Traffic & Transportation

Short-Term Traffic Flow Uncertainty Prediction Based on Novel GM(1,1)

Xu Dong CAO[1], Qin SHI[2], Yi Kai CHEN[3], Chen Chen CHEN[4]

Original Scientific Paper
Submitted: 18 Oct. 2023
Accepted: 4 Jan. 2024

[1] Corresponding author, cxd19910223@126.com, School of Automotive and Transportation Engineering, Hefei University of Technology

[2] shiqin@hfut.edu.cn, School of Automotive and Transportation Engineering & Key Laboratory for Automated Vehicle Safety Technology of Anhui Province, Hefei University of Technology

[3] yikaichen@hfut.edu.cn, School of Automotive and Transportation Engineering, Hefei University of Technology

[4] 982563796@qq.com, College of Civil Engineering, Anhui Jianzhu University

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Anticipating uncertainty in short-term traffic flow is crucial for effective traffic management within intelligent transportation systems. Various methods for predicting uncertainty have been proposed and implemented. However, conventional techniques struggle to provide accurate forecasts when confronted with sparse data. Hence, this study focuses on developing an uncertainty prediction model for short-term traffic flow under limited data conditions. A novel grey model that considers the volatility of the traffic data is proposed, which extends the grey model (GM) by integrating two techniques: smooth pre-processing and background value construction. The performance of the proposed novel grey model is mainly illustrated by comparing the novel grey model with the traditional GM model. Our results, in terms of uncertainty quantification, demonstrate that the proposed model outperforms the GM model regarding mean kick-off percentage (KP), width interval (WI) and width amplitude.

KEYWORDS

intelligent transportation systems; uncertainty quantification; novel GM model; smooth pre-processing; background value construction.

1. INTRODUCTION

Urban traffic congestion has caused problems around the world, which not only impair residents' experience while travelling but also increase travel times [1]. To enhance the operational efficiency of urban road networks, numerous strategies have been devised to alleviate traffic congestion. Among these strategies, traffic management and control systems have gained heightened importance. The categories of traffic management and control systems include reactive systems and proactive systems [2]. In contrast to reactive systems, proactive systems heavily depend on precise predictions to optimise their functionality. Consequently, accurate short-term traffic forecasting stands as a pivotal element within proactive traffic control systems.

Short-term traffic prediction has attracted attention from scholars around the world over the past decades, and many methods have been proposed. Intuitively, most traditional prediction methods implement accurate and effective forecasts with sufficient and complete data [3], however, these methods cannot obtain accurate prediction results under limited data conditions. Therefore, Deng [4] proposed the use of grey prediction methods, which can effectively handle this forecasting problem with limited data. Since grey prediction methods were first proposed, they have been improved to adapt to different domains, e.g. stock prices [5], energy markets [6–8] and transportation [9].

While several grey prediction methods have been devised, their forecasting outcomes are confined to level predictions and fail to capture crucial uncertainty information, such as prediction intervals, essential for decision-makers. To address this challenge, researchers have made efforts to generate interval prediction results through the development of grey interval prediction models. Nevertheless, conventional grey interval prediction models exhibit certain limitations: (i) do not consider the volatility of the traffic data; (ii) subjectively partition the original data into upper and lower sequences. Hence, these grey interval prediction methods lead to increased forecast error. To reduce the forecast error, it is necessary to study a novel grey model that considers the volatility of traffic data and investigate the interval prediction results of short-term traffic flow.

The objective of this study is to yield a prediction interval of traffic flow for reflecting the uncertainty of prediction under limited data conditions. Specifically, a novel grey model, which extends the grey model (GM) by integrating two techniques: the smoothness operators of volatility sequence and background value construction, was proposed in this study. Traffic flow data collected from the road network of Furong District in Changsha were used. To facilitate model comparison, we evaluated and contrasted the traditional GM model with the newly proposed model. Two key indicators, namely, the mean kick-off percentage (KP) and width interval (WI), were employed to gauge the accuracy of traffic flow interval predictions made by these models. The subsequent analysis delves into the performance of the novel grey model in short-term traffic interval prediction. The main contributions of the study are: (a) a novel GM model that considers the volatility of traffic data was proposed by integrating two techniques (smooth pre-processing and background value construction) in this paper and compared to the traditional GM model; and (b) the results of interval prediction can be yielded using the proposed novel GM model.

The paper is structured as follows: Section 2 offers a comprehensive literature review. Section 3 elucidates the specifics of the proposed model. Section 4 presents an empirical study showcasing the efficacy of the model. Lastly, Section 5 encapsulates our study's conclusions.

2. LITERATURE REVIEW

In the past few decades, various traffic prediction methods have been proposed and applied in point prediction [10–15]. Compared with point prediction, the studies of traffic uncertainty quantification are fairly limited [16–18]. Traditionally, the input uncertainty and model uncertainty are two main aspects that are used in uncertainty quantification analysis [19, 20].

Several literatures of input uncertainty usually assumed that the input variables are statistically distributed and then randomly extracted from these distributions [21, 22]. Uncertainty can be quantified by examining the variance across all runs of the input variables employed in model executions. Various methods for quantifying model uncertainty have been proposed, typically relying on analytical expressions to compute the variance of endogenous variables, thereby characterising prediction uncertainty. For instance, the jackknife method creates subsamples from the original dataset by systematically excluding a small fraction of the data, enabling the calculation of standard errors [23]. Additionally, the Bootstrap method involves random sampling with replacement from the original dataset to determine proper standard errors for model coefficients. Recently, the generalised autoregressive conditional heteroskedasticity (GARCH) model has been borrowed from the economy field for uncertainty quantification [24]. Although the above methods have been applied, the forecast results of these models are limited to sufficient and complete data and cannot yield more useful information under limited data conditions.

In light of this, several grey prediction methods have been utilised to derive interval forecasts, serving as representations of prediction uncertainty. These methods encompass the grey straight horn band interval (GPBI) prediction model [25], the grey wrapping band interval (GWBI) prediction model, and the grey envelop prediction model (GEPM) [26]. The GPBI model and GWBI model are very similar. The difference is that the former model divides the sequence into upper group and lower group by using straight lines, whereas the latter uses the exponential line. Furthermore, the GEPM has the capability to determine the upper and lower bounds of the prediction interval based on the maximum and minimum envelope curves generated by the GWBI model. These models can yield interval prediction results. However, these models usually use subjective classification methods to classify the original sequence into upper group and lower group [27]. This will result in a larger prediction interval, affecting the interval prediction accuracy. It's worth noting that these grey interval prediction methods do not account for the inherent volatility within traffic flow data.

