



# Forecasting the Trend of COVID-19 Considering the Impacts of Public Health Interventions: An Application of FGM and Buffer Level

Kai Lisa Lo<sup>1</sup> · Minglei Zhang<sup>1</sup> · Yanhui Chen<sup>1</sup> · Jinhong Jackson Mi<sup>1</sup>

Received: 3 February 2021 / Revised: 23 June 2021 / Accepted: 15 July 2021 /

Published online: 7 September 2021

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

**Purpose** COVID-19 is still showing a tendency of spreading around the world. In order to improve the subsequent control of COVID-19, it is essential to conduct a study on measuring and predicting the scale of the outbreak in the future.

**Methods** This paper uses rolling mechanism and grid search to find the best fractional order of Fractional Order Accumulation Grey Model (FGM). Buffer level is proposed based on the general form of weakening buffer operator to measure the effect of government control measurements on the epidemic. And the buffer level is associated with the Government Response Stringency index and the Mobility Index.

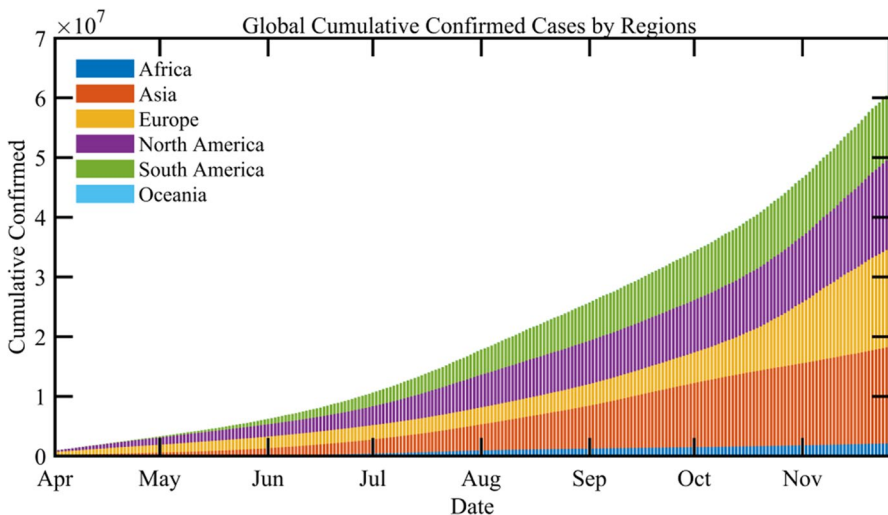
**Results** Firstly, the model proposed in this paper dominates the ARIMA model which has been widely used in predicting the confirmed COVID-19 cases. Secondly, in the process of using the buffer level to modify the FGM, this paper finds that government measurements require the active cooperation of the public and often have a time lag when they are effective. Only when government increase its stringency and the public observe the order can the spread of COVID-19 be slowed down. If there is only the controlling measure and the public does not react actively, it will not slow down the epidemic. Thirdly, according to the Mobility Index and Government Response Stringency Index in December, this paper predicts the cumulative confirmed cases of the end of January in different scenarios according to different buffer levels. The study suggests that the world should continue to maintain high vigilance and take corresponding control measures for the outbreak of COVID-19.

**Conclusions** Government's control measures and public's abidance are both important in this battle with COVID-19. Governments control measures have time-lag effect and the time lag is about 9 days. When the government increases its stringency and the public cooperates with the government, we must consider the weaken buffer operator with proper buffer level in the prediction process. These prediction methods can be considered in the prediction of COVID-19 confirmed cases in the future or the trend of other epidemics.

**Keywords** COVID-19 · Grey system · Fractional order accumulation grey model (FGM) · Buffer level

## 1 Introduction

COVID-19 caused by SARs-CoV-2 continues to spread around the world today, and the earliest appearance of COVID-19 symptoms can be traced back to December 1, 2019 [12]. On January 30, 2020, the World Health Organization declared the outbreak of COVID-19 as PHEIC (Public Health Emergency of International Concern). At this time, a total of 9692 cases were confirmed in China, and there were 98 cases in 18 countries except China. As COVID-19 continues to spread between countries, the World Health Organization announced on March 11, 2020, that COVID-19 may have constituted a pandemic. At this time, a total of 121,190 cases were confirmed globally, spreading across all continents. As of November 26, 2020, a total of 60,110,143 cases were confirmed globally, and the trend continues to rise, as shown in Fig. 1. With the spread of COVID-19 around the world, it has put great pressure on the global health system and the world economy. In order to effectively control the COVID-19, countries adopt methods such as tracking and isolating infected patients to cut off the spread of the virus. The World Health Organization and scientists from various countries work together to develop related vaccines for treating infected patients and enhancing the immunity of those who are susceptible to the virus. With the successful development of the vaccine, people are confident that the virus will be defeated in the long term. For now, however, predicting future trends



**Fig. 1** Global cumulative confirmed cases by regions. Note: The histogram shows the cumulative number of confirmed cases in various regions since April, and the cumulative number of confirmed cases worldwide is mainly distributed in Asia, Europe, and North America

and the scale of confirmed cases is still essential, which can improve the subsequent control of COVID-19.

Regarding the prediction of confirmed cases, it becomes a hot topic since the outbreak of COVID-19. Because accurate prediction of confirmed cases can help the government to arrange appropriate preventive measures and allocate medical resources scientifically, in the past few months, hundreds of related research articles have been published. Generally speaking, the models used in these articles fall into three categories. The first category uses linear regression model, among which ARMA/ARIMA is the most commonly used model. For example, Sharma et al. [27] used ARIMA (2,1,0) to forecast the short-term infected and recovered cases for Saudi Arabia. de Lima et al. [7] integrated ARIMA models for the real-time prediction of cumulative cases of Covid-19 in Brazil. Choi and Ahn [4] forecasted daily imported COVID-19 cases in South Korea with ARIMAX. Malki et al. [21] used the seasonal ARIMA model to predict the spread of COVID-19 in several selected countries. Other similar researches include Abuhasel et al. [1, 28, 29], Singh et al. [28, 29], and Yousaf et al. [36]. The other linear methods include quantile regression [23] and Kalman Filter [28, 29, 37]. The second uses artificial intelligence methods, since the increase of confirmed cases is nonlinear. With this in mind, Hazarika and Gupta [11] proposed a wavelet-coupled random vector functional link networks to predict cumulative number of positive cases for five countries. Chimmula and Zhang [3] adopted LSTM networks to predict the future infections in Canada. Other non-linear growth models constitute the third category. Susceptible-infected-recovered (SIR) model is the most used models in this category. It is very popular in simulating the trend of epidemic. Liu et al. [16] used a modified SIR model to simulate the spread of COVID-19 in the USA. Ibrahim and Al-Najafi [13] developed a generalized SEIR (susceptible-exposed-infected-recovered) model for the spread of COVID-19, taking into account mildly and symptomatically infected individuals. Venkatesan et al. [30] and Sarkar et al. [26] adopted different forms of SIR model for India. Besides SIR model, Mahanty et al. [20] used Gompertz and Verhulst model to predict the number of infected cases for the next 28 days.

With reference to the three mainstream forecasting methods, we find that the first two methods are data-driven methods. Most of them use the number of confirmed cases only, such as Sharma et al. [27], Chimmula and Zhang [3], and Hazarika and Gupta [11]. The researches using the SIR model usually involves the impact of prevention measurements. For example, M. Liu et al. [16] took into account the effect of social distancing. Venkatesan et al. [30] examined the effectiveness of containment and lock-down.

Uncertain systems often have some problems, such as incomplete information and inaccurate data. As a result, constructing complex models often fails to achieve corresponding accurate results. Under the condition of less data environment, the grey system theory makes full use of the known “minimum information,” gives priority to extract the more valuable “new information,” and realizes the modeling of the whole system. It can be used to predict the changing trend of the system in the future (Deng, 1982; [19]). In order to improve the accuracy of prediction, many scholars derived more models from the traditional GM (1,1), such as Grey Verhulst model [8, 32, 38], Grey Bernoulli model [2, 33], and Fractional Order Accumulation Grey Model [35].

Grey system models are also applied to the prediction of COVID-19 epidemic, because it is proved to be effective for the data prediction of exponential growth. Due to the exponential growth in the number of cumulative confirmed cases, Şahin and Şahin [25] employed GM (1,1), nonlinear grey Bernoulli model (NGBM (1,1)) and fractional nonlinear grey Bernoulli model (FANGBM (1,1)) to predict the number of cumulative confirmed cases. Liu et al. [17] suggested that Fractional Order Accumulation grey model can accurately predict the spread of disease. Both studies focused on early-stage trends of COVID-19, but neither took into account the impact of social distancing, home quarantine order, and lock down. However, with the passage of time, the world is deeply involved in the secondary impact of COVID-19. Attention to the epidemic trend of the disease remains crucial. It is of more practical significance to predict the epidemic trend of the disease taking into account the influence of prevention and containment.

When the development of the system is disturbed, the future trend of the system changes accordingly. The problem in forecasting at this time does not lie in the pros and cons of the model, but in the fact that the information provided by the system cannot accurately reflect the future trend. It is difficult to use the information directly to make accurate prediction. In order to deal with this problem in system prediction, Liu [18] proposed the concept of buffer operator in the grey system predictive method. The operator which slows down the growth rate of the sequence is called the weakening buffer operator (WBO). Many scholars have proposed a variety of buffer operators based on different data to accurately predict the system with disturbance [5, 6, 31, 34], among which Wei and Kong [34] summarized the form of buffer operator and proposed a general form. As the COVID-19 spreads across the world, many disturbance factors are added to the prediction of the cumulative confirmed cases, which has an impact on the future trend. With regard to this problem, home quarantine order, gathering restriction, school closure, and international movement restriction issued by the government can effectively slow the spread of the virus, keep the public from being infected, and finally achieve the effect of controlling the spread of the epidemic. However, the implementation of the control measures varies greatly among the people of different countries, and the public's awareness of the COVID-19 and related control measures lags behind. When predicting the future trend of the cumulative number of confirmed cases, the introduction of relevant government control measures and public response should also be taken into consideration.

