Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/taut20

Data grouping and modified initial condition in grey model improvement for short-term traffic flow forecasting

Vincent Birundu Getanda, Peter Kamita Kihato, Peterson Kinyua Hinga & Hidetoshi Oya

To cite this article: Vincent Birundu Getanda, Peter Kamita Kihato, Peterson Kinyua Hinga & Hidetoshi Oya (2023) Data grouping and modified initial condition in grey model improvement for short-term traffic flow forecasting, Automatika, 64:1, 178-188, DOI: 10.1080/00051144.2022.2119500

To link to this article: https://doi.org/10.1080/00051144.2022.2119500

AUTOMATIKA
2023, VOL. 64, NO. 1, 178–188
https://doi.org/10.1080/00051144.2022.2119500

REGULAR PAPER

OPEN ACCESS Check for updates

Data grouping and modified initial condition in grey model improvement for short-term traffic flow forecasting

Vincent Birundu Getanda[a], Peter Kamita Kihato[a], Peterson Kinyua Hinga[a] and Hidetoshi Oya[b]

[a]Department of Electrical and Electronic Engineering, School of Electrical, Electronic and Information Engineering, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya; [b]Department of Computer Science, Tokyo City University, Tokyo, Japan

ABSTRACT
To improve the performance of the conventional grey model, emphasis should be based on the "new information prior using" principle. This paper presents detailed work on improving the precision of the conventional grey model by combining a data grouping technique with modification of initial condition to establish an optimized grey model. The data grouping technique and modification of initial condition methods have the advantage of adhering to the "new information prior using" principle. An empirical example of short-term traffic flow forecasting shows that the proposed optimized grey model, that is the modified initial condition grouped grey model, outperforms the existing models in both fitting and short-term forecasting. Moreover, the results demonstrate our claim that the distribution characteristic of the fitting error influences future short-term forecast accuracy. Now the proposed model can be of help in intelligent transportation systems for optimizing the use of existing infrastructure to enhance urban transportation systems in averting issues such as traffic congestion.

ARTICLE HISTORY
Received 24 November 2021
Accepted 26 August 2022

KEYWORDS
Grey model; data grouping; initial condition; forecasting; intelligent transport systems

1. Introduction

Expansion of road capacity is a common but a "lengthy-term" approach to managing traffic congestion [1,2]. However, intelligent transportation systems (ITS) have solved and managed traffic problems by providing varied and advanced modes of transport [3–5]. For smart implementation of ITS, real-time traffic information is vital [4]. However, the provision of accurate real-time traffic information is an inherent problem in ITS [6]. Consequently, much effort to study on safety and efficiency of ITS has been intensified and devoted [3,7]. To improve the performance of ITS, agent-based evidential reasoning and "ZigBee" approaches have been proposed [6,7]. Furthermore, Amin et al. [8] introduced big data in shaping ITS. In this paper, we propose a combination of the data grouping technique (DGT) with modification of initial condition (MIC) to develop an optimized grey model (GM) which we refer to as the single-variable first-order modified initial condition grouped grey model and denote it as MICGGM(1,1). Where the first 1 stands for first-order and the second 1 stands for 1 variable under consideration in the grey differential equation. This optimized model can find application in advancing ITS. The Original Grey Model (OGM(1,1)), which we optimize in this paper, is well known and has been applied in modelling real-time events in other domains such as hepatitis B incidence and beef consumption forecasting [9,10].

Many intellectuals have investigated on enhancing the accuracy of the conventional GM(1,1) from various perspectives. For example, the function transformation technology, initial condition and background value modification have been considered to improve the accuracy of the OGM(1,1) [11–16]. Additionally, Thành [17] demonstrates that the application of the Fourier series in revising the grey model residual values provides more accuracy than the conventional model. From the grey system theory, to improve the precision of the grey model, emphasis should be based on the "new information prior using" principle [18]. Consequently, we note that most of the authors' efforts to improve the performance of the grey model have not concentrated on modification of the initial condition in adhering to the "new information prior using" principle. This has motivated us in proposing a combination of the DGT and MIC methods, which adheres to this principle, in improving the performance of the OGM(1,1). Therefore, this paper institutes and proves a new perspective of amalgamating the DGT and MIC in optimizing the accuracy of the OGM(1,1) [19,20].

The novelty of this paper is the proposed combining of the DGT with MIC in order to optimize the OGM(1,1). Getanda et al. [21] discussed the DGT and MIC in improving the OGM(1,1)'s accuracy but they never combined the two methods. This new strategy takes the advantage of the two methods in

CONTACT Vincent Birundu Getanda getandav@yahoo.com

adhering to the "new information prior using" principle and in combination the accuracy of the proposed MICGGM(1,1) is generally good compared with existing models as shown later in this paper. And that is the crucial difference between the work of Getanda et al. [21] and the result of this paper. Thus, the proposed strategy is an extension of previous results [19,22]. Further, in this paper, we unveil the insight of the DGT as in how the grouping of raw data is done and how simulated data are superimposed to obtain overall fitted data. More importantly, we clearly show the adherence to the "new information prior using principle" by the DGT. Moreover, we analyze the relative fitting error and make judgment on its influence on future forecast results. Finally, we graphically demonstrate how the accuracy of the OGM(1,1) is improved from a lower accuracy range to an improved higher accuracy range by this new approach.

2. The grey system

The GM(1,1) is based on the grey system theory which was introduced in 1982 and has undergone several modifications being applied in numerous fields for analyzing, estimating, forecasting and modelling grey systems [9,10,23–26]. The system extracts a governing relationship of a system and covers areas such as grey generating space, grey forecasting, grey control etc. [24,26–28].

2.1. Grey generating operations and modelling algorithm

In grey modelling, the grey generating techniques develop a systematic series from a raw data series. In this paper, we present the raw data series as [9,26]:

$$X^{(0)} = \{x^{(0)}_{(1)}, x^{(0)}_{(2)}, x^{(0)}_{(3)}, \cdots, x^{(0)}_{(m)}\} \tag{1}$$

where m is the total number of data points. Accumulation of the series in (1) by:

$$X^{(1)}_{(r)} \hat{=} \left\{ \sum_{i=1}^{r} x^{(0)}_{(i)} \right\}, \; r = 1, 2, \cdots, m, \tag{2}$$

results in the following equation:

$$X^{(1)} \hat{=} \{x^{(1)}_{(1)}, x^{(1)}_{(2)}, \cdots, x^{(1)}_{(m)}\} \tag{3}$$

which is the Accumulated Generating Operation (AGO). The AGO is a pre-processing technique which reduces the randomness of the actual data to improve data regularity and smoothness.