3. METHODOLOGY

This study proposes a novel grey model that considers the volatility of traffic data, which extends the grey model (GM) by integrating two techniques: the smoothness operators of volatility sequence and background value construction to yield an accurate prediction interval. The modelling procedure is shown in *Figure 1*.

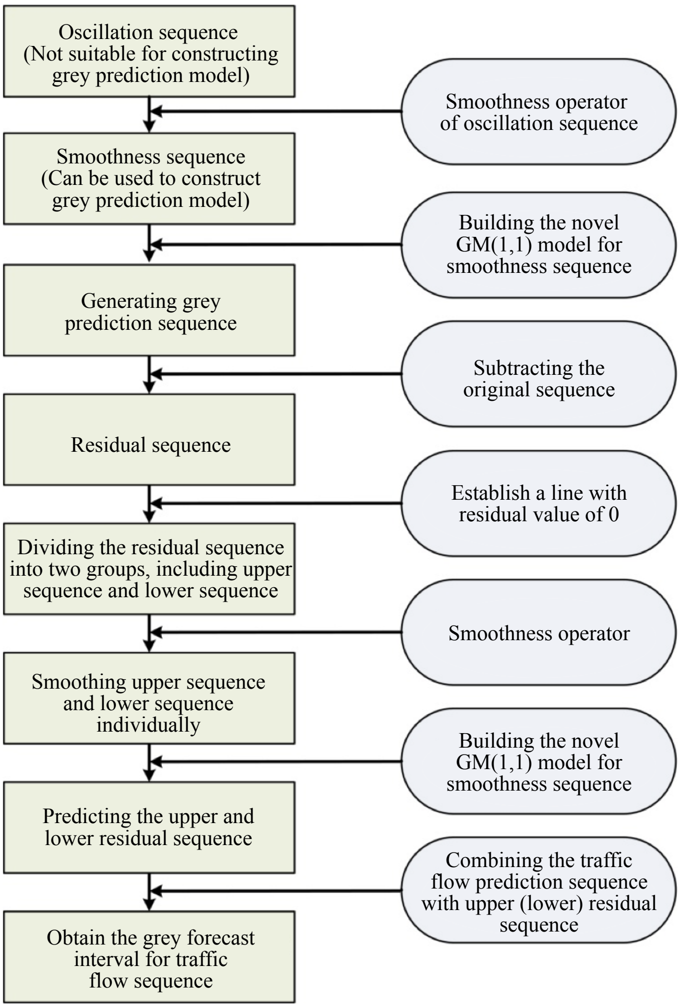


Figure 1 – Modelling procedure of interval prediction

3.1 GM(1,1) revisited

The fundamental structure of the grey model involving a first-order differential equation and a single variable is commonly denoted as GM(1,1). Assume that $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),...,x^{(0)}(n))$ denotes an original sequence and $X^{(1)}=(x^{(1)}(1),x^{(1)}(2),\ldots,x^{(1)}(n))$ is an accumulation sequence of $X^{(0)}$ by the accumulating operations. The basic first-order accumulated generating operation (1-AGO) structure is defined as in *Equation 1*.

$$x^{(1)}(t)=\sum_{i=1}^{t}x^{(0)}(i),\ \ i=1,2,\ldots,t \tag{1}$$

where t is the time index.

The original form of the GM(1,1) is defined as in *Equation 2*

$$x^{(0)}(t)+ax^{(1)}(t)=b \tag{2}$$

where a, b are the coefficients of least-squares estimation.

The mean sequence of $x^{(1)}(t)$ is defined as $z^{(1)}(t)$ for t=2,3,…,n, which can be calculated by *Equation 3*.

$$z^{(1)}(t) = \frac{x^{(1)}(t-1)+x^{(1)}(t)}{2} \tag{3}$$

The basic form of GM(1,1) is given as *Equation 4*.

$$x^{(0)}(t)+az^{(1)}(t)=b \tag{4}$$

Its parameters are estimated by using the least squares estimate method as *Equation 5*.

$$\hat{\alpha}=[a,b]^T=(A^TA)^{-1}A^T\gamma \tag{5}$$

where $\gamma=\left(x^{(0)}(2),x^{(0)}(3),\dots,x^{(0)}(n)\right)^T$; $A\begin{bmatrix} z^{(1)}(2) & 1 \\ z^{(1)}(3) & 1 \\ \vdots & \vdots \\ z^{(1)}(n) & 1 \end{bmatrix}$

The whitenisation equation of the GM(1,1) model is given as *Equation 6*.

$$\frac{dx^{(1)}}{dt}+ax^{(1)}(t)=b \tag{6}$$

Assume that $\hat{x}^{(1)}(t)$ and $\hat{x}^{(0)}(t)$ represent the accumulated forecast sequence and the forecast sequence of GM(1,1) at time t, respectively. Then, the former can be calculated by solving *Equation 7*. The restored values of $\hat{x}^{(0)}(t)$ are obtained according to *Equation 8*.

$$\hat{x}^{(1)}(t+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-at}+\frac{b}{a}, t=1,2,\dots,n \tag{7}$$

$$\hat{x}^{(0)}(t+1)=\hat{x}^{(1)}(t+1)-\hat{x}^{(1)}(t)=(1-e^{a})\left(\hat{x}^{(0)}(1)-\frac{b}{a}\right)e^{-at} \tag{8}$$

where $\hat{x}^{(0)}(t)$ is the forecast sequence; $\hat{x}^{(1)}(t)$ is the accumulated forecast sequence.

3.2 Volatility sequence and smoothness operator

The GM(1,1) model can yield satisfactory prediction accuracy when modelling a monotonic increasing (or decreasing) sequence. However, the GM(1,1) model's predictive accuracy falls short when dealing with sequences exhibiting volatility characteristics. To solve this problem, it becomes imperative to implement a smoothing algorithm to mitigate the amplitude of volatility. The volatility sequence and smoothness operator are defined as follows.

