A Rolling Optimized Nonlinear Grey Bernoulli Model RONGBM(1, 1) and application in predicting total COVID-19 infected cases

Accepted paper at The 2020 International Congress of GSUA

Hoang Anh NGO *
École Polytechnique
Institut Polytechnique de Paris
91120 Palaiseau, FRANCE

Thai Nam HOANG
Department of Mathematics and Computer Science
Beloit College
WI 53511, United States

August 13, 2020

Abstract

The Nonlinear Grey Bernoulli Model NGBM(1, 1) is a recently developed grey model which has various applications in different fields, mainly due to its accuracy in handling small time-series datasets with nonlinear variations. In this paper, to fully improve the accuracy of this model, a novel model is proposed, namely Rolling Optimized Nonlinear Grey Bernoulli Model RONGBM(1, 1). This model combines the rolling mechanism with the simultaneous optimization of all model parameters (exponential, background value and initial condition). The accuracy of this new model has significantly been proven through forecasting Vietnam's GDP from 2013 to 2018, before it is applied to predict the total COVID-19 infected cases globally by day.

Keywords GM(1, 1) · NGBM(1, 1) · Optimized NGBM(1, 1) · rolling mechanism · RONGBM(1, 1)

1 Introduction

At the beginning of 2020, there happened to be an out break of pneumonia, started from an unknown aetiology in Wuhan City, Hubei Province, China. A novel strain of coronavirus (COVID-19) was found from a cluster of patients in January 2020 [22]. Some scientists proposed a hypothesis that this strain of virus might have originated from bats that could be traced back to Huanan South China Seafood Market, a seafood market [43]. There has been evidence about person-to-person transmission since the middle of December 2019 [22].

The number of cases has increased exponentially since then and authorities all over the world have to promulgate laws to alleviate the widespread of disease. Europe made the strict act as to close borders of all Schengen area [5]. At the end of the first quarter, USA surpassed China to become the world's most infected territory at more than $100,000$ confirmed cases [11]. Roughly a month later, USA, again, became the first country to surpass 1 million confirm cases, however, the actually cases could be 10 times higher due to the lack of test kits [20, 33]. Most lately, after half a year of outbreak, the total confirm cases all over the world has reached 10 million, with more than $500,000$ people died. No country has succeeded in producing vaccination or special drugs, while the current situation is extremely complex and unpredictable, the need of data modelling and prediction is a must.

The number of COVID-19 infected cases can be regarded as a time series problem. According to previous research, time-varying process forecasting is one of the most important fields in terms of statistics. By collecting and analyzing different data points, researchers could investigate the situation and surmount current problem. However, most of these problems require a significant amount of data, which might be difficult due to the lack of information, or the ambiguity and non-linearity of data. For example, data of exchange rate between two currencies may fluctuate throughout a day or a period of time. Essentially, there is an urge to develop a method to solve this real-world dilemma.

*Corresponding author, Tel: (+33) 7 61 86 80 94, Email address: hoang-anh.ngo@polytechnique.edu

A lot of methodologies have been developed to solve the time series forecasting problems. Some can be implied, including traditional statistical methods like AR (Autoregressive Model) [13, 10, 36], MA (Moving Average) [19], ARIMA (Autoregressive Integrated Moving Average) [2, 1] with abilities to predict linear trend, but are limited to fuzzy data. Others algorithms that can break through the cyclic of information and predict it can be listed out as LSTM (Long Short-term Memory) [15, 32], SVM (Support Vector Machine) [35], and Sliding Window [28]. Additionally, there are some hybrid models that can outperform others with just limits information and can improve forecasting results.

To deal with the limitation of data, Grey method has been brought into consideration. Grey method was proposed by Deng Ju-Long in March, 1982 [25]. In Grey system theory [24, 25], considered the degree of information, a white system is a system in which all information is known, while a black system is a system contains all unknown parts. Additionally, a system which consists of both known and unknown information is called a Grey system [25]. According to Guo, R. [18], the Grey system outperforms other methods with just a limited number of discrete data, to achieve an insignificant margin of error of prediction versus real value. The system of total COVID-19 infected cases can be treated as a Grey system due to its constraint of data.

Mathematically, the traditional Grey predicting model is based on the least square reduction and the first-order linear ordinary differential equation. For instance, traditional GM(1, 1) can be taken into account [21, 44, 37] to solve this problem. Notably, to reinforce the predicting accuracy, some researchers have studied and combined to create hybrid Grey models, such as grey-Markov [21, 27], grey-Fourier [42], grey-Taguchi [14]. Some more hybrid models such as Grey rolling mechanism [3, 7, 23] or nonlinear Grey Bernoulli method [38, 9, 41, 26] can also be taken into consideration. In this paper, rolling mechanism is combined with nonlinear Grey Bernoulli method to propose a rolling optimized nonlinear Grey Bernoulli model, or RONGBM(1, 1). For benchmarking, AR, Artificial Neural Network (ANN), and LSTM are used to compare the accuracy with RONGBM(1, 1). This hybrid model is used to optimize the accuracy of predicting the total COVID-19 infected cases with just 11 entries of data (from January 28th, 2020 to February 7rd, 2020). The beginning date was chosen due to the fact that, there might have been uncleared data on those previous days.

The remainder of this paper will be organized as follows. In Section 2, we will provide a brief description of the traditional Grey models, including GM(1, 1) and the Nonlinear Grey Bernoulli Model GM(1, 1). Section 3 briefly introduces methods of optimizing parameters simultaneously in the formation of the optimized ONGBM(1, 1) model. In section 4, the application of the rolling mechanism is presented to propose a new RONGBM(1, 1) model. Section 5 provides some statistical and modern machine learning models as a benchmarking tool for the proposed model in the previous section. Section 6 introduces some modern machine learning models as an additional benchmarking tool for the newly proposed model in the empirical examples. Section 7 adopts the example of predicting Vietnam's GDP to demonstrate the accuracy of the new RONGBM(1, 1) model in comparison with models in Section 5, then applies this new model into predicting the total COVID-19 infected cases until `2020-02-18`. Section 8 concludes the paper by drawing some conclusions and future work.

2 Description of the traditional GM(1, 1) and NGBM(1, 1) models

In this section, a brief introduction of the mathematical modelling of two traditional models: GM(1, 1) and Nonlinear Grey Bernoulli Model NGBM(1, 1) is introduced.

2.1 Building the traditional Grey model GM(1, 1)

Assume that $x^{(0)}$ is the non - negative original historical time series of data with m entries as:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(k), ..., x^{(0)}(m)\} \tag{1}$$

Next, we define $x^{(1)}$ using one-time accumulated generating operation (1 - AGO), which is as

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(k), ..., x^{(1)}(m)\} \tag{2}$$

where

$$\begin{cases} x^{(1)}(1) = x^{(0)}(1) \\ x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(k), k = 2, 3, ..., m \end{cases} \tag{3}$$

As $x^{(1)}$ is a monotonic increasing sequence which is similar to the solution of a first-order linear differential equation, one can assume that the solution of the following differential equation

$$\frac{d\hat{x}^{(1)}}{dt} + a\hat{x}^{(1)} = b \tag{4}$$

represents the Grey predicted value complement to the initial condition $\hat{x}^{(1)}(1) = x^{(0)}(1)$ and parameters a and b.

By definition, $\frac{d\hat{x}}{dt}$ can be written as:

$$\frac{d\hat{x}^{(1)}}{dt} = \lim_{\Delta t \to 0} \frac{\hat{x}^{(1)}(t + \Delta t) - \hat{x}^{(1)}(t)}{\Delta t} \tag{5}$$

However, to discretize the differential equation, Δt can be set to be equal to 1, which means that (5) can be rewritten as

$$\frac{d\hat{x}^{(1)}}{dt} = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) = x^{(1)}(t+1) - x^{(1)}(t) = x^{(0)}(t+1) \tag{6}$$

The Grey predicted value is now defined as

$$\hat{x}^{(1)}(t) \approx Px^{(1)}(k) + (1-P)x^{(1)}(k+1) = z^{(1)}(k+1), k = 1, 2, 3, ..., m \tag{7}$$

with P usually set as $\frac{1}{2}$ in the traditional models.

