Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Grey forecasting models based on internal optimization for Novel Coronavirus (COVID-19)

Akash Saxena
Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan, India

Abstract

Pandemic forecasting has become an uphill task for the researchers on account of the paucity of sufficient data in the present times. The world is fighting with the Novel Coronavirus to save human life. In a bid to extend help to the concerned authorities, forecasting engines are invaluable assets. Considering this fact, the presented work is a proposal of two Internally Optimized Grey Prediction Models (IOGMs). These models are based on the modification of the conventional Grey Forecasting model (GM(1,1)). The IOGMs are formed by stacking infected case data with diverse overlap periods for forecasting pandemic spread at different locations in India. First, IOGM is tested using time series data. Its two models are then employed for forecasting the pandemic spread in three large Indian states namely, Rajasthan, Gujarat, Maharashtra and union territory Delhi.

Several test runs are carried out to evaluate the performance of proposed grey models and conventional grey models GM(1,1) and NGM(1,1,k). It is observed that the prediction accuracies of the proposed models are satisfactory and the forecasted results align with the mean infected cases. Investigations based on the evaluation of error indices indicate that the model with a higher overlap period provides better results.

1. Introduction

The month of December in 2019 was a watershed in the history of mankind with a series of cases reported suffering from inexplicable pneumonia at Wuhan, Hubei, China. The clinical investigation of the cases revealed their resemblance with viral pneumonia. Further, deep sequencing analysis from lower respiratory tract samples revealed that the root cause of this pneumonia is a new virus. People's Republic of China (Centre for Disease Control) attributed the cause of this unidentified pneumonia to a novel form of Coronavirus and later World Health Organization (WHO) declared it as COVID-19. Coronavirus came from a family of Coronaviridae and the order Nidovirales. These are enveloped non-segmented positive-sense RNA viruses that are broadly distributed in humans and other mammals [1]. The WHO declared the deadly disease as a global pandemic on March 11, 2020 [2]. The initial cluster was developed in a local seafood market in Wuhan. An epidemiological alert was issued by WHO at the end of April 2020 which spread like a wildfire world over. It ultimately turn out to be the nemesis for the mankind in the form of pandemic. With the spread of this disease, implications of the pandemic appeared as loss of life and several health-related issues. For combating such situations, many scientists have come forward to help the community by forecasting certain parameters so that authorities concerned can frame preventive healthcare policies. The prediction models based on Data science and Machine Learning Methods (MLMs) have assumed importance in providing better understanding about growth and trend of pandemic with respect to time.

The forecasting of pandemic spread is quite challenging due to dependency of the pandemic spread on several factors. These factors are governed by the psychological behaviour of the community, combating strategies deployed by the authorities and of course, volatility and reliability of the data [3]. Moreover, the future does not repeat it in the same way as did in the past. Hence, considering the same policies which were employed in the past for forecasting the pandemic spread may not be applicable directly.

Evolving fear amongst the population due to the enhanced death rate and concerns of politicians to take adequate steps as preventive measures create a strong perception. Hence, this pandemic has emerged as an infodemic as well. However, cut-throat competition among the vaccine manufacturers is a welcome move and it is furthering the cause of protecting the lives of people against COVID. These facts set the strong foundation for discussion and research in the area of forecasting the pandemic spread and healthcare-related parameters.

Recently many prediction approaches have been reported by the researchers for fairly accurate forecasting about the spread of the disease.
of COVID, prediction of recovery rate and death rate. In Ref. [4], Maximum-Hasting (MH) parameter estimation method and the modified Susceptible Exposed Recovered (SEIR) model for prediction of COVID was developed. A study on the worst-hit states of India has been conducted in Ref. [5]. The study was based on system modelling and identification techniques. Time series-based approach in amalgamation with Long Short Term Associated Memory (LSTM) for prediction of COVID has been employed in Canada [6]. A similar approach based on LSTM has been reported in Ref. [7]. SIR model-based prediction has been employed in power demand forecasting in Ref. [32]. The author of such approaches. A good example of this can be found in the development of Novel Grey Prediction Model (NGM) in Ref. [23]. An additional constant term has also been added in Grey system equation in order to overcome shortcomings of NGM and it has been named as NGM(1,1,k,c) [24]. Apparently, classical discrete model of Grey prediction can also be determined by changing the original Grey equations [25]. A novel discrete model was proposed for forecasting the CO₂ emission in China [26]. NGM method has been applied for forecasting the consumption of natural gas in China [27]. Another interesting approach based on integrating heuristic time series and fuzzy theory for forecasting the renewable energy in Taiwan was described in [28].

Optimization-based approaches and especially the approaches which are based on some nature-inspired algorithms have been employed with GPMs in recent years. These approaches are based on the parameter estimation of the Grey models. A novel time delay forecasting model based on a nature-inspired optimizer was performed in [29]. Further, Refs. [12] and [30] are fine examples of such approaches.

In addition to these published approaches, some experiments have been done to alter initial conditions of the Grey model in Ref. [26]. Authors of the paper employed alterable weighted coefficients in initial conditions. Another application of Grey prediction model for predicting sales in global integrated circuit industry is seen in Ref. [31]. Further, the Grey model has also been applied in power demand forecasting in Ref. [32]. The authors employed residual modification with an artificial neural network for the modification of GM(1,1) model.

From the literature review of GPMs, the motivation for employing GPMs in pandemic forecasting is very clear and pragmatic. However, in some studies, it has been reported that GM(1,1) models fail to predict with required accuracy when data mutate swiftly or the associated variables are volatile.

In the past, it has been identified that forecasting accuracy of the grey models can be questionable when the initial conditions

| Nomenclature | Definition |
|--------------|------------|
| a            | Grey Development coefficient |
| b            | Grey control parameter |
| C(1)(W,μ)   | First order accumulated sequence of mean infected cases for a week considering overlap period μ |
| Z(13)(m)    | Background value at mth instant |
| \( \hat{C}_{W,(m)} \) | Forecasted value at mth instant |
| X(0)        | Actual value of mth element of time series |
| δ(m)        | Error in prediction of mth element of time series |
| μ           | Period of overlap |
| \( C_{(m)}(W,μ) \) | Mean infected cases during time span of a week |
| \( \hat{C}_{(m)}(W,μ) \) | Forecasted value at mth instant considering 6 days of overlap |
| \( \hat{C}_{(m)}(W,κj) \) | Forecasted value at mth instant considering 5 days of overlap |

4. Parameter optimization by heuristic and metaheuristic methods.
and starting points are not chosen correctly [15]. From this point of view, it is pragmatic that the involvement of optimization can enhance the forecasting accuracy. This involvement can be done at two levels during grey forecast. First, at the macro level by choosing the external optimization parameters such as data or selection of time series to develop different forecasting strategies. Secondly, by developing an internal optimization routine that integrates a few changes in forecast modelling and try to reduce the error between measured and sample data. Moreover, some researchers have emphasized in applying corrections in the internal parameter aggregation process by modifying the grey equation to achieve better results.

Considering a variable (infectious cases) as a grey variable that increases with every passing day and mutates swiftly, is difficult to forecast. Hence, the presented work primarily focuses on an internal optimization model that aims at enhancing the forecasting accuracy by choosing internal processing parameters through the optimization process. The research objectives of the paper are as follows:

1. To investigate the applicability of the GPMs on variety of benchmark time series data.
2. To propose the prediction model based on internal optimization and analyse the results based on average ranks obtained by stacking the forecaster’s performance chronologically.
3. To employ the proposed internal optimization-based model and other grey models on the real data by taking two different overlap periods and construct two grey prediction models for forecasting pandemic spread in different hot-spots of India.
4. To evaluate the performance of these pandemic forecasting models based on error indices obtained for individual cities, average values of error for a particular city for both models.

Remaining part of the paper is organized as follows: In Section 2, proposal of IOGM is presented and explained. In Section 3, verification of proposed model is conducted on benchmark data series and comparative analysis is presented. Section 4, presents the simulation and results analysis of proposed grey models on different parts of India. Finally, Section 5, presents the concluding remarks of the work and throws light on the future directions of the research work.

2. Grey internal optimization prediction model

In this section, basics of GPM and its application for COVID-19 spread prediction are explained.

