

# Prediction model of COVID-19 based on two dimensional partial differential equation model

Yinyin Liu,

Wuhan Puren Hospital, Wuhan, China

Submitted: 15-10-2021

Revised: 26-10-2021

Accepted: 28-10-2021

## ABSTRACT:

In this paper, the one-dimensional PDE model is extended to the two-dimensional model, so that the COVID-19 prediction model which is in line with the actual two-dimensional regional distribution can be established. Based on the prediction model, the undetermined parameters in the differential equation model are inverted by using the spontaneous perturbation stochastic gradient algorithm (SPSA) according to the actual number of infections, so as to accurately predict the number of infections in the future. The effectiveness of this model is also confirmed by the recent outbreak of local epidemic data in China.

**Keywords:** COVID-19prediction; partial differential equation; spontaneous perturbation stochastic gradient algorithm

## I. INTRODUCTION

COVID-19 has been spreading around the world since 2019. As of now, it has infected more than more than 200000000 people worldwide and nearly 5 million people have died. It is a very dangerous pandemic of world infectious diseases. Some infection prediction models of COVID-19were developed in order to help countries fight against COVID-19. The aim of these models was to predict the number of COVID-19 infections through various mathematical modeling methods.

Ordinary differential equation (ODE) models are often used to describe population growth and infection population in biological or ecological systems. So the ODE model was first used for the transmission of COVID-19. However, the model for COVID-19 infection is difficult to effectively depict the spatial distribution of COVID-19. Therefore, Wang and Yamamoto [1] proposed partial differential equation (PDE) model to predict the number of COVID-19 infection in different regions. However, the PDE model is only an equation for one-dimensional spatial coordinates, which needs to

project the two-dimensional region to one-dimensional coordinates in advance, and the rationality of this projection is difficult to ensure. Considering that the actual region is two-dimensional, it is in line with the actual situation to extend the model to the partial differential equation model of two-dimensional spatial coordinates. We know that based on the differential equation model, the undetermined parameters in the differential equation model are inversely determined by using the existing historical actual data, so as to predict the future situation, which involves the solution of the differential equation model. Due to the complexity of the equation form, the numerical method is generally used to solve the differential equations with different parameters. Finite difference method and finite element method are usually used. However, these methods are based on mesh generation in computing domain, so only two adjacent meshes have direct effects, but this is not reasonable for COVID-19infection, because the point to point traffic tools between non adjacent areas may be transmitted. Artificial intelligence methods such as artificial neural network may be an effective method to solve this problem, but because there are not enough data points and the model parameters fitted and trained may lack sufficient practical significance [2-6], the physical network model that can combine certain physical information often has better performance than the pure statistical learning method. The method based on differential equation model and combined with actual data to retrieve the undetermined parameters in the equation can be regarded as a physics-based data-driven model to a certain extent. This paper aims to continue this idea, extend the one-dimensional PDE model to two-dimensional, make the model conform to the actual two-dimensional regional distribution, and improve the effectiveness and practicability of the model.

## II. METHODOLOGY

### 2.1 partial differential equation model

Wang and Yamamoto proposed the PDE model as follows:

$$\frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x} \left( d(x) \frac{\partial C(x,t)}{\partial x} \right) + r(t)l(x)a(x,t-14)C(x,t) - C \frac{h(x,t-14)C(x,t)}{k + C(x,t)} \quad (1)$$

The specific meaning of each item in the Eq. (1) can be seen in reference [1]. The PDE model has indeed achieved good results, but because the model is a one-dimensional model, it is necessary to determine a reasonable order of the areas to be studied in advance, that is, the areas to be studied are projected onto one-dimensional European space, and

it is difficult to reasonably project the two-dimensional plane areas onto one-dimensional European space, At least we need to sort hard first, which brings obstacles to the practical application of the PDE model. Therefore, this paper aims to expand the above PDE model to a two-dimensional model in line with the actual regional distribution, namely:

$$\frac{\partial C(x,y,t)}{\partial t} = \nabla \cdot (d(x,y) \nabla C(x,y,t)) + r(t)l(x,y)a(x,y,t-14)C(x,y,t) - g \frac{h(x,y,t-14)C(x,y,t)}{k + C(x,y,t)} \quad (2)$$

