), I(0), R(0), T\}$ ) and performing forward iterations (rather than backward iterations) of the difference formula of the generalized SEIR model in (3) (parameters in the generalized SEIR model are kept the same as those in Table 1 in the main text), the shooting method transforms the boundary-value problem into an much easily solved initial-value problem. Now we can again adopt the algorithm of simulated annealing to minimize the gap between the predicted values of  $\{E(T)^*, I(T)^*, R(T)^*\}$  and their actual values evaluated through some properly constructed convex loss function. After successive iterations, the loss function stops decreasing and the correct initial conditions  $\{E(0), I(0), R(0)\}$  and number of days  $T$  are determined. The code is available at <https://github.com/THU-ZCAM/SEIRHD-difference-model>.

---

TABLE S1. Definitions of quantities calculated in Figure 2 in the main text.

Name	Symbol	description	defintion by variables
Cure rate	$\lambda(t)$	ratio of daily new recovered cases to currently confirmed cases	$\frac{R(t+1) - R(t)}{Q(t)}$
Mortality rate	$\kappa(t)$	ratio of daily new deaths to daily confirmed cases	$\frac{D(t+1) - D(t)}{Q(t)}$
Average contacts		ratio of daily new close contacts to daily new confirmed cases	
Average contacts*		ratio of cumulatively close contacts to cumulatively confirmed cases	
Severe condition rate		the ratio of daily cases severe condition to daily confirmed cases	