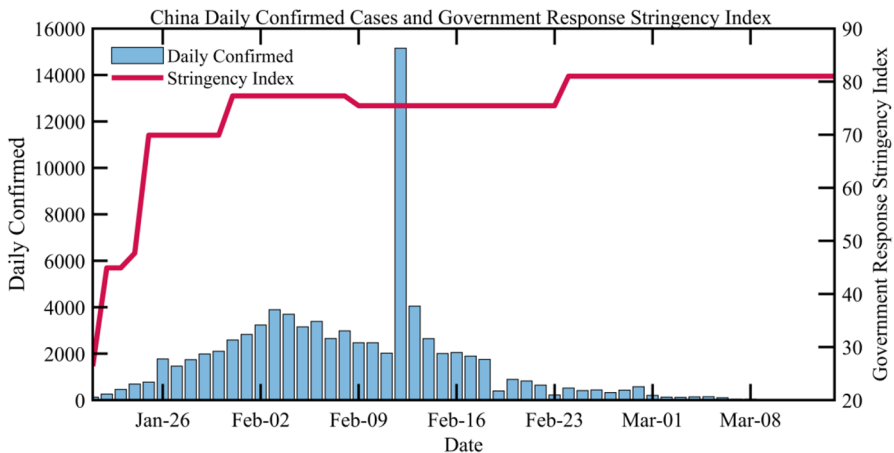
In China, for example, after experts confirmed that the SARs-CoV-2 could be transmitted from person to person, a lockdown is imposed in Wuhan, China, on January 23, 2020. Other control measures are further taken, such as gathering restriction, mandatory wearing masks in public places, partial traffic control and restriction on international movement. Since February 2, home quarantine orders have been comprehensively strengthened nationwide. China is capable of control the COVID-19 epidemic soon because it adopts a comprehensive approach to control the spread of the disease. This is diametrically opposed to the early concept of herd immunity in some western countries, which underestimates the severity of the epidemic and delays control measures such as home quarantine orders. China's COVID-19 epidemic occurred in January, which coincided with the Spring Festival. It is also a time when China's annual Spring Festival travel rush is about to reach its peak.

Visiting relatives and friends during the Spring Festival has been a Chinese tradition for thousands of years. Large-scale interpersonal communication activities accelerated the spread of the disease and many people did not realize the seriousness of the COVID-19 epidemic in the early stage after the lockdown in Wuhan. The Chinese government made a great deal of publicity on the prevention and control of the COVID-19 epidemic, as well as the full implementation of control measures, including persuasion to those who failed to comply with the regulations and accountability for those who seriously violated the regulations. More and more Chinese took the initiative to implement control measures and stay at home. As shown in Fig. 2, this leads to a peak in the number of new daily cases in China on February 12, and the growth rate has slowed significantly since March.

The indicator used to measure the level of government control measures in Fig. 2 comes from the Government Response Stringency Index (hereinafter referred to as “the Stringency Index”) constructed by scholars from University of Oxford.

In this paper, we use rolling fractional order accumulation grey model to predict the cumulative number of confirmed cases in various countries. In order to include control measures and public response in the prediction process to get more accurate results, we propose buffer level  $\Delta$  based on the general form of weakening buffer operator proposed by Wei and Kong [34]. We calculate the buffer level  $\Delta$  from the historical data of countries and introduce appropriate buffer level in the prediction according to the actual control measures and public response of each country. The introduction of different levels of buffer level can not only enhance the forecasting accuracy, but also provide predicted values in different scenarios.

The remainder of the research is arranged as follows: Section 2 introduces the model and methods, including FGM and buffer level. Section 3 shows the model



**Fig. 2** China daily confirmed cases and Government Response Stringency Index (January 21, 2020, to March 14, 2020). Note: The red line depicts the changes in China’s Stringency Index and the blue bars depict the number of daily confirmed cases in China. With the strengthening of the Chinese government’s control measures, the number of daily confirmed cases in China peaked on February 12 and decreased significantly in March

accuracy and application of buffer level. Section 4 presents the predictive results of the cumulative number of confirmed cases in these countries. The conclusions are drawn in the Section 5.

## 2 Models and Methods

### 2.1 Introduction of Fractional Order Accumulation Grey Model

GM (1,1) is the most classical and basic model in grey system theory. Driven by practice, academic research on GM (1,1) has been very active over the past three decades. Scholars derive many models from the traditional GM (1,1) to improve the forecast precision and apply them to more new fields. Fractional Order Accumulation Grey Model (FGM) is one of these derived models.

A standard FGM (1,1) is established as follow.

**Definition 1** It is assumed that the original sequence is.

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

$X^{(0)}$  is a non-negative sequence, while the  $r$ -AGO ( $0 < r < 1$ ) sequence of  $X^{(0)}$  is  $X^{(r)} = (x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(n))$ , in which

$$x^{(r)}(k) = \sum_{i=1}^k C_{k-i+r-1}^{k-i} x^{(0)}(i), k = 1, 2, \dots, n \tag{1}$$

According to the definition of Gamma function,  $X^{(r)}$  can be rewritten in the form of Gamma function as follow.

$$x^{(r)}(k) = \sum_{i=1}^k \frac{\Gamma(k-i+r)}{\Gamma(k-i+1) \cdot \Gamma(r)} x^{(0)}(i), k = 1, 2, \dots, n \tag{2}$$

$Z^{(r)}$  represents the generated mean sequence of consecutive neighbors of  $X^{(r)}$ .

$Z^{(r)} = (z^{(r)}(1), z^{(r)}(2), \dots, z^{(r)}(n))$ , in which

$$z^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k+1)}{2}, k = 1, 2, \dots, n-1 \tag{3}$$

**Definition 2** If  $X^{(0)}, X^{(r)}, Z^{(r)}$  are described as **Definition 1**, the following formula is obtained.

$$x^{(r)}(k) - x^{(r)}(k-1) + az^{(r)}(k) = u \tag{4}$$

This formula is the original form of FGM (1,1), and

$$\frac{dx^{(r)}}{dt} + ax^{(r)} = u \tag{5}$$

is the whitening equation of the FGM (1,1).

**Theorem 1** *If  $X^{(0)}, X^{(r)}$  are described as **Definition 1**, the parameter vector  $\hat{a} = [a, u]^T$  can be estimated by the least square method as follows.*

$\hat{a} = (B^T B)^{-1} B^T Y$ , where

$$B = \begin{bmatrix} -\frac{x^{(r)}(1)+x^{(r)}(2)}{2} 1 \\ -\frac{x^{(r)}(2)+x^{(r)}(3)}{2} 1 \\ \vdots \\ -\frac{x^{(r)}(n-1)+x^{(r)}(n)}{2} 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(r)}(2) - x^{(r)}(1) \\ x^{(r)}(3) - x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) - x^{(r)}(n-1) \end{bmatrix} \tag{6}$$

The solution to the whitening equation of FGM (1,1) is shown as follows

$$\hat{x}^{(r)}(k) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{-a(k-1)} + \frac{u}{a} \tag{7}$$

By substituting the parameter vector  $\hat{a} = [a, u]^T$  into the formula (7), the fitting values sequence  $\hat{X}^{(r)}$  can be obtained. The fitting value sequence  $\hat{X}^{(0)}$  can be calculated by the  $r$ -order inverse accumulation.

$$\begin{aligned} \hat{X}^{(1)} &= (\hat{x}^{(1)}(1), \hat{x}^{(1)}(2), \dots, \hat{x}^{(1)}(n+1), \dots) \\ &= (\hat{x}^{(r)(1-r)}(1), \hat{x}^{(r)(1-r)}(2), \dots, \hat{x}^{(r)(1-r)}(n+1), \dots) \end{aligned} \tag{8}$$

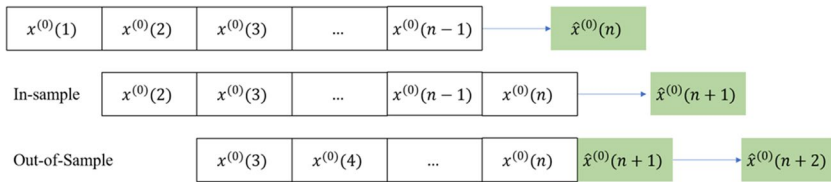
$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{9}$$

$$\hat{X}^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n+1), \dots) \tag{10}$$

### 2.2 Rolling Mechanism

The rolling mechanism updates the initial value of the sequence, excludes the early information that loses its timeliness from the sequence, and adds new information with timeliness to the sequence, which can effectively enhance the predictive accuracy. This is in line with the principle of new information priority, which is one of the basic principles in grey system theory. The effect of new information on cognition is greater than that of old information, which is the information view of the whole grey system theory [19]. Many literatures that use the grey system method to model the confirmed, recovered, and death cases of COVID-19 adopt the rolling mechanism [17, 38]. The example of rolling mechanism used in this paper is shown in Fig. 3.

The original sequence is  $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n-1), x^{(0)}(n)$ . Rolling sequence length is set first, which is  $n-1$  in Fig. 3. The FGM (1,1) is established according to the rolling sequence  $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n-1)$ . In the in-sample prediction, the predicted value  $\hat{x}^{(0)}(n)$  is used to evaluate the predictive accuracy of the model. When rolling sequence is updated, the initial value of the rolling sequence is



**Fig. 3** Rolling mechanism of FGM (1,1). Note: The original sequence is  $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n-1), x^{(0)}(n)$ . The FGM (1,1) is established according to the rolling sequence  $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n-1)$ . In the in-sample prediction, the predicted value  $\hat{x}^{(0)}(n+1)$  is estimated by the updated original sequence. In the out-sample prediction, the predicted value  $\hat{x}^{(0)}(n+2)$  is estimated both by the updated original sequence and the predicted value  $\hat{x}^{(0)}(n+1)$

excluded and the original value  $x^{(0)}(n)$  is added. In the out-of-sample prediction, the real value  $x^{(0)}(n+1)$  remains unknown. So, when rolling sequence is updated, the initial value of the rolling sequence is excluded and the predicted value  $\hat{x}^{(0)}(n+1)$  is added instead of the real value. As it goes on, the new predicted value  $\hat{x}^{(0)}(n+2)$  will be added in the next update. Through the mechanism above, we make full use of all the new information in the original sequence and exclude it when its timeliness weakens, which can effectively improve the predictive accuracy.

### 2.3 Grid Search

Compared with the GM (1,1) model, FGM (1,1) is characterized by the use of fractional order accumulation. According to the research of Wu et al. [35], when the fractional order  $r$  increases, the FGM model increases the weight of the old information. Therefore, when  $r$  is relatively small, FGM (1,1) is suitable for modeling short memory process. Conversely, when  $r$  is relatively large, FGM (1,1) is suitable for modeling the long memory process. In this paper, we use grid search to optimize the fractional order  $r$ .

Grid search is an enumeration method for adjusting parameters. It is very suitable for small-scale and limited parameter adjustment. In the previous literatures, the fractional order  $r$  is accurate to one decimal place. We extend it to three decimal places, which generate 1000 grids on interval (0,1). We compare the in-sample predictive accuracy obtained by 1000 grids, select  $r$  with the highest predictive accuracy, and output its subsequent prediction value. The flow chart of the rolling FGM (1,1) is shown in Fig. 4.