2.1. GM(1,1) model

Based on the above explanation, following mathematical expressions are considered for the proposed grey forecasting model. Following steps are followed for constructing forecasting engine.

The mean values of infected cases is considered in constructing the initial time series. \( C^{(0)}_{[\mu, \nu]} \) is representative denominator of time series. For the evolution of this series, successive elements are calculated based on overlap period of one week (7 days).

In general, consider a time series having an overlap period "\( \nu \)" for the duration "\( \mu \)". The time series comprises of ‘\( k \)’ elements. The representation of this data series can be given as follows:

\[
C^{(0)}_{[\mu, \nu]} = \left( C^{(0)}_{[\mu, \nu]}(1), C^{(0)}_{[\mu, \nu]}(2), C^{(0)}_{[\mu, \nu]}(3), \ldots \ldots \ldots \ldots C^{(0)}_{[\mu, \nu]}(k) \right) \tag{1}
\]

By obtaining a one-time accumulating generation operation, the following series can be generated:

\[
C^{(1)}_{[\mu, \nu]} = \left( C^{(1)}_{[\mu, \nu]}(1), C^{(1)}_{[\mu, \nu]}(2), C^{(1)}_{[\mu, \nu]}(3), \ldots \ldots \ldots \ldots C^{(1)}_{[\mu, \nu]}(k) \right) \tag{2}
\]

\[
C^{(1)}_{[\mu, \nu]}(m) = \sum_{i=1}^{m} C^{(0)}_{[\mu, \nu]}(i) \tag{3}
\]

Where \( m=1, 2, 3, \ldots k \).

\[
c^{(0)}_{[\mu, \nu]}(m) + aZ^{(1)}(m) = b \tag{4}
\]

Where ‘a’ is Grey development coefficient and ‘b’ is Grey control parameter (driving coefficient). It is to be noted that the value of ‘a’ has a potential impact on background values \( Z \) of Grey derivatives, hence, the forecasted values get compromised due to the large value of ‘a’.

\[
Z^{(1)}(m) = (1 - \alpha)c^{(1)}_{[\mu, \nu]}(m) + \alpha c^{(1)}_{[\mu, \nu]}(m - 1) \tag{5}
\]

Where \( m=2, 3, \ldots k \). Here, \( \alpha \) is the background value production coefficient. The values of this coefficient should be optimized between an interval of [0, 1]. Further, the native Grey Model (1, 1) can be derived while keeping \( \alpha = 0.5 \). The following expression is for background values of grey derivatives for the native GM(1,1) model.

\[
Z^{(1)}(m) = (0.5)c^{(1)}_{[\mu, \nu]}(m) + (0.5)c^{(1)}_{[\mu, \nu]}(m - 1) \tag{6}
\]

The expression (4) can be solved with the help of least square estimation method and expression for Grey development coefficient and driving coefficient can be expressed as follows:

\[
\begin{bmatrix}
\hat{a} \\
\hat{b}
\end{bmatrix} = (B^T B)^{-1} B^T Y \tag{7}
\]

\[
B = \begin{bmatrix}
-1 \times ((1 - \alpha) \times c^{(1)}_{[\mu, \nu]}(2) + \alpha \times c^{(1)}_{[\mu, \nu]}(1)) \\
-1 \times ((1 - \alpha) \times c^{(1)}_{[\mu, \nu]}(3) + \alpha \times c^{(1)}_{[\mu, \nu]}(2)) \\
\vdots \\
-1 \times ((1 - \alpha) \times c^{(1)}_{[\mu, \nu]}(m) + \alpha \times c^{(1)}_{[\mu, \nu]}(m - 1)) \\
-1 \times ((1 - \alpha) \times c^{(1)}_{[\mu, \nu]}(k) + \alpha \times c^{(1)}_{[\mu, \nu]}(k - 1))
\end{bmatrix} \tag{8}
\]

\[
Y = \begin{bmatrix}
-Z^{(1)}(2) \\
-Z^{(1)}(3) \\
\vdots \\
-Z^{(1)}(k)
\end{bmatrix} \tag{9}
\]

In simplified form it can be written as

\[
\begin{bmatrix}
\hat{c}^{(1)}_{[\mu, \nu]}(1) \\
\hat{c}^{(1)}_{[\mu, \nu]}(2) \\
\vdots \\
\hat{c}^{(1)}_{[\mu, \nu]}(k)
\end{bmatrix} = \begin{bmatrix}
C^{(0)}_{[\mu, \nu]}(2) \\
C^{(0)}_{[\mu, \nu]}(3) \\
\vdots \\
C^{(0)}_{[\mu, \nu]}(k)
\end{bmatrix} \tag{10}
\]

The solution of Eq. (4) can be written as

\[
\hat{C}^{(1)}_{[\mu, \nu]}(m) = \left( C^{(0)}_{[\mu, \nu]}(1) - \frac{b}{a} \right) e^{-\alpha(m-1)} + \frac{b}{a} \tag{11}
\]

Where \( \hat{C}^{(1)}_{[\mu, \nu]}(m) \) is the \( m \)th associated value. For obtaining predicted values of original time series Inverse Accumulation generation operation is required and can be represented as per following equations.

\[
\hat{C}^{(0)}_{[\mu, \nu]}(1) = C^{(0)}_{[\mu, \nu]}(1) \tag{12}
\]

This expression holds good for \( m=1 \). Generalized equation can be written as

\[
\hat{C}^{(0)}_{[\mu, \nu]}(m) = \hat{C}^{(1)}_{[\mu, \nu]}(m) - \hat{C}^{(1)}_{[\mu, \nu]}(m - 1) \tag{13}
\]
Eq. (13) is the generalized expression for \( m = 2, 3, \ldots k \). After rearranging the expressions one can get a generalized expression for forecasted values at \( m^th \) instance.

\[
\hat{C}_{\text{(W,a)}}(m) = \left( C_{\text{(W,a)}}(1) - \frac{b}{a} \right) \times e^{-\alpha(k-1)} \times (1 - e^\alpha) 
\]

(14)

From these expressions, it can be concluded that the internal optimization of tunable parameters can have a potential impact on the forecast accuracy. Hence, these parameters should be tuned properly. The following subsection presents a discussion on the need of this optimization.

2.2. Discussion

After considering the facts involved in the development of a forecasting engine it appears that there is a potential impact of internal parameters on the accuracy of the forecast. In references [33,34], it has been pointed out that near accurate estimation of the background values can be expressed as per Eq. (5). The relationship between background value production coefficient \( \alpha \) and development coefficient can be defined as follows:

\[
\alpha = \frac{1}{a} - \frac{1}{(e^\alpha - 1)} 
\]

(15)

Further, Chang et al. [35] proved that proper optimization of background value production coefficient can enhance the forecasting accuracy. A detailed explanation regarding this can be found in Ref. [36].

By using the L-Hopital rule as applied in [15] it can be concluded that for diverse values of ‘\( a \)’ the parameter \( \alpha \) revolves around 0.5 value. Moreover, it can be said that higher values of ‘\( a \)’ can lead to erroneous results. As ‘\( a \)’ approaches to zero, \( \alpha \) approaches to 0.5. Fig. 1 exhibits this relationship where 16 samples of ‘\( a \)’ are considered between span [−1, 1] and \( \alpha \) is calculated as per expression (15). As indicated in different researches, larger values of ‘\( a \)’ yield erroneous forecast because of greater difference between \( \alpha \) and \( a \). Hence, in a defined search (objective) space the error function is employed to bridge this difference through optimization.

2.3. Proposed internal optimization model

Based on the discussion in previous subsection, an optimization routine is formulated for predicting COVID-19 infected cases. Following are the steps involved in constructing the model.

**Step 1:** Start the iterative search by taking \( \alpha(0) = 0.5 \), while taking \( \alpha = 0.5 \) the model becomes conventional grey model. Calculate the background values as per Eq. (5) and further calculate values of \( a \) and \( b \).