From adjacent AGO neighbours, a Mean value Generating Operation (MGO) can be obtained as [24,26]:

$$z^{(1)}_{(k)} = 0.5(x^{(1)}_{(k)} + x^{(1)}_{(k-1)}), \; k = 1, 2, \cdots, m, \tag{4}$$

where $z^{(1)}_{(k)}$ is the background value.

The grey modelling algorithm has a unique characteristic of the first-order differential equation given as [26,29,30]:

$$\frac{d}{dt} X^{(1)}(t) + a X^{(1)}(t) = b, \tag{5}$$

where $X^{(1)}(t)$ is a background grey value at time t, and a and b are the developing coefficient and grey input, respectively [9]. The parameters a and b are optimized and obtained as discussed in Section 2.2 [19,26]. The time response function of (5) is deduced as [29]:

$$\hat{x}^{(1)}_{(r+1)} \hat{=} \left(x^{(0)}_{(1)} - \frac{b}{a} \right) e^{-ar} + \frac{b}{a}, \; r = 0, 1, 2, \cdots, m-1, \tag{6}$$

where $\hat{x}^{(1)}_{(r+1)}$ is the prediction of the AGO and $x^{(0)}_{(1)}$ of (6) is the initial condition which causes prediction error [18].

And from (1) and (3), the following equation can be obtained [24]:

$$x^{(0)}_{(k)} + a z^{(1)}_{(k)} = b, \; k = 1, 2, \cdots, m, \tag{7}$$

This is a grey differential model, called "Grey Model (first-order, single-variable)", GM(1,1), where $x^{(0)}_{(k)}$ is a grey derivative [26,28,30].

Now, a systematical sequence of the original series is obtained by retrogressing through inverse accumulated generating operation (IAGO) given by [27,29]:

$$\hat{x}^{(0)}_{(r+1)} = \hat{x}^{(1)}_{(r+1)} - \hat{x}^{(1)}_{(r)}, \; \hat{x}^{(0)}_{(1)} = \hat{x}^{(1)}_{(1)}, \quad r = 1, 2, \cdots, m-1. \tag{8}$$

2.2. Estimating model parameters

The grey model parameter values a and b are such that the best-fit result minimizes the sum of squared errors or residuals which are the differences between the observed or experimental values and the corresponding fitted value given in the model. Thus, the ordinary least-square method best estimates these parameters which are calculated as [19,26,29]:

$$\begin{bmatrix} a \\ b \end{bmatrix} = [A^T A]^{-1} A^T y, \tag{9}$$

where A is the data matrix and y is the measured vector and are given as:

$$A = \begin{bmatrix} -z^{(1)}_{(2)} & 1 \\ -z^{(1)}_{(3)} & 1 \\ \vdots & \vdots \\ -z^{(1)}_{(m)} & 1 \end{bmatrix} \tag{10}$$

and

$$y = \begin{bmatrix} x^{(0)}_{(2)} \\ x^{(0)}_{(3)} \\ \vdots \\ x^{(0)}_{(m)} \end{bmatrix} \tag{11}$$

And as long as the matrix A has a full rank and the inverse of A^TA exists, then it is possible to compute "good" values of the parameters a and b that minimizes the sum of squared errors.

3. Proposed accuracy-improving methods

In this paper, we propose a combination of the DGT and MIC methods in boosting the accuracy of the OGM(1,1). Based on DGT, a Grouped Grey Model (GGM(1,1)) exists [22]. Furthermore, a modified initial condition grey model (MICGM(1,1)) based on MIC also exists [21]. Now in combination, the DGT and MIC methods result in an optimized Modified Initial Condition Grouped Grey Model referred to as MICGGM(1,1). Then the proposed model, MICGGM(1,1), is adopted for modelling short-term traffic flow forecasts and compared with the existing models for validation purposes. An additional existing model considered in this paper is the Modified Background Value Grey Model (MBVGM(1,1)) [23].

3.1. Data grouping technique

In this paper, we adopt the Strong Grouping (SG) technique in which the number of formed groups is given by [19,23]:

$$N = n - [k - 1], \tag{12}$$

where N is the number of groups, n is the total number of data used, and k is the number of data points in a group and it should be consistent throughout the grouping process. Figure 1 illustrates the SG in 4s (i.e. four data points per group) to form 19 groups as can be identified by 19 rectangles. Because of the dropping of old data and adding of new data in the grouping process, the SG technique has the advantage of adhering to the "new information prior using" principle [20,23]. The procedure of grouped grey modelling entails the repeated application of the conventional grey model on each group of data and averaging the predicted values at points of overlap [21]. Thus, grouped grey modelling inherently modifies the initial condition of (6) and this promotes accuracy improvement.

3.2. Modification of the initial condition (MIC)

As mentioned earlier, the initial condition of (6) is a cause of the precision error of the conventional grey model [18]. Thus, modification of this initial condition to enhance the model's accuracy is important and it is achieved as follows. As in Getanda et al. [20,21], MIC is adopted and by applying IAGO on $\hat{x}^{(1)}_{(r)}$ of (8), the restored (predicted) value of the raw data is obtained as:

$$\begin{aligned} \hat{x}^{(0)}_{(r)} &= \hat{x}^{(1)}_{(r)} - \hat{x}^{(1)}_{(r-1)} \\ &= C*(1 - e^{a})*e^{-ar} \\ &= C*(e^{-ar} - e^{a}*e^{-ar}) \\ &= C*(e^{-ar} - e^{-a(r-1)}),\ r = 2, 3, \ldots m, \end{aligned} \tag{13}$$

where Cis the optimized initial condition given as:

$$C = \frac{\sum_{r=2}^{m}(e^{-ar} - e^{-a(r-1)}) * x^{(0)}_{(r)}}{\sum_{r=2}^{m}(e^{-ar} - e^{-a(r-1)})^2} \tag{14}$$

Now by incorporation of this MIC into the OGM(1,1), an optimized model is established which has the advantage of improved accuracy compared with the OGM(1,1).

Thus, the process of time series grey prediction based on MIC can be outlined in the following four steps:

(a) Calculation of the background value from the AGO sequence of (3) by (4),
(b) Computation of the developing coefficient and grey input parameters by (9),
(c) Computation of the optimized value of C by (14) and finally,
(d) Computation of the restored (predicted) values by (13).