### 2.4 Buffer Operator and Buffer Level

In order to include control measures and public response in the prediction process to improve the forecasting accuracy, we propose buffer level  $\Delta$  based on the general form of weakening buffer operator proposed by Wei and Kong [34]. They summarized the form of buffer operator and proposed a general form as follows.

**Proposition 1** *Let the original sequence is assumed to be*



$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$$

$\omega = (\omega_1, \omega_2, \dots, \omega_n)$  is its corresponding weight vector, in which  $\omega_i > 0, i = 1, 2, \dots, n$ . The sequence with buffer operator  $X^{(0)}D = (x^{(0)}(1)d, x^{(0)}(2)d, \dots, x^{(0)}(n)d)$ , in which

$$x^{(0)}(k)d = x^{(0)}(k) \cdot \left[ x^{(0)}(k) / \frac{\sum_{i=k}^n w_i x^{(0)}(i)}{\sum_{i=k}^n w_i} \right] \tag{11}$$

When  $\alpha < 0$ ,  $D$  is a weakening buffer operator (WBO) for both monotonically increasing sequence and monotonically decreasing sequence.

When  $\alpha > 0$ ,  $D$  is a strengthening buffer operator (SBO) for both monotonically increasing sequences and monotonically decreasing sequences.

When  $\alpha = 0$ ,  $D$  is an identity operator, which means there is no buffering.

When the control measures are reinforced and the public respond actively, the growth trend of the cumulative number of confirmed cases slow down. Therefore, we only consider the WBO and identity operator in the general form, when  $\alpha \leq 0$ . We want to obtain the buffer operator through historical data, so we propose buffer level  $\Delta$  based on the general form of weakening buffer operator.

We can obtain  $\Delta$  from the formula  $\Delta = -\alpha$ . As  $\alpha \leq 0$ , it means that  $\Delta \geq 0$ . For simplicity, it is assumed that  $\omega_1 = \omega_2 = \dots = \omega_n$ . Then, formula (11) can be simplified to

$$\begin{aligned} x^{(0)}(k)d &= x^{(0)}(k) \cdot \left( \frac{x^{(0)}(k)+x^{(0)}(k+1)+\dots+x^{(0)}(n)}{(n-k+1) \cdot x^{(0)}(k)} \right)^\Delta \\ &= x^{(0)}(k) \cdot \left( \frac{\sum_{i=k}^n x^{(0)}(i)}{(n-k+1) \cdot x^{(0)}(k)} \right)^\Delta \end{aligned} \tag{12}$$

When  $\Delta=0, x^{(0)}(k)d = x^{(0)}(k)$ ,  $D$  is an identity operator.

When  $\Delta=1, x^{(0)}(k)d = \frac{\sum_{i=k}^n x^{(0)}(i)}{n-k+1}$ ,  $D$  is an average weakening buffer operator (AWBO) [18].

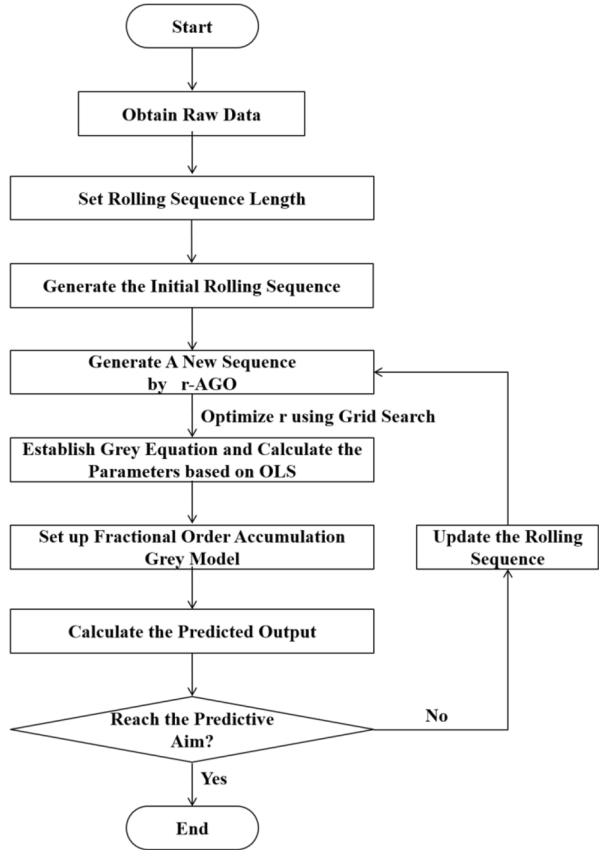
**Proposition 2**  $\Delta$  is defined as buffer level. When  $x^{(0)}(k) > 1$ , as  $\Delta$  increases, the weakening effect on the sequence will be enhanced.

**Proof** Let  $\Delta_1 < \Delta_2$ , which means the original sequency will be buffered at different levels. The buffer operators are  $D_1$  and  $D_2$  separately. The sequency with buffer operator  $D_1$  is.

$$X^{(0)}D_1 = (x^{(0)}(1)d_1, x^{(0)}(2)d_1, \dots, x^{(0)}(n)d_1)$$

And the sequency with buffer operator  $D_2$  is

**Fig. 4** Flow chart of rolling FGM (1,1). Note: This is the flow chart of the rolling FGM (1,1) model. After obtaining the raw data, we set the rolling sequence length and generate the initial rolling sequence. After the optimal parameter is calculated by grid search, we establish the FGM (1,1) model and update the rolling sequence until the predictive aim is reached



$$X^{(0)}D_2 = (x^{(0)}(1)d_2, x^{(0)}(2)d_2, \dots, x^{(0)}(n)d_2)$$

In which,

$$x^{(0)}(k)d_1 = x^{(0)}(k) \cdot \left( \frac{\sum_{i=k}^n x^{(0)}(i)}{(n - k + 1) \cdot x^{(0)}(k)} \right)^{\Delta_1}$$

$$x^{(0)}(k)d_2 = x^{(0)}(k) \cdot \left( \frac{\sum_{i=k}^n x^{(0)}(i)}{(n - k + 1) \cdot x^{(0)}(k)} \right)^{\Delta_2}$$

When  $X^{(0)}$  is monotonically increasing,

$$\frac{x^{(0)}(k)d_2}{x^{(0)}(k)d_1} = \left[ \sum_{i=k}^n x^{(0)}(i) \right]^{\Delta_2 - \Delta_1} \cdot (n - k + 1)^{\Delta_1 - \Delta_2} \cdot x^{(0)}(k)^{\Delta_1 - \Delta_2}$$

$$\geq [(n - k + 1)x^{(0)}(k)]^{\Delta_2 - \Delta_1} \cdot (n - k + 1)^{\Delta_1 - \Delta_2} \cdot x^{(0)}(k)^{\Delta_1 - \Delta_2} = 1$$

So

$$x^{(0)}(k)d_2 \geq x^{(0)}(k)d_1$$

When  $X^{(0)}$  is monotonically decreasing,

$$\begin{aligned} \frac{x^{(0)}(k)d_2}{x^{(0)}(k)d_1} &= \left[ \sum_{i=k}^n x^{(0)}(i) \right]^{\Delta_2 - \Delta_1} \cdot (n - k + 1)^{\Delta_1 - \Delta_2} \cdot x^{(0)}(k)^{\Delta_1 - \Delta_2} \\ &\leq [(n - k + 1)x^{(0)}(k)]^{\Delta_2 - \Delta_1} \cdot (n - k + 1)^{\Delta_1 - \Delta_2} \cdot x^{(0)}(k)^{\Delta_1 - \Delta_2} = 1 \end{aligned}$$

So

$$x^{(0)}(k)d_1 \geq x^{(0)}(k)d_2$$

To be specific, the sequence curve of  $X^{(0)}D_2$  is flatter than  $X^{(0)}D_1$ . In other words, through buffering, the increasing or decreasing speed is reduced. The larger the value of  $\Delta$  is, the larger the speed is reduced.

Because  $\Delta$  is the negative value of  $\alpha$ , the weakening operator  $D$  constructed from  $\Delta$  satisfies three axioms of buffer operator proposed by S. Liu [18].

The calculation of buffer level takes China and Japan as an example. The early epidemic control in both China and Japan achieved certain results. After lockdown was implemented in Wuhan, China, on January 23, control measures were strengthened comprehensively nationwide. China's Stringency Index rose rapidly from 26.39 before the lockdown of Wuhan to 81.02 in early March. The increase in the number of daily confirmed cases has been significantly slow since March, and the spread of the disease has been initially controlled. At the beginning of March, the buffer level is estimated by constructing a corresponding weakening buffer operator to minimize the MAPE (mean absolute percentage error) of the sequence in a period of time after March. It can be estimated that the buffer level  $\Delta$  of the China's epidemic control in early March is about 1.5, which is bigger than the buffer level of AWBO.

The early epidemic of COVID-19 in Japan is shown in Fig. 5. With the emergence of a small-scale spread of COVID-19 on Feb. 6, 2020, the Japanese government gradually began to take measures such as gathering restriction, international movement restriction, and information campaigns. On April 7, the Japanese government issued the home quarantine order. In the following week, Japan's Stringency Index reached 47.22, which was the peak of the index as of November 31.

We use the Google COVID-19 mobility dataset to construct a Mobility Index. This dataset contains the mobility trend of people in various places compared to the baseline (Jan 3 to Feb 6). We choose retail and recreation mobility and residential mobility. Generally speaking, reduction in the mobility of retail and recreation and increase in the mobility of residential means that people respond actively to government control measures. However, the raw data of mobility has obvious weekly periodicity. Therefore, we process the raw data in a seven-day moving average in

order to smooth the unnecessary periodicity. As the retail and recreation mobility is almost all negative, we change it into additive inverse value and accumulate it with residential mobility. We construct an index namely Mobility Index, which can reflect public response to COVID-19 and government control measures.

Since Google does not have relevant services in China, we calculate the buffer level of China on the basis of the Stringency Index. For Japan and other countries, we can take both Stringency Index and Mobility Index into consideration. Since the implementation of the home quarantine order in April, mobility has increased significantly and the epidemic has been effectively controlled. The buffer level of Japan was estimated from late April to mid-May, which is about 0.9. It is slightly smaller than 1, the buffer level of AWBO.

### 3 Empirical Analysis

#### 3.1 Brief Introduction to Datasets

The raw data of Stringency Index is collected from publicly available sources such as news articles and government press releases and sorted by the team from Oxford University. The raw data of Mobility Index is collected and sorted by Google. The data of cumulative confirmed cases used in this paper are collected from Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The data of cumulative confirmed cases used in this paper for training and testing is from January 2 to November 26, 2020, and the total length of the sequence is 330. Nine countries are contained in our model and prediction, including China, Japan, the USA, France, the UK, Germany, Brazil, India, and Argentina.

