



Short Communication

Global prediction system for COVID-19 pandemic

Jianping Huang^{a,*}, Li Zhang^a, Xiaoyue Liu^a, Yun Wei^a, Chuwei Liu^a, Xinbo Lian^a, Zhongwei Huang^a, Jifan Chou^a, Xingrong Liu^b, Xun Li^c, Kehu Yang^d, Jinguo Wang^e, Hongbin Liang^a, Qianqing Gu^a, Pengyue Du^a, Tinghan Zhang^a

^a Collaborative Innovation Center for Western Ecological Safety, Lanzhou University, Lanzhou 730000, China

^b School of Public Health, Lanzhou University, Lanzhou 730000, China

^c The First Clinical Medical School, Lanzhou University, Lanzhou 730000, China

^d School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

^e School of Politics and International Relations, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:

Received 18 May 2020

Received in revised form 24 June 2020

Accepted 28 June 2020

Available online 2 August 2020

© 2020 Science China Press. Published by Elsevier B.V. and Science China Press. All rights reserved.

The outbreak of a novel coronavirus (SARS-CoV-2) has resulted in a worldwide pandemic infecting over 5.9 million people [1]. This positive-strand RNA virus can cause severe respiratory distress syndrome in humans (COVID-19), with over 364,000 deaths between December 2019 and May 30, 2020 [1,2]. To combat this pandemic, the World Health Organization (WHO) is coordinating global efforts on surveillance, epidemiology, mathematical modeling, diagnostics, treatment and control, and has issued interim guidance to countries. Nevertheless, this is a difficult situation and the number of cases is rapidly increasing globally. The temporal evolution and the spatial spread of this virus have also raised serious concerns about the future trajectory of this outbreak.

In the current epidemic period, when the number of reported COVID-19 cases is growing quasi-exponentially, nowcasting and forecasting are essential for public health planning and control [3–6]. Due to high population connectivity, there is a high risk that a localized outbreak may evolve into a global pandemic. Therefore, a system for global pandemic prediction is urgently needed to provide important scientific data to the WHO and local governments to help with public decision-making and allocation of medical resources.

An important feature of modern epidemiological responses involves the use of all available data to provide real-time response information [7]. Although it is difficult to establish an accurate epidemiological model describing the spread of a pandemic, real global pandemic data contain particular solutions to the mathematical equations incorporated in epidemiological models. Therefore, it is theoretically possible to improve the credibility of

prior epidemiological models by introducing the latest pandemic data. In addition, there is spatiotemporal heterogeneity in the occurrence of COVID-19, which may be related to the meteorological conditions and intervention measures implemented by local governments in different regions of the world [8]. Meteorological factors influence the spread of many diseases, and temperature and relative humidity may interact with the incidence of COVID-19. Hence, it is necessary to establish a continental or even global early warning system for epidemics that incorporate weather prediction and climate analysis as independent variables to improve the overall accuracy of the prediction [8,9]. Many countries also implemented epidemic prevention and control measures during the outbreak, and it is important to consider the role of national interventions in the spread of COVID-19. These factors are important in revising and improving prediction models [5,6,10]. Currently, however, no studies have considered the influence of meteorological factors and interventions on global COVID-19 prediction models.

In this article, we integrated the epidemic prediction model with real global pandemic data and considered the influence of environmental factors (temperature and humidity) as well as the implementation of control measures, to establish our own global prediction system, which is the first of global COVID-19 prediction system. This system predicts COVID-19 country by country and day by day over globe. The results of predictions using this system are now available online (<http://covid-19.lzu.edu.cn/>).

Our prediction system is a modified epidemiological susceptible–infectious–recovery (SIR) model [6,11] that incorporates real global pandemic data, meteorological factors, and quantified quarantine measures [9]. In this model, it is assumed that the total population (N) in the region remains unchanged during the outbreak,

* Corresponding author.

E-mail address: hjp@lzu.edu.cn (J. Huang).

