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 envelope prediction model (GEMP) [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 GEMP 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.

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 $X^{(1)}=(x^{(1)}(1), x^{(1)}(2),..., x^{(1)}(n))$ 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.

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

where $t$ is the time index.

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

$$x^{(0)}(t) + ax^{(1)}(t) = b$$ (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 \( t=2,3,\ldots,n \), which can be calculated by Equation 3.

\[
z^{(1)}(t) = \frac{x^{(1)}(t-1) + x^{(1)}(t)}{2}
\]

(3)

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

\[
x^{(0)}(t) + az^{(1)}(t) = b
\]

(4)

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

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

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

(5)

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

\[
\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b
\]

(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.

\[
\hat{x}^{(1)}(t+1) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}t, \quad t = 1,2,\ldots,n
\]

(7)

\[
\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) = (1 - e^{-a})(\hat{x}^{(0)}(1) - \frac{b}{a})e^{-at}
\]

(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 \( X^{(0)} = (x^{(0)}(1),x^{(0)}(2),\ldots,x^{(0)}(n)) \), then

a) If for \( \forall t=2,3,\ldots,n, x^{(0)}(t)-x^{(0)}(t-1)>0 \) then the original sequence is named a monotonic increasing sequence;

b) If for \( \forall t=2,3,\ldots,n, x^{(0)}(t)-x^{(0)}(t-1)<0 \) then the original sequence is named a monotonic decreasing sequence;

c) If for \( \exists t, i=2,3,\ldots,n, x^{(0)}(t)-x^{(0)}(t-1)>0 \) and \( x^{(0)}(i)-x^{(0)}(i-1)<0 \) 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 \( X^{(0)} = (x^{(0)}(1),x^{(0)}(2),\ldots,x^{(0)}(n)) \), then the structure of smoothness operator is as Equation 9, and \( d \) is named a first-order smoothness operator of \( X \).

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

(9)

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

**Proof.** Set \( x^{(0)}(p) = \max \{x^{(0)}(t) \mid t=1,2,\ldots,n \} \) and \( x^{(0)}(q) = \min \{x^{(0)}(t) \mid t=1,2,\ldots,n \} \), then \( T(X) = x^{(0)}(p)-x^{(0)}(q) \).

Set \( x^{(0)}(i)d = \max \{x^{(0)}(t)d \mid t=1,2,\ldots,n-1 \} \) and \( x^{(0)}(j)d = \max \{x^{(0)}(t)d \mid t=1,2,\ldots,n-1 \} \), then \( T(XD) = x^{(0)}(i)d-x^{(0)}(j)d \).

According to the Equation 9, then
We apply the background value construction of $f$, the smoothness sequence of $12() () ()01 11++$ where calculated by sequence and the forecast sequence of the novel GM(1,1) at time squares estimate method as

where $z$ means new background value sequence.

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

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.

\[
z^{(1)}(t) = \frac{(y^{(1)}(t)+y^{(1)}(t-1)+y^{(1)}(t-2))}{3}, \quad t = 3,4,\ldots,n
\]

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

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

\[
y^{(0)}(t)+\frac{1}{3}a(y^{(1)}(t)+y^{(1)}(t-1)+y^{(1)}(t-2)) = tb + c
\]

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.

\[
\hat{\beta} = [a,b,c]^T = (B^TB)^{-1}B^T\delta
\]

where $\delta = (y^{(0)}(3),y^{(0)}(4),...,y^{(0)}(n))^T$;

\[
B = \begin{bmatrix}
-\frac{1}{3}(y^{(1)}(3)+y^{(1)}(2)+y^{(1)}(1)) & 3 & 1 \\
-\frac{1}{3}(y^{(1)}(4)+y^{(1)}(3)+y^{(1)}(2)) & 4 & 1 \\
\vdots & \vdots & \vdots \\
-\frac{1}{3}(y^{(1)}(n)+y^{(1)}(n-1)+y^{(1)}(n-2)) & n & 1
\end{bmatrix}
\]

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.

\[
\hat{y}^{(0)}(t) = \hat{y}^{(1)}(t)-\hat{y}^{(1)}(t-1), \quad t = 3,4,\ldots,n
\]

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.
\[tb + c - \frac{1}{3}a(t - 1)y(t) = y(t) - y(t - 1)\]  

Equation 15

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

\[
\begin{align*}
\hat{y}(t) &= \frac{3 - a}{3 + a} y(t - 1) - \frac{a}{3 + a} y(t - 2) + \frac{3b}{3 + a} t + \frac{3c}{3 + a}
\end{align*}
\]

Equation 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.

\[R(0) = \hat{X}(0) - X(0)\]  

Equation 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](image)

**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 \([R_U(0)(t) \quad R_L(0)(t)]\).
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

\[ I = \{ (t, \hat{X}^{(0)}(t)) | \hat{X}^{(0)}(t) \in [\hat{X}^{(0)}(t) - \hat{R}_L^{(0)}(t), \hat{X}^{(0)}(t) + \hat{R}_U^{(0)}(t)] \} \]  

(18)

where \( \hat{X}^{(0)} \) is the original sequence; \( \hat{X}^{(0)} \) is the prediction sequence by using the novel GM(1,1); \( \hat{R}^{(0)} \) 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](image)

**Table 1 – Data overview**

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

*Note: See [http://www.openits.cn/openPaper/567_jhtml](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:

\[ KP = \frac{KN}{N} \]  
\[ WI = U - L \]

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 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

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

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 \(R^0((0))=0\). 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


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