In addition, it should also be noted that non adjacent regions (e.g. provinces) also have certain diffusion capacity, and the diffusion can be realized without passing through adjacent provinces. For example, flights between different provinces can realize virus transmission between non adjacent regions. Therefore, the traditional grid method cannot be used for the partial differential equation in Eq. (1) (e.g. finite difference method, finite volume method,

finite element method, etc.) The meshless method with physical network characteristics should be used to solve, because in the meshless method, the node influence domain can contain more nodes by setting different influence domain sizes, so as to characterize the virus diffusion and propagation between geometrically non adjacent nodes. That is, the PDE model in this paper is

$$\frac{\partial C(x,t)}{\partial t} = \nabla \cdot (d(x,y) \nabla C(x,y,t)) + r(t)l(x,y)a(x,y,t-14)C(x,y,t) - g \frac{h(x,y,t-14)C(x,y,t)}{k + C(x,y,t)} \quad (3)$$

In this paper, the finite volume method is used to effectively solve the two-dimensional PDE [7]. The specific discrete format is deduced as follows. The control volume V is obtained by integrating both ends of the equation at the same time:

$$\frac{\partial C(x,y,t)}{\partial t} = \nabla \cdot (d(x,y) \nabla C(x,y,t)) + r(t)l(x,y)a(x,y,t-14)C(x,y,t) - g \frac{h(x,y,t-14)C(x,y,t)}{k + C(x,y,t)} \quad (4)$$

Integrating the control volume and time at both ends of the above formula, the left end of the above formula obtains:

$$\int_{V_i} \int_t^{t+\Delta t} \frac{\partial C(x, y, t)}{\partial t} dt d\Omega = \int_{V_i} \int_t^{t+\Delta t} [C(x, y, t + \Delta t) - C(x, y, t)] d\Omega = \Delta t V_i [C(x_i, y_i, t + \Delta t) - C(x_i, y_i, t)] \quad (5)$$

For the first term at the right end of the above equation, the following is obtained by using Oko formula:

$$\int_t^{t+\Delta t} \int_{V_i} \nabla \cdot (d(x, y) \nabla C(x, y, t)) d\Omega dt = \int_t^{t+\Delta t} \int_{\partial\Omega} d(x, y) \nabla C(x, y, t) \cdot \bar{n} d\Gamma dt \quad (6)$$

The second term at the right end of Eq.(6) is obtained:

$$\int_t^{t+\Delta t} \int_{V_i} r(t) l(x, y) a(x, y, t - 14) C(x, y, t) d\Omega dt = \Delta t V_i r(t + \Delta t) l(x_i, y_i) a(x_i, y_i, t + \Delta t - 14) C(x_i, y_i, t + \Delta t) \quad (7)$$

The third item at the right end of the above equation is obtained:

$$\int_t^{t+\Delta t} \int_{V_i} g \frac{h(x, y, t - 14) C(x, y, t)}{k + C(x, y, t)} d\Omega dt = \frac{\Delta t V_i g h(x_i, y_i, t + \Delta t - 14) C(x_i, y_i, t + \Delta t)}{k + C(x_i, y_i, t + \Delta t)} \quad (8)$$

It can be obtained that the fully implicit discrete scheme is as follows:

$$\Delta t V_i [C(x_i, y_i, t + \Delta t) - C(x_i, y_i, t)] = \int_t^{t+\Delta t} \int_{\partial\Omega} d(x, y) \nabla C(x, y, t) \cdot \bar{n} d\Gamma dt + \Delta t V_i r(t + \Delta t) l(x_i, y_i) a(x_i, y_i, t + \Delta t - 14) C(x_i, y_i, t + \Delta t) + \frac{\Delta t V_i g h(x_i, y_i, t + \Delta t - 14) C(x_i, y_i, t + \Delta t)}{k + C(x_i, y_i, t + \Delta t)} \quad (9)$$

When the parameters in Eq. (2) are determined, the differential equation can be solved by Eq. (9). It should be pointed out that since the number of infected people is an integer, we calculate the calculated number of infected people by rounding.