**Step 2:** Now by substituting the values of ‘\( a \)’, in expression (15) new value of \( \alpha \) that is designated as \( \alpha(1) \) is obtained. Calculate the absolute error between obtained \( \alpha(1) \) and initial value i.e. (0.5), if the value of error is greater than tolerance then stop the loop otherwise perturb the value of \( \alpha \). The absolute error can be defined for two successive iterations by \( \Delta(t) \), where ‘\( t \)’ denotes current iteration.

\[
\Delta(t) = |\alpha(t) - \alpha(t - 1)| 
\]

(16)

Now, update \( \alpha \) as

\[
\tilde{\alpha} = \alpha(t) + \Delta(t) 
\]

(17)

Where \( \tilde{\alpha} \) is perturbed value, again calculate the values of \( a \) and \( b \) from the expression (5)-(7) and compute the absolute error between \( \tilde{\alpha} \) and \( \alpha(t) \).

**Step 3:** If the error reduces then increase the loop counter by 1 and accept the perturbation vector \( \Delta(t) \) and append \( \alpha(t + 1) \) as \( \tilde{\alpha} \) as in same direction. Otherwise reject the perturbation value and assign opposite perturbation (−\( \Delta(t) \)). Now, update the alpha as

\[
\tilde{\alpha} = \alpha(t) - \Delta(t) 
\]

(18)

and \( \alpha(t + 1) = \tilde{\alpha} \). Repeat the process, till the error between the successive iterations of the \( \alpha \) becomes less than tolerance value.

**Step 4:** Now the optimized model can be realized with the modified \( \alpha \) as represented by Eq. (14). For simulating the time series and prediction, \( \epsilon \) is taken 10E-8 in this work.

**Algorithm 1:** Pseudo code of Proposed IOGM

1. Initialize \( \alpha \) and construct the original GM(1,1) i.e. by setting \( \alpha = 0.5 \)
2. Calculate the background values, \( a \) and \( b \) as per Eq. (5).
3. Calculate updated \( \alpha \) by substituting the values of \( a \), in Eq. (15).
4. Calculate the absolute error between \( \alpha(1) \) and initial value i.e. \( (\alpha(0)=0.5) \)
5. While \( \Delta(t) < \epsilon \) do
6. Perturb \( \alpha \) by \( \Delta(t) \), where \( \Delta(t) = \alpha(t) - \alpha(t - 1) \).
7. Perturb \( \alpha \) to obtain \( \alpha(t + 1) \) and apply condition for choosing perturbation value by calculating absolute error between perturbed \( \alpha \) and previous value.
8. \( \alpha(t + 1) = \begin{cases} 
\alpha(t) + \Delta(t) & \text{if } \Delta(t + 1) < \Delta(t) \\
\alpha(t) - \Delta(t) & \text{Otherwise}
\end{cases} 
\)
9. Check feasibility of new positions and evaluate these positions
10. End while
11. Print the values of \( a \), \( b \) and \( \alpha \), and construct IOGM for forecast.

In this work, an internal optimization scheme is employed for forecasting COVID-19 cases. The flow of algorithm along with data stacking process are shown in Figs. 2 and 3 respectively. Following steps are considered for framing GPMs for forecasting the pandemic:

- For stacking data into Model-I and Model-II, two different overlap periods (\( m=5 \) and 6) are considered. The data of three different states and Delhi are depicted in the result Section 4.
- Further, the data stacked in model array are segregated into two parts simulation and validation (forecasted) parts.
- On the basis of tolerance value ‘\( \epsilon \)’ obtained from simulated data, grey models have been constructed and coefficients ‘\( a \)’ and ‘\( b \)’ are calculated.

However, it is quite necessary to judge the forecasting performance of the IOGM in comparison to classical GM model and Novel Grey Model (NGM) on benchmark time series. Following section depicts this analysis in depth.

3. Verification of proposed optimization model with benchmark time series data

For understanding the impact of internal optimization, let us consider a homogeneous Geometric Progression data series. This series is defined as follows:

\[
A = (1, 2, 4, 8, 16, 32, 64, 128) 
\]

(19)

While forecasting the next value of series from GM(1,1) model, \( \alpha \) is set to 0.5. From this value of \( \alpha \), undermentioned series is
obtained.

\[ A_{GM(1,1)} = (0.97, 1.89, 3.69, 7.19, 14.00, 27.27, 53.13, 103.48, 201.56) \]  (20)

For this experiment obtained value of \( a \) is -0.667. The forecasted value of the series is written in bold face. From this forecasted value (201.56), it can be observed that the error is very high. A huge difference exists between actual value of the series (256) and the forecasted value of GM(1,1) model. Further, using optimized model, following series is obtained.

\[ A_{IOGM(1,1)} = (0.99, 1.99, 3.99, 7.99, 15.99, 31.99, 63.99, 127.99, 255.9963) \]  (21)

For this model, value of \( a \) is 0.5573 and value of \( a \) is -0.6931. It is observed that forecasted value is 255.9963. This value is quite close to the actual value as compared to non-optimized model. This fact validates the necessity of internal optimization for improving the forecasting accuracy of the conventional GM(1,1). Further, for verification of the internal optimized model of grey forecasting, certain benchmarks time series are used here from [37] and same are defined as under:

1. **Homogeneous Exponential Sequence (B1):** The series can be identified with the help of following formula:
   \[ X_{B1} : x_{B1}^{(0)}(i) = 0.8 \times 1.5^i, i = 1, 2, 3, ..., 15 \]  (22)

2. **Non-homogeneous Exponential Sequence (B2):** The series can be identified with the help of following formula:
   \[ X_{B2} : x_{B2}^{(0)}(i) = 1.3 \times 1.8^i + 3.5, i = 1, 2, 3, ..., 15 \]  (23)

3. **Approximate Non-homogeneous Exponential Sequence (B3):** The series can be identified with the help of following formula:
   \[ X_{B3} : x_{B3}^{(0)}(i) = 1.3 \times 1.4^i + 1.6, i = 1, 2, 3, ..., 15 \]  (24)

4. **Random Number Sequence (B4):** The series can be identified as per following sequence:
   \[ X_{B4} : x_{B4}^{(0)} = [78.3571, 35.0894, 40.8045, 48.9120, 58.0352, 66.4530, 77.8831, 68.2308, 73.0020, 79.3541, 79.4816, 105.474, 95.2562, 119.923, 133.4256] \]

For evaluation of the performance of the proposed IOGM, simulations are carried out on B1–B4. For a better understanding of the computed forecasting accuracy, the whole process is subdivided into two parts. In the first part, 10 out of 15 samples are considered for building the grey architecture and for calculation of internal parameters of the forecaster. In the second part, remaining five samples of each series are employed as Out-Samples for evaluating the performance of the forecaster on the basis of error and Mean Absolute Percentage Error (MAPE).

Both of these indices are defined as under:

\[ \delta(m) = \frac{|x^{(0)}(m) - \hat{x}^{(0)}(m)|}{x^{(0)}(m)} \times 100 \]  (26)

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x^{(0)}(m) - \hat{x}^{(0)}(m)}{x^{(0)}(m)} \right| \times 100 \]  (27)

**Table 1** shows the results of B1. Similar to previously reported results on Geometric Progression Series, it is observed that the IOGM model is better as far as simulated value error analysis is concerned. It is further observed that NGM is not suitable for this kind of time series. Grey models are developed with current series having 10 data points. Values of ‘a’ and ‘b’ are provided along with the name of models in respective rows. Further, as per the analysis conducted on forecasted values, MAPE of models have been calculated and it is observed that the IOGM model gives the best results as MAPE values are optimal for this model.

Further, **Table 2** shows results of B2. High errors are there for the non-homogeneous exponential model. The coefficients and internal parameters have been calculated based on In-Samples and the rest five Out-Samples are simulated with the grey equations of associated models. It is observed that MAPE for simulated values is optimal for IOGM. MAPE for forecasted values is also very competitive. However, NGM model possesses the optimal value of MAPE for forecasted values.

The B3 benchmark consists of an approximate non-homogeneous model, the simulated results are shown in Table 3. From the Table, it can be observed that MAPE value is optimal for IOGM (7.63). However, it is worth mentioning here that other competitive models NGM and GM possess high simulation errors for this sequence. This fact indicates that the IOGM model
suits well for this kind of time series. Further, inspecting the forecasted results of Out-Samples, it can be concluded that the IOGM model possesses least MAPE. On the other hand, the NGM model possesses highest simulated and forecasted errors.