Volatility sequence. Suppose that the original sequence is $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\dots,x^{(0)}(n))$, then

a) If for $\forall t$=2,3,…,n, $x^{(0)}(t)$-$x^{(0)}(t$-1)>0 then the original sequence is named a monotonic increasing sequence;
b) If for $\forall t$=2,3,…,n, $x^{(0)}(t)$-$x^{(0)}(t$-1)<0 then the original sequence is named a monotonic decreasing sequence;
c) If for $\exists\, t, t'$=2,3…,n, $x^{(0)}(t)$-$x^{(0)}(t$-1)>0 and $x^{(0)}(t')$-$x^{(0)}(t''$-1)<0 then the original sequence is named a volatility sequence. Suppose that M=max$\{x^{(0)}(t)\}$, m=min$\{x^{(0)}(t)\}$, and the amplitude of volatility sequence can be calculated as T=M-m.

Smoothness operator. Suppose that the volatility sequence is $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\dots,x^{(0)}(n))$, then the structure of smoothness operator is as *Equation 9*, and d is named a first-order smoothness operator of X.

$$x^{(0)}(t)d=\frac{[x^{(0)}(t)+T]+[x^{(0)}(t+1)+T]}{4}, \quad t=1,2,\dots,n \tag{9}$$

Thus, the sequence $X^{(0)}D=(x^{(0)}(1)d,x^{(0)}(2)d,\dots,x^{(0)}(n-1)d)$ is called smoothness sequence of $X^{(0)}$.

Proof. Set $x^{(0)}(p)$=max$(x^{(0)}(t)\,|\,t$=1,2…,$n)$ and $x^{(0)}(q)$=min$(x^{(0)}(t)\,|\,t$=1,2…,$n)$, then $T(X)$=$x^{(0)}$(p)-$x^{(0)}(q)$. Set $x^{(0)}(i)d$=max$(x^{(0)}(t)d\,|\,t$=1,2,…,n-1) and $x^{(0)}(j)d$=max$(x^{(0)}(t)d\,|\,t$=1,2,…,n-1), then $T(XD)$=$x^{(0)}(i)d$-$x^{(0)}(j)d$.

According to the *Equation 9*, then

$$\begin{cases} x^{(0)}(i)d = \{[x^{(0)}(i)+T]+[x^{(0)}(i+1)+T]\}/4 \\ x^{(0)}(j)d = \{[x^{(0)}(j)+T]+[x^{(0)}(j+1)+T]\}4 \end{cases}$$

So, $T(XD) = \{[x^{(0)}(i) - x^{(0)}(j)] + [x^{(0)}(i+1) - x^{(0)}(j+1)]\}4.$

For

$$T(X) = x^{(0)}(q) - x^{(0)}(q) > x^{(0)}(i) - x^{(0)}(j) \Rightarrow T(X) = x^{(0)}(q) - x^{(0)}(q) > x^{(0)}(i+1) - x^{(0)}(j+1) \tag{10}$$

So, $2T(X) > 4T(XD) = [x^{(0)}(i) - x^{(0)}(j)] + [x^{(0)}(i+1) - x^{(0)}(j+1)]$

So, $T(X) > 2T(XD)$

The proof process reveals that the application of smoothing operators can effectively reduce the amplitude of volatile sequences. Consequently, enhancing the smoothness of the volatile sequence allows for the development of a more rational grey prediction model.

3.3 Proposed novel GM(1,1) model

We cannot immediately build a grey prediction model based on the volatility sequence due to the poor degree of smoothness of this volatility sequence. To solve this problem, this paper proposed a novel GM(1,1) model, which improves the GM(1,1) model by integrating two techniques, i.e. the smoothness operators of volatility sequence and background value construction.

Generating a smoothness sequence. Suppose that the volatility sequence is $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\ldots,x^{(0)}(n))$, according to *Equation 9*, the smoothness sequence of $X^{(0)}$ is $X^{(0)}D=(x^{(0)}(1)d,x^{(0)}(2)d,\ldots,x^{(0)}(n\text{-}1)d)$. Set $y^{(0)}(k)=x^{(0)}(k)d$, so the smoothness sequence converts to $Y^{(0)}=(y^{(0)}(1),y^{(0)}(2),\ldots,y^{(0)}(\mathrm{n}))$. Through the 1-AGO structure processing, the accumulation sequence of $Y^{(0)}$ is $Y^{(1)}=(y^{(1)}(1),y^{(1)}(2),\ldots,y^{(1)}(n))$.

Constructing background value of the grey model. We apply the background value construction of three-parameter to alleviate the volatility by *Equation 11*. The method extends the background value from two to three, which improves the smoothness of the grey model.

$$z^{(1)}(t) = \frac{(y^{(1)}(t) + y^{(1)}(t-1) + y^{(1)}(t-2))}{3}, \quad t = 3,4,,n \tag{11}$$

where $z^{(1)}(t)=(z^{(1)}(3),z^{(1)}(4),\ldots,z^{(1)}(t))T$ means new background value sequence.

The grey differential equation of the novel GM(1,1) model is given as *Equation 12*

$$y^{(0)}(t) + \frac{1}{3}a\left(y^{(1)}(t) + y^{(1)}(t-1) + y^{(1)}(t-2)\right) = tb + c \tag{12}$$

where a, b, c are the coefficients of least-squares estimation; its parameters are estimated by using the least-squares estimate method as *Equation 13*.

$$\hat{\beta} = [a,b,c]^T = (B^T B)^{-1} B^T \delta$$
$$\text{where } \delta = (y^{(0)}(3), y^{(0)}(4), \ldots, y^{(0)}(n))^T;$$
$$B = \begin{bmatrix} -\frac{1}{3}\left(y^{(1)}(3)+y^{(1)}(2)+y^{(1)}(1)\right) & 3 & 1 \\ -\frac{1}{3}\left(y^{(1)}(4)+y^{(1)}(3)+y^{(1)}(2)\right) & 4 & 1 \\ \vdots & \vdots & \vdots \\ -\frac{1}{3}\left(y^{(1)}(n)+y^{(1)}(n-1)+y^{(1)}(n-2)\right) & n & 1 \end{bmatrix}. \tag{13}$$

Deducing the novel GM(1,1) model. Suppose that $\hat{y}^{(1)}(t)$ and $\hat{y}^{(0)}(t)$ represent the accumulated forecast sequence and the forecast sequence of the novel GM(1,1) at time t, respectively. Then, the latter can be calculated by *Equation 14*.

$$\hat{y}^{(0)}(t) = \hat{y}^{(1)}(t) - \hat{y}^{(1)}(t-1), \quad t = 3,4,\ldots,n \tag{14}$$

where $\hat{y}^{(0)}(t)$ is the prediction sequence; $\hat{y}^{(1)}(t)$ is the 1-AGO of prediction sequence.