The differential equation can now be discretized as:

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{8}$$

By the least-squared method, the coefficients a and b can be determined by

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y \tag{9}$$

with

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ ... & \\ -z^{(1)}(m) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ ... \\ x^{(0)}(m) \end{bmatrix} \tag{10}$$

The particular solution of equation (4) with the initial condition is:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} + \frac{b}{a}, k = 1, 2, 3, ..., m-1 \tag{11}$$

The prediction of the historical time series of data at point $k+1$ can now be deduced by:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{-a}) \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} \tag{12}$$

with

$$\begin{cases} \hat{x}^{(0)}(k) = x^{(0)}(k), k = 1, 2, ..., m \text{ fitted values,} \\ \hat{x}^{(0)}(m+1), \hat{x}^{(0)}(m+2), ..., x^{(0)}(m+h) \text{ predicted values} \end{cases} \tag{13}$$

2.2 The Nonlinear Grey Bernoulli Model NGBM(1, 1)

In order to obtain higher accuracy in predicting comparing to the original GM(1, 1) model, Professor Chen [4] proposed the Nonlinear Bernoulli Grey Model NBGM(1, 1) as follows:

Similarly to the traditional Grey model GM(1, 1), assume that $x^{(0)}$ is the non - negative original historical time series of data with m entries as:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(k), ..., x^{(0)}(m)\} \tag{14}$$

Next, we define $x^{(1)}$ using one-time accumulated generating operation (1 - AGO), which is as

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(k), ..., x^{(1)}(m)\} \tag{15}$$

where

$$\begin{cases} x^{(1)}(1) = x^{(0)}(1) \\ x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(k), k = 2, 3, ..., m \end{cases} \tag{16}$$

As indicated previously, equation (4) is a linear differential equation. A similar form of this equation, which is nonlinear and has the form of:

$$\frac{d\hat{x}^{(1)}}{dt} + a\hat{x}^{(1)} = b\left[\hat{x}^{(1)}\right]^n \tag{17}$$

where $n \in \mathbb{R}$ any real number is called a Bernoulli equation, or the with then differential equation of the NGBM(1, 1) model.

The background value is now also defined as:

$$z^{(1)}(k+1) = (1-P)x^{(1)}(k) + Px^{(1)}(k+1) \tag{18}$$

with $P = \frac{1}{2}$ for the traditional model. Then, discretizing the ODE, one obtains

$$x^{(0)}(k) + az^{(1)}(k) = b\left[x^{(1)}(k)\right]^n \tag{19}$$

which is called the basic Grey differential equation of the NGBM(1, 1) model. It can easily be recognized that for $n = 0$, this equation turns to equation (4), which is the traditional GM(1, 1) model; for $n = 2$, the equation turns to the Grey-Verhulst equation.

By the least square method, the parameters a and b can be determined by:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y \tag{20}$$

with

$$B = \begin{bmatrix} -z^{(1)}(2) & \left[z^{(1)}(2)\right]^n \\ -z^{(1)}(3) & \left[z^{(1)}(3)\right]^n \\ \dots & \\ -z^{(1)}(m) & \left[z^{(1)}(m)\right]^n \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(m) \end{bmatrix} \tag{21}$$

The particular solution of equation (17), or the discrete time function, with the initial condition is:

$$\hat{x}^{(1)}(k) = \left[\left(x^{(0)}(1)^{(1-n)} - \frac{b}{a}\right)e^{-a(1-n)(k-1)} + \frac{b}{a}\right]^{\frac{1}{1-n}}, k = 1, 2, 3, ..., m \tag{22}$$

The prediction of the historical time series of data at point k can now be deduced by:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \tag{23}$$

3 The Optimized Nonlinear Grey Bernoulli Model ONGBM(1, 1)

3.1 Optimization of the exponential parameter

Based on the basic principle of the information overlapping of the Grey system, Wang et al. (2009) [37] proposed a formula determining the value of the exponential parameter by:

$$n = \frac{1}{m-2}\sum_{k=2}^{m-1}\gamma(k) \tag{24}$$

with

$$\gamma(k) = \frac{\begin{array}{c}\left[x^{(0)}(k+1) - x^{(0)}(k)\right] \times z^{(1)}(k+1) \times z^{(1)}(k) \times x^{(0)}(k) \\ -\left[x^{(0)}(k) - x^{(0)}(k-1)\right] \times z^{(1)}(k+1) \times z^{(1)}(k) \times x^{(0)}(k+1)\end{array}}{\left[x^{(0)}(k+1)\right]^2 \times z^{(1)}(k) \times x^{(0)}(k) - \left[x^{(0)}(k)\right]^2 \times z^{(1)}(k+1) \times x^{(0)}(k+1)} \tag{25}$$

However, with the existence of modern programming languages, this formula yields the problem of inflexibility. Moreover, as further investigated in **Section 7**, this formula yield a result which is significantly less effective, comparing to the method of iterating on a pre-defined range of the exponential parameter of $[-1, 1)$.

3.2 Optimization of the background value estimation

According to Wang et al. (2011) [38], the precision of any Grey model is highly dependent on the value of P in the background value $z^{(1)}(k)$. Change et al. (2005) [7] produced an even stronger statement by stating that forecasting accuracy of a Grey prediction model can be improved by optimizing parameter P.

The background value $z^{(1)}(k)$ of any (derived) Grey model is considered to be an approximation of the integral region generated by the curve of the function $x^{(1)}$ and the abscissa axis within the interval $[k-1, k]$. Mathematically, this value is expressed as:

$$z^{(1)}(k) = \int_{k-1}^{k} x^{(1)} dx \tag{26}$$

Applying the mean value theorem, it can be derived that:

$$z^{(1)}(k) = Px^{(1)}(k) + (1-P)x^{(1)}(k-1), P \in [0,1] \tag{27}$$

which are exactly equations (8) and (18). Mentioned beforehand, $P = \frac{1}{2}$ is the usually chosen values for traditional Grey models. However, it is trivial that the accuracy of the model depends heavily on the parameter P, which means that fixing this parameter at $\frac{1}{2}$ as usual will reduce massively the effectiveness when there are values fluctuating and deviating from the overall trend.

Zhang (1993) [45] has proven that the parameter P determining the background values and parameter a satisfy the following condition:

$$P = \frac{1}{a} - \frac{1}{e^a - 1} \tag{28}$$

From this formula, we derive:

$$\begin{aligned}
\lim_{a\to 0} P &= \lim_{a\to 0} \frac{1}{a} - \frac{1}{e^a-1} = \lim_{a\to 0} \frac{e^a - 1 - a}{a(e^a-1)} \\
&= \lim_{a\to 0} \frac{e^a - 1}{ae^a + e^a - 1} \left(\text{L'Hôpital rule on } \frac{f(x)}{g(x)}\right) \\
&= \lim_{a\to 0} \frac{e^a}{ae^a + e^a + e^a} \left(\text{L'Hôpital rule on } \frac{f'(x)}{g'(x)}\right) \\
&= \frac{e^0}{0 \times e^0 + e^0 + e^0} \\
&= \frac{1}{2}
\end{aligned} \tag{29}$$

which is a valid explanation for the setting of the value of P to $\frac{1}{2}$ for traditional Grey prediction models. However, when the value of a is significant (the absolute value of a is large enough), customarily setting the coefficient of P to $\frac{1}{2}$ is incorrect, reducing the efficiency of the model.

As a result, a new calculating formula was proposed by Tan (2000) [34] as:

$$z^{(1)}(k) = \frac{1}{2q}\left[(q+1)x^{(1)}(k-1) + (q-1)x^{(1)}(k)\right], k = 2, 3, ..., m \tag{30}$$

with q being given empirically by [26]:

$$q = \left(\sum_{i=2}^{n} \frac{x^{(1)}(k)}{x^{(1)}(k-1)}\right)^{\frac{1}{m-1}} + (m-1) \tag{31}$$

This means that, we have defined the value of P by:

$$P = \frac{q+1}{2q} = \frac{1}{2} + \frac{1}{2q} \tag{32}$$

Once again, this formula yields the problem of inflexibility. As 1-AGO sequence is a strictly increasing sequence,

$$\frac{x^{(1)}(k)}{x^{(1)}(k-1)} > 0, k = 2, 3, ..., m \tag{33}$$

which means that

$$q = \left(\sum_{i=2}^{m} \frac{x^{(1)}(k)}{x^{(1)}(k-1)} \right)^{\frac{1}{m-1}} + (m-1) > 0 \tag{34}$$

leading to

$$P = \frac{1}{2} + \frac{1}{2q} > \frac{1}{2} \tag{35}$$

which has already omitted the range of $\left[0, \frac{1}{2}\right]$ for consideration. Moreover, as applied later, upon applying this formula, the algorithm yields higher error than the empirically derived value of the same parameter. The optimal solution for this problem would also be iterating on a well-defined range of values of P ($[0, 1]$) with a small enough step size.