To test the applicability of the IOGM model further, a random sequence time series is considered. Comparative analysis of the performance of different grey models is depicted through errors in simulated and forecasted values of the series in Table 4. In-Samples are employed for building the grey prediction models, by observing the individual error component and accumulated MAPE for the simulation model. It is observed that the NGM model is not suitable for forecasting the random samples. Further, the GM model also possesses higher MAPE as compared to the IOGM model. By comparing the MAPEs of these models it can be concluded that the IOGM model possesses optimal MAPE.

3.1. Discussion

By comparing the results, on all benchmarks, on the basis of MAPE and arranging ranks on the basis of performance, one can calculate the average rank obtained from models. It can be concluded that IOGM secured 1.25 average rank as the performance of IOGM is better than other models on three out of four benchmarks. NGM performance is very weak on these benchmarks as the average rank possessed by this model is 2.5. Except, non-homogeneous model, performance of NGM is comparatively weak. Further, the average rank obtained by the GM model is 2.25. The same analysis is depicted in Fig. 4, where (a) segment shows the average rank and the rank obtained on the basis of MAPE of simulated and forecasted data of benchmark time series. Segment (b) shows the graphical representation of rank-based comparison of forecasters for simulated data. Likewise, segment (c) and (d) show rank-based analysis of forecasted data and MAPEs of forecasting engines respectively. From this analysis, it can be concluded that for all types of time series, the performance of IOGM is satisfactory. Following points support the argument for choosing IOGM for COVID-19 forecasting:

1. It is observed that the overall performance of IOGM is competitive with conventional GM and NGM models. This...
Fig. 3. Proposed Grey Forecasting Models based on Internal Optimization.

Table 1
Simulated and Forecasted results of different Grey Models on B1.

| Sample   | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM  | Error |
|----------|-----------|--------------|-------|-----------------|-------|-------|-------|
|          |           | a = -0.400, b = 0.9600 | Simulated value | a = -0.3867, b = 0.2266 | Simulated value | a = -0.4055, b = 0.9731 | Simulated value |
| X_{B1}^i (1) | 1.2 | 1.2 | 0 | 1.2 | 0 | 1.2 | 0 |
| X_{B1}^i (2) | 1.8 | 1.770568912 | 1.63506 | 0.972690805 | 45.96162 | 1.8 | 0 |
| X_{B1}^i (3) | 2.7 | 2.641378431 | 2.171169 | 1.70861309 | 36.71803 | 2.7 | 0 |
| X_{B1}^i (4) | 4.05 | 3.940473579 | 2.704356 | 2.791985048 | 31.0621 | 4.05 | 0 |
| X_{B1}^i (5) | 6.075 | 5.878495806 | 3.234637 | 4.386847474 | 27.7852 | 6.075 | 0 |
| X_{B1}^i (6) | 9.1125 | 8.769685228 | 3.762028 | 6.734689447 | 26.09394 | 9.1125 | 0 |
| X_{B1}^i (7) | 13.66875 | 13.08283301 | 4.286544 | 10.19101386 | 25.44297 | 13.66875 | 0 |
| X_{B1}^i (8) | 20.503125 | 19.51729341 | 4.808202 | 15.27916653 | 25.47884 | 20.503125 | 0 |
| X_{B1}^i (9) | 30.7546875 | 29.11638033 | 5.327016 | 22.76957965 | 25.96387 | 30.7546875 | 0 |
| X_{B1}^i (10) | 46.13203125 | 43.46353529 | 5.843003 | 33.79642812 | 26.73978 | 46.13203125 | 0 |
| **MAPE** | 3.752446129 | 30.13885312 | 0 |

Forecasted results

| Sample   | Out-sample | Simulated value | Error | Simulated value | Error | Simulated value | Error |
|----------|------------|----------------|-------|----------------|-------|----------------|-------|
| X_{B1}^o (1) | 69.198046888 | 64.7998 | 6.356027 | 50.02936286 | 27.70119 | 69.19720902 | 0.001211 |
| X_{B1}^o (2) | 103.7970703 | 96.6699 | 6.866447 | 73.92632408 | 28.77802 | 103.7957088 | 0.001312 |
| X_{B1}^o (3) | 155.6956055 | 144.2145 | 7.374072 | 109.105149 | 29.9237 | 155.6934061 | 0.001413 |
| X_{B1}^o (4) | 233.5434082 | 215.1427 | 7.878924 | 160.8942887 | 31.10733 | 233.5398735 | 0.001514 |
| X_{B1}^o (5) | 350.3151123 | 320.9551 | 8.381029 | 237.1337093 | 32.0846 | 350.3094568 | 0.001614 |
| **MAPE** | 7.37 | 29.96374116 | **0.001412603** |

is on the basis of average rank obtained in simulated data (In-Samples) and forecasted data (Out-Samples).

2. Application of IOGM on previously reported approaches, motivated the author to employ this prediction theory for forecasting the pandemic growth in terms of reported infected cases. As it is indicated in the results of the benchmark data series that IOGM is compatible for all types of data and it can give fruitful results.

3. Further, it is apparent from the results that the average rank method is a suitable criterion for evaluating the performance of different prediction models.

Based on the results on known time-series data, the following section presents an application of conventional GM, NGM and proposed IOGMs for forecasting the spread of pandemic at different locations in India.
### Table 2
Simulated and Forecasted results of different Grey Models on B2.

| Sample | In-sample | GM(1,1) [14] | NGM(1,1,k) [23] | IOGM | Error |
|--------|-----------|-------------|------------------|------|-------|
|        |           | $a = -0.5544, b = -2.1190$ | $a = -0.5622, b = -0.7212$ | $a = -0.5693, b = -2.1749$ |
| $X_{21}^{(1)}$ | 5.84 | 5.84 | 0 | 5.84 | 0 |
| $X_{21}^{(2)}$ | 7.712 | 1.495361505 | 5.84 | 2.999639095 | 61.10426 |
| $X_{21}^{(3)}$ | 11.0816 | 2.603735408 | 7.712 | 4.295021234 | 61.24186 |
| $X_{21}^{(4)}$ | 17.1469 | 4.532391325 | 11.0816 | 6.567789434 | 61.69693 |
| $X_{21}^{(5)}$ | 28.0644 | 7.890744863 | 17.1469 | 10.55539082 | 62.38868 |
| $X_{21}^{(6)}$ | 47.7159 | 13.73752839 | 28.0644 | 8.549750751 | 69.53524 |
| $X_{21}^{(7)}$ | 83.0886 | 23.91658702 | 47.7159 | 15.10814402 | 76.83737 |
| $X_{21}^{(8)}$ | 146.7595 | 41.63799475 | 83.0886 | 26.69738831 | 76.88777 |
| $X_{21}^{(9)}$ | 261.3671 | 72.49038524 | 146.7595 | 47.1765785 | 76.8545 |
| $X_{21}^{(10)}$ | 467.66 | 126.2033867 | 261.3671 | 65.89076 | 69.10422 |

### Table 3
Simulated and Forecasted results of different Grey Models on B3.

| Sample | In-sample | GM(1,1) [14] | NGM(1,1,k) [23] | IOGM | Error |
|--------|-----------|-------------|------------------|------|-------|
|        |           | $a = -0.3026, b = 1.7313$ | $a = -0.2810, b = 0.4217$ | $a = -0.3049, b = 1.7449$ |
| $X_{31}^{(1)}$ | 3.32 | 3.32 | 0 | 3.32 | 0 |
| $X_{31}^{(2)}$ | 4.248 | 3.194908005 | 3.32 | 24.7903 | 67.60781 |
| $X_{31}^{(3)}$ | 5.1672 | 4.323835724 | 4.248 | 74.65351 | 68.43192 |
| $X_{31}^{(4)}$ | 6.5041 | 5.851672516 | 5.1672 | 83.5193931 | 69.68354 |
| $X_{31}^{(5)}$ | 8.6217 | 7.919377775 | 6.5041 | 162.51625 | 70.01061 |
| $X_{31}^{(6)}$ | 11.4084 | 10.71770179 | 8.6217 | 28.57526 | 70.54208 |
| $X_{31}^{(7)}$ | 15.4138 | 14.50482511 | 11.4084 | 8.0468383 | 71.07175 |
| $X_{31}^{(8)}$ | 20.8853 | 19.63013673 | 15.4138 | 24.23483 | 69.21202 |
| $X_{31}^{(9)}$ | 28.5994 | 26.5664884 | 20.8853 | 5.9295971 | 70.01061 |
| $X_{31}^{(10)}$ | 39.3031 | 35.95381507 | 28.5994 | 12.90571 | 70.01061 |