To obtain the $\hat{y}^{(1)}(t)$ sequence, *Equation 14* combines with *Equation 12*, then *Equation 15* as shown below.

$$tb + c - \frac{1}{3}a\left(\hat{y}^{(1)}(t) + \hat{y}^{(1)}(t-1) + \hat{y}^{(1)}(t-2)\right) = y^{(1)}(t) - y^{(1)}(t-1) \tag{15}$$

Then, the values of $\hat{y}^{(1)}(t)$ are further obtained based on the formulation in *Equation 16*. $y^{(0)}(1)$ and $y^{(0)}(2)$ are called the initial value of novel grey prediction model.

$$\begin{cases} \hat{y}^{(1)}(t) = \frac{3-a}{3+a}\hat{y}^{(1)}(t-1) - \frac{a}{3+a}\hat{y}^{(1)}(t-2) + \frac{3b}{3+a}t + \frac{3c}{3+a} \\ \hat{y}^{(1)}(1) = \hat{y}^{(0)}(1) \\ \hat{y}^{(1)}(2) = \hat{y}^{(0)}(2) \end{cases} \tag{16}$$

3.4 Forecasting interval

The purpose of this section is to obtain forecast interval based on the point prediction results, which can be obtained by using the novel GM(1,1) model as mentioned previously. The initial step involves generating a residual sequence by computing the difference between the predicted sequence and the original one. Subsequently, the residual sequence is divided into two groups and the proposed model is used for the prediction of the two groups of residual sequences individually. Finally, upper and lower forecast boundaries are established for both groups of residual sequences, forming the basis for constructing the grey residual prediction interval. The modelling procedure is outlined as follows.

Step 1: Generating residual prediction sequences.

Suppose that the original sequence is $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\ldots,x^{(0)}(n))$, and the prediction sequence of $X^{(0)}$ by using the novel GM(1,1) model is $\hat{X}^{(0)}=(\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\ldots,\hat{x}^{(0)}(n-1))$. So, the residual prediction sequence is calculated as *Equation 17*.

$$R^{(0)} = \hat{X}^{(0)} - X^{(0)} \tag{17}$$

Step 2: Dividing the residual sequence into two groups.

We use the line of $R^{(0)}=0$ as the dividing line to divide the residual sequence into two groups just as shown in *Figure 2*.

a) If $r^{(0)}(t)>0$, then the residual sequence is named upper residual sequence, and expresses as $R_U^{(0)}=(r_U^{(0)}(1),r_U^{(0)}(2),\ldots,r_U^{(0)}(n-1))$;
b) If $r^{(0)}(t)<0$, then the residual sequence is named lower residual sequence, and expresses as $R_L^{(0)}=(r_L^{(0)}(1),r_L^{(0)}(2),\ldots,r_L^{(0)}(n-1))$;

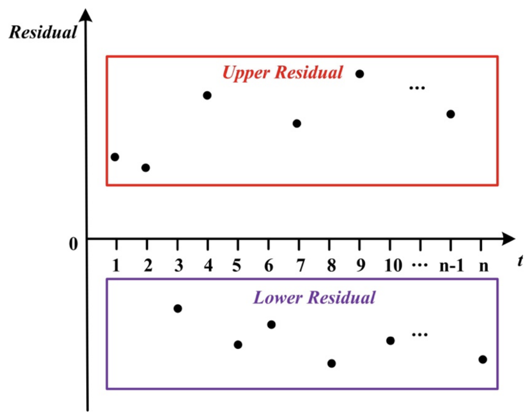


Figure 2 – Classification demonstration

Step 3: Forecasting the two groups of residual sequences.

The two groups of residual sequences are used for the novel GM(1,1) modelling and prediction individually. The forecast values of the two groups of residual sequences are obtained by the same procedure as in the above section 3.3. The forecast values are called $R_U^{(0)}$ and $R_L^{(0)}$. Therefore, the prediction interval of the residual is $[\hat{R}_U^{(0)}(t) \quad \hat{R}_L^{(0)}(t)]$.

Step 4: Determining the forecast interval of the novel GM(1,1) model.

The forecast interval of the proposed model is created as *Equation 18*

$$I=\left\{\left(t,\hat{X}^{(0)}(t)\right)\middle|\,\hat{X}^{(0)}(t)\in\left[\left(\hat{X}^{(0)}(t)-\hat{R}_L^{(0)}(t)\right),\left(\hat{X}^{(0)}(t)+\hat{R}_U^{(0)}(t)\right)\right]\right\} \tag{18}$$

where $X^{(0)}$ is the original sequence $\hat{X}^{(0)}$ is the prediction sequence by using the novel GM(1,1); $R^{(0)}$ is the residual sequence; $\hat{R}_L^{(0)}(t)$ is the lower prediction residual sequence; $\hat{R}_U^{(0)}(t)$ is the upper prediction residual sequence; I is the prediction interval of the novel GM(1,1) model.

4. EMPIRICAL STUDY

4.1 Study site and data collection

Real traffic flow data in this study were collected on the road network of Furong District in Changsha (*Figure 3*). There are eleven roads selected on the studied network. The road network has installed the loop detector on each road cross-section. These detectors collect traffic flow, every 5 min and output this data via the Traffic Reporter of SCATS for research. *Figure 3* shows the specific location of eleven road segments. Complete traffic flow data sets from the loop detectors were available for 5 consecutive days (2013.09.23-2013.9.27) just as shown in *Table 1*. To assess the model's accuracy, we utilised traffic flow data from 27 September 2013. The analysis of traffic flow characteristics involved dividing the study period into two distinct segments: the morning peak (7-10 a.m.) and the afternoon peak (5-8 p.m.).