COVID-19 is only spread via human-to-human infection, and there is no difference in immunity among individuals. The total population of each country ($N = S + I + R$) is divided into three categories, namely S (susceptible), I (infected), and R (recovered + dead). The SIR model is described by [1–2,7]

$$dS/dt = -\frac{r\beta IS}{N}, \tag{1}$$

$$dI/dt = r\beta IS/N - \mu I, \tag{2}$$

$$dR/dt = \mu I. \tag{3}$$

Based on the SIR model, we developed a model that includes the potential influence of temperature, humidity, urban population density, and intensity of control measures on COVID-19 infection. Our model is defined by the following equations:

$$dI/dt = r\beta I/N - \varepsilon I^2 - \eta IR, \tag{4}$$

$$dR/dt = \mu I, \tag{5}$$

where $\beta = \beta_0 + \beta_1 F(T_{2m}) + \beta_2 F(RH_{2m})$ and $\varepsilon = \varepsilon_0 + \varepsilon_c F(NO_2)$. $F(T_{2m})$ and $F(RH_{2m})$ are the probability distribution functions (PDFs) of local temperature and relative humidity at 2 m above ground level obtained by Huang et al. [9], who found that 60.0% of confirmed COVID-19 cases occurred in places where the air temperature ranged from 5 to 15 °C. The parameters (β_0, β_1 , and β_2) are obtained using non-linear fitting. $F(NO_2)$, the anomaly of the local NO_2 concentration, was introduced to quantify the effectiveness of quarantine measures. We chose NO_2 because it is a major atmospheric pollutant from vehicle exhaust and is highly correlated with traffic volume. Thus, a large decline in the local NO_2 concentration is associated with reduced traffic flows and is correlated with the efficiency of the quarantine measures.

With over 5.9 million confirmed cases globally [1], the coefficients in Eqs. (4) and (5) (β, μ , and η) are inversed [12] and calibrated using pandemic, meteorological, and quantified quarantine data that are updated in real-time for each country or region. Thus, this global prediction system incorporates various modules instead of using a simple model. The residual term (Q), which is predicted using the EEMD-ARMA (Ensemble Empirical Mode Decomposition-Autoregressive moving Average Model) method [13], was introduced to help correct the bias between

the hindcast and the actual reported data. Because the parameters included in this system are inversed from the latest data repository, they self-adapt to the reported cases for each region.

Fig. 1 shows the predicted number of new cases of COVID-19 per day on June 20, 2020. Implementation of the quarantine measures and the rise in temperature are expected to restrain the spread of the virus in the northern hemisphere to some extent, and the number of new cases is expected to decline. However, with gradual lifting of the blockade in the United States of America, it is predicted that the number of new cases per day will remain above 20,000. The outbreak in European countries will be contained initially and the number of newly confirmed cases will decline. Regions such as India and Russia will still be in the outbreak stage, in which the number of new cases per day exceeds 10,000. In the southern hemisphere, however, the declining temperature due to the switch in season to autumn and the loose quarantine measures will provide favorable conditions for the virus to spread. It is foreseeable that the epidemics in Brazil and other southern hemisphere countries will deteriorate. Effective quarantine measures are urgently required as a precaution against further outbreak to save lives before it is too late.

Fig. 2 compares the reported number of confirmed cases between January 22 and May 14, 2020, with the results of our simulation for six countries (USA, Italy, UK, Russia, Saudi Arabia, and Brazil). The figure also shows a 10-day hindcast for May 15 to May 24, 2020. Our system captures the systematic variation in number of new cases per day, and reproduces the epidemic curves for Italy before and after the peaks. The system also successfully emulates the epidemic curves for Saudi Arabia and Brazil, where the epidemic is still in its rapidly increasing phase, and the USA, Russia, and UK, where the curve is oscillating. For countries in the post-peak stages, nationwide alerts and quarantine methods are still necessary as a precaution against potential recurrent outbreaks.

The COVID-19 pandemic is one of the most severe global crises since the Second World War, with a devastating impact on human health and the global economy. Over 4 billion people worldwide have been affected by severe restrictions on their movements and social relationships [14]. Governments, social policies, and health systems around the world need to be prepared for the pandemic to transform passive prevention into active prevention. In this digital, globalized world, new data and information on the