### 4. Simulation results

The proposed internal optimization based model is implemented to predict the infected cases in different states (Gujarat, Rajasthan and Maharashtra) and union territory Delhi. The data for this study has been taken from [38], [39]. For better understanding, it is to be noted here that variable $c^{(0)}_{W,b}(1)$ indicates the mean values of infected cases of COVID-19 for the duration of the first week of April to the Second week of May 2020. Time series is constructed by excluding two values for Model-1 and excluding only one entry for Model-2. For considering the uncertainty and unavailability of the data in a few cases, the mean of available data is considered for the forecast. A few points may be noted here:
Table 4
Simulated and Forecasted results of different Grey Models on B4.

| Sample   | In-sample | Simulated value | Error  | NGM(1,1,k)         | Error  | IOGM          | Error  |
|----------|-----------|-----------------|--------|--------------------|--------|---------------|--------|
|          |           |                 | a = -0.0854, b = 33.9309 | a = 0.3757, b = 29.7504 | a = -0.0854, b = 33.9549 | a = -0.0854, b = 33.9309 |
| X_{in}^0(1) | 78.3571 | 78.3571 | 0 | 78.3571 | 0 | 78.3571 | 0 |
| X_{in}^0(2) | 35.0894 | 42.40830908 | 20.85789 | 13.43485632 | 61.71249 | 42.43763141 | 20.94146 |
| X_{in}^0(3) | 40.8045 | 46.18944512 | 3.19694 | 34.02779312 | 16.60774 | 46.22328777 | 13.27988 |
| X_{in}^0(4) | 48.912 | 50.30770825 | 2.85309 | 48.17121952 | 1.514517 | 50.34664428 | 2.933113 |
| X_{in}^0(5) | 58.0352 | 54.79315681 | 5.386339 | 57.88506036 | 0.258704 | 54.83782553 | 5.509371 |
| X_{in}^0(6) | 66.453 | 59.67825915 | 10.19438 | 64.55661937 | 2.853177 | 59.72964339 | 10.11746 |
| X_{in}^0(7) | 77.831 | 64.99948257 | 16.54225 | 69.13871006 | 11.2258 | 65.05783672 | 16.47373 |
| X_{in}^0(8) | 68.2308 | 70.79485361 | 3.757912 | 72.28573387 | 5.942967 | 70.86133247 | 3.855345 |
| X_{in}^0(9) | 73.002 | 77.10694146 | 5.623051 | 74.47739999 | 1.97959 | 77.18253008 | 5.726597 |
| X_{in}^0(10) | 79.3541 | 83.98181674 | 5.83173 | 75.93161443 | 4.312928 | 84.06761124 | 5.726597 |

Forecasted results

| Out-sample | Simulated value | Error  | Simulated value | Error  | Simulated value | Error  |
|------------|-----------------|--------|-----------------|--------|-----------------|--------|
| X_{out}^0(1) | 79.4816 | 91.4696577 | 15.08281 | 76.95116571 | 3.183673 | 91.56687727 | 15.20513 |
| X_{out}^0(2) | 105.7574 | 99.62511655 | 5.798444 | 77.6514033 | 26.57591 | 99.73511664 | 5.694432 |
| X_{out}^0(3) | 95.2562 | 108.5077183 | 13.91145 | 78.13233319 | 17.97664 | 108.6320052 | 14.04193 |
| X_{out}^0(4) | 119.923 | 118.1822951 | 1.451519 | 78.46264045 | 34.57324 | 118.3225423 | 1.334571 |
| X_{out}^0(5) | 133.4256 | 128.7194598 | 5.327164 | 78.6894865 | 41.02369 | 128.8775255 | 3.408697 |

MAPE 9.382667169 11.82336045 8.75894572

Forecasting of such time series is quite difficult due to the unavailability of data. Also, at the initial stage, the value of the variable is quite small and it takes large abrupt changes at a later stage. This change sometimes is quite higher than the previous value. Also, it is to be noted here that the forecasting of the day ahead cases are merely meaningful as this short time forecast will give very less time to authorities for taking any preventive action. On the basis of this fact, the work reported in this paper addresses two representative models that can forecast weekly mean infected cases.

4.1. Results of Model-I

On the basis of the discussion and steps presented in Section 2.3, two models are constructed. These are named as Model-I...
and Model-II. As shown in Fig. 3, the period of overlap is 5 days, which indicates that new time series element $C_{(w,5)}^{(0)}(m)$ consists of 5 same values and last two are replaced by new values of infected case. The results of Model-I are shown in Tables 5–8 for Rajasthan, Maharashtra, Delhi and Gujarat respectively.

4.1.1. Forecasted results of Rajasthan

Forecasted results of model-I for the state of Rajasthan are depicted in Table 5. The data of reported infected cases have been segregated into two parts. First, 11 samples are taken as simulated data and the remaining five samples are considered for validation of grey models and for generating forecasts. The data of 6th April 2020–12th April 2020 is considered as $C_{(w,5)}^{(0)}(1)$ and the data of mean infected cases during 26th April 2020 to 5th May 2020 is denoted as $C_{(w,5)}^{(0)}(11)$. Simulation process with defined time series with 11 data points provides the coefficient ‘a’ and ‘b’ for representative models of GM, NGM and IOGM. These values are depicted with corresponding grey models. From the obtained results, it can be concluded that proposed model gives competitive performance as per Lewis’ criterion for model evaluation [30]. This criterion has been used for evaluating the performance of the grey models on the basis of calculated MAPE. Simulated and forecasted results of these grey models are depicted in Fig. 5. It is observed that the NMG model exhibits better results as the increment in the forecasted values are at a comparatively low exponential rate. Further, errors in the forecasting and simulation process have been encapsulated in Fig. 6.

4.1.2. Forecasted results of Maharashtra

Table 6 shows the comparative analysis of different grey forecasters with the proposed IOGM. The analysis is being depicted through the calculation of MAPE for simulated data and forecasted data. The infected data for the state of Maharashtra are employed for the construction of IOGM and other forecasters. The first 11 samples are taken for constructing grey architectures and the internal parameters (a, b) of these, are depicted with respective grey forecasters. For simplification, it is to be noted that the data sample from 6th April 2020 to 12th April 2020 is considered as $C_{(w,5)}^{(0)}(1)$ while $C_{(w,5)}^{(0)}(11)$ represents the data of mean infected cases during 26th April 2020 to 5th May 2020. After careful evaluation of the data, it can be concluded that for this particular data, IOGM’s performance indicator i.e. MAPE is competitive (4.1345972666). However, it can be concluded from this result that further improvement in the performance of IOGM may be possible, if the overlap period is varied. It is observed that the NGM model gives better results in this case, as the forecasted values increase at a comparatively lower exponential rate. Graphical representation of simulated and forecasted results along with errors in forecasting and simulation process of this model for Maharashtra state have been encapsulated in Figs. 5 and 6 respectively.

4.1.3. Forecasted results of Delhi

Results of grey models along with proposed IOGM (model-I) for Delhi are depicted in Table 7. Data of infected cases have been segregated into two parts, the first 10 samples are taken as simulated data and the remaining 6 samples have been considered for validation and forecasting purposes. The data from 7th April 2020 to 13th April 2020 are depicted in Table 7 as $C_{(w,5)}^{(0)}(1)$ and the data of mean infected cases during 25th April 2020 to 1st May 2020 are denoted as $C_{(w,5)}^{(0)}(27)$. Coefficients of constructed grey models are depicted along with the models. In addition to that error in the prediction of each sample is also shown in this analysis. Careful observation of Table 7 yields the fact that the proposed model exhibits competitive performance as values of MAPEs are competitive (3.278769158) and (2.160679097) for simulated and forecasted data. Depiction of forecasting performance and errors is exhibited in Figs. 5 and 6. It is observed that the IOGM model gives better results in this case, as the forecasted values increase with a comparatively higher exponential rate.