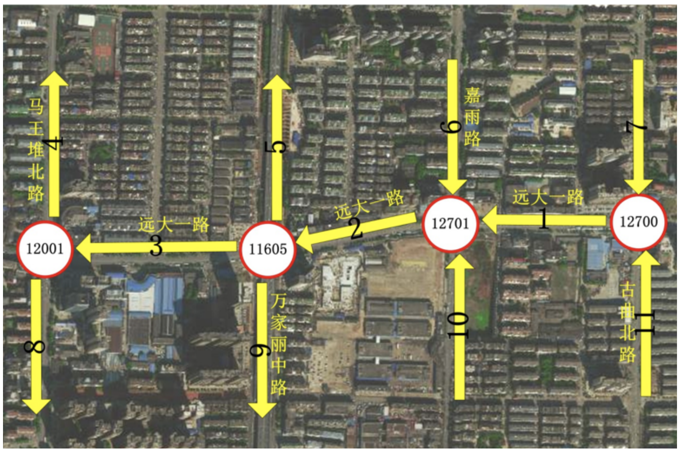


Figure 3 – Study site

Table 1 – Data overview

ID	Road name	Flow direction	Start	End	AM (Time)	PM (Time)
1	Yuanda One Road	East → West	23/9/2013	27/9/2013	7-11	5-9
2	Yuanda One Road	East → West	23/9/2013	27/9/2013	7-11	5-9
3	Yuanda One Road	East → West	23/9/2013	27/9/2013	7-11	5-9
4	Mawangdui North Road	South → North	23/9/2013	27/9/2013	7-11	5-9
5	Wanjiali Middle Road	South → North	23/9/2013	27/9/2013	7-11	5-9
6	Jiayu Road	North → South	23/9/2013	27/9/2013	7-11	5-9
7	Guqu North Road	North → South	23/9/2013	27/9/2013	7-11	5-9
8	Mawangdui North Road	North → South	23/9/2013	27/9/2013	7-11	5-9
9	Wanjiali Middle Road	North → South	23/9/2013	27/9/2013	7-11	5-9
10	Jiayu Road	South → North	23/9/2013	27/9/2013	7-11	5-9
11	Guqu North Road	South → North	23/9/2013	27/9/2013	7-11	5-9

Note: See http://www.openits.cn/openPaper/567.jhtml for more information

4.2 Experimental design

To evaluate the performance of the novel GM(1,1) model, we compared it with the traditional GM(1,1) model. For each model, two measures are evaluated for uncertainty quantification, which includes the kick-off percentage (*KP*) and the width interval (*WI*). The *KP* is computed as the ratio of the total number of original true values outside the prediction intervals to the total number of original true values. Meanwhile, the *WI* measures the width of the prediction interval. Ideally, the *KP* and *WI* are expected to be small. The relevant formulas of the two measures are shown as follows:

$$KP = \frac{KN}{N} \tag{19}$$

$$WI = U - L \tag{20}$$

where *KN* is the number of original true values lying outside the predicted interval; *N* is the total number of original true values; *U* is the upper prediction value; *L* is the lower prediction value; *R* is the real measure value.

4.3 Model performance comparison

A traditional GM model is chosen to compare the performance of interval prediction with the proposed novel GM model. *Figure 4* displays the observed flow alongside the interval prediction flow generated by various models for the eleven segments during the morning peak hours. Additionally, *Figure 5* showcases the observed flow and the interval prediction flow for these segments during the afternoon peak hours. These visual representations highlight that the interval prediction flow produced by the novel GM model closely aligns with the field-measured flow, in contrast to the traditional GM model. This observation suggests that the novel GM model excels at capturing the fluctuation patterns present in the field-measured flow. Moreover, most of the filed-measured flow falls within the range determined by the novel GM model.

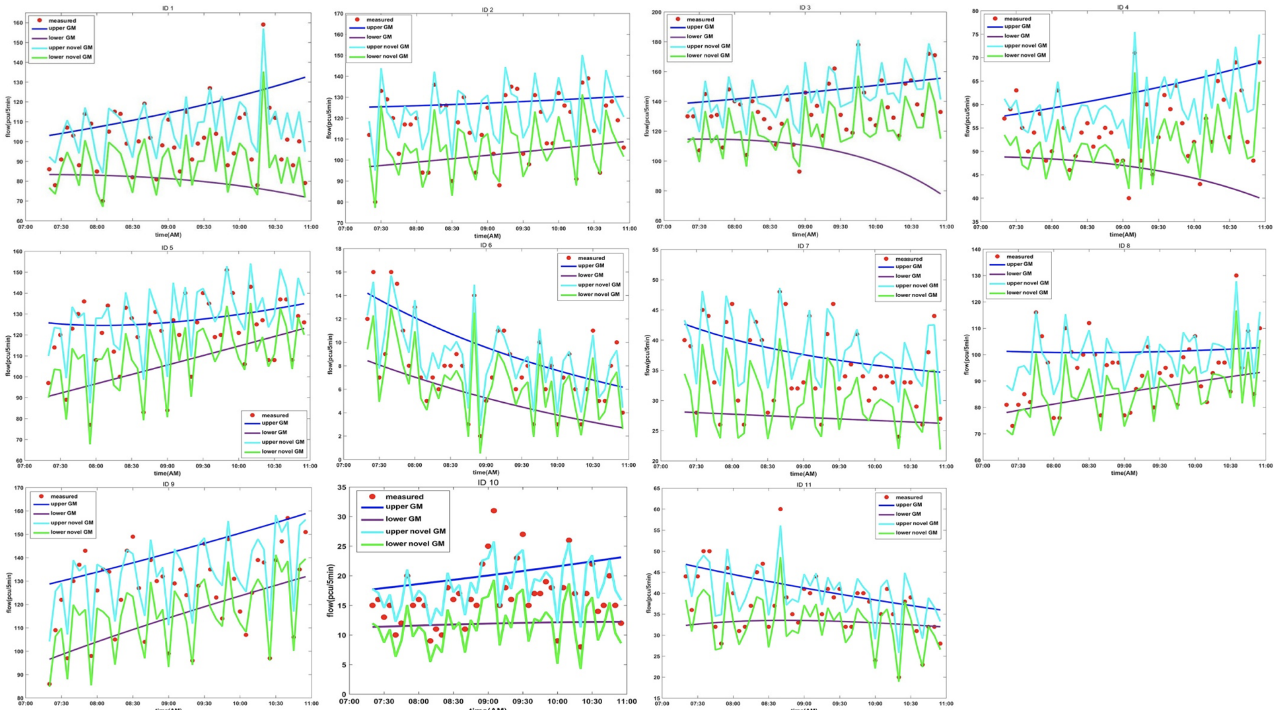


Figure 4 – Flow interval prediction by using novel GM model and GM model for AM, 27 September 2013