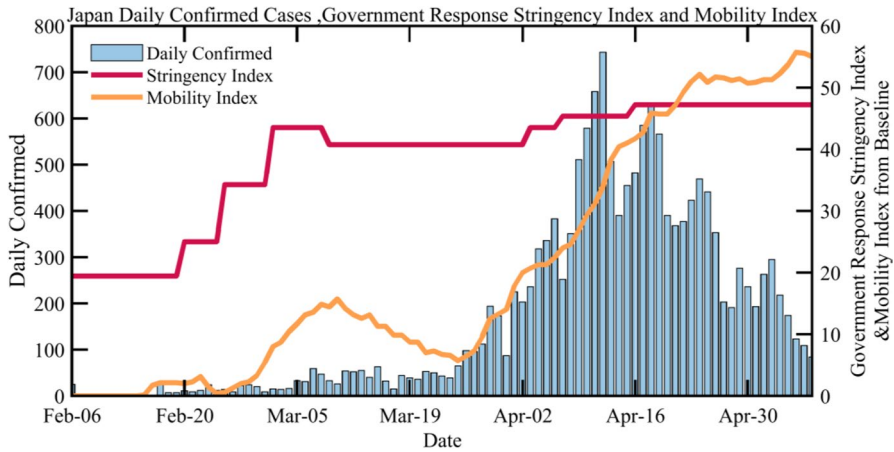
#### 3.2 Rolling Sequence Length Selection

We first select the optimal rolling sequence length for each country. Table 1 reports the MAPE of different rolling sequence lengths in nine countries. Except for China, which has large outliers in the early stage when the length is 7 and 8, the MAPEs of these countries are all less than 10.

It is initiatively proved that rolling FGM (1,1) can well fit the cumulative number of confirmed cases in these countries. In this paper, the length with the minimum MAPE is selected as the optimal length of the corresponding country. The actual value and predicted value by the optimal rolling FGM (1,1) in various countries are shown in Fig. 6. The other eight countries show an obvious growth trend, while the cumulative number of confirmed cases in China is stable.

#### 3.3 Comparison of Model Accuracy

In this paper, rolling FGM (1,1) is compared with the ARMA, which is the most commonly used time series model in modeling the cumulative confirmed cases. The



**Fig. 5** Japan daily confirmed cases, Government Response Stringency Index, and Mobility Index (February 6, 2020, to May 8, 2020). Note: The red line depicts the changes in Japan’s Stringency Index. The orange line depicts the changes in Japan’s Mobility Index, and the bar chart depicts the number of daily confirmed cases in Japan. With the strengthening of the Japan government’s control measures and the public’s awareness to the epidemic of COVID-19, the number of daily confirmed cases in Japan peaked in the early April and decreased significantly on May

specifications of ARMA are different for different countries. We use AIC and BIC to choose the best order of ARMA. The specifications of ARMA are shown in the Appendix (Table 7). We use the data from January 2 to October 31 as the training set and the data from November 1 to November 26 as the test set. Table 2 reports the predictive accuracy of a 14-day test set (from November 1 to November 14), and Table 3 reports the predictive accuracy of a 26-day test set (from November 1 to November 26). The predictive accuracy is calculated in both MAPE and RMSE.

According to Tables 2 and 3, the predictive accuracy of the rolling FGM (1,1) is better than that of the ARMA in 14-day and 26-day test set. The MAPEs of the 14-day test sets in all countries are less than 10%, and the MAPEs of the 26-day test sets are less than 15%. This indicates that the FGM (1,1) model can effectively predict the cumulative number of confirmed cases. Figure 7 compares the performance of ARMA and FGM (1,1) in the 26-day test set. With the exception of Brazil and India, the FGM (1,1) model shows good predictive ability.

With reference to the forecast of cumulative confirmed cases in Brazil and India in November, our model is not better than ARMA; our model predicts that there will be more cumulative confirmed cases in Brazil and India. In fact, related studies point out that the official cumulative confirmed cases of Brazil and India may be understated. Marinho et al. [22] pointed out that there was little statistical correlation in the number of confirmed cases in Brazil, because as of October 29, 2020, only less than 7% of the population had been tested. The number of confirmed cases in Brazil was underestimated due to the limitation of testing capacity and speed. Mukherjee [24] pointed out that the test speed in India was lower than that in the early stage of the outbreak, and the extensive use of only RAT (Rapid Antigen Test) instead of RAT + RT-PCR testing

protocol led to a large number of false negatives in the test results, which seriously affected the reliability of Indian official data. As a result, FGM (1,1) which proves to be effective for the data prediction of exponential growth does not achieve excellent predictive results.

### 3.4 Application of Buffer Level

As the increase of control measures and public response have a weakening effect on the cumulative confirmed number in the future, it is possible that using all the new information and excluding the old information timely may not predict the future trend accurately. This feature appears in the prediction of cumulative confirmed cases in Italy, of which MAPE reaches 25.2%. The problem comes down to the ignorance of the control measures and public response. Applying the buffer level to adjust the rolling sequence of Italy, the result with better predictive accuracy is obtained.

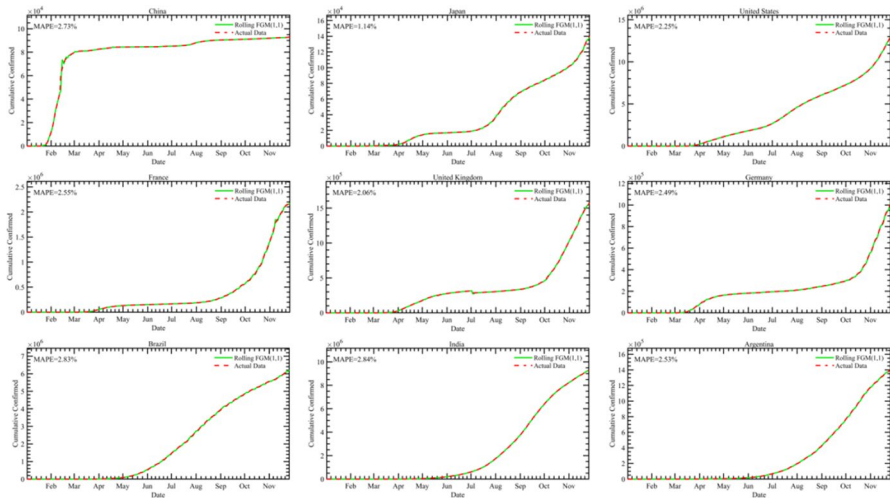
Since the end of October, Italy has strengthened control measures such as traffic control and home quarantine order, making the Stringency Index gradually increase from 50 to 66.7 in early November. As shown in Fig. 8, the Mobility Index increases at the same time. This indicates that people actively respond to these policies and their lifestyles change partly. These factors have a weakening effect on the future trend. Using FGM (1,1) only to predict produces a relatively large error.

In the middle and late March of 2020, the Italy Government implemented relatively strict control measures and the public complied with these measures. This allows the growth trend of COVID-19 epidemic to be controlled, so we use the historical data to estimate the buffer level to be 0.55 at that time. By applying the buffer level to the data in early November (9 or 10 days after the increase of the

**Table 1** MAPE of different rolling sequence lengths for FGM

Country		Rolling sequence length					
		$L=5$	$L=6$	$L=7$	$L=8$	$L=9$	$L=10$
China	MAPE (%)	<b>2.73</b>	7.307	44.048	213.25	3.089	7.694
Japan	MAPE (%)	<b>1.138</b>	1.175	1.215	1.278	1.35	1.398
USA	MAPE (%)	2.248	<b>2.246</b>	2.404	2.511	2.618	2.638
France	MAPE (%)	2.573	<b>2.553</b>	2.579	2.73	2.975	2.905
UK	MAPE (%)	2.499	<b>2.063</b>	2.28	2.278	2.361	2.361
Germany	MAPE (%)	<b>2.493</b>	2.613	2.838	2.927	3.048	2.899
Brazil	MAPE (%)	<b>2.825</b>	2.975	3.073	3.222	3.96	4.388
India	MAPE (%)	<b>2.837</b>	4.464	6.263	7.392	7.924	7.988
Argentina	MAPE (%)	<b>2.527</b>	2.65	3.834	3.978	4.2	4.416

Note: We calculate the MAPE of the FGM (1,1) model with a rolling sequence length of 5–10 to select the optimal rolling sequence length for each country. Except for the USA, Britain, and Germany, the optimal rolling sequence length is 6, and the optimal rolling sequence length for other countries is 5



**Fig. 6** Predictive results of the cumulative cases in different countries using FGM (1,1). Note: The picture shows the predictive results of the cumulative cases in different countries using the FGM (1,1), and the MAPE of these models is less than 3%

Stringency Index), the predictive accuracy is greatly improved, with MAPE decreasing from 25.2 to 2.84%. The MAPE of FGM in the 14-day test set is the same as that of ARMA, both at 0.89%. The Stringency Index increased from 66.67 to 79.85 in November 6, and the Mobility Index continued to increase at the same time. By applying the buffer level to the data in November 15 (9 or 10 days after the increase of the Stringency Index), the predictive accuracy of FGM (1,1) is better than that of ARMA in the 12-day test set (from Nov 15, 2020, to Nov 26, 2020). The comparison of model accuracy is shown in Table 4. From the example of Italy, we find that government control measures require the active cooperation of the public and often have a time lag when they are effective. Applying the buffer level at that time can enhance predictive accuracy of FGM (1,1).

The result in France can provide a similar example. As shown in Fig. 9, the French government strengthened the control measures on October 30, 2020. The Mobility Index increased in the same time. However, it has only been implemented for 2 days. So, FGM is better than ARMA in predicting cumulative cases in 26-day test set (from Nov 1, 2020, to Nov 26, 2020).

We use the historical data in the middle and late March of 2020 to estimate the buffer level to be 0.25 at that time. Applying the buffer level to the data of November 9 (9 or 10 days after the increase of the Stringency Index) can enhance the predictive accuracy in 18-day test set, which is shown in Table 5.

The example of the USA also supports our finding. As is shown in Fig. 10, the Stringency Index of the USA rose twice on November 8 and November 16. However, there is no significant increase in the Mobility Index, which means that Americans do not respond actively to the control measures. As a result, there is no weakening effect on the cumulative number of confirmed cases in the USA.