4.1.4. Forecasted results of Gujarat

Forecasted results of model-I for Gujarat state are depicted in Table 8. Similar to the results reported in the previous subsection for different states, the data of infected cases have been subdivided into two parts. First 5 samples are taken as simulated data and the remaining 5 samples are considered for validation and forecasting purpose. The time series of this model can be identified as $C_{(w,5)}^{(0)}(1) - C_{(w,5)}^{(0)}(5)$. The mean value of infected cases from 18th April 2020 to 24th April 2020 is considered as the first element of the time series and the mean value of infected cases from 21st April 2020 to 27th April 2020 is considered as the last element of the time series. Likewise, the data employed for simulation, yield the coefficients ‘a’ and ‘b’ for representative models of GM, NGM and IOGM respectively. The entries of these parameters are also depicted. Inspecting the obtained results, it is easily predictable that the proposed model gives a competitive performance as values of MAPEs are quite competitive. Proposed IOGM exhibits superior performance with simulated MAPE (0.4525292188769158) and forecasted MAPE (4.831822784). Simulated and forecasted results of these models are depicted in Fig. 5. It is observed that the IOGM model gives competitive results in this case as the forecasted values increase with a comparatively higher exponential rate. Errors in the forecasting and simulation process of this model for the state of Gujarat have been exhibited in Fig. 6.

4.2. Results of Model-II

The results of proposed Model-II are presented in this section. The difference between this model and the first model is that it employs an extended overlap period. The compilation of forecasted results is presented in Tables 9–12 for Delhi, Maharashtra, Rajasthan and Gujarat respectively.

4.2.1. Forecasted results of Delhi

Table 9 depicts results of forecasters in terms of In-Samples and Out-Samples for three grey forecasting models as described in the previous section also. Internal parameters of grey models (‘a’ and ‘b’) have also been depicted along with the errors. For constructing the time series, data from 6th April 2020 to 12th April 2020 is considered as $C_{(w,6)}^{(0)}(1)$ and $C_{(w,6)}^{(0)}(27)$ represents the data of mean infected cases from 2nd May 2020 to 8th May 2020. Time series is divided into two parts. First 27 samples are considered for constructing the grey model and for obtaining parameters (‘a’, ‘b’). The remaining, 5 samples are employed to evaluate the performance of constructed model.

As observed from the Table 9, IOGM outperforms other opponents, when MAPE of Out-Sample is considered. It is observed that values of MAPEs are quite competitive for the proposed IOGM. The MAPE values are quite competitive for simulated (4.441512436) and forecasted data (2.316478228). Pictorial representation of all these models is depicted in Fig. 7. It is also worth mentioning here that IOGM model gives competitive results in this case, as the forecasted value increases with every sample at a higher exponential rate. In addition to that, the graphical representation of simulation and forecasting errors for Delhi Model-II is presented in Fig. 8.
### Table 5
Simulated and Forecasted results for Rajasthan on Model-I.

| Sample | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM | Error |
|--------|-----------|--------------|-------|-----------------|-------|------|-------|
|        |           | a = −0.1411, b = 596.2662 |        |                  |       |      |       |
|        |           | Simulated value | 355.5714286 | 0 | 355.5714286 | 0 | 355.5714286 | 0 |
|        |           | Forecasted results | 3437.714286 | 4.134597266 | MAPE 3.888420598 | 13.60796909 | 0.096631 |

### Table 6
Simulated and Forecasted results for Maharashtra on Model-I.

| Sample | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM | Error |
|--------|-----------|--------------|-------|-----------------|-------|------|-------|
|        |           | a = −0.1884, b = 1316.9084 |        |                  |       |      |       |
|        |           | Simulated value | 1028.142857 | 0 | 1028.142857 | 0 | 1028.142857 | 0 |
|        |           | Forecasted results | 8690.571429 | 19.99368955 | MAPE 18.79731828 | 10.08917147 | 0.099982 |

---

A. Saxena
Applied Soft Computing 111 (2021) 107735

8.08522394
9.614796160
8.109403844
5.685926932
30.28745404
3.888420598
13.60796909
0.096631

MAPE 3437.714286 2471.787853 8.479718 5.36696 2480.611805 8.866976

Out-sample Simulated value Error Simulated value Error Simulated value Error

1028.142857 355.5714286 3437.714286 1028.142857 355.5714286 3437.714286

1028.142857 355.5714286 3437.714286 1028.142857 355.5714286 3437.714286

W(1) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(2) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(3) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(4) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(5) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(6) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(7) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(8) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(9) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(10) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857

W(11) 4274.857143 3522.428571 2587.285714 1386.285714 2352.571429 1028.142857
Simulated and Forecasted results for Delhi on Model-I.

Table 7
Simulated results

| Sample | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM | Error |
|--------|-----------|--------------|-------|----------------|-------|------|-------|
|        |           |              |       |                |       |      |       |
| C(0)W(1) | 744.8571429 | 0 | 0 | 744.8571429 | 0 | 744.8571429 | 0 |
| C(0)W(2) | 968.4285714 | 15.71825 | 576.8271009 | 40.43679 | 1122.089215 | 15.86701 |
| C(0)W(3) | 1239 | 1.970327 | 949.8658806 | 23.36909 | 1265.216358 | 2.115929 |
| C(0)W(4) | 1459.857143 | 2.431317 | 1282.342684 | 12.15971 | 1426.599962 | 2.278112 |
| C(0)W(5) | 1698.857143 | 5.476537 | 1578.667982 | 7.074707 | 1608.568716 | 5.314657 |
| C(0)W(6) | 1865.428571 | 2.950446 | 1842.77267 | 1.214515 | 1813.748341 | 2.770421 |
| C(0)W(7) | 2066.285714 | 1.222625 | 2078.16022 | 0.574679 | 2045.09948 | 1.025329 |
| C(0)W(8) | 2286.142857 | 0.651481 | 2287.953159 | 0.079186 | 2305.960418 | 0.866856 |
| C(0)W(9) | 2563.571429 | 1.193775 | 2474.93449 | 3.457557 | 2600.095253 | 1.424724 |
| C(0)W(10) | 2899.142857 | 0.88005 | 2641.58469 | 8.883945 | 2931.748209 | 1.124655 |

MAPE: 3.249480357 9.721718244 3.278769158

Forecasted results

| Sample | Out-sample | Simulated value | Error | Simulated value | Error | Simulated value | Error |
|--------|------------|-----------------|-------|----------------|-------|----------------|-------|
| C(0)W(11) | 3236.714286 | 1.86998 | 2790.114207 | 13.79795 | 3305.704877 | 2.131501 |
| C(0)W(12) | 3683.571429 | 0.915337 | 2922.4936 | 20.66141 | 3727.361273 | 1.188788 |
| C(0)W(13) | 4195 | 0.098965 | 3040.478863 | 27.52136 | 4202.801694 | 0.185976 |
| C(0)W(14) | 4846.142857 | 2.50206 | 3145.631526 | 35.08992 | 4738.886517 | 2.212231 |
| C(0)W(15) | 5560.428571 | 4.204512 | 3239.357336 | 41.74267 | 5343.351187 | 3.903969 |
| C(0)W(16) | 6233.142857 | 3.65661 | 3322.887863 | 46.69 | 6024.917838 | 3.34061 |

MAPE: 2.2084349 30.9172172 2.160679097

Out-sample SimulatedValue Error SimulatedValue Error SimulatedValue Error

Table 8
Simulated and Forecasted results for Gujarat on Model-I.