Within the previous session, the parameters P and n are not optimized simultaneously.

3.3 Optimization of the initial condition

Considering the particular solution of equation (17), the initial condition is set to be $x^{(0)}(1)$, the first point of the original time series of data. Using the principle of new information prior choosing, the last item of the 1 - AGO series, $x^{(1)}(m)$, can be used to enhance the accuracy of the model. Dang (2004) [12] has confirmed this theory by testing on two different Grey models: GM(1, 1) and the Grey-Verhulst model. Replacing $x^{(1)}(m)$ for ${}^{(0)}(1)$ as the initial condition, the time response function for the whitening equation of the NGBM(1, 1) model can be rewritten as:

$$\hat{x}^{(1)}(k) = \left[\left(x^{(1)}(m)^{(1-n)} - \frac{b}{a} \right) e^{-a(1-n)(k-m)} + \frac{b}{a} \right]^{\frac{1}{1-n}} \tag{36}$$

Furthermore, since the model parameters a and b are obtained with least square method, there will be high chances that the model will not pass through the point $x^{(0)}(n)$, As a result, the new initial condition $x^{(1)}(m)$ will require correction in order to maximize optimization. This means that the new initial term, instead, would be $x^{(1)}(m) + c$.

The correction parameter c would be chosen in order to minimize the following function (proposed by Lu et al. 2016 [26]):

$$f(c) = \sum_{k=1}^{m} \left\{ \left[\hat{x}^{(1)}(k) \right]^{1-n} - \left[x^{(1)}(k) \right]^{1-n} \right\}^2 \tag{37}$$

Substituting the discrete time response function deduced from **Theorem 1**, we obtain:

$$f(c) = \sum_{k=1}^{m} \left\{ \frac{b}{a} + \left[\left(x^{(1)}(m) + c \right)^{1-n} - \frac{b}{a} \right] e^{-a(1-n)(k-m)} - \left[x^{(1)}(k) \right]^{1-n} \right\}^2 \tag{38}$$

Let

$$\begin{cases} E(k) = e^{-a(1-n)(k-m)} \\ A(k) = \left[x^{(1)}(k) \right]^{1-n} - \frac{b}{a} \left(1 - E(k) \right) \end{cases} \tag{39}$$

Then, $f(c)$ will be rewritten as:

$$f(c) = \sum_{k=1}^{m} \left\{ \left[x^{(1)}(k) + c \right]^{1-n} E(k) - A(k) \right\}^2 \tag{40}$$

To find the minimum value of $f(c)$, we set the first - order condition (FOC) to 0, which means that $\frac{\partial f(c)}{\partial c} = 0$, or:

$$2(1-n) \left[x^{(1)}(k) + c \right]^{-n} \sum_{k=1}^{m} \left\{ \left[x^{(1)}(k) + c \right]^{1-n} E(k)^2 - A(k)E(k) \right\} = 0 \tag{41}$$

We deduce

$$\left[x^{(1)}(m) + c \right]^{1-n} = \frac{\sum_{k=1}^{m} A(k)E(k)}{\sum_{k=1}^{m} E(k)^2} \tag{42}$$

which means that the discrete time response function of the NGBM(1, 1) model can be now rewritten as:

$$\hat{x}^{(1)}(k) = \left[\left(\frac{\sum_{k=1}^{m} A(k)E(k)}{\sum_{k=1}^{m} E(k)^2} - \frac{b}{a} \right) e^{-a(1-n)(k-m)} + \frac{b}{a} \right]^{\frac{1}{1-n}} \tag{43}$$

3.4 Proposing algorithm

Based on the simultaneous optimization of three parameters: the exponential parameter, the initial condition and the background value, we propose a further optimized algorithm of the Optimized NGBM(1, 1) given by Wu et al. (2019) [41]. This algorithm is more practical and less time-consuming upon implementing in the currently available computer languages (R, Python or Matlab):

1. Given the original time series of the data $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(m)\}$, calculate the 1 - AGO sequence $x^{(1)}$.
2. Indicate the range of values for P and n. According to the mean value theorem, P must belong to the range $[0, 1]$, while n is usually considered within the range of $[-1, 1)$. Considering currently available computational tools, the step size is usually set at $0.001 - 0.01$.

   Remark: Empirically, for $n \notin [-1, 1)$, the model has been proven to yield higher error on average, reducing the efficiency of the model. As a result, the values of n outside the interval is not within our interest.
3. For the pair of values of P and n, calculate the matrices B and Y, which yields parameters a and b.
4. Calculate RPE and ARPE for each a and b. Then, choose the parameters which yield the minimum ARPE.
5. Calculate the values of $E(k)$ and $A(k)$ to deduce the error correction first initial condition $x^{(1)}(m) + c$. Then, compute $\hat{x}^{(1)}$ based on the proposed discrete time response function (44).
6. From $\hat{x}^{(1)}$, using the inverse AGO, compute $\hat{x}^{(0)}$.

4 The Rolling Optimized Nonlinear Grey Bernoulli Model RONGBM(1, 1)

As mentioned by Akay and Atak (2007) [3], the rolling mechanism is an efficient technique to increase forecasting accuracy of Grey prediction when there are exponential increasing/decreasing or chaotic data. This mechanism provides the ability to update input data by removing the oldest data points for each loop before applying the prediction algorithm. Its purpose is that, for each rolling step, the set of data points used for the upcoming forecasting is the most recent.

Wu et al. (2019) [41] is the first paper to ever mention the application of rolling mechanism to the NGBM(1, 1) to enhance accuracy. However, the proposed model did not reach the maximum efficiency due to the following reasons:

1. In the proposed algorithm, the optimization of the parameters P and n are determined by running a loop through a set of pre-defined value by the user. In this case, the author defined the set of values of P to be between 0 and 1, with the step of 0.01; while n is considered between -1 and 0.99, with the step of 0.01. This step size for both parameters is still large enough to miss the optimal value.
2. In **Algorithm 1**, the background value is set as

   $$z^{(1)}(k) = P(x^{(1)}(k) + x^{(1)}(k+1)), k = 1, 2, ..., m \tag{44}$$

   which is not the optimal approximation and not compatible with the mean value theorem.
3. As proposed, re-initiating the initial value to $x^{(1)}(m)$ will enhance accuracy of the algorithm. However, this can further be improved by correcting this initial value itself with the error correction parameter c proposed in the previous section.
4. The rolling mechanism implemented in **Algorithm 1** of [41] is incorrect. The rolling mechanism requires the insertion of the predicted value into the new array, while removing the oldest data point. In the algorithm, however, the author inserted the actual value instead of the predicted value, which develops biases for this model.

Assume that $x^{(0)}$ is a non - negative input of time series of data with m entries as:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(k), ..., x^{(0)}(m)\} \tag{45}$$

and the output sequence of predicted data would be:

$$\hat{x}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), ..., \hat{x}^{(0)}(k), ..., \hat{x}^{(0)}(m), \hat{x}^{(0)}(m+1), ..., \hat{x}^{(0)}(M)\}, M \geq m+1, M \in \mathbb{N} \tag{46}$$

We propose a rolling Optimized NGBM(1, 1) model with the following algorithm:

1. Set $i = 1$.
2. Set the training data to $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(m)\}$. Use this training data to predict the following data points:
$$\{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(m), x^{(0)}(m+1)\} \tag{47}$$
3. While $i \leq M - m$:
   (a) Set the new value of i as $i := i + 1$
   (b) Set the training data to $x^{(0)} = \{x^{(0)}(i), x^{(0)}(i+1), ..., x^{(0)}(m+i-1)\}$.
   (c) Apply the Optimized NGBM(1, 1) model on this new training data to predict $\hat{x}^{(0)}(m+i)$.
   Remark: For each i, there will be different values of the set of the model parameters $\{a, b, P, n\}$.
   (d) Return to the beginning of this step.
4. When $i = M - m$, we have had all the predicted values $\{\hat{x}^{(0)}(m+1), \hat{x}^{(0)}(m+1), ..., \hat{x}^{(0)}(M)\}$.
5. Our output sequence of predicted data would now be:
$$\hat{x}^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(k), ..., x^{(0)}(m), \hat{x}^{(0)}(m+1), ..., \hat{x}^{(0)}(M)\} \tag{48}$$

5 Performance evaluation

In order to examine the precision of each model, it is necessary to evaluate the difference between the fitted values and the actual values. For each specific time stamp k, the relative percentage error (RPE) calculates the relative difference of the forecast and recorded value by:

$$RPE = \epsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \times 100\%, k = 1, 2, 3, ..., m \tag{49}$$

where $x^{(0)}(k)$ is the actual value and $\hat{x}((0)(k)$ is the fitted value.