**Table 2** Fourteen-day model accuracy (from Nov 1, 2020, to Nov 14, 2020)

Country	USA		France		UK		Germany		China		Japan		Brazil		India		Argentina	
	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA
MAPE (%)	0.76	2.29	4.79	5.96	0.96	11.13	7.49	9.25	0.01	2.2	0.74	1.16	1.08	0.33	1	0.17	0.2	2.82
RMSE	109,496.02	312,309.05	96,411.48	121,936.26	14,579.85	159,173.79	64,252.72	81,729.42	14.79	2326.87	1123.27	1883.51	66,051.26	21,242.99	10,6662.97	17,105.80	2889.28	41,221.05

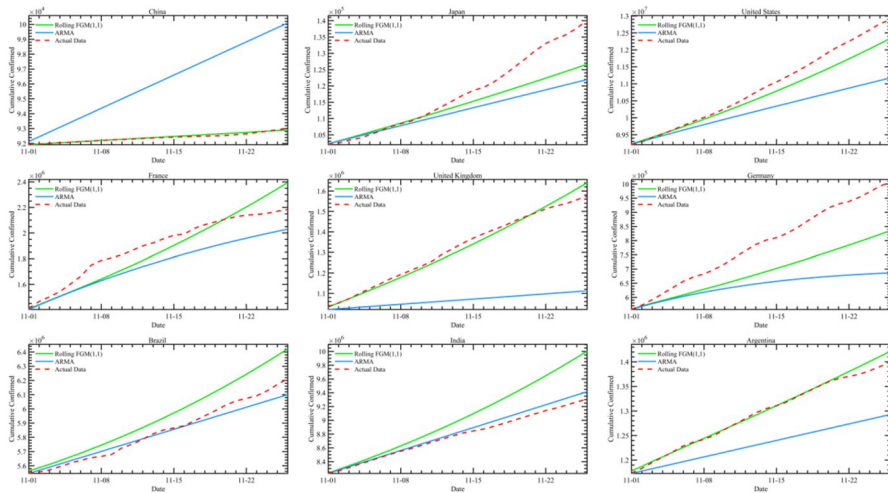
Note: This table reports the predictive accuracy of FGM (1,1) and ARMA in the 14-day test set. Calculated in both MAPE and RMSE, the predictive accuracy of FGM (1,1) is better than that of ARMA model except India and Brazil. We believe that this may be due to the limited COVID-19 testing speed of Brazil and India. As a result, FGM (1,1) which proves to be effective for the data prediction of exponential growth does not achieve excellent predictive results



**Table 3** Twenty-six-day model accuracy (from Nov 1, 2020, to Nov 26, 2020)

Country	USA		France		UK		Germany		China		Japan		Brazil		India		Argentina	
	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA	FGM	ARMA
MAPE (%)	2.13	5.91	4.33	7.01	1.3	17.98	11.36	16.93	0.05	4.03	3.26	4.64	1.78	0.57	2.87	0.52	0.38	4.60
RMSE	324,198.53	893,624.98	99,094.62	147,697.03	23,966.15	288,782.91	110,615.55	176,618.23	61.5	4,282.96	6103.02	8427.28	119,640.48	44,223.57	335,865.31	58,827.99	8204.87	69,199.27

Note: This table reports the predictive accuracy of FGM (1,1) and ARMA in the 26-day test set. Calculated in both MAPE and RMSE, the predictive accuracy of FGM (1,1) is better than that of ARMA model except India and Brazil. We believe that this may be due to the limited COVID-19 testing speed of Brazil and India. As a result, FGM (1,1) which proves to be effective for the data prediction of exponential growth does not achieve excellent predictive results



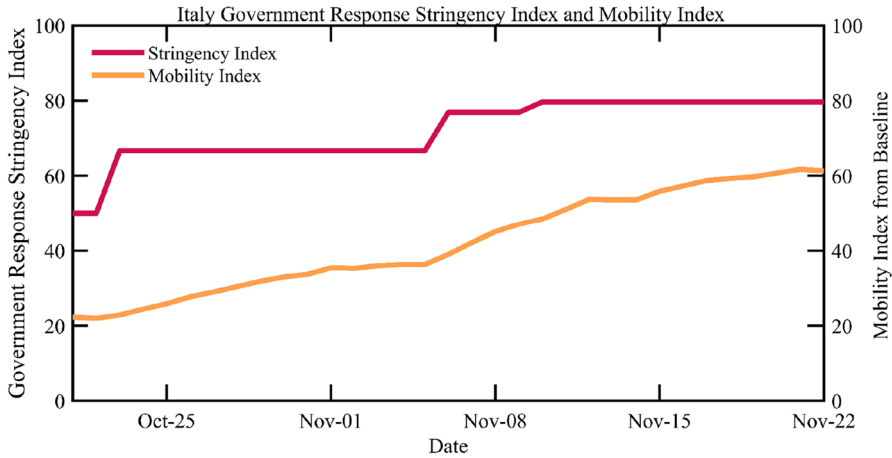
**Fig. 7** Comparison of model accuracy in different countries. Note: The picture shows the predictive results of the cumulative cases in different countries using the FGM (1,1) and ARMA. The prediction accuracy of FGM (1,1) is better than that of ARMA model except India and Brazil

The cumulative number of confirmed cases in the USA continued to grow exponentially throughout November. Better result is obtained by using FGM (1,1) with buffer level of 0.

The Stringency Index and Mobility Index of Brazil and India in the end of October and in November remain stable, as there is no weakening effect on the cumulative number of confirmed cases. We do not use buffer level to improve our predictive accuracy. Our model predicts that the cumulative number of confirmed cases in Brazil and India is much more than the reported number, mainly due to bias between actual number and reported number mentioned above (Figs. 11 and 12).

#### 4 Predictive Results in Different Scenarios

As the buffer level can be used to improve predictive accuracy in the test set, it can also be used to predict cumulative confirmed cases in different scenarios. In this section, we predict the cumulative confirmed cases from January 1 to 31, 2021, in Japan, the USA, France, Britain, Germany, Italy, Brazil, India, and Argentina based on the data as of December 31, 2020 in different scenarios. Different scenarios mean using different buffer levels to represent the mitigation effect of the coordination of control measures and public responses on the spread of the epidemic. We use the buffer levels of 0.55 and 0.9 as the benchmark, which is based on the early data in Italy and Japan.  $\Delta = 0.55$  means that



**Fig. 8** Italy Government Response Stringency Index and Mobility Index. Note: The red line depicts the changes in Italy’s Stringency Index and the orange line depicts the changes in Italy’s Mobility Index. Since the end of October, Italy has strengthened control measures and people have actively responded to these policies

the country has effective epidemic control while  $\Delta = 0.9$  means that the country has better epidemic control.

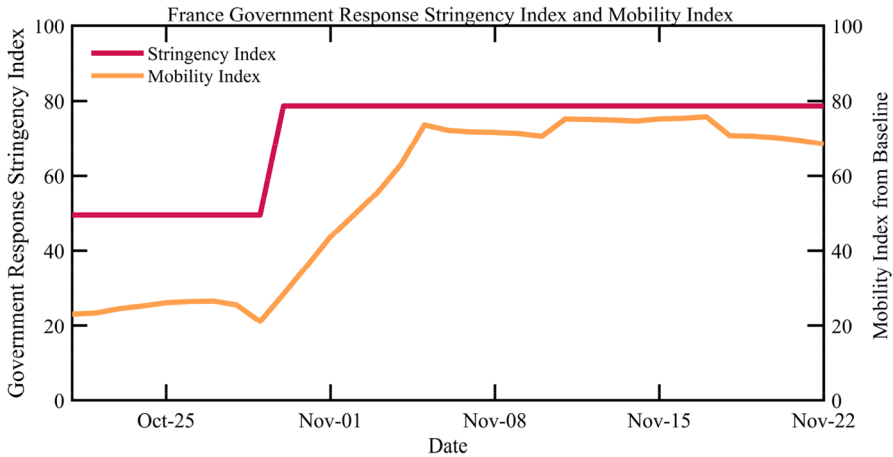
### 4.1 Japan

The cumulative number of confirmed cases in Japan as of December 31 was 235,788. The actual number of confirmed cases in Japan in January 31 fall in the range of the predicted value with a buffer level of 0 to 0.9, and the it is most close to the predictive result with a buffer level of 0.55 (Fig. 13).

**Table 4** Comparison of model accuracy in Italy

Model	FGM	FGM $\Delta=0.55$	ARMA
26-day prediction (Nov 1, 2020, to Nov 26, 2020)			
MAPE	25.2%	2.84%	0.89%
RMSE	429,147.3	59,699.84	11,607.61
14-day prediction (Nov 1, 2020, to Nov 14, 2020)			
MAPE	10.67%	0.95%	0.95%
RMSE	131,542	10,785.85	9696.08
12-day prediction (Nov 15, 2020, to Nov 26, 2020)			
MAPE	4.89%	1.12%	2.51%
RMSE	298,989.9	64,079.85	149,981.4

Note: Applying a buffer level in the FGM (1,1) can enhance the predictive accuracy of the model, and the predictive accuracy of FGM (1,1) is better than that of ARMA model



**Fig. 9** France Government Response Stringency Index and Mobility Index. Note: The red line depicts the changes in France’s Stringency Index and the orange line depicts the changes in France’s Mobility Index. Since the end of October, France has strengthened control measures and people have actively responded to these policies

## 4.2 USA

The cumulative number of confirmed cases in the USA as of December 31 was 20,098,800. Judging from the past trend, the cumulative number of confirmed cases in the USA has shown a trend of exponential growth since the end of October. The actual number of confirmed cases in the USA in January 31 fall in the range of the predicted value with a buffer level of 0 to 0.9, and it is most close to the predictive result with a buffer level of 0.55 (Fig. 14).