Simulated results

| Sample | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM | Error |
|--------|-----------|--------------|-------|----------------|-------|------|-------|
|        |           |              |       |                |       |      |       |
| C(0)W(1) | 1784.714286 | 0 | 0 | 1784.714286 | 0 | 1784.714286 | 0 |
| C(0)W(2) | 2234.142857 | 0.700285 | 1463.378541 | 34.49933 | 2254.19028 | 0.897321 |
| C(0)W(3) | 2650.857143 | 1.016826 | 2317.289441 | 12.58339 | 2629.492419 | 0.805955 |
| C(0)W(4) | 3077.142857 | 0.549699 | 2885.858555 | 6.216296 | 3067.278943 | 0.320554 |
| C(0)W(5) | 3569.428571 | 0.00895 | 3264.435335 | 8.544596 | 3577.952935 | 0.238816 |

MAPE: 0.455152029 12.36872234 0.452529218

Forecasted results

| Sample | Out-sample | Simulated value | Error | Simulated value | Error | Simulated value | Error |
|--------|------------|-----------------|-------|----------------|-------|----------------|-------|
| C(0)W(6) | 4125.142857 | 0.9083 | 3516.507377 | 14.75429 | 4173.649492 | 1.175878 |
| C(0)W(7) | 4751.285714 | 2.1788 | 3684.347344 | 22.45578 | 4868.524097 | 2.467509 |
| C(0)W(8) | 5467.51429 | 3.557973 | 3796.102121 | 30.5706 | 5679.088991 | 3.868583 |
| C(0)W(9) | 6224.428571 | 6.092475 | 3870.513069 | 37.81738 | 6624.605553 | 6.429136 |
| C(0)W(10) | 7011.142857 | 9.850317 | 3920.058939 | 44.08816 | 7727.542066 | 10.21801 |

MAPE: 4.517573198 29.93724099 4.831822784

4.2.2. Forecasted results of Maharashtra

Comparative analysis of the forecasting results for the state of Maharashtra is presented in Table 10. For constructing the time series for grey models like previous case studies, 27 such indicators have been considered for simulation and the remaining 5 samples are considered as Out-Samples for evaluating the constructed grey models. The internal parameters of grey forecasting systems such as (‘a’ and ‘b’) are shown in the respective columns. The data from 5th April 2020 to 11th April 2020 is considered as C(0)W(1) and C(0)W(27) represents the data of mean infected cases from 2nd May 2020 to 8th May 2020. A careful inspection of obtained MAPEs for simulated data In-Samples and Out-Samples indicate that MAPEs are quite competitive for proposed IOGM for simulated results (9.881179166)
and competitive for forecasted results (12.36321992). Graphical representation of the forecasting performance along with errors for each model are depicted in Figs. 7 and 8 respectively. It is concluded that Model-II provides quite competitive results with the proposed IOGM.

4.2.3. Forecasted results of Rajasthan
Forecasting results of Rajasthan state for all three grey models are depicted in Table 11 and graphical representation of the forecasting results is presented in Fig. 7. For construction of the time series, from 6th April 2020 to 12th April 2020 is considered as \( C^{(0)}(W,6) \) and data of mean infected cases from 3rd May 2020 to 9th May 2020 is considered as \( C^{(0)}(W,6) \) (27). Inspecting the forecasting results from Table 11 and Fig. 7, it is observed that for this particular case NGM model provides better results. Similar to previously reported results all internal parameters of the grey system are shown in Table 11.

4.2.4. Forecasted results of Gujarat
Table 12 shows the comparative analysis of different grey models along with the proposed IOGM (Model-II). The analysis is depicted through calculated error for each sample and the same is shown along with the sample. All parameters obtained for an internal grey mechanism such as ‘a’ and ‘b’ are shown in Table 12.
that the rise in infected cases in this particular state is swift forecasted results (4.338970394). It is also worth mentioning here (0.889304425) performance is better than competitors as the MAPE values obtained from the NGM model (11.25642155) for simulated data. The MAPE is very high (20.02449253) for forecasted data. Graphical representation of the forecasted results is presented in Fig. 7.

For constructing the time series for the state of Gujarat infected cases, 15 samples are taken for constructing the model and the remaining five samples are kept for testing the efficacy of the proposed IOGM.

The data from 18th April 2020 to 24th April 2020 represent \( C(W,G)(1) \), and \( C(W,G)(15) \) represent the data of mean infected cases from 2nd May 2020 to 8th May 2020. After assessment of results, it can be concluded that for this particular data IOGM’s performance is better than competitors as the MAPE values obtained for simulated data is optimal (0.889304425) and competitive for forecasted results (4.338970394). It is also worth mentioning here that the rise in infected cases in this particular state is swift comparatively, hence the NGM model provides pessimistic results. The same fact can be observed from higher MAPEs obtained from the NGM model (11.25642155) for simulated data. The MAPE is very high (20.02449253) for forecasted data. Graphical representation of the forecasted results is presented in Fig. 7.

Further, the analysis depicting simulation and forecasting errors for Gujarat Model-II is shown in Fig. 8. From this analysis, it can be concluded that proposed IOGM performs satisfactorily.
4.3. Comparative analysis of the proposed IOGM models

To showcase the efficacy of the proposed approach, the analysis based on errors reported in simulated and forecasted data have already been discussed in the previous section. On the basis of MAPE values of forecasting models, it can be concluded that proposed models based on different overlap periods and mean infected cases in the duration of a span of 6–7 days can be a potential tool for alignment of medical facilities and policy decisions. Further, to have a clear insight, comparative analysis of these proposed grey models are depicted in Fig. 9. The following points can be concluded from this analysis:

1. Fig. 9(a) and (c) show the MAPE of forecasted results of IOGM models. It can be easily concluded that the proposed IOGM (Model-II) provides competitive results as compared to the results obtained by conventional GM and NGM models.

2. Fig. 9(b) depicts the average rank analysis. For conducting this analysis, developed models have given rank as per the performance. The evaluation of performance is based on the calculated MAPE for simulated and forecasted samples. After taking the mean of the MAPE obtained from forecasted and simulated models, it has been observed that the average rank of IOGM (Model-II) is 1 (1.5) as compared to...
other grey models i.e. 2 and 2.5. However, the results for Model-I are quite comparable with the original GM model. Hence, it is to be noted that for Model-II proposed IOGM based methodology provides very competitive results. This method is suitable for forecasting as it produces meaningful results without knowing the pattern of variables. It can generate a reliable forecast for planning combating strategies.

3. Fig. 9(d) depicts the average MAPE analysis obtained by IOGM models (state-wise). As shown, the average MAPE obtained by models I and II are (11.76 and 7.82), (16.29, 11.37), (13.09, 9.54) and (21.95, 13.25) for Delhi, Maharashtra, Gujarat and Rajasthan respectively. It can be concluded that the proposed IOGM model-II exhibits superior performance as the average MAPE calculated for forecasting is optimal.

4.4. Discussion

From the results reported in this section, it can be concluded that the estimated results are always higher than the actual infected cases. This indicator is sufficient enough to spark an
Table 12
Simulated and Forecasted results for Gujarat on Model-II.

| Sample | In-sample | GM(1,1) [14] | Error | NGM(1,1,k) [23] | Error | IOGM | Error | Simulated value | Simulated value | Simulated value |
|--------|-----------|--------------|-------|-----------------|-------|------|-------|-----------------|----------------|-----------------|
| \( C_{\text{GM}(1,1)}^{(1)} \) | 1784.714286 | 1784.714286 | 0     | 1784.714286     | 0     |      |       |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(2)} \) | 2013.714286 | 2112.805827 | 4.920834 | 834.784594     | 58.54503 | 2113.801078 | 4.970258 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(3)} \) | 2234.142857 | 2274.533853 | 1.807897 | 1427.202548     | 36.11856 | 2275.681355 | 1.859259 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(4)} \) | 2443.714286 | 2448.641604 | 0.201632 | 1951.998846     | 20.12164 | 2449.958835 | 0.255535 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(5)} \) | 2650.857143 | 2636.076705 | 0.557572 | 2416.892175     | 8.826012 | 2419.7202175 | 0.500752 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(6)} \) | 2862.571429 | 2837.85932 | 0.863284 | 2828.72017 | 1.182547 | 2839.575755 | 0.803322 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(7)} \) | 3077.142857 | 3055.0877 | 0.716741 | 3193.539984 | 3.782636 | 3057.037707 | 0.653371 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(8)} \) | 3316.428571 | 3288.944166 | 0.828735 | 3516.717373 | 6.039292 | 3291.153449 | 0.762119 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(9)} \) | 3569.428571 | 3540.701541 | 0.804808 | 3803.00562 | 6.543822 | 3543.198372 | 0.734857 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(10)} \) | 3841.714286 | 3811.730078 | 0.78049 | 4056.615444 | 5.593887 | 3814.545537 | 0.707204 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(11)} \) | 4125.142857 | 4103.504918 | 0.524538 | 4281.276927 | 3.784937 | 4106.726627 | 0.447735 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(12)} \) | 4429 | 4417.61412 | 0.257076 | 4480.294385 | 1.158148 | 4421.726627 | 0.176729 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(13)} \) | 4751.285714 | 4755.767302 | 0.094324 | 4656.594959 | 1.99295 | 4759.757337 | 0.178302 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(14)} \) | 5104.285714 | 5119.804948 | 0.304043 | 5182.248007 | 16.74339 | 5124.27169 | 0.391553 |                 |                 |                 |
| \( C_{\text{GM}(1,1)}^{(15)} \) | 5467.571429 | 5511.708425 | 0.80725 | 4951.121564 | 9.44659 | 5516.701481 | 0.898572 |                 |                 |                 |
| MAPE | 0.897948273 | 11.25642155 | 0.889304425 | 4.281084185 | 20.02449253 | 4.338970394 | 0.897948273 | 11.25642155 | 0.889304425 | 4.281084185 | 20.02449253 | 4.338970394 |