The overall precision of the model can be measured by the average relative percentage error (ARPE), which is determined by:

$$ARPE = \epsilon(avg) = \sum_{k=1}^{m} |\epsilon(k)| = \sum_{k=1}^{m} \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\% \tag{50}$$

The classification of the model precision based on ARPE is described in Table 1 below [41].

Table 1: ARPE classification of model precision

ARPE (%)	≤ 10	$10 \sim 20$	$20 \sim 50$	≥ 50
Classification	Excellent	Good	Reasonable	Unacceptable

Besides ARPE, the root-mean-square error (RMSE) or root-mean-square deviation (RMSD) is also used to evaluate the precision of the model:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{m} \left(x^{(0)}(k) - \hat{x}^{(0)}(k)\right)^2}{m}} \tag{51}$$

Based on the concept of RMSE, the posterior error ratio (c) is established and is also to evaluate the efficiency of a model:

$$c = \frac{\sqrt{\frac{1}{n}\sum_{k=1}^{m}(\epsilon(k) - \epsilon(avg))^2}}{\sqrt{\frac{1}{n}\sum_{k=1}^{m}(x^{(0)}(k) - \bar{x}^{(0)})^2}} \tag{52}$$

The precision classification with respect to different levels of posterior error ratio is described in the following table 2 [6].

Table 2: Posterior error ratio and precision rank

Posterior error ratio (c)	≤ 0.35	$0.35 \sim 0.5$	$0.5 \sim 0.65$	≥ 0.65
Classification	1 (Highly accurate)	2 (Qualified)	3 (Marginal)	4 (Disqualified)

Within the scope of this paper, RPE and ARPE will be used as the main tool of benchmarking and comparing efficiency between different models. This approach is the most popular and widely-used upon considering empirical results of any newly-proposed Grey model.

6 Comparison with alternative algorithms

In this section, we will indicate some alternative machine learning algorithms based on neural network. For this paper, we used 2 Neural Network models: Artificial Neural Network (ANN) and Long Short-term Memory (LSTM). Historically, recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) were first introduced to handle such heavy-load tasks, like Image Processing or Natural Language Processing, and thus they were too complex for typical time series problems. However, researchers have started to adopt these deep learning methods, and turned them into perfect fits for a time-varying process. There are 2 main properties that make NN a competitive candidate for time series. NNs have general nonlinear function mapping capability that can approximate any continuous function. This gives NNs a capability of handling any complex problems with limited data. Moreover, NN is a non-parametric data-driven model, hence it does not require (mostly) any constraints. This is the key point for solving nonlinear, non-pattern problems that may have unique characteristics that cannot be captured by parametric model [36].

6.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a brain-inspired system, which simulates the way human learns. A single layer neural network is called a Perceptron, which can give a single output.

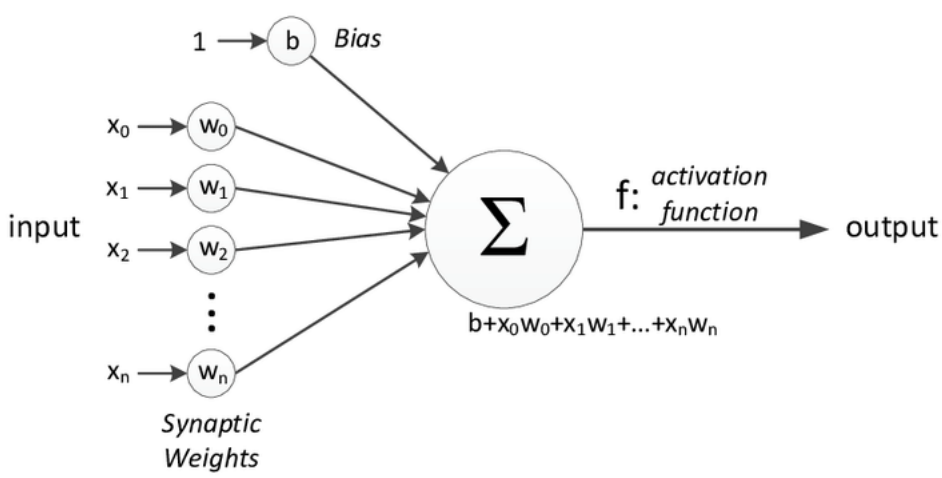


Figure 1: A Perception [8]

To use ANN, we have to decide an Activation function, which converts an input signal of a node in an ANN to an output signal. This output signal is used as input to the next layer in the stack [8]. There are 4 Activation functions that can be taken into consideration.

- Threshold Activation function (Binary Step function) A threshold-based binary activation function. For classification only (because of activated cases only 1 or 0).

$$f(x) = \begin{cases} 0, x < 0 \\ 1, x \geqslant 0 \end{cases} \tag{53}$$

- Sigmoid Activation function (Logistic function) A S-shaped curve or sigmoid, ranges between 0 and 1. For predicting the probability

$$\phi(z) = \frac{1}{1 + e^{-z}} \tag{54}$$

- Hyperbolic Tangent function (tanh) A similarity to sigmoid, but better in performance, ranges between -1 and 1. Perfect fit for layer stacking.

$$f(x) = tanh(x) \tag{55}$$

- Rectified Linear Units (ReLU) The most common Activation function in both CNN and ANN, ranges from zero to infinity $[0, \infty)$.

Based on the difference between the actual value and the predicted value, an error value called Cost Function is computed and sent back through the system. Between an input layer and an output layer is at least a hidden layer. The hidden layers perform computations on the weighted inputs and produce net input which is then applied with activation functions to produce the actual output. ANN uses a procedure called Back-propagation cyclically until the error value is kept at minimum [8].

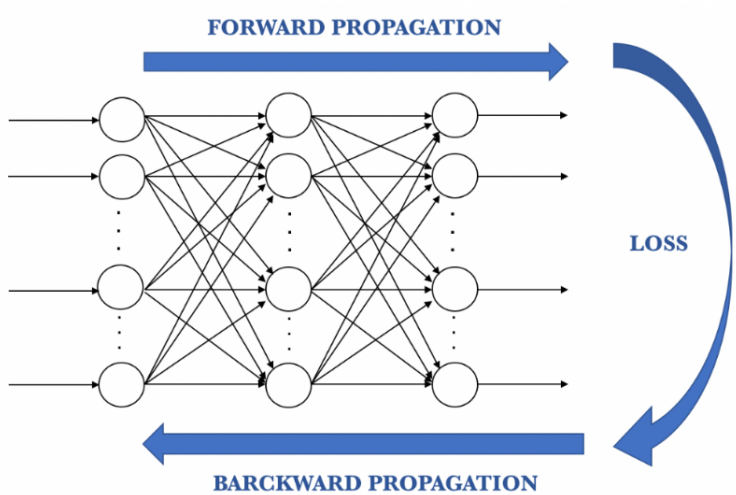


Figure 2: Back-Propagation procedure [8]

6.2 Long Short-term Memory (LSTM)

Long Short-term Memory (LSTM) is a RNN-based algorithm with the capability of learning long-term dependencies, which was proposed by Sepp Hochreiter and Jürgen Schmidhuber in 1997. Vanilla RNN method suffers from short-term memory during feed-forward. This means that, if processing data which has long time piece, it will be hard for RNN to carry (or remember) data from earlier time steps to later ones. Thus, RNN may leave out important information from the beginning. It also suffers from vanishing gradient problem in back-propagation. If a gradient value is too smal, it does not contribute much in learning process.