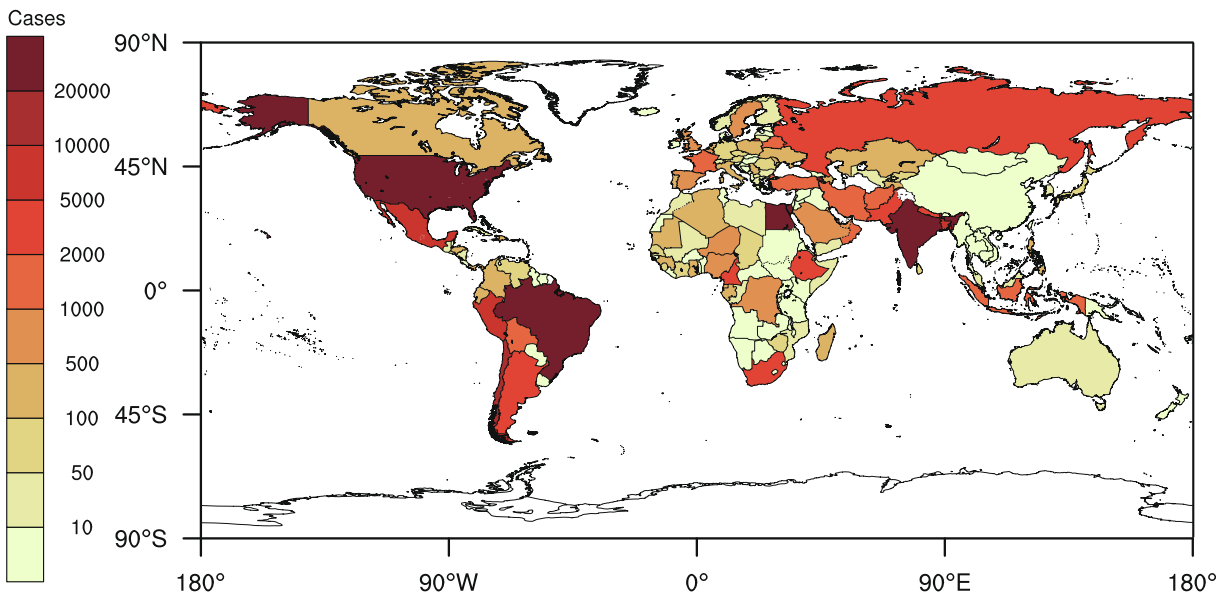


Fig. 1. Global prediction of new cases of COVID-19 per day on June 20, 2020.

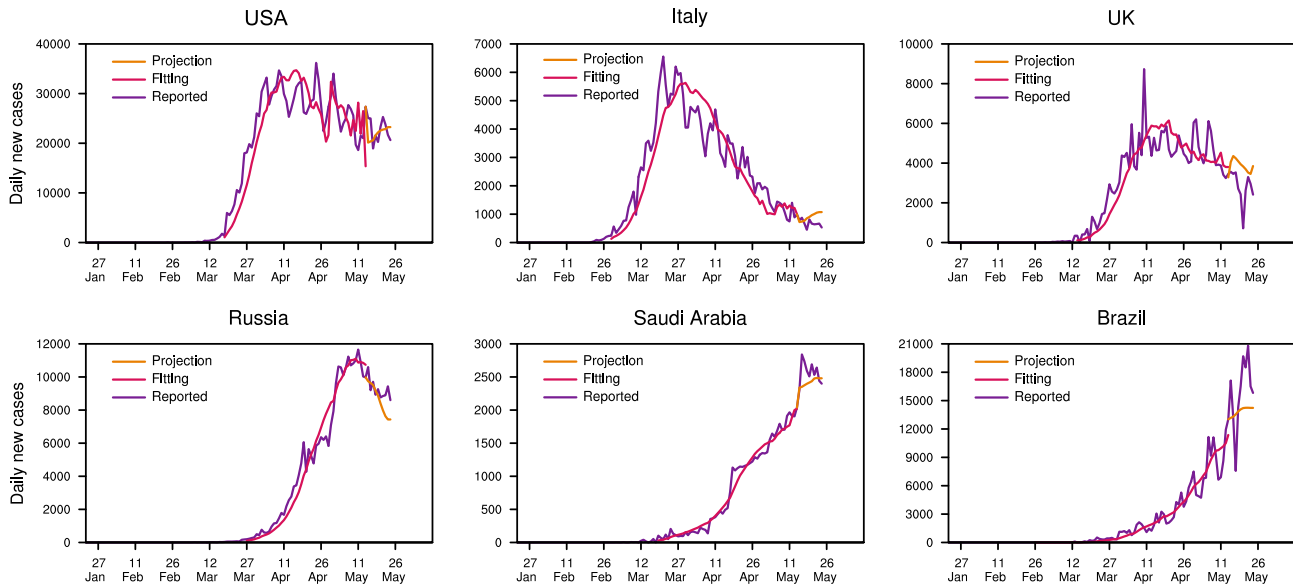


Fig. 2. Time series of the number of new cases per day in six countries (USA, Italy, UK, Russia, Saudi Arabia, and Brazil) between January 22 and May 24, 2020. The numbers of reported confirmed cases for the same period are shown by violet lines, whereas the simulated cases (January 22–May 14, 2020) and hindcast cases (May 15–24, 2020) are shown by dark pink and orange lines, respectively.