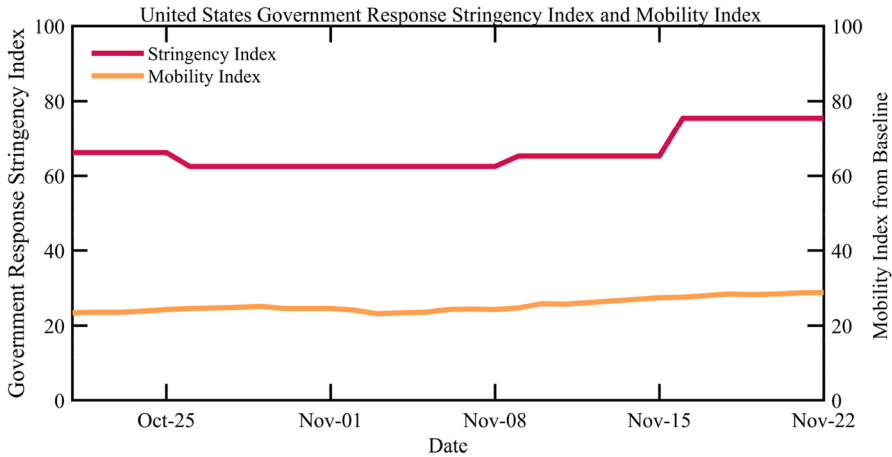
## 4.3 France

The cumulative number of confirmed cases in France as of December 31 was 2,620,425. Judging from the past trend, the growth rate of cumulative confirmed

**Table 5** Comparison of model accuracy in France (from Nov 9, 2020, to Nov 26, 2020)

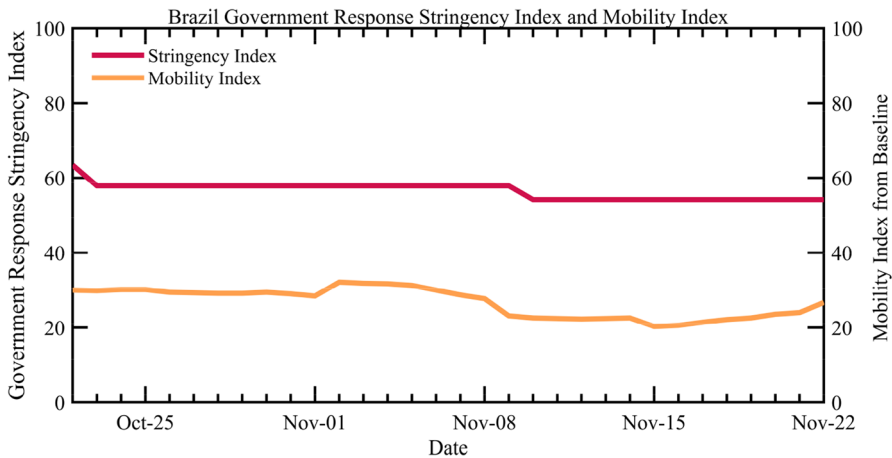
Model	FGM	FGM $\Delta = 0.25$	ARMA
18-day prediction (Nov 9, 2020, to Nov 26, 2020)			
MAPE	6.68%	3.3%	4.75%
RMSE	691,477.58	351,298.14	461,733.24

Note: Applying a buffer level in the FGM (1,1) can enhance the predictive accuracy of the model, and the predictive accuracy of FGM (1,1) is better than that of ARMA model

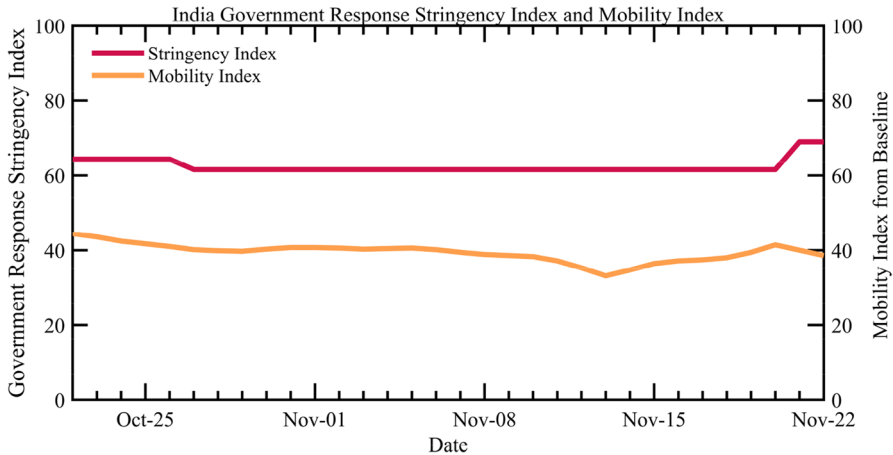


**Fig. 10** United States Government Response Stringency Index and Mobility Index (October 22, 2020, to November 22, 2020). Note: The red line depicts the changes in America’s Stringency Index and the orange line depicts the changes in America’s Mobility Index. Since the end of October, there is no significant increase in the Mobility Index, which means that Americans do not respond actively to the control measures

cases in France has decreased since the mid-late November. The actual number of confirmed cases in the France falls in the range of the predicted value with a buffer level of 0 to 0.9 (Fig. 15).



**Fig. 11** Brazil Government Response Stringency Index and Mobility Index (October 22, 2020, to November 22, 2020). Note: The red line depicts the changes in America’s Stringency Index and the orange line depicts the changes in Brazil’s Mobility Index. Since the end of October, the Stringency Index and Mobility Index of Brazil remain stable



**Fig. 12** India Government Response Stringency Index and Mobility Index (October 22, 2020, to November 22, 2020). Note: The red line depicts the changes in America's Stringency Index and the orange line depicts the changes in India's Mobility Index. Since the end of October, The Stringency Index and Mobility Index of India remain stable

#### 4.4 UK

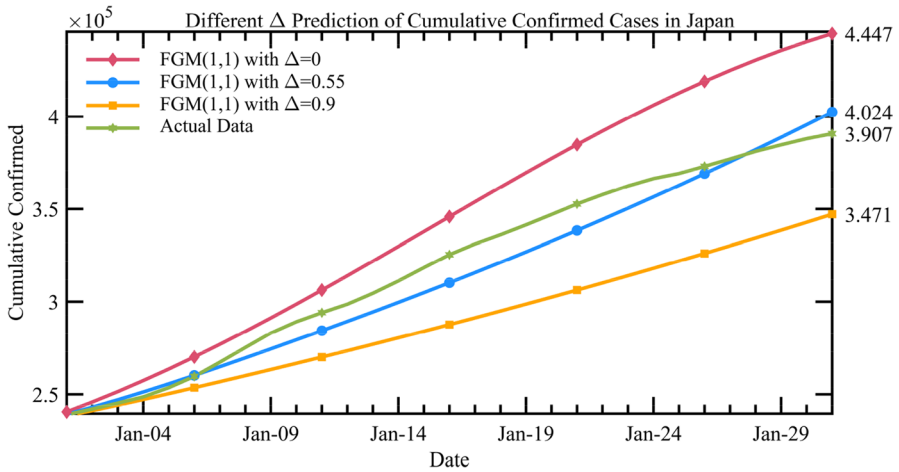
The cumulative number of confirmed cases in the UK as of December 31 was 2,488,780. Judging from the past trend, the growth rate of cumulative confirmed cases in the UK has increased since the mid-December. The actual number of confirmed cases in the UK in January 31 fall in the range of the predicted value with a buffer level of 0 to 0.9, and it is most close to the predictive result with a buffer level of 0.9 (Fig. 16).

#### 4.5 Germany

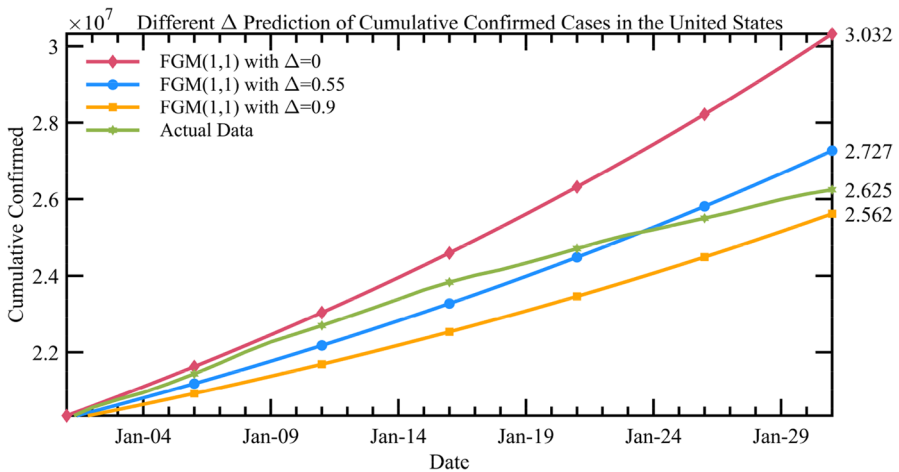
The cumulative number of confirmed cases in Germany as of December 31 was 1,756,248. Judging from the past trend, the cumulative number of confirmed cases in Germany has shown a trend of exponential growth since the end of October. The actual number of confirmed cases in Germany in January 31 is most close to the predictive result with a buffer level of 0.9, which is consistent with our finding (Fig. 17).

#### 4.6 Italy

The cumulative number of confirmed cases in Italy as of December 31 was 2,107,166. Judging from the past trend, the growth rate of cumulative confirmed cases in Italy has decreased since the end of October. The actual number of confirmed cases in Italy in January 31 falls in the range of the predicted values, and it is most close to the predictive result with a buffer level of 0.9 (Fig. 18).



**Fig. 13** Predictive result of cumulative confirmed cases in Japan. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



**Fig. 14** Predictive result of cumulative confirmed cases in the USA. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)

### 4.7 Brazil

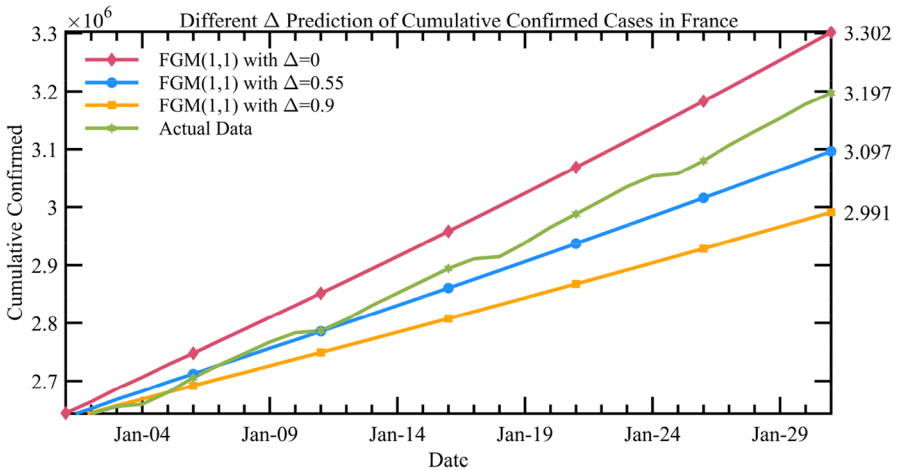
The cumulative number of confirmed cases in Brazil as of December 31 was 7,681,032. Judging from the past trend, the cumulative number of confirmed cases in Brazil has shown a trend of exponential growth since the end of October. The cumulative number of confirmed cases officially announced in Brazil in January 31 is most close to the predictive result with a buffer level of 0.9. However, due to the limited testing ability in the early stage, the actual number of the cumulative confirmed cases may be closer to the predicted value provided by FGM (1,1) with a buffer level of 0 (Fig. 19).