LSTM was created as a solution to this problem. LSTM has feedback connection, which allows it to process the whole data sequence. A general LSTM consists of multiple cells, with forget gates, an input and an output. They have the ability to choose which information to carry on or leave behind. [32, 15, 30].

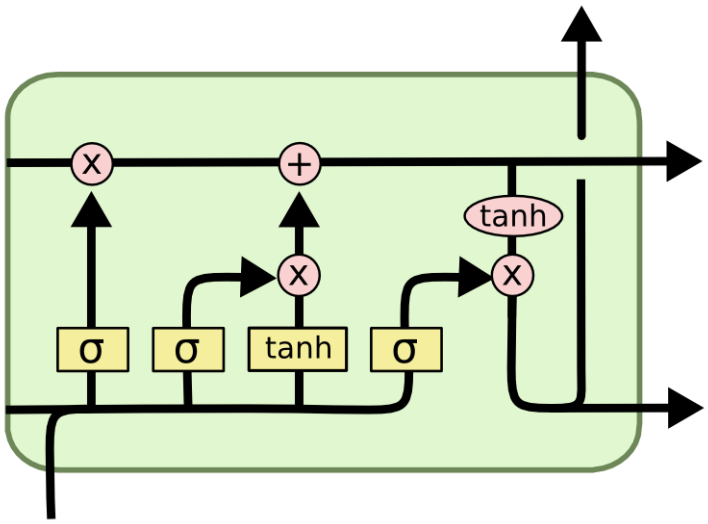


Figure 3: A LSTM cell [30]

7 Empirical applications

7.1 Efficiency testing for the Rolling Optmizied Nonlinear Grey Bernoulli Model RONGBM(1, 1)

7.1.1 Context

After the end of the Vietnam War in Spring 1975, Vietnam's economic status started to slow down and decline. Furthermore, the embargo of the United States on Vietnam and the collapse of the Soviet Union led to the plummet of the Vietnam economy, hit rock-bottom in 1989 - 1990 [16]. Since then, the economic status has revived significantly, with the annual growth rate of 6-8% [17].

As mentioned above, we used GDP data between 2005 and 2018, as from 2005 and after, Vietnam's GDP has some certain growth in economic status and annual growth rate. Moreover, Vietnam is on the verge of development and integration into the world, using this range of data will enhance the forecasting accuracy, as it can avoid the inaccurate that those previous world events bring.

In this section, we will regard the most important macroeconomics variable as our testing data: the annual Gross Domestic Product (GDP) of Vietnam, between 2005 and 2018. These data can be retrieved from World Bank Open Data (`https://data.worldbank.org/`)

7.1.2 Methodology

In order to demonstrate the accuracy and effectiveness of the newly proposed Rolling Optimized NGBM(1, 1) (abberivated as RONGBM(1, 1)), we will proceed to compare this model with the existing Grey models, including:

- The traditional Grey model GM(1, 1)
- The traditional Nonlinear Grey Bernoulli Grey Model NGBM(1, 1)
- The Optimized Nonlinear Grey Bernoulli Model ONGBM(1, 1). This model has adapted all possible optimization mentioned previously, including optimization of the background value, initial condition and exponential parameter.

The data from 2004 to 2013 will be used for training, and data from 2014 to 2018 will be used for forecasting and testing models.

Table 3: Results of GM(1, 1), NGBM(1, 1) and ONGBM(1, 1) in predicting Vietnam's GDP $2004 - 2018$

Year	Actual Data	GM(1, 1)		Formula-optimized NGBM(1, 1) $n = 0.126$		Nash NGBM(1, 1) $n = 0.13, P = 0.495$	
		Prediction	RPE (%)	Prediction	RPE (%)	Prediction	RPE (%)
2004	45.42785	45.42785	0.00	45.42785	0.00	45.42785	0.00
2005	57.63326	61.43522	6.60	57.62228	−0.02	57.55257	−0.14
2006	66.37166	70.01275	5.48	68.73623	3.56	68.75453	3.59
2007	77.41443	79.78786	3.07	79.99618	3.33	80.07765	3.44
2008	99.13030	90.92776	−8.27	91.99635	−7.20	92.12421	−7.06
2009	106.01466	103.62301	−2.26	105.05247	−0.91	105.21288	−0.75
2010	115.93175	118.09075	1.86	119.40416	3.00	119.58390	3.15
2011	135.53944	134.57846	−0.71	135.27036	−0.20	135.45557	−0.06
2012	155.82000	153.36817	−1.57	152.87099	−1.89	153.04632	−1.78
2013	171.22203	174.78129	2.08	172.43780	0.71	172.58566	0.79
2014	186.20465	199.18408	6.97	194.22121	4.30	194.32111	4.35
2015	193.24111	226.99396	17.47	218.49546	13.07	218.52332	13.08
2016	205.27617	258.68664	26.02	245.56317	19.63	245.49057	19.59
2017	223.77987	294.80421	31.74	275.75975	23.23	275.55313	23.13
2018	245.21369	335.96448	37.01	309.45795	26.20	309.07767	26.04
ARPE			**10.07**		**7.15**		**7.13**
Classification			Good		Excellent		Excellent

7.1.3 Results and analysis

From Table 3 and Table 4, it can easily be observed that the ARPE generated by the newly proposed Rolling Optimized NGBM(1, 1) model is significantly smaller than that of comparing models. This is due to the fact that the initial condition $x^{(1)}(m) + c$ and the original training data have been updated, which yields closer result to the actual data.

This empirical result also explains the conclusion driven in the previous sections that the estimating formula for the exponential parameter n and the background value P is not optimal, namely (25) and (31). Upon using the equations above, the values of P and n would be:

$$\begin{cases} P = 0.5484042 \\ n = -0.1387509 \end{cases} \tag{56}$$

which yield higher ARPE (2.8314%) for the first 10 GDP values between $2004 - 2013$ comparing to the parameters derive empirically below with ARPE of 1.3855%.

Moreover, as mentioned above, the empirical value of $P = 0.495$ cannot be captured by equation (31).

7.2 Applying and benchmarking RONGBM(1, 1) in predicting 2019 - nCoV total infected cases

7.2.1 Context

Right before the beginning of 2020, a new strain of novel coronavirus (COVID-19) was found from of patients in the area of Wuhan City, Hubei Province, China. Full-length genome sequences were obtained from patients at early state showed that this new type of disease shared 79.5% sequence identify to SARS-CoV in 2003. This COVID-19 genome could also be traced back to bats coronavirus [31, 43, 22]. Researchers ared scared that this could be a new epidemic similar to the SARS-CoV and MERS-CoV epidemic in 2003 and 2012 separately that terrified worldwide.

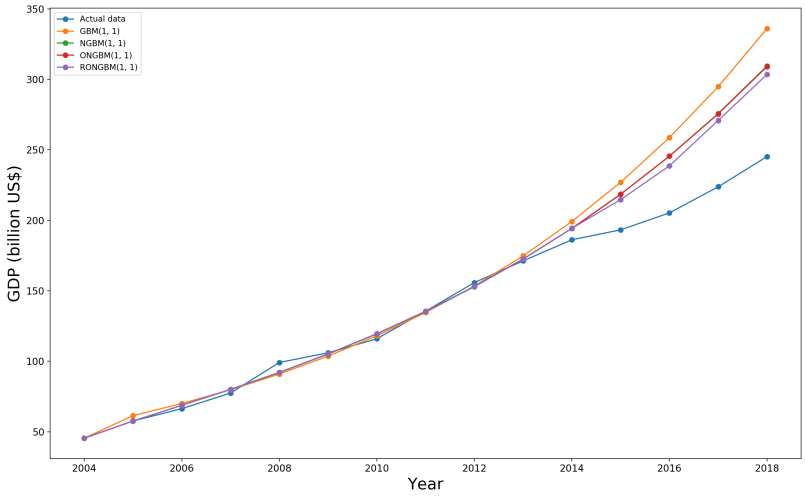
7.2.2 Methodology

The total infected cases were retrieved free of charge from Worldometer (`https://www.worldometers.info/coronavirus/coronavirus-cases/`). We decided to use `2020-01-28` as the starting data point, as it would be more precise in terms of medical data collecting and analysis.