For a more quantitative assessment of model prediction accuracy, we present the model performance metrics in *Table 2*. According to the results of performance measures, the novel GM model performs better with the lowest *KP*, *WI* and *WI* amplitude. For example, the average *KP*, *WI*, max(*WI*), min(*WI*) and *WI* amplitude are approximately 0.44, 20, 31, 13 and 18, respectively for all segments by using a novel GM model under morning peak hours. By comparison, the segment forecasts have a mean *KP*, *WI*, max(*WI*), min(*WI*)

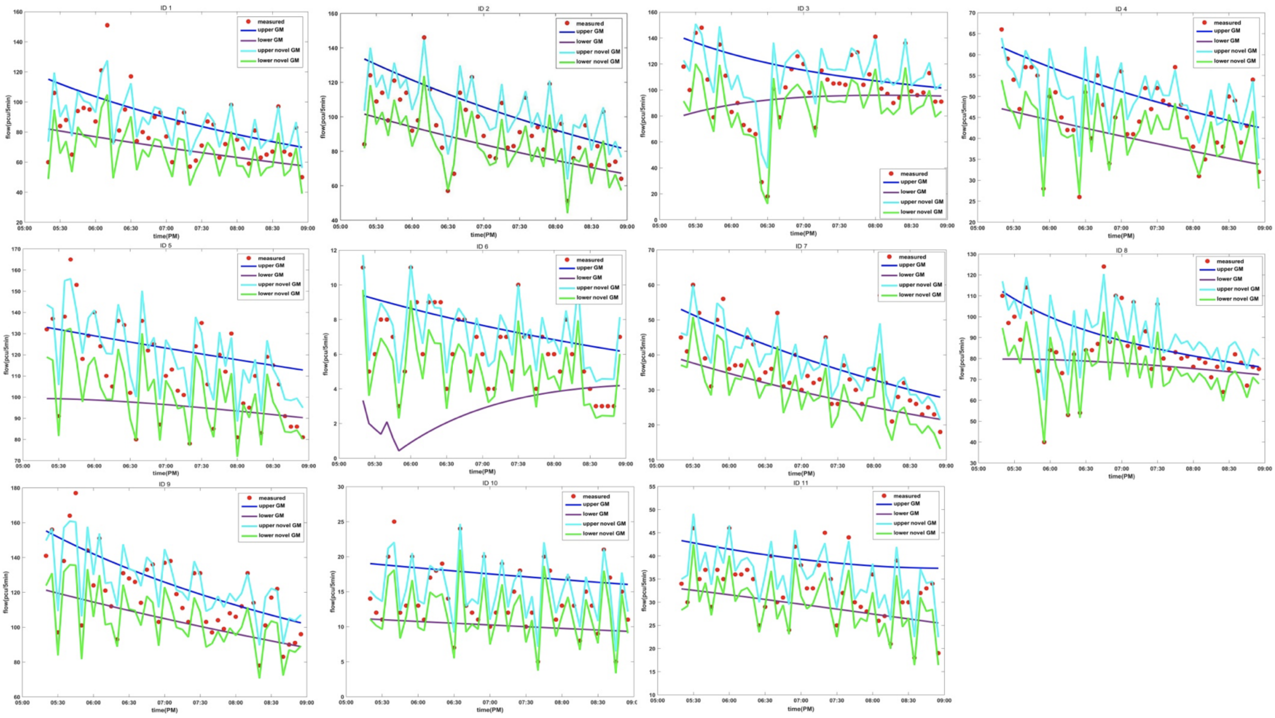


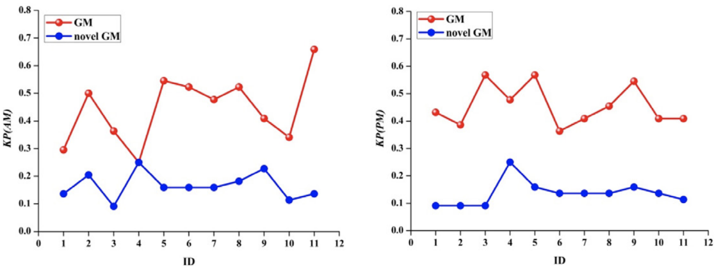
Figure 5 – Flow interval prediction by using novel GM model and GM model for PM, 27 September 2013

Table 2 – Comparison of WI at different peak hours

Time	ID	GM(1,1)					Novel GM(1,1)				
		KP	mean WI	max WI	min WI	max-min WI	KP	mean WI	max WI	min WI	max-min WI
AM	1	0.30	36	60	19	41	0.14	19	23	16	7
	2	0.50	25	29	22	7	0.20	18	19	18	1
	3	0.36	41	78	24	54	0.09	23	26	20	6
	4	0.25	16	29	8	21	0.25	9	10	8	2
	5	0.55	22	37	12	25	0.16	19	20	19	1
	6	0.52	4	6	3	3	0.16	2	3	2	1
	7	0.48	11	15	8	7	0.16	8	9	8	1
	8	0.52	16	24	9	15	0.18	14	17	11	6
	9	0.41	28	33	27	6	0.23	18	19	17	2
	10	0.34	8	11	6	5	0.11	7	7	6	1
	11	0.66	8	16	4	12	0.14	7	8	7	1
	Mean	0.44	20	31	13	18	0.17	13	15	12	3
PM	1	0.43	21	35	12	23	0.09	21	25	17	8
	2	0.39	22	33	15	18	0.09	20	22	19	3
	3	0.57	26	66	6	60	0.09	26	31	21	10
	4	0.48	11	15	9	6	0.25	8	10	7	3
	5	0.57	28	54	23	31	0.16	19	25	15	10
	6	0.36	5	8	2	6	0.14	2	2	2	0
	7	0.41	10	15	6	9	0.14	9	9	8	1
	8	0.45	14	36	4	32	0.14	17	22	14	8
	9	0.55	22	36	14	22	0.16	22	26	18	8
	10	0.41	7	8	7	1	0.14	4	4	3	1
	11	0.41	10	12	10	2	0.11	6	7	6	1
	Mean	0.46	16	29	10	19	0.14	14	17	12	5