### 4.8 India

The cumulative number of confirmed cases in India as of December 31 was 10,286,234. Judging from the past trend, the cumulative number of confirmed cases in India has shown a trend of exponential growth since the end of October. However, the actual data in India did not fall within our forecasts, mainly due to the limited COVID-19 testing speed in India mentioned above. It can be observed that in late January, the number of daily confirmed cases is growing at an almost fixed rate, probably due to limited testing speed, resulting in an underestimation of confirmed cases (Fig. 20).

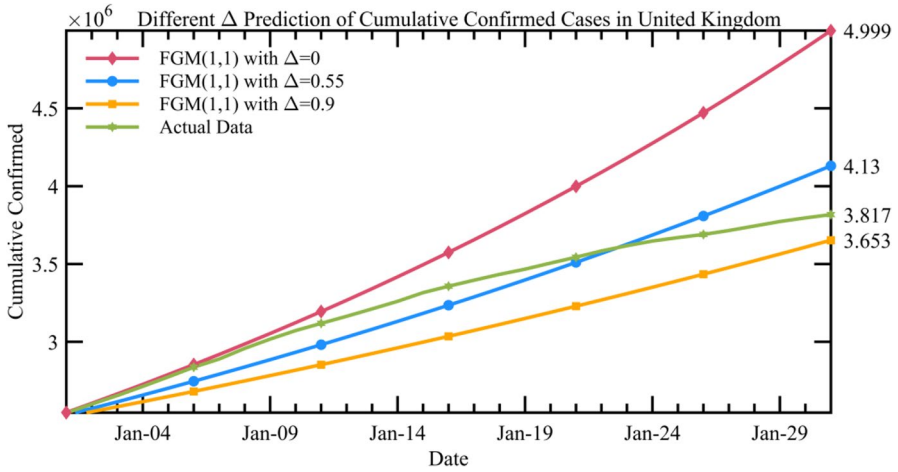
### 4.9 Argentina

The cumulative number of confirmed cases in Argentina as of December 31 was 1,625,514. Judging from past trends, the cumulative number of confirmed cases in

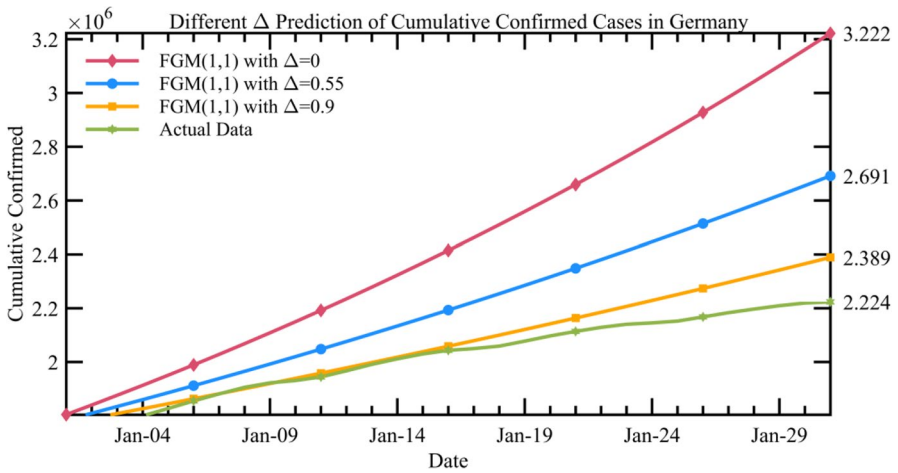


**Fig. 15** Predictive result of cumulative confirmed cases in France. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)

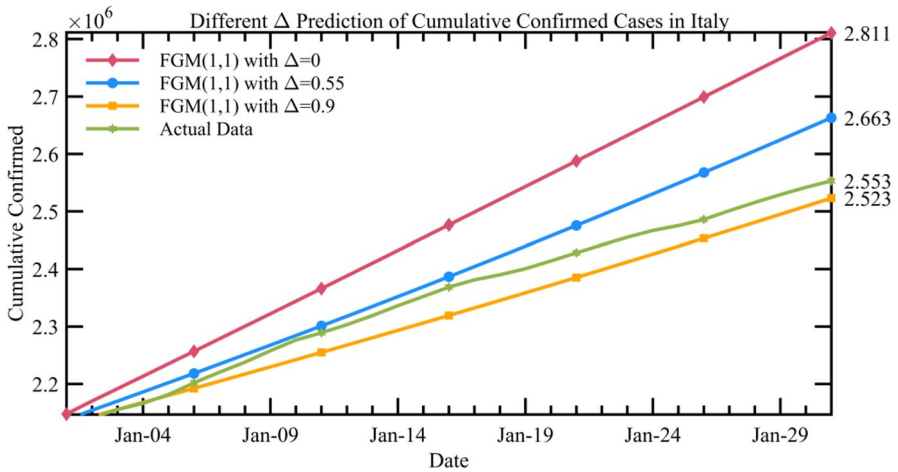




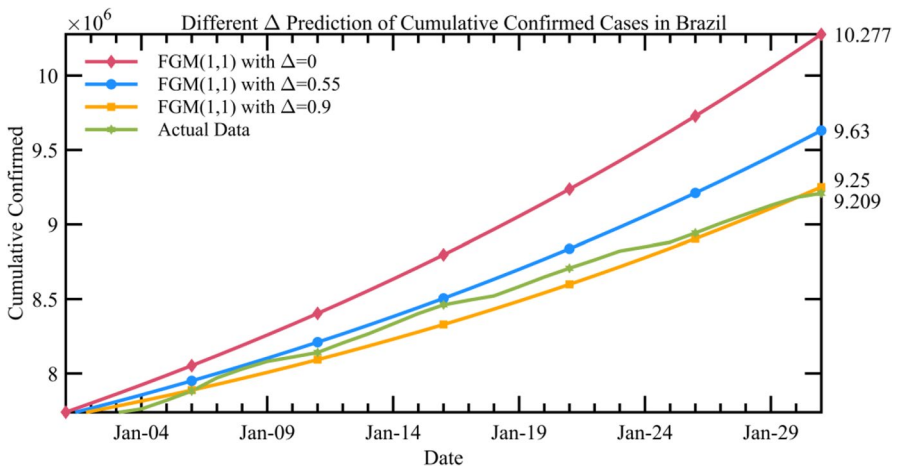
**Fig. 16** Predictive result of cumulative confirmed cases in the UK. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



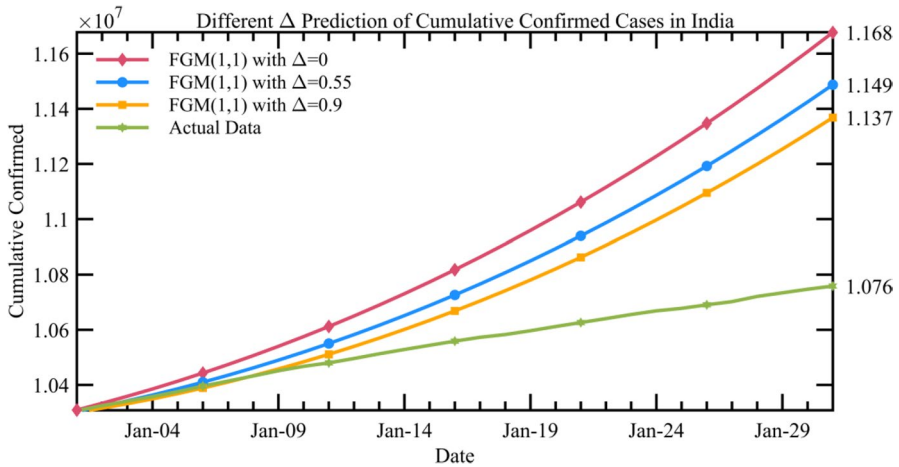
**Fig. 17** Predictive result of cumulative confirmed cases in Germany. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



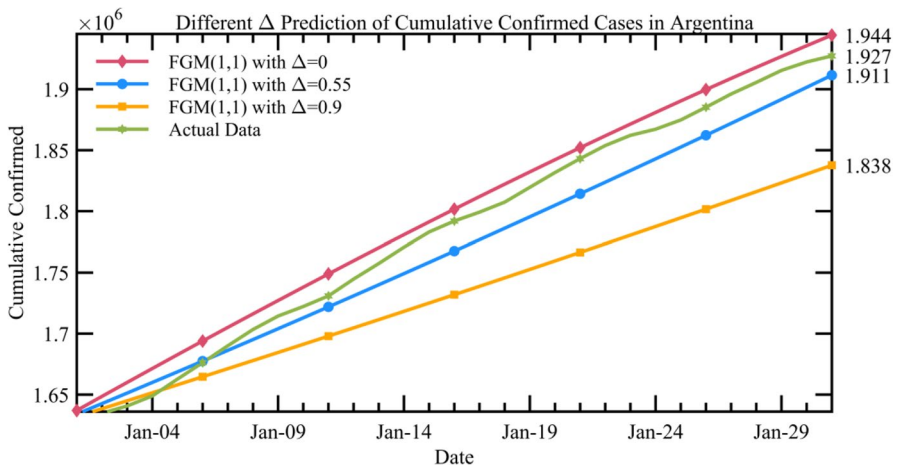
**Fig. 18** Predictive result of cumulative confirmed cases in Italy. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



**Fig. 19** Predictive result of cumulative confirmed cases in Brazil. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



**Fig. 20** Predictive result of cumulative confirmed cases in India. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)



**Fig. 21** Predictive result of cumulative confirmed cases in Argentina. Note: The green line depicts the changes of actual cumulative confirmed cases and the other three lines depicted the changes of predictive cumulative confirmed cases using different FGM (1,1) model. FGM (1,1) with buffer level of 0 means that we do not apply the buffer level in the FGM (1,1). FGM (1,1) with buffer level of 0.55 means that we apply a buffer level of 0.55 in the FGM (1,1). FGM (1,1) with buffer level of 0.9 means that we apply a buffer level of 0.9 in the FGM (1,1)

Argentina has increased slowly since the end of October. The actual number of confirmed cases in Brazil in January 31 fall in the range of the predicted value with buffer levels from 0 to 0.9, and it is most close to the predictive result with a buffer level of 0 (Fig. 21).

#### 4.10 Model Accuracy for Different Countries

We forecast the cumulative number of confirmed cases in Japan, the USA, France, the UK, Germany, Italy, Brazil, India, and Argentina on January 31, including about 402,417 in Japan, 27,271,051 in the USA, 3,096,515 in France, 4,130,429 in the UK, 2,389,240 in Germany, 2,523,042 in Italy, 9,250,362 in Brazil, 11,368,243 in India, and 1,944,242 in Argentina. Table 6 reports the predictive accuracy of the January test set (from January 1 to January 14) for different countries. The predictive accuracy is calculated in both MAPE and RMSE. FGM (1,1) with buffer levels show a strong capacity to predict the cumulative confirmed cases in these countries. The MAPEs of the test sets in these countries are less than 5%.