Table 4: Parameters and results of RONGBM(1, 1) in predicting Vietnam's GDP $2004 - 2018$

Year	Actual Data	Rolling Optimized NGBM(1, 1)			
		Parameter P	Parameter n	Prediction	RPE (%)
2004	45.42785	0.495	0.13	45.42785	0.00
2005	57.63326	0.495	0.13	57.55257	−0.14
2006	66.37166	0.495	0.13	68.75453	3.59
2007	77.41443	0.495	0.13	80.07765	3.44
2008	99.13030	0.495	0.13	92.12421	−7.06
2009	106.01466	0.495	0.13	105.21288	−0.75
2010	115.93175	0.495	0.13	119.58390	3.15
2011	135.53944	0.495	0.13	135.45557	−0.06
2012	155.82000	0.495	0.13	153.04632	−1.78
2013	171.22203	0.495	0.13	172.58566	0.79
2014	186.20465	0.495	0.13	194.32111	4.35
2015	193.24111	0.525	0.165	214.67979	11.09
2016	205.27617	0.48	0.155	238.67530	16.27
2017	223.77987	0.495	−0.02	270.80266	21.01
2018	245.21369	0.47	0.03	303.5345	23.78
ARPE Classification					**6.48** Excellent

Figure 4: Vietnam's GDP prediction $(2004 - 2018)$ with GM(1, 1), NGBM(1, 1), ONGBM(1, 1), and RONGBM(1, 1)



Apart from comparing the Grey models, alternative modern machine learning models are also taken into account to compare accuracy and efficiency with Grey models.

In this section, the Rolling Optimized NGBM(1, 1) will not be taken into account considering the ARPE. As the training data sample is relatively small, this model will only be used to predict from the available data, based on the significant efficiency it has proven comparing to the Optimized NGBM in the previous example.

Considering the press release given be Reuters [39] stating that the peak of the COVID-19 outbreak will happen at mid- or late-February, we will proceed to predict the total infected cases everyday globally, starting from 2020-02-09 to 2020-02-18.

Figure 5: Comparison between COVID-19, SARS-CoV, MERS-CoV and other outbreaks [29]

Table 1. Pathogenicity and Transmissibility Characteristics of Recently Emerged Viruses in Relation to Outbreak Containment.

Virus	Case Fatality Rate (%)	Pandemic	Contained	Remarks
2019-nCoV	Unknown*	Unknown	No, efforts ongoing	
pH1N1	0.02–0.4	Yes	No, postpandemic circulation and establishment in human population	
H7N9	39	No	No, eradication efforts in poultry reservoir ongoing	
NL63	Unknown	Unknown	No, endemic in human population	
SARS-CoV	9.5	Yes	Yes, eradicated from intermediate animal reservoir	58% of cases result from nosocomial transmission
MERS-CoV	34.4	No	No, continuous circulation in animal reservoir and zoonotic spillover	70% of cases result from nosocomial transmission
Ebola virus (West Africa)	63	No	Yes	

* Number will most likely continue to change until all infected persons recover.

Table 5: Results of ANN and LSTM in predicting COVID-19 infected cases by day

Date	Actual Data	ANN		LSTM	
		Prediction	RPE (%)	Prediction	RPE (%)
2020-01-28	6061	6061	0.00	6061	0.00
2020-01-29	7816	8836	13.05	8707	11.40
2020-01-30	9821	10184	3.70	10213	4.00
2020-01-31	11948	12395	3.74	12486	4.50
2020-02-01	14551	14481	−0.48	14621	0.48
2020-02-02	17387	17371	0.09	17464	0.44
2020-02-03	20626	21294	3.23	21269	3.11
2020-02-04	24553	23061	−6.07	22928	−6.62
2020-02-05	28276	27808	−1.65	27837	−1.55
2020-02-06	31439	31649	0.67	31717	0.88
2020-02-07	34875	34727	−0.43	34757	−0.34
2020-02-08	37552	38285	1.95	38197	1.72
ARPE			**2.92**		**2.92**

7.2.3 Results and analysis

Considering 5 different models predicting within the time range 2020-01-28 to 2020-02-08, the Optimized NGBM(1, 1) and even the traditional NGBM(1, 1) perform significantly better than two modern machine learning model (ANN and LSTM). This is due to the fact illustrated previously that NGBM(1, 1) yields relatively high accuracy upon handling short one - dimensional time series data.

Within the next 10 days, from 2000-02-09, similar to the example of predicting Vietnam's GDP $2013 - 2018$, the predictions of the rolling optimized NGBM(1, 1) are lower than that of ONGBM(1, 1). This is due to the update of new data, reducing the increasing rate. However, both of the two predictions confirm that the peak of this COVID-19 outbreak will still not be happening within the next 10 days. From plot 9, it can be easily derived that the new infected cases by day is increasing, and there has been no signs of stopping.

Figure 6: Total COVID-19 infected cases prediction by ANN and LSTM (from 2020-01-28 to 2020-02-08)

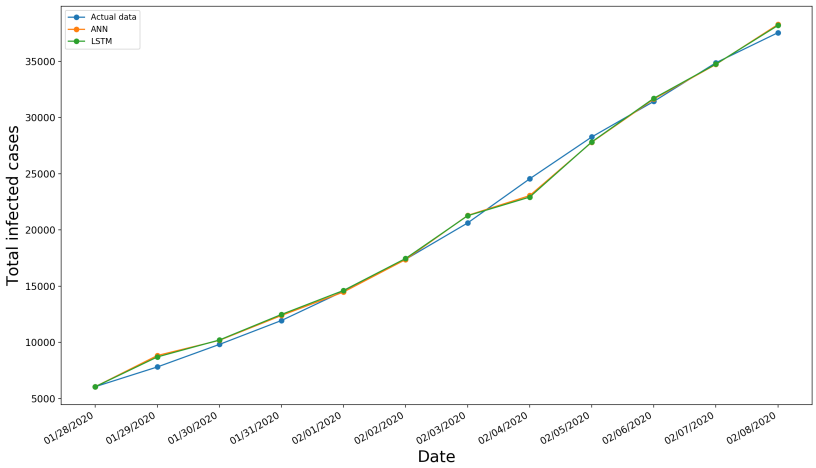


Table 6: Results of GM(1, 1), NBGM(1, 1) and ONGBM(1, 1) in predicting COVID-19 infected cases by day

Year	Actual Data	GM(1, 1)		NGBM(1, 1) $n = 0.41$		Optimized NGBM(1, 1) $n = 0.505, P = 0.7$	
		Prediction	RPE (%)	Prediction	RPE (%)	Prediction	RPE (%)
2020-01-28	6061	6061	0.00	6061	0.00	6061	0.00
2020-01-29	7816	9946	27.25	7258	−7.14	7130	−5.44
2020-01-30	9821	11451	16.60	9822	0.01	9824	0.03
2020-01-31	11948	13185	10.35	12418	3.93	12378	3.60
2020-02-01	14551	15181	4.33	15098	3.76	15056	3.47
2020-02-02	17387	17479	0.53	17898	2.93	17860	2.72
2020-02-03	20626	20125	−2.43	20842	1.05	20793	0.81
2020-02-04	24553	23172	−5.62	23953	−2.44	23862	−2.82
2020-02-05	28276	26679	−5.64	27251	−3.62	27068	−4.27
2020-02-06	31439	30719	−2.29	30755	−2.17	30417	−3.25
2020-02-07	34875	35369	1.42	34483	−1.12	33915	−2.75
2020-02-08	37552	40724	8.45	38455	2.40	37564	0.03
ARPE			**7.08**		**2.55**		**2.43**

Figure 7: Total COVID-19 infected cases predicted by GM(1, 1), NGBM(1, 1) and ONGBM(1, 1) (2020-01-28 - 2020-02-08)