## 2.2 training data and model prediction method

This paper needs to fit the actual data to

$$O(\bar{m}) = \|g(\bar{m}) - \bar{n}\| \quad (10)$$

where  $g(m)$  represents the differential equation model in Section 2.1,  $\bar{m}$  is the vector of undetermined parameters in the differential equation, and  $\bar{n}$  is the observed local cumulative number of

determine the unknown parameters in Eq. (2). In order to reduce the parameters in the model training process and reduce the multi solution of model inversion, we know that this fitting problem is actually an optimization problem. Therefore, this paper uses the spontaneous perturbation stochastic gradient algorithm to solve this optimization [8]. The specific solution is: First, consider defining the optimization objective function,

infections in each province.

In the optimization process, the undetermined parameters in the differential equation will meet the limiting conditions:

$$m^{low} \leq m \leq m^{up} \tag{11}$$

Where  $m^{low}$  is the vector of lower limits of the model parameters,  $m^{up}$  is the vector of upper limits of the model parameters.

The simultaneous perturbation stochastic approximation (SPSA) algorithm is used to optimize the objective function in Eq. (11). Then, based on SPSA, suppose it is at the kth iteration, through multiple simultaneous perturbations of  $m^k$ , then the gradient of  $O(m^k)$  is approximated as

$$\bar{g}(m^k) = \frac{1}{n} \sum_{i=1}^n g_i(m^k) = \sum_{i=1}^n \frac{O(m^k + c_k \Delta_i) - O(m^k - c_k \Delta_i)}{2c_k} \Delta_i^{-1} \tag{12}$$

Therefore, the model parameters at the k+1th step  $m^{k+1}$  can be updated as

$$m^{k+1} = m^k - \alpha_k \bar{g}_k(m^k) \tag{13}$$

where  $\alpha_k$  denotes the kth iteration step size. Until the maximum number of iterative steps is reached or the convergence condition is given.

### III. A PRACTICAL APPLICATION

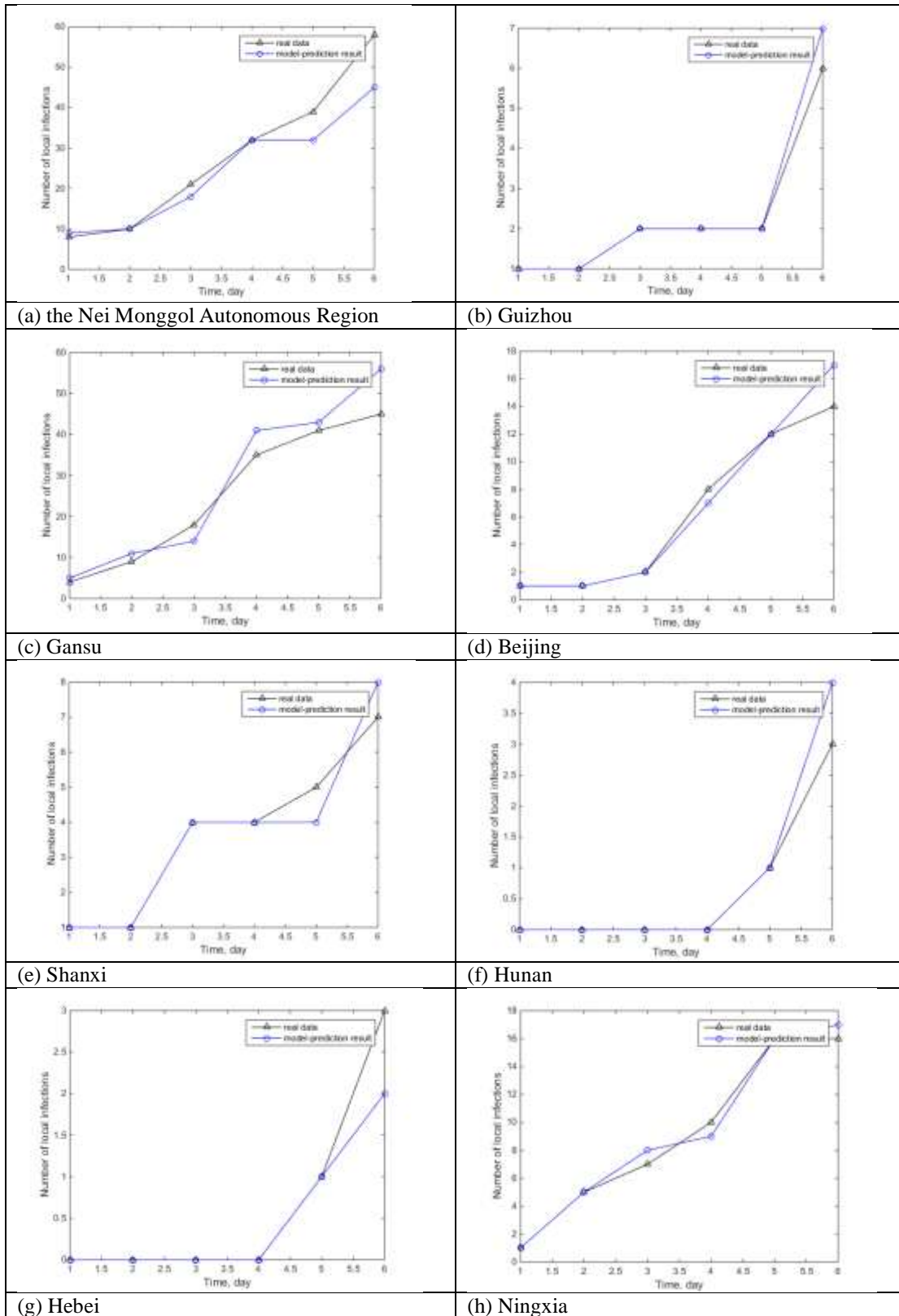
The training data selected in this paper use the relevant data of local outbreaks in Gansu, Shaanxi and Inner Mongolia. The specific data are as follows:

Table. 1 the data of infections during the local outbreak in Northwest China

Cities	New confirmed population of COVID-19 (local)					
	October 19th	October 20th	October 21th	October 22th	October 23th	October 24th
Inner Mongolia Autonomous Region	8	2	11	11	7	19
Guizhou	1	0	1	0	0	4
Gansu	4	5	9	17	6	4
Beijing	1	0	1	6	4	2
Shanxi	1	0	3	0	1	2
Hunan	0	0	0	0	1	2
Hubei	0	0	0	0	1	2
Ningxia	1	4	2	3	6	0
Yunnan	1	0	0	1	0	0
Qinghai	0	0	1	0	0	0
Hubei	0	2	0	0	0	0

The historical data of the first five days in Table 1 are used to train the differential equation model and predict the cumulative number of confirmed patients on the sixth day. The fitting and prediction results are shown in Figure 1. It can be

seen that the fitting accuracy of the number of historical infections is high, the average error is less than 10%, and the predicted number of infections on the sixth day is not much different from the actual data, and the error is also less than 10%.



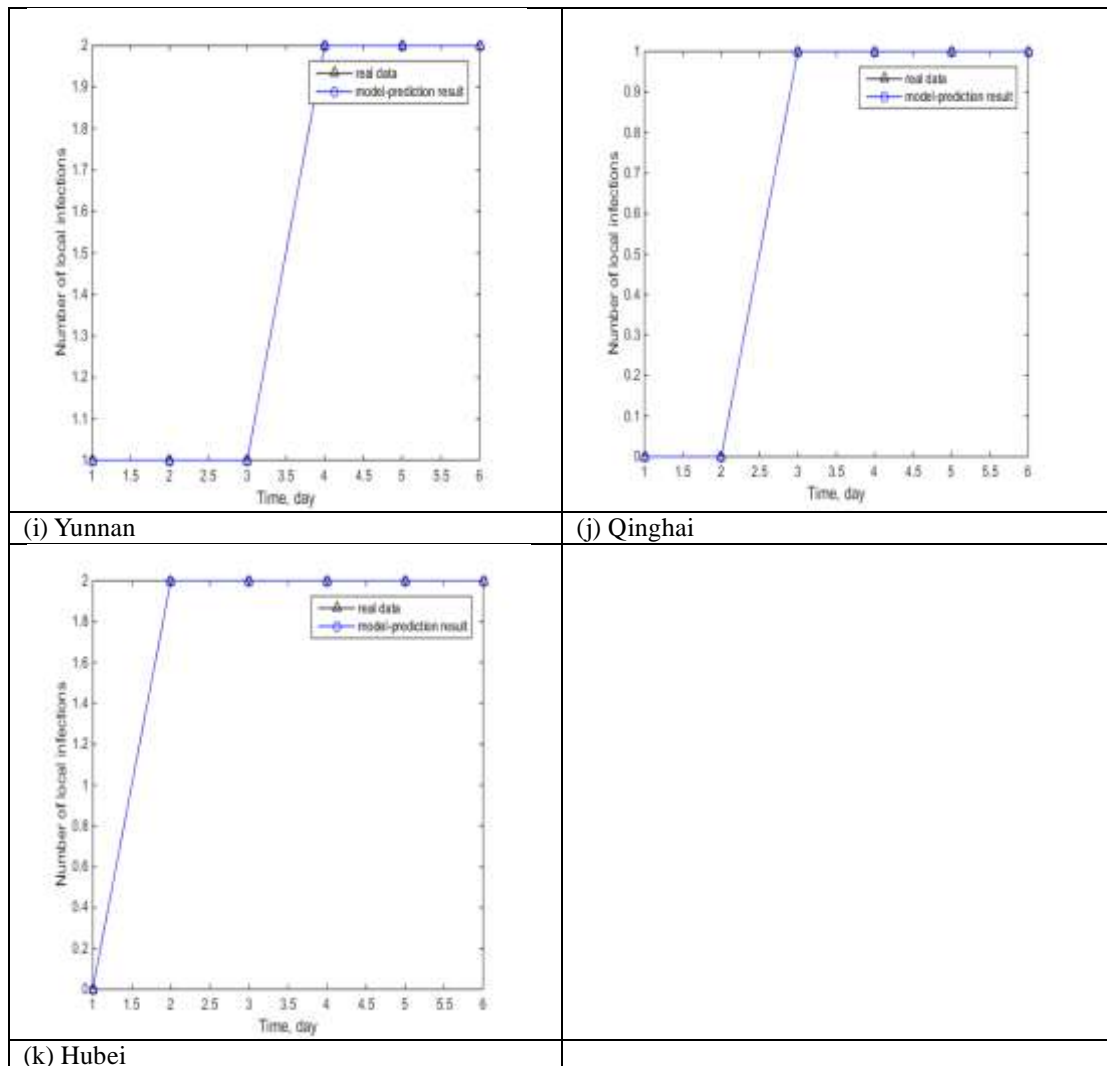


Fig. 1 comparisons of model-prediction results and real data of some different provinces

#### IV. CONCLUSION

This paper presents a novel COVID-19 prediction model based on two-dimensional partial differential equation (PDE) model. The model was fitted with training data from recent outbreak in Northwest China, and the parameters of the differential equation model were determined to predict the number of COVID-19 infection. The actual calculation results also verified the validity of the model.