3.3. The proposed MICGGM(1,1)

Now to improve the OGM(1,1)'s accuracy in short-term traffic flow forecast, we combine the DGT and the MIC methods in conventional grey modelling. The combination process involves the incorporation of MIC in each formed group data set. In other words, the raw time series data of Figure 2 is grouped before fitting and forecasting and the four-step process outlined above is adopted in the simulation processes. As illustrated by Figure 1, the DGT drops an old data point and adds a new data point in the formation of a

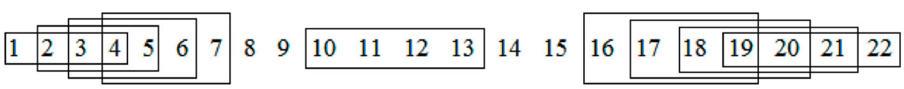

Figure 1. Strong grouping (SG). The 22 data points have been grouped in 4s to form 19 groups.

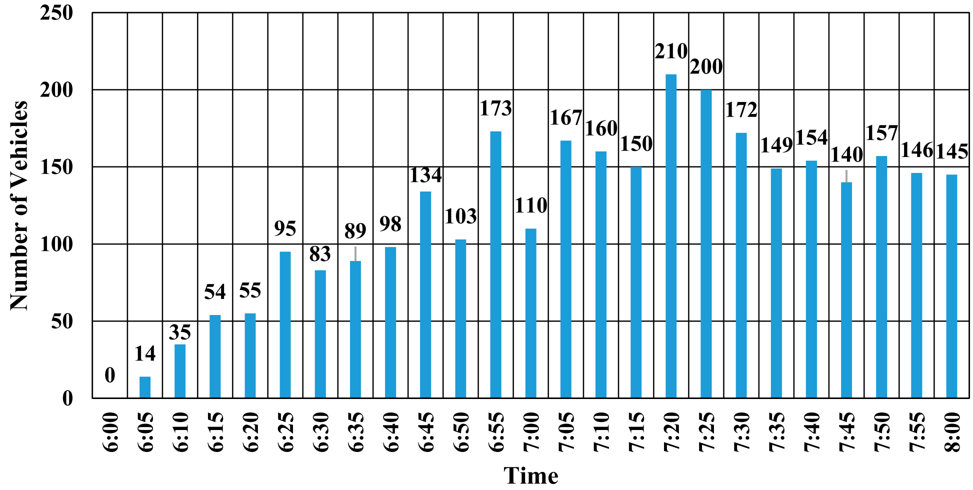

Figure 2. Time series traffic flow raw data. This is the number of vehicles passing a point of study every 5 min.

new group. This process keeps on modifying the initial condition of each formed group. Overall, the two methods have the advantage of modifying the initial condition as a requirement by the "new information prior using" principle, and in combination, we establish the proposed model denoted as MICGGM(1,1). Moreover, MICGGM(1,1) can achieve high accuracy compared with the existing models as demonstrated in Sections 5–7 of this paper. With this high accuracy, the MICGGM(1,1) can be used in advancing ITS by optimizing the use of existing infrastructure to enhance urban transportation systems in averting issues such as traffic congestion.

The proposed model can be applied in both univariate and multivariate data prediction capturing the non-linear features of the traffic data, unlike other statistical methods such as the Autoregressive Integrated Moving Average (ARIMA) model which is univariate and does not support seasonal data. However, Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. Neural networks such as the recurrent neural networks have a short-term memory and this "forgetfulness" makes model training more difficult and time-consuming. With the proposed model, no data gets lost even though the modification of the initial condition may literally mean loss of information, but this is compensated by the adoption of the "new information prior using" principle. Generally, grey models require small amount of data to predict a system. Other models such as the ARIMA model and the long short-term memory neural network require large amount of historical traffic data to achieve short-term traffic flow forecasting [31]. Further, the predictive effect of the ARIMA model depends on its parameters and orders. Thus, the proposed model is more advantageous in time-series data modelling and forecasting.

4. Vehicle traffic forecast data

4.1. The origin of the data set

The two-hour vehicle traffic data set used in this paper was obtained from Ref. [19] as simulated in MITRAM [32,33]. The data are graphically presented in Figure 2. The number of vehicles is seen to have a trending pattern from six o'clock in the morning to the peak hour at around 7:20 AM. This is the morning peak when the traffic jam is high. Thereafter, the flow of traffic decreases to a lower value at around eight o'clock AM. These types of data are suitable for this research as our objective is to forecast by the proposed model for the purpose of its incorporation in ITS to avert traffic congestion. Thus, it is desirable that we subject our new model to such traffic congestion data.

4.2. Training and validation data sets

The original traffic flow data set of Figure 2 was subdivided into training and validation data sets as presented in Figure 3 for estimating and evaluating the proposed grey models, respectively [34]. The goal is to develop a trained (fitted) model for good generalization to new, unknown data. Then the fitted model is evaluated using "new" data from the held-out dataset (validation dataset) to estimate the model's accuracy in classifying the new data. The reason for sub-dividing the original data set is to reduce the risk of issues such

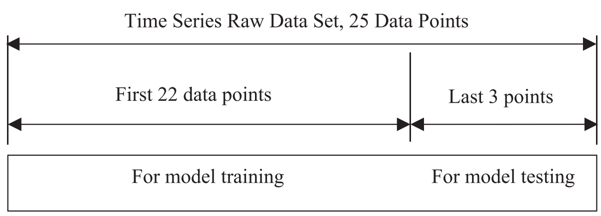

Figure 3. Data split in two parts.

as overfitting and, therefore, the data in the validation dataset are not used to train the models. Thus, in this paper, we employed the classical hold-out method.

5. Empirical application review

For analysis of the performance of the proposed grey model, we simulated a numerical example in MATLAB R2014a. The time series data of Figure 2 was subdivided into two parts as illustrated in Figure 3. Data from 6:00 AM to 7:45 AM are used for training the grey models, whilst data from 7:50 AM to 8:00 AM are used for testing the models.

5.1. Training the grey models

Based on the grey generating operations, we trained the OGM(1,1), MBVGM(1,1) and the MICGM(1,1) and tabulated their simulation values in Table 1. Figure 4 shows the plots for the real, simulated and residual data curves. The residual curve indicates the difference between the simulated and actual values.

In training for the proposed model, we adopt the DGT and MIC in the simulation as follows. First, we group the first 22 data points shown in Figure 2 into 19 groups of 4s based on (12) and Figure 1. These groups

Table 1. Original and modified grey models' simulation data.