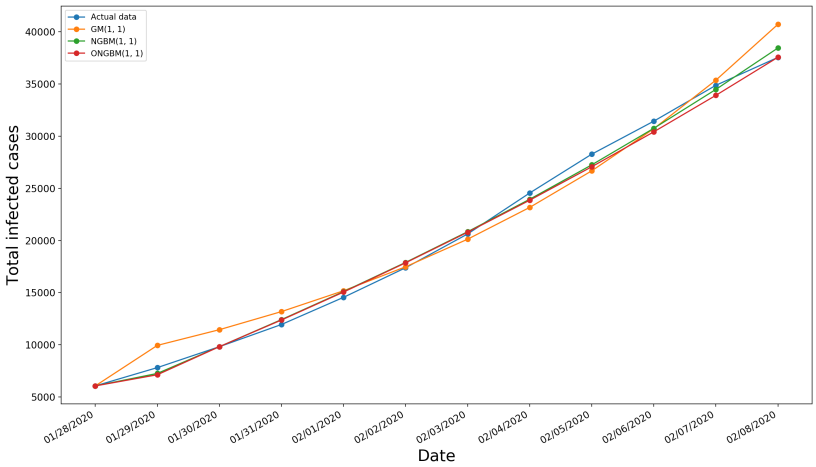


Table 7: Predictions by ONGBM(1, 1) and RONGBM(1, 1) of COVID-19 infected case by day

Year	Optimized NGBM(1, 1) $n = 0.505, P = 0.7$	Rolling Optimized NGBM(1, 1)		
		Parameter P	Parameter n	Prediction
2000-02-09	41373	0.7	0.505	41373
2000-02-10	45344	0.55	0.47	45050
2000-02-11	49484	0.505	0.36	48461
2000-02-12	53799	0.5	0.285	52176
2000-02-13	58295	0.49	0.22	56045
2000-02-14	62978	0.47	0.14	60493
2000-02-15	67854	0.49	0.095	65041
2000-02-16	72930	0.485	0.065	70001
2000-02-17	78214	0.49	0.045	75238
2000-02-18	83711	0.49	0.035	80848

8 Conclusion and future suggestions

In this paper, a new rolling optimized NGBM(1, 1) model has been proposed. There are two following major advantages of this model in comparison with the existing ones:

- Within the proposed model, all of the parameters are optimized simultaneously, including the exponential parameter, the background value and the initial condition.
- The first application of the rolling mechanism upon the combination of all currently available methods of parameter optimization.

The efficiency of this newly proposed model has also been tested with the other Grey models (GM(1, 1), NGBM(1, 1) and ONGBM(1, 1)), which showed significant improvement. Later, while this model is applied to predict the total COVID-19 infected cases, it has also been put into comparison with the latest and most effective machine learning models, including Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM). Within the context of extremely limit amount of data and time constrain, RONGBM(1, 1) outperforms both of the two other models, which additionally confirms the effectiveness of Grey models upon dealing with small amount of data.

Figure 8: Total COVID-19 infected cases predicted by ONGBM(1, 1) and RONGBM(1, 1) (2020-01-28 to 2020-02-18)

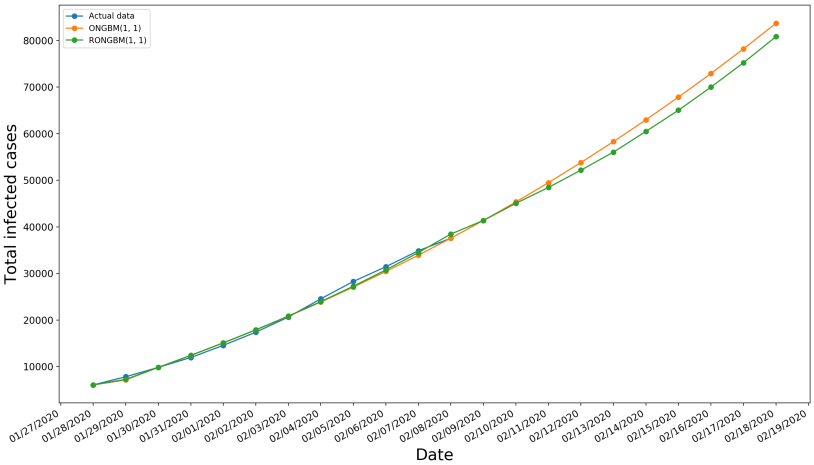
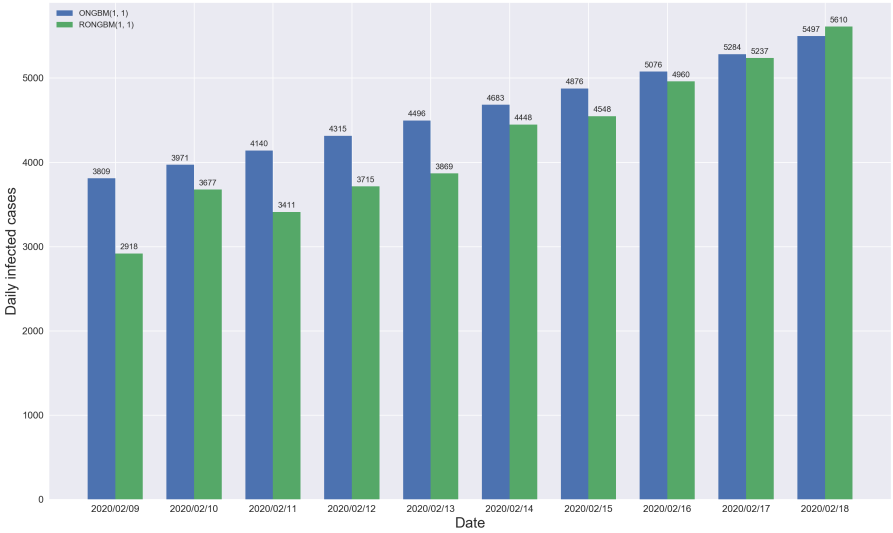


Figure 9: COVID-19 newly infected cases by day predicted by ONGBM(1, 1) and RONGBM(1, 1) (2020-02-09 to 2020-02-18)



From the results obtained by applying RONGBM(1, 1) on available data until the final day of retrieval (2020-02-10), there are now signs that the outbreak of COVID-19 will reach its peak within the next 10 days as expected. The number of newly infected patients reported everyday will even be higher than that of the present, increasing day by day. However, as Grey models are only applied on one-dimensional time series data, it is impossible to detect or predict changes due to any other related or outside effects (medical advancement, natural conditions, etc.). As a result, with enough amount of background knowledge into the field, applying a multivariate Grey model would yield more accurate and medically meaningful results.

Considering the proposed model RONGBM(1, 1) itself, there are two major problems that should be discussed and investigated in the future:

- **Training data size:** As the training data size has a significant impact on any Grey model [40], the amount of data used in the COVID-19 infected case estimation has been manually chosen. In the future, for Grey models to reach maximum optimality, there should be a formal, mathematically derived procedure to define the optimal size of training data.
- **Parameter optimization:** Within the proposed model, in order to obtain the optimization of the two parameters P and n, two loops have been run through. Due to the resource constrain of the current computational tools, the step size of these loops are still large enough to omit actual optimal values, leading to the model not perfectly optimized. Moreover, using the proposed formula, as demonstrated within the practical section, does not yield the optimal error. As a result, in the future, there should be a mathematically exact way to derive the parameter values optimally.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Ayodele A. Adebiyi, Aderemi O. Adewumi, and Charles K. Ayo. "Stock Price Prediction Using the ARIMA Model". In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. 2016.

[2] Meyler Aidan, Kenny Geoff, and Quinn Terry. "Forecasting Irish inflation using ARIMA models". In: (1998).

[3] Diyar Akay and Mehmet Atak. "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey". In: *Energy* 32 (9 2007), pp. 1670–1675.

[4] "Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate". In: *Chaos, Solitons & Fractals* (2008). vol. 37, no. 1, pp. 278–287.

[5] Michael Birnbaum. "Europe is closing borders amid coronavirus outbreak. They may be hard to reopen.t". In: *Washington Post* (Mar. 2020). URL: https://www.washingtonpost.com/world/europe/europe-closing-borders-coronavirus/2020/03/17/131a6f56-67c8-11ea-b199-3a9799c54512_story.html.

[6] Hsia-Ching Chang and Chen-Ya Wang. "Grey Forecasting Models for Social Media Adoptions of Transnational Enterprises". In: *2nd International Symposium on Computer, Communication, Control and Automation (3CA 2013)*.