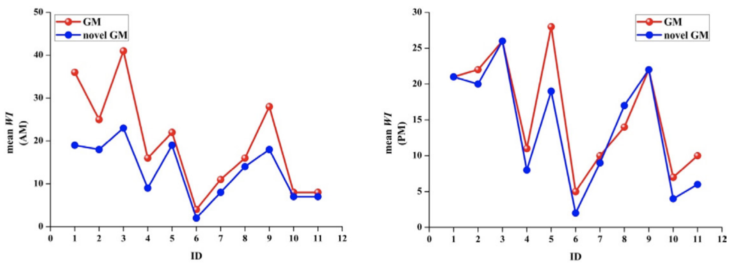
and *WI* amplitude of approximately 0.17, 13, 15, 12 and 3 for the GM model. During both the morning and afternoon peak hours, the novel GM model demonstrates a substantial improvement in the average prediction interval coverage, with increases of 27% and 32% compared to the GM model. This outcome underscores the significance of considering volatility characteristics within the traffic flow sequence, achieved through the application of smooth pre-processing and background value construction.

Figure 6 shows the *KP* value of the two methods for each segment under different peak hours, the volatility amplitude of *KP* by using the GM model is stronger than that by using novel GM. The results indicate that traffic flow interval prediction is more accurate by using a novel GM model for each segment. In addition, we compare the average *WI* and *WI* amplitude of the two models as shown in *Figure 7* and *Figure 8*, the similar results can be also found for the two models across the volatility amplitude of measures during morning and afternoon peak hours. This aligns with the notion that the smoothness operator applied to the volatile sequence effectively diminishes traffic flow fluctuations, resulting in greater stability and predictability within the flow series.



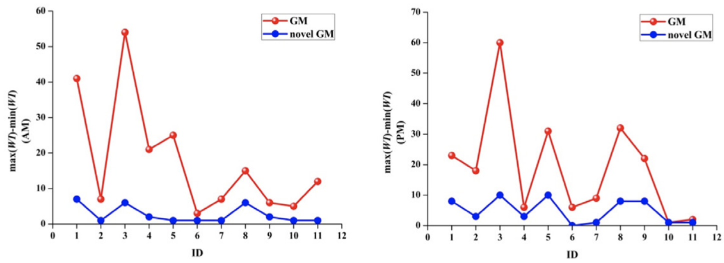
a) AM *b) PM*

Figure 6 – Comparison of KP values using the two models in different peak hours



a) AM *b) PM*

Figure 7 – Comparison of mean WI values using the two models in different peak hours



a) AM *b) PM*

Figure 8 – Comparison of WI amplitude using the two models in different peak hours

5. CONCLUSION

We proposed a novel GM(1,1) model which extends the GM(1,1) by integrating two techniques (smooth pre-processing and background value construction) to forecast the uncertainty quantification of short-term traffic flow. Smooth pre-processing uses the smoothness operator, which can compress the volatility amplitude of traffic flow. Background value construction applies the three-parameter construction method, which extends the background value from two to three, to further alleviate the volatility in traffic flow data. Moreover, the upper sequence and lower sequence of forecast interval are determined by the line of R^((0))=0. In order to evaluate the performance of the proposed novel GM model, this study used the real traffic flow data of the Furong District in Changsha.

The interval prediction results verified that the novel GM model outperforms the original GM model, as evidenced by the volatility trend of upper sequence (lower sequence) prediction. The novel GM model showed high prediction accuracy by compressing the volatility amplitude of traffic flow. We further compare the proposed model with the original GM model by calculating two performance measures: *KP* and *WI*. The performance outcomes demonstrate the superiority of the novel GM model, as evidenced by its lower *KP* and *WI* values. These results affirm that the novel GM model excels at capturing the inherent variations within field-measured traffic flow data, primarily due to the effective implementation of smooth pre-processing and background value construction.

Additionally: (i) future research should aim at the development of more background value construction methods to improve the smoothness of the traffic data; (ii) the proposed model is a parametric model with a fixed structure: the predicted results can only be obtained off-line, and do not provide an on-line forecast, so future research should focus on new adaptation mechanisms via which the proposed model could yield real-time predictions; (iii) the kick-off percentage (*KP*) and the width interval (*WI*) can gauge the length of the prediction interval and coverage; however, the two measures are not uniform performance measures. Thus, future research should further investigate uniform performance measures with which to evaluate the prediction intervals.

ACKNOWLEDGEMENTS

The authors acknowledge support from the Hefei University of Technology. We also thank the Bengbu Public Security Bureau for its support in algorithm testing work. This work was funded by the China Postdoctoral Science Foundation (Grant No. 2022M720980), the National Natural Science Foundation of China (Grant No. 72371094), Anhui Provincial Key Research and Development Project (Grant No. 2022k07020007) and Young Teacher Research Innovation Special Initiation Project of Hefei University of Technology (Grant No. JZ2022HGQB0210). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

[1] Shaygan M, et al. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation Research Part C: Emerging Technologies.* 2022;145:103921. DOI: 10.1016/J.TRC.2022.103921.

[2] Papageorgiou M, et al. Review of road traffic control strategies. *Proceedings of the IEEE.* 2004;91(12):2041–2042. DOI: 10.1109/JPROC.2003.819606.

[3] Medina-Salgado B, et al. Urban traffic flow prediction techniques: *A review. Sustainable Computing: Informatics and Systems.* 2022;35:100739. DOI: 10.1016/j.suscom.2022.100739.