Fig. 7. Forecasted Results of Proposed Grey Model-II for different states and Delhi.

alarm to the authorities. However, the authenticity of the forecast largely depends upon the removal of potential uncertainties in the data. Another problem with the forecasting of epidemic and pandemic is that the data of confirmed cases multi-folds with time. Apart from these issues, forecasting is immensely valuable as it allows us to foresee many preventive and corrective measures in health care. Model-I and II give a sufficient amount of accuracy in the prediction of mean weekly infected cases. Higher values of MAPE can be justified with the larger population in three states and Delhi. Following recommendations can be drawn from this forecast:
1. It can be seen that the performance of the models relies on the mean infected cases of a duration of more than five days. It is also a known fact that by taking the average of infected cases, the forecaster can easily deal with the randomness in data. This randomness is due to environment, policies, strategic decisions, sentiments and medical conditions.

2. Considering a large population and the density of population in the major states of India, it may be concluded that based on the predictions of pandemic spread in these states, authorities can take decisions on the availability of Intensive Care Units and for severe cases, more ventilators can be procured.

3. It is also empirical to spread awareness of this deadly disease in rural areas. Here, the authorities can plan online/offline campaigns to educate people before it hits the masses. Also, the strategies can be framed to impose lockdown in certain states and the period of lockdown can also be calculated based on this forecast.

4. In addition to the above-cited recommendations, special care is to be taken of those patients who are already suffering from other diseases and taking regular treatment from hospitals. Based on the infected cases forecast, local hospitals, schools and some unused official buildings can be converted into COVID relief and cure centres. Based on the prediction results, the supply of required first aid...
treatment programme and medicines can be foreseen. An awareness programme for the first line of medication can be developed. The knowledge of these programmes can be disseminated at different levels.

5. Conclusion

Novel Coronavirus poses a threat to human beings. This has significantly changed the way of thinking towards life. As the pandemic hit masses, the prediction methods offer help to medical practitioners, policymakers, and leaders of the states and countries to combat this disease effectively. The work reported in this paper has discussed difficulties in the forecasting of spread of pandemic and it has offered a probable solution in the form of the proposed grey mathematics-based optimization model. Following are the major conclusions of this work:

- This paper has presented theoretical aspects of GPMs. Also, it has discussed how the accuracy of these models is compromised due to the inherent nature of the models. An Internal Optimization-Based Model has been proposed for addressing these issues. This model has been validated on benchmark time series data. After the validation of this model, two sub-models have been developed with the help of different overlap periods to conduct the forecast of pandemic spread.
- These models are based on the mean values of the infected cases in the three major states and Delhi consisting of different overlap patterns i.e. 5 and 6 days respectively.
- The proposed prediction models are based on internal optimization and also on the hypothesis that performance can substantially be enhanced with the help of a careful selection of the grey model’s internal parameters. Further, both models have been tested on three major states and Delhi and forecasting of the infected cases has been done. It is observed that the values of error indices are optimal as compared with non-optimized models.
- The comparison of optimized models and non-optimized conventional models such as GM and NGM has been done in terms of evaluation of error indices. Further, this analysis has been extended to the evaluation of the average rank associated with these models. It has been observed that the proposed models perform satisfactorily as ranks obtained by these models are optimal in comparison to other grey models. Further, it is stated that the results of the proposed models are closely aligned with the actual data.
- For extending the analysis, the average MAPE of proposed IGM models (place wise) have been evaluated. Moreover, it is observed that the proposed model II (with a higher overlap period) yields satisfactory results. Based on prediction results, certain suggestions and recommendations have been framed. These recommendations can be further utilized for framing the policies and preventive strategies for the COVID-19 by the Government of India.

The comparative analysis of performance of other grey models for forecasting Corona spread can be a future research direction. It will be interesting to develop new grey models with the application of nature-inspired optimizers in future.

CRediT authorship contribution statement

Akash Saxena: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Supervision, Software, Validation, Writing – review & editing.
[29] X. Ma, X. Mei, W. Wu, X. Wu, B. Zeng, A novel fractional time delayed grey model with Grey Wolf Optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China, Energy 178 (2019) 487–507.

[30] P. Zhang, X. Ma, K. She, A novel power-driven grey model with whale optimization algorithm and its application in forecasting the residential energy consumption in China, Complexity 2019 (2019).

[31] L.-C. Hsu, Applying the grey prediction model to the global integrated circuit industry, Technol. Forecast. Soc. Change 70 (6) (2003) 563–574.

[32] C.-C. Hsu, C.-Y. Chen, Applications of improved grey prediction model for power demand forecasting, Energy Convers. Manage. 44 (14) (2003) 2241–2249.

[33] H. Zhuan, Mechanism and application of GM (1, 1) and improvement methods, Syst. Eng. Theory Practice 2 (1993) 56–62.

[34] Z. Hengyang, Modeling mechanism and prerequisites for GM (1, 1) and the revised methods, Syst. Eng.-Theory Methodol. Appl. 2 (3) (1993) 56–62.

[35] S.-C. Chang, H.-C. Lai, H.-C. Yu, A variable P value rolling grey forecasting model for Taiwan semiconductor industry production, Technol. Forecast. Soc. Change 72 (5) (2005) 623–640.

[36] N. Xu, Y. Dang, S. Ding, Optimization method of background value in GM (1, 1) model based on least error, Control Decis. 30 (2) (2015) 283–288.

[37] B. Zeng, C. Li, Forecasting the natural gas demand in China using a self-adapting intelligent grey model, Energy 112 (2016) 810–825.

[38] kaggle data, COVID-19 corona virus India dataset, 2020, https://www.kaggle.com/imdevskp/covid19-corona-virus-india-dataset.

[39] M. of Health, F. Welfare, MHFQ COVID-19 India dataset, 2020, https://www.mohfw.gov.in/.

Dr. Akash Saxena received his Ph.D. and Master of Technology Degree (with Honours) in Electrical Engineering from Malviya National Institute of Technology, Jaipur in 2015 and 2008 respectively and a Bachelor of Engineering Degree (with Honours) in Electrical Engineering from Engineering College Kota in 2001. Dr. Akash is the proud recipient of Chartered Management Institute (Level 3) Diploma in Management which represents a global benchmark for excellence. Also, he is a senior member of IEEE. Dr. Akash Saxena has been working as a Professor in Electrical Engineering Department of SKIT, Jaipur, India since 2014 with almost two decades of teaching.

As an academician, he has been associated with numerous institutions as Member, Doctoral Research Committee and Member, Board of Studies. He has presented his research work at various platforms to showcase his findings to the fellow researchers and providing a pathway to the future researchers. His intellectual work has been published in leading journals in the form of short communication/letters/articles/research papers. A number of prestigious professional organizations have taken him on board as an editor, reviewer and adviser. Dr. Akash Saxena has to his credit a number of accomplished sponsored projects from the prestigious government organizations. Dr. Akash Saxena who has been a passionate researcher is currently immersed into extensive research in the areas of Computational Intelligence, Application of Artificial Intelligence in the Power System, and Smart Grid. He has published more than 70 quality research papers in international refereed journals and conferences and supervised 22 Master theses hitherto.