[7] S. C. Chang, H. C. Lai, and H. C. Yu. "A variable P value rolling grey forecasting model for Taiwan semiconductor industry production". In: *Technol. Forecast Social Change* (2005). vol. 72, issue 5, pp. 623–640.

[8] Nagesh Singh Chauhan. *Introduction to Artificial Neural Networks(ANN)*. Accessed: 2020-02-09. URL: https://towardsdatascience.com/introduction-to-artificial-neural-networks-ann-1aea15775ef9.

[9] Chun-I Chen. "Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate". In: *Chaos, Solitons & Fractals* 37 (1 July 2008), pp. 278–287. DOI: 10.1016/j.chaos.2006.08.024.

[10] Chung Chen and Lon-Mu Liu. "Forecasting Time Series with Outliers". In: *Journal of Forecasting* (1993). vol. 12, issue 1, pp. 13–35. DOI: 10.1002/for.3980120103.

[11] "Coronavirus: US becomes first country to exceed 100,000 cases". In: *The Guardian* (Mar. 2020). URL: https://www.theguardian.com/world/2020/mar/28/coronavirus-us-becomes-first-country-to-exceed-100000-cases.

[12] Y. G. Dang, S. F. Liu, and K. J. Chen. "The GM model that x(n) be taken as initial value". In: *Kybernetes* (2004). vol. 33, issue 2, pp. 247–254.

[13] Jie Ding, Lili Han, and Xiaoming Chen. "Time series AR modeling with missing observations based on the polynomial transformation". In: *Mathematical and Computer Modelling* 51 (5-6 2010), pp. 527–536. DOI: 10.1016/j.mcm.2009.11.016.

[14] Md. Israr Equbala et al. "A Grey-based Taguchi Method to Optimize Hot Forging Process". In: *Procedia Material Science* 6 (2014), pp. 1495–1504. DOI: 10.1016/j.mspro.2014.07.129.

[15] Gers Felix, Schmidhuber Jürgen, and Cummins Fred. "Learning to Forget: Continual Prediction with LSTM". In: *Neural computation* 12 (Oct. 2000), pp. 2451–71. DOI: 10.1162/089976600300015015.

[16] *GDP (current US$) - Vietnam*. Accessed: 2020-02-09. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=VN.

[17] *GDP growth (annual %) - Vietnam*. Accessed: 2020-02-09. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=VN.

[18] Renkuan Guo. "A Repairable System Modelling: Combining Grey System Theory with Interval-valued Fuzzy set". In: *International Journal of Reliability, Quality and Safety* 12 (Feb. 2005), pp. 241–266. DOI: 10.1142/S0218539305001811.

[19] Sang Hansun. "A new approach of moving average method in time series analysis". In: 2013. DOI: 10.1109/conmedia.2013.6708545.

[20] Lynsey Jeffery. "U.S. Surpasses 1 Million Coronavirus Cases". In: *National Public Radio* (Apr. 2020). URL: https://www.npr.org/sections/coronavirus-live-updates/2020/04/28/846741935/u-s-surpasses-1-million-coronavirus-cases.

[21] Guo-Dong Li, Daisuke Yamaguichi, and Masatake Nagai. "A GM(1, 1)-Markov chain combined model with an application to predict the number of Chinese international airlines". In: *Technological Forecasting and Social Change* 74 (July 2006), pp. 1465–1481. DOI: 10.1016/j.techfore.2006.07.010.

[22] Qun Li, Xuhua Guan, Peng Wu, et al. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia". In: *The New Englang Journal of Medicine* (Jan. 2020). DOI: 10.1056/NEJMoa2001316.

[23] Li Liu, Qianru Wang, et al. "A Rolling Grey Model Optimized by Particle Swarm Optimization in Economic Prediction". In: *Computational Intelligence* 0 (0 2014), pp. 391–419. DOI: 10.1111/coin.12059.

[24] Sifeng Liu and Yi Lin. "Grey Information: Theory and Practical Applications". In: (2006).

[25] Deng Ju-Long. "Control problems of grey systems". In: *Systems & Control Letters* 1 (5 1982), pp. 288–294.

[26] Jianshan Lu, Weidong Xie, et al. "An optimized nonlinear grey Bernoulli model and its application". In: *Neurocomputing* 17 (2016), pp. 206–214.

[27] Zhan-li Mao and Jin-hua Sun. "Application of Grey-Markov Model in Forecasting Fire Accidents". In: *Procedia Engineering* 11 (2011), pp. 314–318. DOI: 10.1016/j.proeng.2011.04.663.

[28] Ladan Mozaffari, Ahmad Mozaffari, and Nasser L. Azad. "Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on San Francisco urban roads". In: *Engineering Science and Technology, an International Journal* 18 (2 June 2015). DOI: 10.1016/j.jestch.2014.11.002.

[29] Vincent J. Munster, Marion Koopmans, et al. "A Novel Coronavirus Emerging in China — Key Questions for Impact Assessment". In: *The New Englang Journal of Medicine* (Jan. 2020). DOI: 10.1056/NEJMp2000929.

[30] Christopher Olah. "Understanding LSTM Networks". In: (2015). URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[31] Jonathan M. Read, Jessica R.R. Bridgen, et al. "Novel coronavirus 2019-nCoV: early estimation of epidemiological parameter and epidemic predictions". In: (Jan. 2020). DOI: 10.1101/2020.01.23.20018549.

[32] Hochreiter Sepp and Schmidhuber Jürgen. "Long Short-term Memory". In: *Neural computation* 9 (Dec. 1997), pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735.

[33] Michael D. Shear, Abby Goodnough, et al. "The Lost Month: How a Failure to test blinded the U.S. to Covid-19". In: *New York Times* (Mar. 2020). URL: https://www.nytimes.com/2020/03/28/us/testing-coronavirus-pandemic.html.

[34] G. T. Tan. "The structure method and application of background value in grey system GM model (I)". In: *Syst. Eng.-Theory Pract.* (2000). vol. 4, pp. 98–103.

[35] U. Thissen, R. van Brakel, et al. "Using support vector machines for time series prediction". In: *Chemometrics and Intelligent Laboratory Systems* 69 (1,2 Nov. 2003), pp. 35–49. DOI: 10.1016/S0169-7439(03)00111-4.

[36] Oskar Triebe, Nikolay Laptev, and Ram Rajagopal. "AR-Net: A simple Auto-Regressive Neural Network for time-series". In: (2019).

[37] Z. W. Wang, Y. G. Dang, et al. "Solution of GM(1,1) power model and its properties". In: *Sys. Eng. Electrion* 31 (10 2009), pp. 2380–2383.

[38] Zheng-Xin Wang et al. "An optimized NGBM model for forecasting the qualified discharge rate of industrial waterwaste in China". In: *Applied Mathematical Modelling* (2011). DOI: 10.1016/j.apm.2011.05.022.

[39] "When will coronavirus peak? Epidemic forecasts are often wrong but can be useful". In: *Reuters* (Feb. 2020). URL: https://www.japantimes.co.jp/news/2020/02/07/asia-pacific/science-health-asia-pacific/when-will-coronavirus-peak/.

[40] L. Wu et al. "The effect of sample size on the grey system model". In: *Applied Mathematical Modelling* (2013). Vol. 37, no. 9, pp. 6577–6583.

[41] Wen-Ze Wu, Tao Zhang, and Chengli Zheng. "A Novel Optimized Nonlinear Grey Bernoulli Model for Forecasting China's GDP". In: *Complexity* (2019). Article ID 1731262, 10 pages. DOI: 10.1155/2019/1731262.

[42] Lin Y.-T et al. "Hybrid grey modification model with a Fourier series and exponential smoothing optimization technique". In: *Journal of Grey System* 26 (Jan. 2014), pp. 108–121.

[43] Peng Zhou, Xing-Lou Yang, et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin". In: *Nature* (Feb. 2020). DOI: 10.1038/s41586-020-2012-7.

[44] Wei Zhou and Jian-Min He. "Generalized GM (1, 1) model and its application in forecasting of fuel production". In: *Applied Mathematical Modelling* 13 (Jan. 2013), pp. 6234–6243. DOI: 10.1016/j.apm.2013.01.002.

[45] H. Y. Zhuang. "Mechanism and application of GM(1,1) and improvement methods". In: *Syst. Eng. Theor. Methodol. Appl.* (1993). vol. 3, issue 3, pp. 56–62.