Rawdata		Grey models			
K	Real	OGM(1,1)	MBVGM(1,1)	MICGM(1,1)	MICGGM(1,1)
Training		Simulated Values			
1	0	0	0	0	0
2	14	71.2956	71.4918	69.5654	15.9281
3	35	75.0718	75.2683	73.2499	35.1083
4	54	79.0479	79.2444	77.1295	50.8452
5	55	83.2346	83.4305	81.2147	60.9350
6	95	87.6431	87.8377	85.5162	88.8117
7	83	92.2851	92.4777	90.0455	86.2782
8	89	97.1729	97.3628	94.8147	87.3272
9	98	102.3196	102.5060	99.8365	102.5789
10	134	107.7389	107.9209	105.1242	122.9481
11	103	113.4452	113.6218	110.6921	119.5043
12	173	119.4538	119.6239	116.5548	153.5545
13	110	125.7805	125.9430	122.7281	128.7178
14	167	132.4424	132.5959	129.2283	156.4133
15	160	139.4572	139.6003	136.0728	158.9242
16	150	146.8434	146.9747	143.2798	159.0757
17	210	154.6209	154.7386	150.8685	202.3187
18	200	162.8103	162.9127	158.8592	200.9595
19	172	171.4335	171.5185	167.2731	171.9366
20	149	180.5133	180.5790	176.1326	153.0167
21	154	190.0741	190.1180	185.4613	148.3897
22	140	200.1413	200.1610	195.2842	143.2855
Testing		Short-term Forecasted Values			
23	157	210.7416	210.7345	205.6273	122.4716
24	146	221.9034	221.8666	216.5182	130.0599
25	145	233.6564	233.5866	227.9860	131.0204

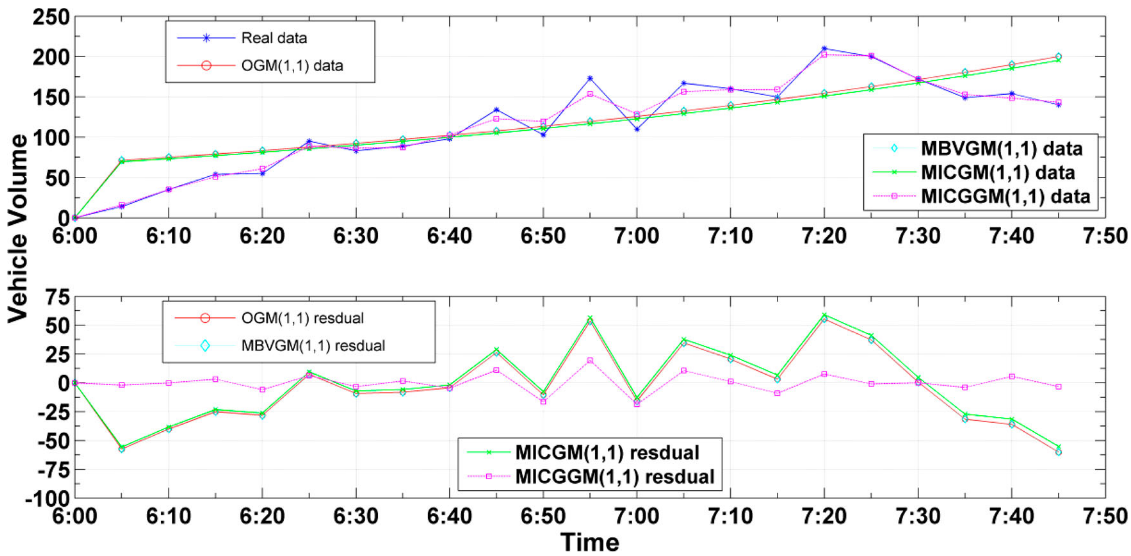

Figure 4. Vehicle flow fitting. The first 22 data points have been used for training the grey models.

Table 2. Actual data (AD) grouping based on SG in 4s.

		Groups 1–19																		
DP	AD	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17	G 18	G 19
1	0	0																		
2	14	14	14																	
3	35	35	35	35																
4	54	54	54	54	54															
5	55		55	55	55	55														
6	95			95	95	95	95													
7	83				83	83	83	83												
8	89					89	89	89	89											
9	98						98	98	98	98										
10	134							134	134	134	134									
11	103								103	103	103	103								
12	173									173	173	173	173							
13	110										110	110	110	110						
14	167											167	167	167	167					
15	160												160	160	160	160				
16	150													150	150	150	150			
17	210														210	210	210	210		
18	200															200	200	200	200	
19	172																172	172	172	172
20	149																	149	149	149
21	154																		154	154
22	140																			140

are formed and tabulated as in Table 2 where DP stands for the data point.

Second, we introduce MIC in the OGM procedure [9] and apply it to each group to obtain fitted data (FD) for each group as tabulated in Table 3. Note that each group will have unique values of the parameters a and b [19] which have not been provided in this paper. Consequently, the corresponding time functions are different. Also note that in Table 3, the FD for groups 6–14 is not shown because of wanting to reduce the size of Table 3.

Third, the overall simulation data sequence is obtained by superimposing the group simulation data at points of overlaps and this overall (final) sequence is as indicated in Table 3. For instance, groups 1 and 2 overlap once at data point 2 and, thus, the final FD is obtained as $(17.8561 + 14)/2 = 15.9281$ and groups 1, 2 and 3 overlap twice at data point 3 resulting to a final fitted value computed as $(31.2975 + 39.0274 + 35)/3 = 35.1083$. Also, groups 1, 2, 3 and 4 overlap three times at data point 4 and the final fitted value is obtained as $(54.8572 + 47.4416 + 47.082 + 54)/4 = 50.8452$. This sequence of computations is continued to generate the final simulation sequence (i.e. final FD) which is also tabulated in Table 1.

Lastly, this proposed model, MICGGM(1,1), resulted in simulation values tabulated in Table 1 and plotted in Figure 4. A close observation of the error curves in Figure 4 reveals that the proposed MICGGM(1,1) is the most accurate model in this fitting.

Table 3. Group and final fitted data.