In this paper, the one-dimensional PDE model is extended to the two-dimensional model, so that the COVID-19 prediction model which is in line with the actual two-dimensional regional distribution can be established. Based on the prediction model, the undetermined parameters in the differential equation model are inverted by using the synchronous perturbation random gradient algorithm (SPSA) according to the actual number of infections,

so as to accurately predict the number of infections in the future. The effectiveness of this model is also confirmed by the recent outbreak of local epidemic data in China.

#### REFERENCES

- [1] Wang Haiyan, Yamamoto Nao. Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. *Mathematical Biosciences and Engineering*. 17(5): 4891-4904.
- [2] S. He, S. Tang, L. Rong, A discrete stochastic model of the COVID-19 outbreak: Forecast and control, *Math. Biosci. Eng.*, 17 (2020), 2792–2804.
- [3] Y. Huang, L. Yang, H. Dai, F. Tian, K. Chen, Epidemic situation and forecasting of COVID-19 in and outside China, *Bull. World Health Organ.*, (2020). E-pub: 16 March 2020.

- [4] L. Wang, J. Wang, H. Zhao, Y. Shi, K. Wang, P. Wu, et al., Modelling and assessing the effects of medical resources on transmission of novel coronavirus (COVID-19) in Wuhan, China, *Math. Biosci. Eng.*, 17 (2020), 2936–2949.
- [5] Fma, B , et al. "Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran - ScienceDirect." *Biomedical Journal* (2021).
- [6] Ahmad, I. , and S. M. Asad . "Predictions of coronavirus COVID-19 distinct cases in Pakistan through artificial neural network." *Epidemiology and Infection* 148(2020):1-17.
- [7] asak, H. . "Error Analysis and Estimation for the Finite Volume Method With Applications to Fluid Flows." *Imperial College London m.8(1996):A385*.
- [8] Spall, J. C. . "A one-measurement form of simultaneous perturbation stochastic approximation." *Automatica* 33.1(1997):109-112.