evolution of the COVID-19 epidemic are rapidly updated. In this article, we describe a more realistic and practical global prediction system that incorporates large amounts of data. Except for Africa and some regions, our predictions are consistent with the total number of confirmed cases worldwide. Our predictions indicate that countries in the western hemisphere and Africa are high-risk regions, and countries with initial control of pandemic should remain alert for secondary outbreaks (e.g., China and South Korea). Nevertheless, because the prediction system is still at an initial stage of development, it still has some flaws that warrant further improvements. The quality control module of input data should be improved to reduce the system sensitivity to the initial condition.

The uncertainty of epidemic prediction is not only related to the prediction system being used, but also to local medical conditions, the degree of population aging, and other factors. Our prediction system may be more useful for assessing qualitative trends and evaluating intervention options than for accurately predicting the number of cases. Establishment of a COVID-19 global prediction system will help us to understand and mitigate the impact of this virus and future pandemics, and provide a valuable reference for policymakers. In the face of a global pandemic, we are all responsible for combating the virus, protecting vulnerable members of society, and ensuring that national security systems can respond to the pandemic and provide sufficient care for the population. Joint efforts of all countries and people worldwide will help us to overcome this epidemic and restore normal life as soon as possible.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (41521004) and the Gansu Provincial Special Fund Project for Guiding Scientific and Technological Innovation and Development (2019ZX-06). The authors acknowledge the Center for Systems Science and Engineering at Johns Hopkins University for providing the COVID-19 data, and we would like to

acknowledge NASA and ECMWF for making the OMNO2d and ERA-interim data publicly accessible.

Author contributions

Jianping Huang designed the study and contributed to the ideas, interpretation and manuscript writing. Li Zhang, Xiaoyue Liu, Yun Wei, Chuwei Liu and Xinbo Lian contributed to the data analysis, interpretation and manuscript writing. All of the authors contributed to the discussion and interpretation of the manuscript. All of the authors reviewed the manuscript.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at <https://doi.org/10.1016/j.scib.2020.08.002>.

References

- [1] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University 2020. <https://github.com/CSSEGISandData/COVID-19> (accessed May 30, 2020).
- [2] Yan R, Zhang Y, Li Y, et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367:1444–8.
- [3] Cori A, Donnelly CA, Dorigatti I, et al. Key data for outbreak evaluation: building on the Ebola experience. *Philos Trans R Soc B Biol Sci* 2017;372:20160371.
- [4] Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020;20:553–8.
- [5] Flaxman S, Mishra S, Gandy A, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. *Imp Coll London* 2020. <https://doi.org/10.25561/77731>.
- [6] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020;395:689–97.
- [7] Polonsky JA, Baidjoe A, Kamvar ZN, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos Trans R Soc B Biol Sci* 2019;374:20180276.
- [8] Qi H, Xiao S, Shi R, et al. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. *Sci Total Environ* 2020;728:138778.
- [9] Huang Z, Huang J, Gu Q, et al. Optimal temperature zone for the dispersal of COVID-19. *Sci Total Environ* 2020;736:139487.
- [10] Woolhouse M. How to make predictions about future infectious disease risks. *Philos Trans R Soc B Biol Sci* 2011;366:2045–54.

- [11] Giles P, Bailey NTJ. The mathematical theory of infectious diseases and its applications. *Oper Res Q* 1977;28:479.
- [12] Huang J, Yi Y. Inversion of a nonlinear dynamical model from the observation. *Sci China B Chem* 1991;34:1246–51.
- [13] Wang W, Chau K, Xu D, et al. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour Manag* 2015;29:2655–75.
- [14] Boccaletti S, Ditto W, Mindlin G, et al. Modeling and forecasting of epidemic spreading: the case of Covid-19 and beyond. *Chaos Solitons Fractals* 2020;135:110151.



Jianping Huang is a professor at College of Atmospheric Sciences and a director of Collaborative Innovation Center for Western Ecological Safety, Lanzhou University. He has long been dedicating to the study of long-term climate prediction, dust-cloud interaction and semi-arid climate change by combining field observations and theoretical study. Since the COVID-19 pandemic, he and his team have established a COVID-19 global prediction system with statistical-dynamic climate prediction methods.