## 5 Conclusion

In this paper, we adopt the rolling FGM (1,1) to predict the cumulative number of confirmed cases. FGM is one of the forefront methods in grey system modeling. It makes full use of known information under the condition of less data environment and has an excellent performance in predicting the future trend. We have a marginal contribution on proposing the concept of buffer level based on the general form of WBO. Combining buffer level with FGM in an appropriate way can improve the forecasting accuracy. The example of Italy, France, and the USA shows that government measures require the active cooperation of the public and often have a time lag when they are effective. Only when a government increases its stringency and the public observe the order can the spread of COVID-19 be slowed down. The practical significance of this paper lies in the comprehensive consideration of government control measures and public response in the prediction.

Our main marginal contribution lies in the concept of buffer level, which takes into account both government control measures and public response in predicting cumulative confirmed cases. In practice, we find that only when government control measures are in place and people response actively to these measures can the spread of COVID-19 be

**Table 6** Model accuracy (from Jan 1, 2021, to Jan 31, 2021)

Country	USA	France	UK	Germany	Japan	Italy	Brazil	India	Argentina
MAPE (%)	1.626	1.23	3.204	2.229	2.444	1.27	0.753	1.716	0.762
RMSE	453,012.06	48,079.37	128,784.87	61,606	9546.47	34,108.64	73,211.53	263,899.79	13,966.52

Note: This table shows our predictive accuracy for the test set from January 1 to January 31, and the MAPE of the test set in these countries is less than 5%

**Appendix**

**Table 7** The results of augmented Dickey-Fuller Test and the ARMA specifications

	USA	France	Italy	UK	Germany	China	Japan	Brazil	India	Argentina
	I(1)	I(0)	I(1)	I(0)	I(0)	I(0)	I(2)	I(0)	I(2)	I(0)
Trend and Intercept	Trend and Intercept	Trend and Intercept	Trend and Intercept	Trend and Intercept	Trend and Intercept	Trend and Intercept	None	Trend and Intercept	None	Trend and Intercept
<i>t</i> -Statistic	-6.0607	-3.8878	-7.2911	-3.2128	-3.2019	-3.3575	-4.5078	-4.1812	-1.8813	-4.6810
Prob.*	0.0000	0.0137	0.0000	0.0839	0.0859	0.0591	0.0000	0.0055	0.0573	0.0009
Model	AR(1) with trend	ARMA(2,1) with trend	ARMA(1,1) with trend	ARMA(1,1) with trend	ARMA(2,1) with trend	ARMA(1,1) with trend	ARMA(1,1)	ARMA(1,1) with trend	ARMA(1,1)	ARMA(2,1) with trend

slowed down. Appropriate use of buffer level according to government stringency and mobility can improve the predictive accuracy of FGM (1,1). Buffer levels are obtained from countries' historical data, which can be used to calculate the predicted value in various scenarios. We predict the cumulative confirmed cases in various countries with buffer levels of 0.55 and 0.9. Based on the Government Response Stringency Index and Mobility Index in December, we forecast the most likely results at the end of January. Compared with the actual data in January, FGM (1,1) with buffer level has high predictive accuracy.

Based on rolling FGM (1,1) with different buffer levels and indices in December, we forecast the cumulative number of confirmed cases in Japan, the USA, France, the UK, Germany, Italy, Brazil, India, and Argentina from January 1 to 31. The conclusion indicates that countries around the world should cooperate to control the second wave of the COVID epidemic as soon as possible.

**Author Contribution** Lo, Kai Lisa wrote the paper.

Zhang, Minglei programmed with Matlab.

Chen, Yanhui designed the research and compared the results.

Mi, Jinhong Jackson designed the research and collected the data.

**Funding** This work is supported by the National Natural Science Foundation of China [No. 71701127, 72073086], National Statistical Science Research [No. 2020LY048], and Shanghai Soft Science Research Funds [No. 20692191100].

**Availability of Data and Materials** All the data are open data.

We collected them from GitHub (<https://covid19datahub.io/articles/doc/data.html>).

**Code Availability** In this paper, we coded with Matlab R2013b.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

## References

- Abuhasel KA, Khadr M, Alquraish MM (2020) Analyzing and forecasting COVID-19 pandemic in the Kingdom of Saudi Arabia using ARIMA and SIR models. *Comput Intell* <https://doi.org/10.1111/coin.12407>
- Chen CI (2008) Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate. *Chaos Solitons Fractals* 37(1):278–287
- Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* 135:109864. <https://doi.org/10.1016/j.chaos.2020.109864>
- Choi SB, Ahn I (2020) Forecasting imported COVID-19 cases in South Korea using mobile roaming data. *PLOS ONE* 15(11):e0241466. <https://doi.org/10.1371/journal.pone.0241466>
- Dang Y, Liu S, Liu B, Tang X (2004) Study on the buffer weakening operator. *Chin J Manag Sci* 12(2):109–102
- Dang Y, Liu S, Mi C (2007) Study on characteristics of the strengthening buffer operators. *Control Decis* 22(7):730–734
- de Lima CL, da Silva CC et al (2020) COVID-SGIS: a smart tool for dynamic monitoring and temporal forecasting of Covid-19. *Front Public Health* 8. <https://doi.org/10.3389/fpubh.2020.580815>
- Deng J (2002) *The basis of grey theory*. Press of Huazhong University of Science & Technology, Wuhan
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20(5):533–534
- Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C (2020) Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* 584(7821):420–424

- Hazarika BB, Gupta D (2020) Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks. *Appl Soft Comput* 96:106626. <https://doi.org/10.1016/j.asoc.2020.106626>
- Huang C, Wang Y et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395(10223):497–506
- Ibrahim MA, Al-Najafi A (2020) Modeling, control, and prediction of the spread of COVID-19 using compartmental, logistic, and Gauss models: a case study in Iraq and Egypt. *Processes* 8(11):1400
- Ju-Long D (1982) Control problems of grey systems. *Syst Control Lett* 1.1(5):288–294
- Listings of WHO's response to COVID-19 (2020), available at: <https://www.who.int/news/item/29-06-2020-covidtimeline>
- Liu L, Chen Y, Wu L (2020a) Forecasting confirmed cases, deaths, and recoveries from COVID-19 in China during the early stage. *Math Problems Eng* 2020. <https://doi.org/10.1155/2020/1405764>
- Liu M, Thomadsen R, Yao S (2020b) Forecasting the spread of COVID-19 under different reopening strategies. *Sci Rep* 10(1):1–8
- Liu S (1991) The three axioms of buffer operator and their application. *J Grey Syst* 3(1):39–48
- Liu S, Dang Y, Fang Z, Xie N (2013) Grey system theory and its application (in Chinese). Science Press
- Mahanty C, Kumar R, Mishra BK, Hemanth DJ, Gupta D, Khanna A (2020) Prediction of COVID-19 active cases using exponential and non-linear growth models. *Expert Syst* e12648. <https://doi.org/10.1111/exsy.12648>
- Malki Z, Atlam E et al (2020) ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Comput Appl* 1–20
- Marinho PRD et al (2020) Covid-19 in Brazil: A sad scenario. *Cytokine Growth Factor Rev* 10:51–54
- Mavragani A, Gkillas K (2020) COVID-19 predictability in the United States using Google Trends time series. *Sci Rep* 10(1):1–12
- Mukherjee K (2021) COVID-19 situation in India: Artifact or Fact?. Working Paper, available at: <https://www.cambridge.org/engage/coe/article-details/604cacfb4a6f4a0aa3339c7>
- Şahin U, Şahin T (2020) Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fraction. *Chaos Solitons Fractals* 138:109948. <https://doi.org/10.1016/j.chaos.2020.109948>
- Sarkar K, Khajanchi S, Nieto JJ (2020) Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals* 139:110049. <https://doi.org/10.1016/j.chaos.2020.110049>
- Sharma SK, Bhardwaj S, Bhardwaj R, Alowaidi M (2021) Nonlinear time series analysis of pathogenesis of COVID-19 pandemic spread in Saudi Arabia. *CMC-Comput Mater Continua* 66(1):805–825
- Singh KK, Kumar S, Dixit P, Bajpai MK (2020a) Kalman filter based short term prediction model for COVID-19 spread. *Appl Intell* 1–13
- Singh S, Sundram BM et al (2020b) Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *J Infect Dev Ctries* 14(9):971–976
- Venkatesan M, Mathivanan SK et al (2020) Forecasting of the SARS-CoV-2 epidemic in India using SIR model, flatten curve and herd immunity. *J Ambient Intell Humaniz Comput* 1–9
- Wang Z, Dang Y, Liu S (2009a) Study on buffer operators with variable weights and their effect strength to original sequence. *Control Decis* 24(8):1218–1222
- Wang Z, Dang Y, Liu S (2009b) Unbiased grey Verhulst model and its application. *Syst Eng Theory Pract* 29(10):138–144
- Wang Z, Hipel KW, Wang Q, He S (2011) An optimized NGBM (1, 1) model for forecasting the qualified discharge rate of industrial wastewater in China. *Appl Math Model* 35(12):5524–5532
- Wei Y, Kong X (2010) Constructing methods of several kinds of strengthening and weakening buffer operators and their inner link. *Control Decis* 2(25):196–202
- Wu L, Liu S, Yao L, Yan S, Liu D (2013) Grey system model with the fractional order accumulation. *Commun Nonlinear Sci Numer Simul* 18(7):1775–1785
- Yousaf M, Zahir S, Riaz M, Hussain SM, Shah K (2020) Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos Solitons Fractals* 138:109926. <https://doi.org/10.1016/j.chaos.2020.109926>
- Zeng X, Ghanem R (2020) Dynamics identification and forecasting of COVID-19 by switching Kalman filters. *Comput Mech* 66(5):1179–1193
- Zhao Y, Shou M, Wang Z (2020) Prediction of the number of patients infected with covid-19 based on rolling grey verhulst models. *Int J Environ Res Public Health* 17(12):4582

## Authors and Affiliations

Kai Lisa Lo<sup>1</sup>  · Minglei Zhang<sup>1</sup> · Yanhui Chen<sup>1</sup>  · Jinhong Jackson Mi<sup>1</sup>

✉ Yanhui Chen  
chenyh@shmtu.edu.cn

<sup>1</sup> School of Economics and Management, Shanghai Maritime University, No. 1550 Haigang Avenue, Pudong, Shanghai, People's Republic of China