	Groups 1–5 and 15–19 fitted data										
AD	G1	G2	G3	G4	G5	G15	G16	G17	G18	G19	Final FD
0	0										0
14	17.8561	14									15.9281
35	31.2975	39.0274	35								35.1083
54	54.8572	47.4416	47.082	54							50.8452
55		57.6698	65.4726	65.5975	55						60.9350
95			91.0465	77.0809	92.1196						88.8117
83				90.5745	88.9628						86.2782
89					85.9141						87.3272
98											102.5789
134											122.9481
103											119.5043
173											153.5545
110											128.7178
167											156.4133
160						160					158.9242
150						163.8034	150				159.0757
210						185.7543	212.9778	210			202.3187
200						210.6468	193.412	199.7791	200		200.9595
172							175.6437	172.4208	167.6817	172	171.9366
149								148.809	158.1476	152.0936	153.0167
154									149.1556	147.6238	148.3897
140										143.2855	143.2855

In Sections 6 and 7 of this paper, we further evaluate the fitting errors and accuracies of the models as shown in Figures 7–9.

5.2. Validating the fitted grey models by short-term forecast

For forecasting in the near term (6) is extrapolated three points into the future. In testing the grey models, these short-term forecasting results are recorded in Table 1 and their accuracies are shown in Figure 9. Plotted in Figure 5 are the real, simulated and residual curves for the three grey models. The last three time sample points in Figure 5 (i.e. at 7:50, 7:55 and 8:00 AM) are the extrapolated points of focus. Hence to clearly show this short-term forecast, Figure 6 is an extracted part of Figure 5.

6. Error analysis

Our goal is to find the model that has the best performance on new data and the simplest approach is to evaluate the time-series fitting error involved in the models. The error of any particular model in forecasting future trends of a system is much influenced by its error in fitting the past data of that particular system. To validate this claim, we examine the distribution characteristic of the relative fitting errors over the full range of the data set under consideration.

Table 4 is a record of OGM(1,1)'s, MBVGM(1,1)'s, MICGM(1,1)'s and MICGGM(1,1)'s relative fitting errors. MICGGM(1,1)'s relative errors are smaller. However, its large error part concentrates at the centre (middle) of the data set, whereas the errors of the OGM(1,1) are larger and distributed over the full range of the data set (Figure 7). At either ends of the data

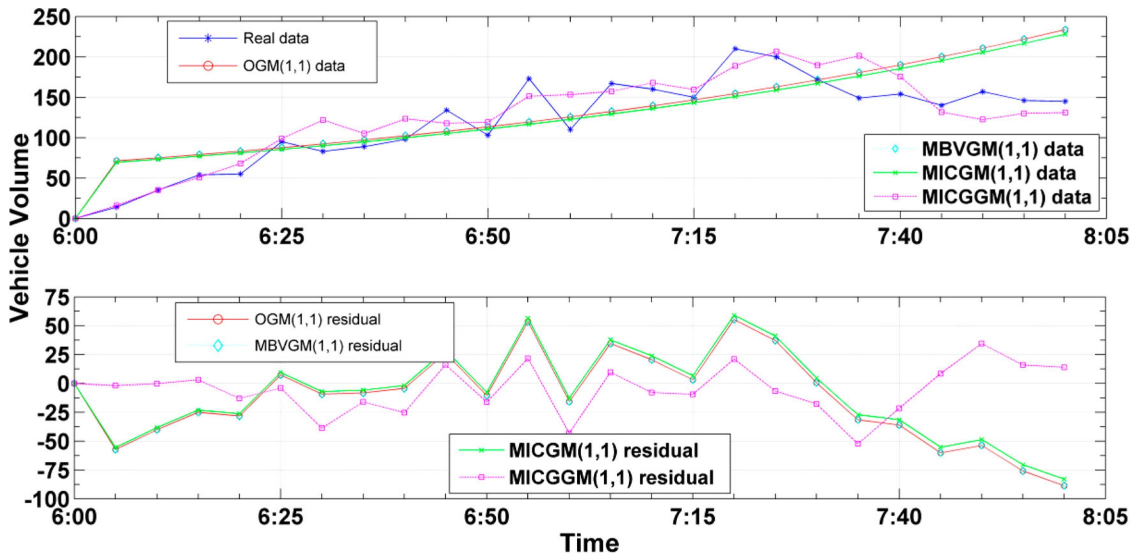

Figure 5. Short-term vehicle flow forecast-extrapolation. The last three instants show traffic flow estimation.

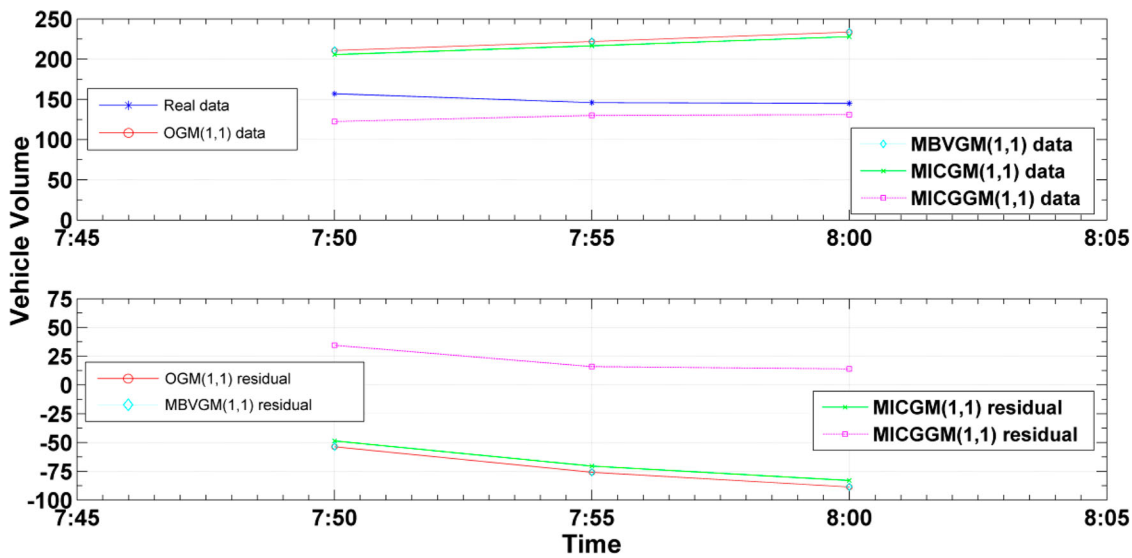

Figure 6. Short-term vehicle flow forecast. This figure clearly illustrates three extrapolated points of focus.

Table 4. Model simulated relative errors.

	OGM(1,1)		MBVGM(1,1)		MICGM(1,1)		PROPOSED MICGGM(1,1)	
Real Data	Fitted Data	Relative Error	Fitted Data	Relative Error	Fitted Data	Relative Error	Fitted Data	Relative Error
0	0	0	0	0	0	0	0	0
14	71.2956	−57.2956	71.4918	−57.4918	69.5654	−55.5654	15.9281	−1.9281
35	75.0718	−40.0718	75.2683	−40.2683	73.2499	−38.2499	35.1083	−0.1083
54	79.0479	−25.0479	79.2444	−25.2444	77.1295	−23.1295	50.8452	3.1548
55	83.2346	−28.2346	83.4305	−28.4305	81.2147	−26.2147	60.935	−5.935
95	87.6431	7.3569	87.8377	7.1623	85.5162	9.4838	88.8117	6.1883
83	92.2851	−9.2851	92.4777	−9.4777	90.0455	−7.0455	86.2782	−3.2782
89	97.1729	−8.1729	97.3628	−8.3628	94.8147	−5.8147	87.3272	1.6728
98	102.3196	−4.3196	102.5060	−4.5060	99.8365	−1.8365	102.5789	−4.5789
134	107.7389	26.2611	107.9209	26.0791	105.1242	28.8758	122.9481	11.0519
103	113.4452	−10.4452	113.6218	−10.6218	110.6921	−7.6921	119.5043	−16.5043
173	119.4538	53.5462	119.6239	53.3761	116.5548	56.4452	153.5545	19.4455
110	125.7805	−15.7805	125.9430	−15.9430	122.7281	−12.7281	128.7178	−18.7178
167	132.4424	34.5576	132.5959	34.4041	129.2283	37.7717	156.4133	10.5867
160	139.4572	20.5428	139.6003	20.3997	136.0728	23.9272	158.9242	1.0758
150	146.8434	3.1566	146.9747	3.0253	143.2798	6.7202	159.0757	−9.0757
210	154.6209	55.3791	154.7386	55.2614	150.8685	59.1315	202.3187	7.6813
200	162.8103	37.1897	162.9127	37.0873	158.8592	41.1408	200.9595	−0.9595
172	171.4335	0.5665	171.5185	0.4815	167.2731	4.7269	171.9366	0.0634
149	180.5133	−31.5133	180.5790	−31.5790	176.1326	−27.1326	153.0167	−4.0167
154	190.0741	−36.0741	190.1180	−36.1180	185.4613	−31.4613	148.3897	5.6103
140	200.1413	−60.1413	200.1610	−60.1610	195.2842	−55.2842	143.2855	−3.2855

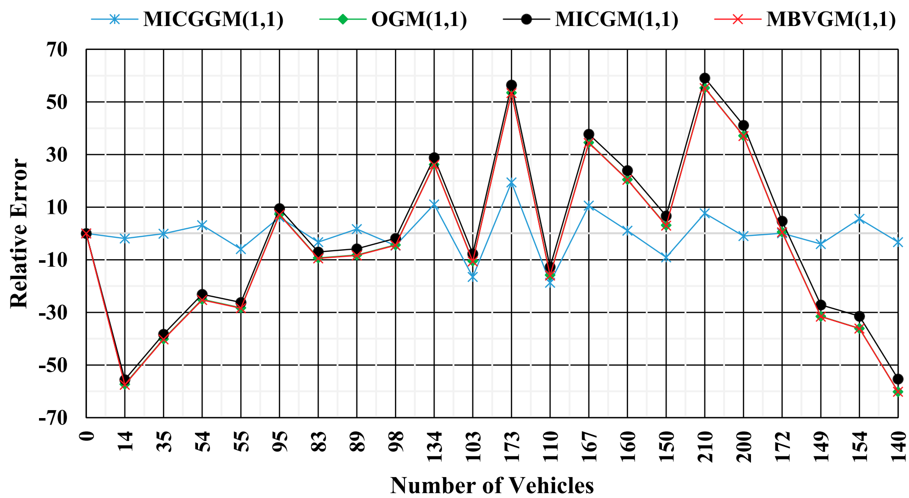

Figure 7. Error analysis. These are the distribution characteristics of the relative fitting errors.

set range, the accuracy of the proposed MICGGM(1,1) is high meaning that it can forecast the future trend of a system with good accuracy. For the OGM(1,1), MBVGM(1,1) and MICGM(1,1), the Lager error trend is likely to distribute into the future forecasting of a system and these models can forecast the future with lower accuracy. The role played by the DGT is to concentrate the fitting errors in the middle of the time series rather than spreading it along the entire data set. The fitting accuracy at the middle is lower, whereas at the extreme ends, it is good and this is advantageous in forecasting the system behaviour at the ends. This is why the MICGGM(1,1) has a high fitting and good/reasonable forecasting accuracies compared with OGM(1,1), MBVGM(1,1) and MICGM(1,1).

Similarly, we considered the relative fitting errors of the MICGM(1,1). Observe Figure 7 and note that MICGM(1,1), MBVGM(1,1) and OGM(1,1) have similar error distribution characteristics. Thus, they have almost the same fitting accuracies, see Figure 8. Although MICGM(1,1) and MBVGM(1,1) are improved grey models, their accuracies are much lower compared with that of the proposed model.

7. Accuracy improvement analysis

Measures of model performance, namely root mean square percentage error (RMSPE), mean absolute percentage deviation (MAPD), root mean square error (RMSE) and the mean absolute error (MAE) are adopted to evaluate the accuracy improvement of the proposed grey model [19].

In MATLAB R2014a, we computed the simulation errors, and in Excel, we plotted the fitting accuracies of each model as shown in Figure 8. Similarly, we plotted the short-term forecasting accuracies in Figure 9. It is evident from Figure 8 that the OGM(1,1) has lower fitting accuracy compared with MICGGM(1,1).

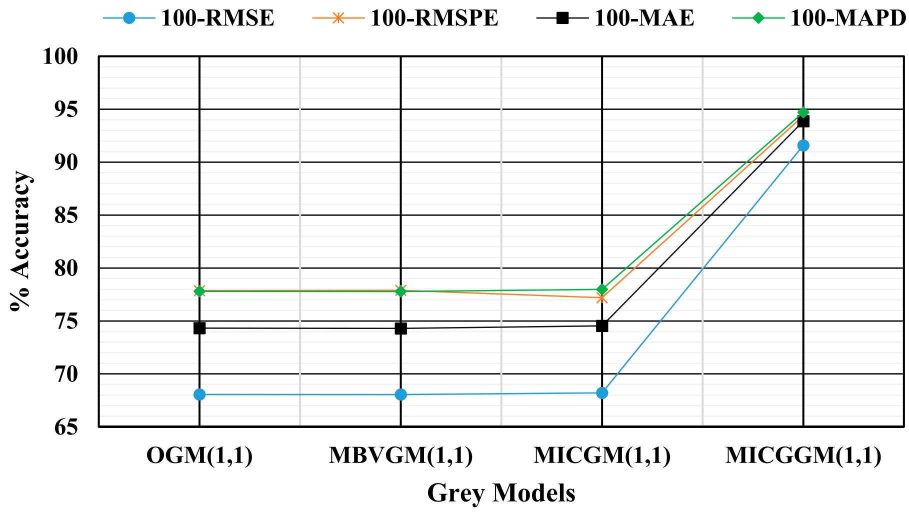

Figure 8. Model fitting accuracy analysis.

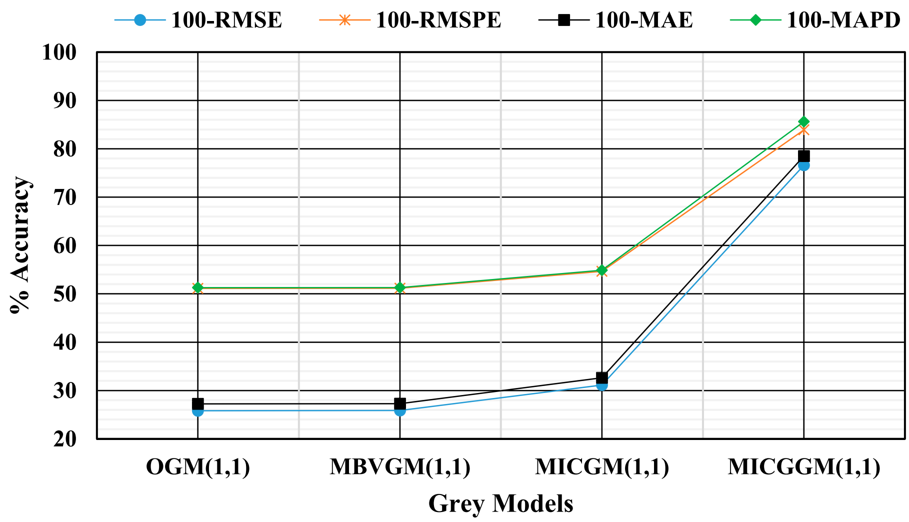

Figure 9. Model shot-term forecasting accuracy analysis.

Note that the proposed MICGGM(1,1) has the highest fitting accuracy of 94.6987%. Thus, the OGM(1,1)'s accuracy of 77.802% has been improved to an accuracy of 94.6987% by MICGGM(1,1) as indicated by MAPD. See Figure 8. On the other hand, MICGM(1,1) and MBVGM(1,1) have fitting accuracies of 77.9812% and 77.7807 respectively.

Short-term forecast accuracy analysis is shown in Figure 9 in which the OGM(1,1) has lower short-term forecast accuracy. On the other hand, the proposed MICGGM(1,1) has good forecast accuracy at 85.6143%, as computed by MAPD. Worthy is the improvement of the OGM(1,1)'s fitting and short-term forecast accuracies as shown in Figures 8 and 9, respectively.

Table 5. Criteria for MAPD and RMSPE.

MAPD and RMSPE	Forecasting power
Less than 10%	High accuracy
10–20%	Good
20–50%	Reasonable
More than 50%	Inaccurate

The criteria of MAPD and RMSPE are as tabulated in Table 5 [19,35,36] and it shows that the fitting accuracy of the proposed MICGGM(1,1) is high, whereas its short-term forecasting accuracy is good.

8. Computation time

We did the simulations in MATLAB R2014a environment and to measure the performance of our MATLAB code, we used the *tic* and *toc* performance functions to time how long the code takes to run. We ran the code multiple times and averaged and recorded the time in seconds as tabulated in Table 6. The fitting and forecasting time for the MICGGM(1,1) is longer compared with those of the OGM(1,1). This is because of the increased number of statements to be executed in the proposed model's MATLAB code. Nevertheless, the MICGGM(1,1) is more accurate and the increased computation time is insignificant. Moreover, timing is merely a performance metric, not a system correctness criterion. Thus, the proposed model can perform well in automated systems.

Table 6. MATLAB code computing time.

	Computing time	
Grey model	Fitting	Forecasting
OGM(1,1)	0.3791	0.3586
MBVGM(1,1)	0.3824	0.3698
MICGM(1,1)	0.3834	0.3739
MICGGM(1,1)	0.4092	0.3924

9. Conclusions and future focus

In this paper, we have analyzed and presented a new approach of amalgamating the DGT and MIC in optimizing the conventional GM(1,1) to establish a newly optimized model denoted as MICGGM(1,1). Further, with a numerical example, we did a performance analysis and the results have shown that the newly improved model, i.e., the MICGGM(1,1) has high fitting accuracy and good short-term forecasting accuracy as compared with the existing models. Moreover, we have shown that when large errors tend to concentrate at the middle of the data set, the near future can forecast with high accuracy meaning the new model has the best performance on new data. Thus, the distribution characteristic of the fitting error influences the future forecast accuracy of the model. We note that the proposed MICGGM(1,1) combines the advantage of DGT and MIC by adhering to the "new information prior using" principle in boosting the accuracy of the OGM(1,1).

The proposed optimized GM can enhance the design of useful ITS's architecture through grey control and this is of help to transport management decision-makers. Thus, it enriches and widens the application scope of the grey model. For instance, it can be useful in real-time proactive traffic control for large-scale adaptive traffic networks (such as those of city or mega-city scale) and this calls for multiple measurement points of data to be processed.

Our future work is to continue refining and developing the GM by combining the DGT with Fourier Series to ascertain their capability in enhancing the OGM(1,1)'s forecasting accuracy.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

We wish to acknowledge the African Development Bank Group for the financial support given to this research.

Data availability statement

We, the authors of this paper, confirm that the data supporting the findings of this research are available from the corresponding author [V. B. G.] and can be shared upon reasonable request.

References

[1] Nugmanova A, Arndt WH, Hossain MA, et al. Effectiveness of ring roads in reducing traffic congestion in cities for long run: big Almaty ring road case study. Sustainability. 2019;11(18):4973.

[2] Tennøy A, Tønnesen A, Gundersen F. Effects of urban road capacity expansion – experiences from two Norwegian cases. Transp Res D Transp Environ. 2019;69(2019):90–106. doi:10.1016/j.trd.2019.01.024.

[3] Alam M, Ferreira J. Editorial of the special issue: intelligent transportation systems. Information. 2017;8(4):146.

[4] Singh B, Gupta A. Recent trends in intelligent transportation systems: a review. J Transp Lit. 2015;9(2):30–34.

[5] Stawiarska E, Sobczak P. The impact of intelligent transportation system implementations on the sustainable growth of passenger transport in EU regions. Sustainability. 2018;10(5):1318.

[6] Benalla M, Achchab B, Hrimech H. Improving driver assistance in intelligent transportation systems: an agent-based evidential reasoning approach. J Adv Transp. 2020;2020:1–14.

[7] Iqbal R, Yukimatsu K. Intelligent transportation systems using short range wireless technologies. J Transp Technol. 2011;1(4):132–137.

[8] Amin MA, Hadouej S, Darwish TS. Big data role in improving intelligent transportation systems safety: a survey. International conference on emerging internetworking, data & Web technologies; Cham: Springer; 2019. p. 187–199.

[9] Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM (1, 1) models for prediction of hepatitis B in China. PloS One. 2018;13(9):e0201987.

[10] Zeng B, Li S, Meng W, et al. An improved gray prediction model for China's beef consumption forecasting. PloS One. 2019;14(9):e0221333.

[11] Cao C, Fan C, Hu Z. Grey forecasting model and its application based on the sine function transformation. J Math. 2013;33(4):697–701.

[12] Liu X, Peng H, Bai Y, et al. Tourism flows prediction based on an improved grey GM (1, 1) model. Procedia-Soc Behav Sci. 2014;138(2014):767–775.

[13] Mahdi MH, Mohamed N. An improved GM(1,1) model based on modified background value. Inf Technol J. 2017;16(1):11–16.

[14] Wei X, Aiping Y, Wenzhan D. Improvement of background value and its application in non-equidistance GM (1, 1) modelling. 2010 8th world congress on intelligent control and automation; IEEE; 2010. p. 5919–5923.

[15] Ying L, Min-an T, Tao L. Study on optimization for grey forecasting model. AASRI international conference on industrial electronics and applications (IEA 2015); Atlantis Press; 2015. p. 275–279.

[16] Zhang J, Lü X, Ran M, et al. DGM model based on anti-cotangent function and its application. J Grey Syst. 2016;28(3):63–74.

[17] Thành PV. Enhanced accuracy of grey forecasting model: a case by tourism industry in Vietnam. Hue Univ J Sci (HU JOS). 2016;113(14):157–167.

[18] Jong Y, Liu S. Grey power models based on optimization of initial condition and model parameters. Grey Syst Theory Appl. 2014;4(2):370–382.

[19] Getanda VB, Oya H, Kubo T. Improving the accuracy of the GM(1,1) by data grouping technique and its application to forecast vehicle volume and CO_2 emission in

Tokushima city, Japan. Proceedings of the 36th IASTED international conference on modelling, identification and control (MIC 2017); ACTA Press; 2017. p. 39–46.
[20] Mahdi MH, Mohamed N. Optimized GM(1,1) model based on the modified initial condition. J Appl Sci. 2017;17(2):90–96.
[21] Getanda VB, Kihato PK, Hinga PK, et al. Improved grey models in vehicular CO_2 emission modelling and forecasting to decarbonize urban areas – sustainable environment. Proceedings of the 9th international conference on appropriate technology (9th ICAT), Pretoria, South Africa; 2020. p. 88–100. https://www.scribd.com/document/485488919/9th-ICAT-Technology-Exchange-and-Employment-Creation-for-Community-Empowerment-Cross-Pollinating-Innovative-Models.
[22] Getanda VB, Oya H, Kubo T, et al. Data grouping techniques' performance analysis in GM(1,1)'s prediction accuracy improvement for forecasting traffic parameters. Mechatron Syst Control. 2020;48(1):25–34.
[23] Getanda VB, Kihato PK, Hinga PK, et al. Improved grey model, GM(1,1), in short-term traffic flow forecasting-smart transportation systems. In 2019 IEEE AFRICON; IEEE; 2019. p. 1–9.
[24] Li Q, Lin Y. Review paper: a briefing to grey systems theory. J Syst Sci Inf. 2014;2(2):178–192.
[25] Liu S, Forrest J, Yang Y. A brief introduction to grey systems theory. Grey Syst Theory Appl. 2012;2(2):89–104.
[26] Lu M. Grey system: theory, methods, applications and challenges. Bucharest: Leverhulme Trust Workshop on Grey Systems and Applications; 2015.
[27] Shen Y, He B, Qin P. Fractional-order grey prediction method for non-equidistant sequences. Entropy . 2016;18(6):227.
[28] Slavek N, Krmpotić D, Blažević D. Grey system theory approach to quality of intranet. Int J Adv Res Electr Electron Instrum Eng. 2015;4(10):8223–8230.
[29] Chang CJ, Li DC, Dai WL, et al. Utilizing an adaptive grey model for short-term time series forecasting: a case study of wafer-level packaging. Math Probl Eng. 2013;2013:1–6.
[30] Iqelan BM. Forecasts of female breast cancer referrals using grey prediction model GM (1, 1). Appl Math Sci. 2017;11(54):2647–2662.
[31] Lu S, Zhang Q, Chen G, et al. A combined method for short-term traffic flow prediction based on recurrent neural network. Alexandria Eng J. 2021;60(1):87–94. DOI:10.1016/j.aej.2020.06.008
[32] Ishikawa R, Honda N, Kazama H. The wide-area road traffic analysis by microscopic road traffic simulator MITRAM. IPSJ Trans Math Model Appl. 2005;46 (SIG17):46–55.
[33] Mori S, Tada K, Idu R, et al. Traffic jam analysis and CO_2 emission models in Tokushima city via MITRAM. Proceedings of 2013 annual conference of electronics, information and systems society; IEEJ; 2013. p. 1046–1051. Japanese.
[34] Liu H, Cocea M. Semi-random partitioning of data into training and test sets in granular computing context. Granular Comput. 2017;2(4):357–386.
[35] Hsu LC, Wang CH. Applied multivariate forecasting model to tourism industry. Tour Int Interdiscip J. 2008;56(2):159–172.
[36] Lou P, Wang L, Zhang X, et al. Modelling seasonal brucellosis epidemics in Bayingolin Mongol autonomous Prefecture of Xinjiang, China, 2010–2014. BioMed Res Int. 2016;2016:1–17.