[4] Deng JL. Control problems of grey systems. *Systems & Control Letters*. 1982;1(5):288–294. DOI: 10.1016/S0167-6911(82)80025-X.

[5] Chen CI, Hsin PH, Wu CS. Forecasting Taiwan's major stock indices by the Nash nonlinear grey Bernoulli model. *Expert Systems with Applications*. 2010;37(12):7557–7562. DOI: 10.1016/j.eswa.2010.04.088.

[6] Wang Y, et al. A novel structure adaptive fractional derivative grey model and its application in energy consumption prediction. *Energy.* 2023;282:128380. DOI: 10.1016/J.ENERGY.2023.128380.

[7] Wu W, et al. Application of the novel fractional grey model FAGMO(1,1,k) to predict China's nuclear energy consumption. *Energy.* 2018;165:223-234. DOI: 10.1016/J.ENERGY.2018.09.155.

[8] Duan H, Liu Y, Wang G. A novel dynamic time-delay grey model of energy prices and its application in crude oil price forecasting. *Energy.* 2022;251:123968. DOI: 10.1016/J.ENERGY.2022.123968.

[9] Duan H, Wang G. Partial differential grey model based on control matrix and its application in short-term traffic flow prediction. *Applied Mathematical Modelling.* 2023;116:763–785. DOI: 10.1016/J.APM.2022.12.012.

[10] Lippi M, Bertini M, Frasconi P. Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems.* 2013;14(2):871–82. DOI: 10.1109/TITS.2013.2247040.

[11] Kumar SV, Vanajakshi L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*. 2015;7(3):1–9. DOI: 10.1007/s12544-015-0170-8.

[12] Alghamdi T, et al. Forecasting traffic congestion using ARIMA modeling. *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC).* 2019. p. 1227–1232. DOI: 10.1109/IWCMC.2019.8766698.

[13] Vázquez JJ, et al. A comparison of deep learning methods for urban traffic forecasting using floating car data. *Transportation Research Procedia*. 2020;47:195–202. DOI: 10.1016/J.TRPRO.2020.03.079.

[14] Rajalakshmi V, Ganesh Vaidyanathan S. Hybrid time-series forecasting models for traffic flow prediction. *Promet – Traffic&Transportation*. 2022;34(4):537–549. DOI: 10.7307/ptt.v34i4.3998.

[15] Lee K, Rhee W. DDP-GCN: Multi-graph convolutional network for spatiotemporal traffic forecasting. *Transportation Research Part C: Emerging Technologies*. 2022;145:103466. DOI: 10.1016/J.TRC.2021.103466.

[16] Ashley DJ. Uncertainty in the context of highway appraisal. *Transportation.* 1980;9(3):249–267. DOI: 10.1007/bf00153867.

[17] Brundell K. Sampling, specification and estimation as sources of inaccuracy in complex transport models-some examples analysed by Monte Carlo simulation and bootstrap, *Proceedings of seminar F of the European Transport Conference 2000, 11–13 Sept.2000, Cambridge, UK*. 2000. p. 225–237. http://worldcat.org/isbn/086050333X.

[18] Hugosson MB. Quantifying uncertainties in a national forecasting model. *Transportation Research Part A: Policy and Practice.* 2005;39(6):531–547. DOI: 10.1016/J.TRA.2005.02.010.

[19] Klir GJ. *Uncertainty and information: Foundations of generalized information theory*. New York, USA: John Wiley& Sons; 2006.

[20] Zhao Y, Kockelman KM. The propagation of uncertainty through travel demand models: An exploratory analysis. *Annals of Regional Science.* 2002;361:145–163. DOI: 10.1007/s001680200072.

[21] Rodier CJ, Johnston RA. Uncertain socioeconomic projections used in travel demand and emissions models: Could plausible errors result in air quality nonconformity? *Transportation Research Part A: Policy and Practice.* 2002;36(7):613–631. DOI: 10.1016/S0965-8564(01)00026-X.

[22] Boyce, AM, Bright, MJ. Reducing or managing the forecasting risk in privately-financed projects. *Proceedings of the European Transport Conference 2003, 8–10 Oct.2003, Strasbourg, FRANCE*. 2003. p. 19. http://worldcat.org/isbn/0860503429.

[23] Armoogum J. Madre JL, Bussiere Y. Measuring the impact of uncertainty in travel demand modelling with a demographic approach. *IATSS Research*. 2009;33(2):9–20. DOI: 10.1016/S0386-1112(14)60241-7.

[24] Yang M, Liu Y, You Z. The reliability of travel time forecasting. *IEEE Transactions on Intelligent Transportation Systems.* 2010;11(1):162–171. DOI: 10.1109/TITS.2009.2037136.

[25] Liu S, Lin Y. *Grey information: Theory and practical applications*. London, UK: Springer London; 2006. DOI: 10.1007/1-84628-342-6.

[26] Huang YL, Lin CT. Developing an interval forecasting method to predict undulated demand. *Quality & Quantity*. 2011;45:513–524. DOI: 10.1007/s11135-010-9317-9.

[27] Chen YY, Liu HT, Hsieh HL. Time series interval forecast using GM(1,1) and NGBM(1, 1) models. *Soft Computing.* 2019;23(5):1541–1555. DOI: 10.1007/s00500-017-2876-0.

曹旭东，石琴，陈一锴，陈晨辰

基于新型GM(1,1)模型的短期交通流不确定性预测

摘要：

在智能交通系统中，预测短期交通流的不确定性对于有效的交通管理至关重要。各种预测不确定性的方法已经提出并实施。然而，在面对稀疏数据时，传统方法往往

难以提供准确的预测。因此，本研究侧重于在有限数据条件下开发短期交通流不确定性预测模型，提出了一种考虑交通数据波动性的新型灰色模型，该模型通过整合两种技术：平滑预处理和背景值构建，对灰色模型（GM）进行了扩展。所提出的新型灰色模型的性能主要通过与传统GM模型的比较来说明。我们的结果在不确定性量化方面表明，所提出的模型在平均偏离百分比（KP）、宽度区间（WI）和宽度幅度方面优于GM模型。

关键词：
智能交通系统；非确定性量化；新型GM模型；平滑预处理；背景值构建