

Regional Prediction of COVID-19 in the United States Based on the Difference Equation Model*

Ceyu Lei¹ and Xiaoling Han^{1,†}

Abstract The novel coronavirus pneumonia 2019 (COVID-19) has swept the globe in just a few months with negative social and psychological consequences for public health. So far, the United States has been one of the countries most affected by the epidemic. In this study, 51 states in the United States are divided into 10 state clusters according to relevant factors, and a difference equation model with spatio-temporal dynamic characteristics is established to predict the transmission dynamics of COVID-19 in the 10 state clusters and obtain data on regional aggregation levels (the United States). The study showed that the Pearson Correlation Coefficient between the actual data and the predicted data in the 10 state clusters is between 0.6 and 0.96 (mean $R^2=0.8448$), and the mean absolute error (MAE) of the newly confirmed cases in each cluster is between 300 and 1650 (mean MAE=878) and the average forecasting error rate (AFER) of the total confirmed cases in each cluster is between 0.9% and 3% (mean AFER=1.57%). These results show that the difference equation model can well predict the changes in the recent confirmed cases of infectious diseases such as COVID-19.

Keywords COVID-19, Prediction, Difference equation, Modeling, Mean absolute error.

MSC(2010) 35R02, 97M10.

1. Introduction

Since the first case appeared in Wuhan, China in December 2019, the COVID-19 has aroused people's attention. Since March 2020, the COVID-19 has rapidly spread around the world. As of December 2020, more than 200 countries and regions have been affected by the COVID-19. Among them, the United States, India, Brazil, Russia and France are the five countries affected by the pandemic most. In particular, the United States has become the center of the global pandemic with about 100,000 newly confirmed cases every day. The COVID-19 is disrupting the lives of people around the world in a variety of ways, and has a negative impact on global economic development. However, in the process of epidemic prevention and control, the prediction of COVID-19 is particularly important.

[†]the corresponding author.

Email address: 714327480@qq.com (C. Lei), hanxiaoling@nwnu.edu.cn (X. Han)

¹Department of Mathematics, Northwest Normal University, Lanzhou, Gansu 730070, China

*The authors were supported by National Natural Science Foundation of China (No. 11561063) and Natural Science Foundation of Gansu Province (No. 20JR10RA086).

With the development of the dynamic theory of infectious diseases, the methods and theories for studying infectious diseases are becoming more abundant, such as SI model, SIR model, SEIR model, partial differential equation model, machine learning technology and other methods. In the process of the continuous spread of COVID-19, many scholars in the world have used different methods to obtain abundant and classic research results [1, 2, 4–13, 15–20, 23, 27]. Tang et al., [21] established a random discrete epidemic model with case input, analyzed the effectiveness of China's Shaanxi Province epidemic prevention policy and conducted a predictive analysis of multiple outbreaks caused by economic recovery. Wang et al., [24] predicted the epidemic in Arizona, the United States by establishing a partial differential equation model with temporal and spatial factors, and analyzed the impact of human preventive measures on the reduction of COVID-19 cases. O. Torrealba-Rodriguez et al., [22] studied the outbreak in Mexico by using the Gompertz, Logistic and artificial neural network models, the results showed that these models had good predictive performance and the R^2 of each model was 0.9998, 0.9996 and 0.9999 respectively.

This work aims to explore the spread of COVID-19 between state clusters and to make further predictions of the changes in the epidemic. Specifically, we divided the United States as a whole into 10 different state clusters according to the relevant regulations of the United States, and abstractly divided the spread of COVID-19 into two processes: local spread and global spread. On this basis, we established a difference equation model with time and space factors is used to describe the spatiotemporal dynamic propagation process of COVID-19 and predict the development of the epidemic. The results show that the model not only has good predictive ability, but also provides a policy introduction to a certain extent, and provide a more scientific theoretical basis for controlling the development of the epidemic.

The rest of this article is structured as follows: In Section 2, the data sources are introduced in details, and the established model is described in details. Section 3 gives experiments and results analysis. Finally, the results of our research are discussed in Section 4.

2. Methods

2.1. Data set

We use the motif clustering algorithm in [3] to divide the United States into several clusters, and then use the difference equation model to predict the confirmed cases of each cluster of COVID-19. To make modeling and reporting more feasible and easier for the public to understand the situation in the United States according to the definition of the U.S. Department of Health and Human Services (HHS), 51 states in the United States are divided into 10 regions [25,26], as shown in Figure 1. The COVID-19 data repository adopted in this study is obtained from the World Health Organization (WHO) website (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>). In this study, the number of confirmed cases and the number of deaths in 51 states in the United States from 10 March, 2020 to 24 November, 2020 for a total of 260 days are used. We calculated the relevant data of 10 clusters based on the collected data of each state, as is shown in Figure 2.

The states contained in each cluster are as follows: