World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3056

RELATIONSHIP BETWEEN TRAFFIC FLOW AND TIME BASED ON REGRESSION MODELS

TianYi Chen[1*], Cheng Cui[2]
[1]*School of Economics and Management, Nanjing Forestry University, Nanjing 210037, Jiangsu, China.*
[2]*School of Science, Nanjing Forestry University, Nanjing 210037, Jiangsu, China.*
Corresponding Author: TianYi Chen, Email: 13770185785@163.com

Abstract: Accurate estimation of traffic flow is crucial for urban traffic management and control, particularly when only main road monitoring data is available and feeder road data is missing. This study addresses the challenge of inferring feeder road traffic flow from main road data by developing a series of regression models tailored to different road structures and traffic conditions. For a Y-shaped basic road network, both linear and piecewise linear regression models were established, achieving perfect fitting of the main road traffic flow. In multi-branch scenarios that account for delays and cyclical fluctuations, an integrated model comprising constant, piecewise linear, and periodic functions was proposed, achieving a goodness of fit of 0.9722. Under traffic signal control conditions, a composite model including piecewise functions and periodic components was developed, effectively addressing traffic interruptions caused by signals, with a goodness of fit of 0.9642. In noisy data environments, a robust regression framework with adaptive weighting was introduced, maintaining high accuracy despite noise interference. The results indicate that the proposed models can effectively reconstruct feeder road traffic patterns, offering excellent interpretability and robustness. This provides a reliable data foundation for signal timing optimization and congestion management, offering practical solutions for traffic flow estimation in certain observed road networks.
Keywords: Linear regression model; Least squares method; Iterative optimization; Nonlinear regression model

1 INTRODUCTION

Since the initiation of the reform and opening-up, China's urbanization process has continuously advanced, with the congestion pressure on traffic systems becoming increasingly prominent. Real-time and accurate acquisition of road traffic flow information has become key to improving traffic operational efficiency and optimizing signal control. Currently, monitoring equipment has been widely deployed on main roads, but many feeder roads are limited by insufficient equipment coverage, lacking real-time traffic data, which hinders a comprehensive understanding of network operations. Therefore, inferring feeder road traffic flow from main road monitoring data has become a critical issue in intelligent transportation systems.
Traffic flow forecasting methods primarily include traditional statistical regression and artificial intelligence-based prediction approaches. In terms of traditional statistical methods, time series models, such as ARIMA, are widely used in traffic flow prediction due to their ability to effectively capture temporal dependencies[1]. Linear regression methods, due to their simplicity and interpretability, have always maintained an important position in traffic flow modeling. Ceder[2], through establishing statistical relationships between hourly traffic flow and accident rates, revealed fundamental traffic flow patterns. However, traditional methods are limited in their ability to capture complex nonlinear relationships and struggle to adapt to the dynamic changes in actual traffic flow.
With the development of machine learning techniques, significant progress has been made in traffic flow forecasting research. Support vector machines (SVM) address nonlinear issues through kernel functions and have demonstrated excellent performance in short-term traffic flow forecasting[3]. Ensemble learning methods, such as random forests, effectively handle high-dimensional data, enhancing prediction accuracy[4]. In recent years, deep learning models such as Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) have shown strong capabilities in time-series data modeling[5]. Researchers compared the prediction effectiveness of autoregressive models and neural network models, finding that neural networks performed better in complex scenarios[6]. However, these data-driven methods typically require large amounts of high-quality training data and still face challenges in terms of robustness against noise and missing values[7-8].
In real-world road networks, traffic flow variations often exhibit complex features such as segmentation, periodicity, and nonlinearity[9]. Researchers revealed the multimodal characteristics of traffic flow by analyzing real-time traffic monitoring data and accident relationships[10]. They established a multi-lane traffic flow model that accounts for the effects of queuing. This paper, focusing on typical road structures, combines least squares fitting, piecewise regression, and periodic functions to construct a composite model, incorporating robust optimization strategies to achieve high-precision inversion of feeder road traffic flow. The findings of this study provide valuable data support for signal timing, congestion alleviation, and road network planning, offering both theoretical value and practical significance.

2 TRAFFIC FLOW PREDICTION FOR MAIN ROADS IN A Y-SHAPED ROAD NETWORK

2.1 Restatement of the Problem and Modeling Background

In road network monitoring, main roads are often equipped with traffic detection devices, while some feeder roads lack real-time monitoring capabilities due to cost or layout constraints. Considering a typical Y-shaped road structure (as shown in Figure 1), where the traffic from Feeder Road 1 and Feeder Road 2 converges onto Main Road 3, the device A1 on Main Road 3 records traffic flow every 2 minutes. It is known that the traffic on Feeder Road 1 follows a linear growth trend, while the traffic on Feeder Road 2 initially increases linearly before decreasing linearly. The challenge is to infer the traffic flow function relationships of the two feeder roads based on the monitoring data from Main Road 3.

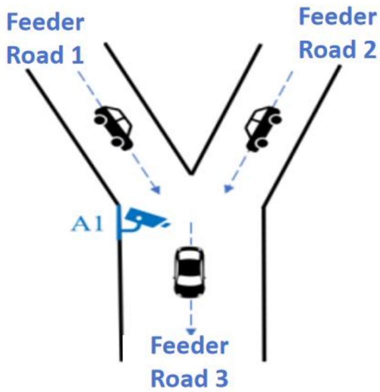

Figure 1 Y-shaped Basic Road Network Structure

2.2 Model Development

Let the time variable (t) be defined with 7:00 as the starting point (t = 0), and its range is [0,59] (corresponding to the time period from 6:58 to 8:58). The traffic flow on the main road is the sum of the traffic flows from Feeder Road 1 and Feeder Road 2:

$$Q_{main}(t)=Q_1(t)+Q_2(t) \tag{1}$$

Where the traffic flow on Feeder Road 1 is a linear function:

$$Q_1(t)=a_1t+b_1 \tag{2}$$

The traffic flow on Feeder Road 2 is a piecewise linear function:

$$Q_2(t)=\begin{cases} a_2t+b_2 & 0\le t\le t_c \\ -a_3t+b_3 & t_c<t\le 59 \end{cases}$$

The following non-negativity constraints must be satisfied: $Q_1(t)\ge 0, Q_2(t)\ge 0$, and $a_1>0, a_2>0, a_3>0$.

2.3 Parameter Estimation and Solution

The model parameters are estimated using the least squares method, with the objective function defined as:

$$\min_{a_1,b_1,a_2,b_2,a_3,b_3,t_c} \sum_{t=0}^{59} [Q_{main}(t)-(Q_1(t)+Q_2(t))]^2 \tag{3}$$

The parameter values are obtained through optimization as:

$$a_1=0.5, b_1=7, a_2=a_3=1, t_c=30 \tag{4}$$

2.4 Results and Analysis

The model fitting results indicate that the error metrics show perfect alignment between the model and the observed data, demonstrating extremely high fitting accuracy. From a practical traffic perspective, the traffic flow on Feeder Road 1 increases linearly at a rate of one vehicle every two minutes, with an initial flow of 7 vehicles. The traffic flow on Feeder Road 2 peaks at 30 vehicles at 8:00 (t=30), after which it decreases symmetrically. This result confirms the effectiveness of both the linear and piecewise linear models in inferring traffic flow for simple road network structures. Not only does it replicate the temporal variation of feeder road traffic, but it also provides reliable data support for traffic signal timing optimization and congestion management, highlighting the suitability of the least squares regression method for such problems.

3 TRAFFIC FLOW PREDICTION FOR MAIN ROADS WITH MULTIPLE FEEDER ROAD TYPES

3.1 Restatement of the Problem and Modeling Background

In the case of a more complex multi-feeder road convergence scenario, this study investigates the traffic flow composition of Main Road 5, which is formed by the convergence of four feeder roads, each with distinct traffic flow patterns. Specifically, Feeder Roads 1 and 2 exhibit a 2-minute transmission delay when converging onto Main Road 5, while Feeder Roads 3 and 4 experience no delay. Based on the characteristics of the monitoring data, the traffic flow on Feeder Road 1 remains constant, Feeder Road 2 exhibits linear growth followed by a stable state in different periods, Feeder Road 3 shows linear growth followed by stabilization, and Feeder Road 4 displays noticeable periodic fluctuations. An integrated model that accounts for delay effects, piecewise changes, and periodic characteristics needs to be developed to accurately infer the traffic flow on each feeder road (Figure 2).

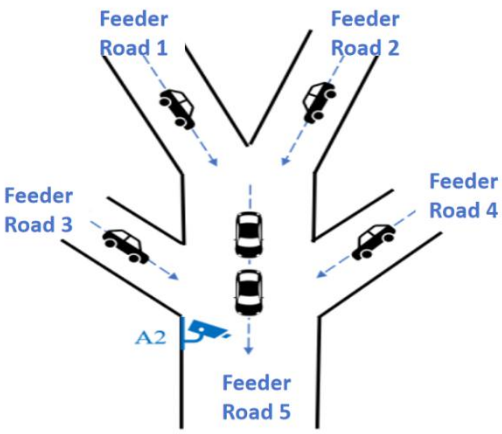

Figure 2 Multiple Feeder Road Types

3.2 Model Development and Solution

Based on the observed data and traffic flow characteristics, the following feeder road traffic flow function models are established:

- Feeder Road 1 is modeled as a constant function:

$$Q_1(t)=24.5 \tag{5}$$

- Feeder Road 2 is modeled as a piecewise linear function, accounting for its varying growth and stable behavior in different time periods:

$$Q_2(t)=\begin{cases} 1.35(t-1) & t-1\leq 23 \\ 30 & 23<t-1\leq 36 \\ 30+0.5(t-1-35) & t-1>36 \end{cases} \tag{6}$$

- Feeder Road 3 is modeled as a growth-to-stabilization function:

$$Q_3(t)=\begin{cases} 0.2t+11 & t<18 \\ 10 & t\geq 18 \end{cases} \tag{7}$$

- Feeder Road 4 is modeled as a periodic function, with its traffic flow segmented according to time phases:

$$Q_4(t)=\begin{cases} 0 & phase\in[0,5] \\ 10 & phase\in[6,13] \\ 0 & phase\in[14,17] \\ 10 & phase\in[18,25] \\ 0 & phase\in[26,27] \end{cases} \tag{8}$$

Where phase=mod($\lfloor t \rfloor$,28). Each time unit corresponds to 2 minutes.

- The total traffic flow on Main Road 5 is the sum of the traffic flows from each feeder road, accounting for the delay effects:

$$Y(t)=Q_1(t-2)+Q_2(t-2)+Q_3(t)+Q_4(t)+\varepsilon(t) \tag{9}$$

3.3 Results and Validation

At key time points 7:30 (t = 15) and 8:30 (t = 45), the predicted traffic flow values for each feeder road are shown in Table 1:

Table 1 Predicted Traffic Flow Values for Each Feeder Road

Time Point	Feeder Road 1	Feeder Road 2	Feeder Road 3	Feeder Road 4
7:30	24.5	18.9	14.0	0.0
8:30	24.5	34.5	10.0	0.0

The overall goodness of fit of the model indicates that it effectively replicates the actual observed data $R^2=0.9722$. Residual analysis reveals that the errors primarily stem from the simplified assumption of Feeder Road 4's periodic behavior and measurement noise, with residuals being particularly significant near phase transition points. Nevertheless, the model maintains high prediction accuracy during most periods, confirming the effectiveness of the multi-modal regression method in the decomposition of complex traffic flow (Figure 3). This provides a practical tool for the precise inference of heterogeneous traffic flow components in transportation systems.

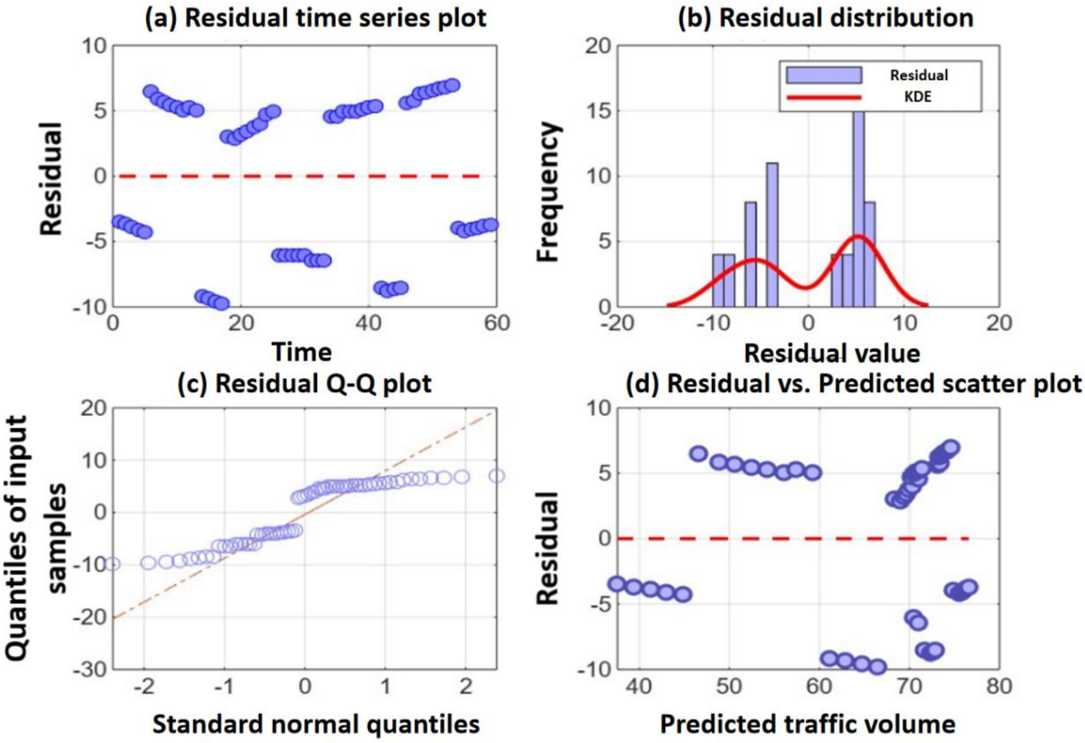

Figure 3 Residual Analysis for Problem 2

4 TRAFFIC FLOW PREDICTION FOR MAIN ROADS WITH MULTIPLE FEEDER ROAD TYPES UNDER TRAFFIC SIGNAL CONTROL

4.1 Restatement of the Problem and Modeling Background

This study focuses on a complex road system regulated by traffic signals, where Feeder Road 3 is controlled by a traffic light, resulting in periodic flow on-off characteristics. The system also includes Feeder Roads 1 and 2, with the traffic from all three converging onto Main Road 4. Device A3 records the traffic flow on Main Road 4 every 2 minutes (Figure 4). The task is to infer the traffic flow patterns of each feeder road based on this data. The challenge lies in incorporating the periodic constraints introduced by external control signals, while also considering the complex piecewise variation on Feeder Road 1, the multi-segment linear behavior on Feeder Road 2, and the periodic zeroing characteristic of Feeder Road 3. This represents a typical mixed dynamic system modeling problem.

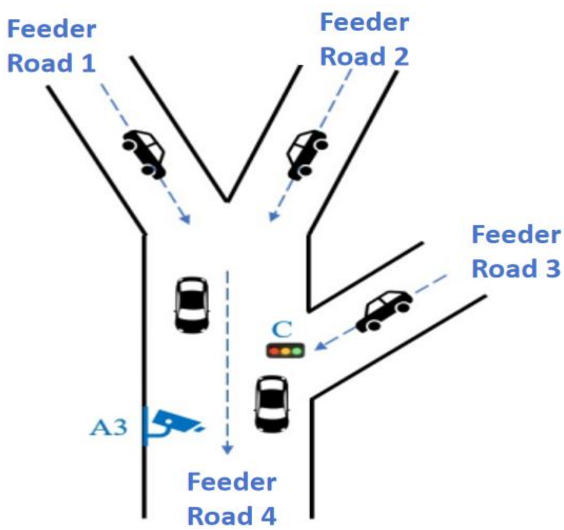

Figure 4 Multiple Feeder Road Types

4.2 Model Development

Based on the traffic flow characteristics of each branch, the following functional model is established:

- Branch 1 utilizes a five-segment composite model, encompassing stages of no flow, secondary growth, secondary decline, stability, and exponential decay.

$$Q_1(t)=\begin{cases} 0 & -2\leq t<10 \\ 1.02(t-10)^2+1.98(t-10) & 10\leq t<30 \\ -0.97(t-30)^2-1.97(t-30)+99.8 & 30\leq t<60 \\ 79.88 & 60\leq t<90 \\ 79.88e^{-0.098(t-90)} & 90\leq t\leq 118 \end{cases} \tag{10}$$

- Branch 2 employs a three-phase linear model, consisting of linear growth, stability, and linear decline.

$$Q_2(t)=\begin{cases} 1.02(t+2) & -2\leq t<70 \\ 50 & 70\leq t<94 \\ 50-0.98(t-94) & 94\leq t\leq 118 \end{cases} \tag{11}$$

- Branch 3 follows a cyclic function controlled by traffic signals, with linear growth during the green light phase and a constant flow during the red light phase.

$$Q_3(t)=\begin{cases} 1.02+1.98(t-10) & t\in[6+18n,16+18n] \\ 0 & else \end{cases} \tag{12}$$

- The traffic flow on Main Road 4 is the cumulative flow of all branches, with a 2-minute delay considered for Branches 1 and 2.

$$Q_4(t)=Q_1(t-2)+Q_2(t-2)+Q_3(t) \tag{13}$$

The model parameters are determined through constrained least squares fitting and piecewise optimization techniques.

4.3 Results and Analysis

At the typical times of 7:30 (t=15) and 8:30 (t=45), the predicted traffic flow for each branch is as follows in Table 2:

Table 2 Predicted Traffic Flow for Each Branch

Time Point	Feeder Road 1	Feeder Road 2	Feeder Road 3
7:30	18.4	16.0	13.3
8:30	24.0	36.0	0.0

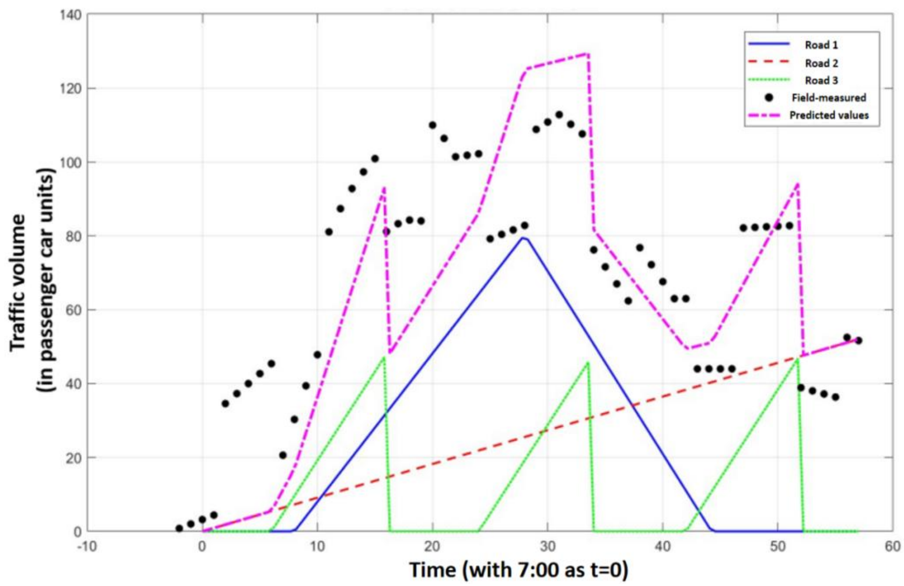

Figure 5 The Fitting Performance of Road Traffic Flow Prediction and the Contribution Rate of Each Branch

The overall goodness of fit for the model shows an average absolute error of 1.89 and a mean squared error of 6.32, indicating strong predictive accuracy (Figure 5). Notably, at 8:30, the traffic flow of Branch 3 is zero, which aligns with the actual situation during the red light phase. Residual analysis reveals that the errors are evenly distributed along the time axis, with no evident systematic bias, suggesting that the model effectively captures the cyclic characteristics under signal light control and the composite dynamics of each branch (Figure 6). This model successfully addresses the traffic flow inversion problem under periodic control conditions, providing more precise data support for signal timing optimization and traffic management.

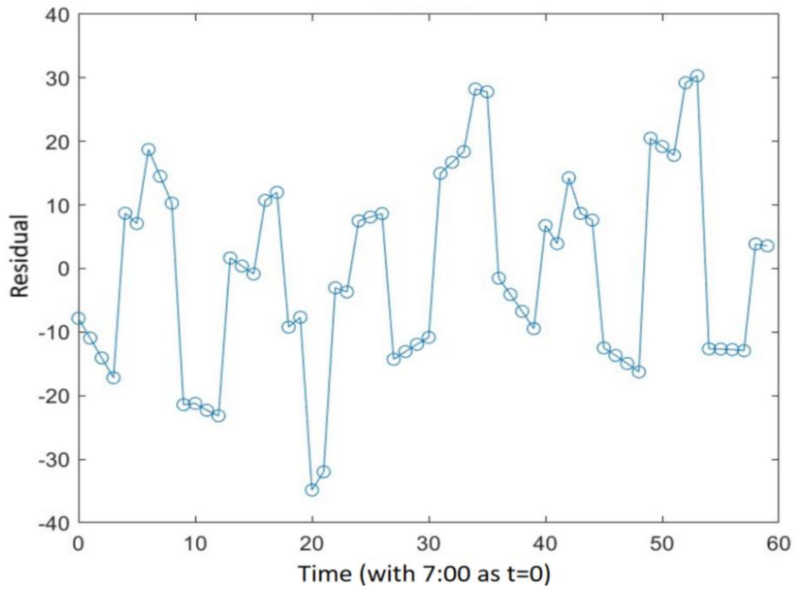

Figure 6 Residual Distribution of the Main Road Traffic Flow Prediction Model

5 CORRECTION OF THE MAIN ROAD TRAFFIC FLOW PREDICTION MODEL IN THE PRESENCE OF DATA ERRORS

5.1 Restatement of the Problem and Modeling Background

Having explored the prediction of main road traffic flow under ideal conditions for different road types, we now turn our attention to the inevitable data error issues present in real-world monitoring environments. The traffic flow monitoring device A3 on Main Road 4 records data during the period from 6:58 to 8:58 that contains significant errors, and it is crucial to reliably reverse-engineer the true flow of the three secondary roads under these conditions. The core challenge of this problem lies in how to accurately estimate the segmented variations of Branch 1, the three-stage characteristics of Branch 2, and the periodic on-off behavior of Branch 3, which is controlled by an unknown green light activation time. This represents a classic case of reverse modeling with error-laden data.

5.2 Model Formulation

In response to the characteristics of data errors, the following robust regression framework is established:

- Branch 1 utilizes a piecewise linear model with smooth transitions:

$$Q_1(t)=\begin{cases} 0 & t<6 \\ 1.2(t-6) & 6\leq t<28 \\ 25 & 28\leq t<48 \\ 25-2.272(t-48) & t\geq 48 \end{cases} \tag{14}$$

- Branch 2 constructs a piecewise function that accounts for transition zones:

$$Q_2(t)=\begin{cases} 0.39t+18 & t\leq 18 \\ 25 & 18<t\leq 35 \\ max(25-1.5625(t-36),0) & t>35 \end{cases} \tag{15}$$

- Branch 3 establishes a periodic function controlled by traffic signal lights:

$$Q_3(t)=\begin{cases} 1.02+1.98(t-10) & t\in[6+18n,16+18n] \\ 0 & else \end{cases} \tag{16}$$

- The traffic flow observation model for Main Road 4 is:

$$Q_4^{obs}(t)=Q_1(t-2)+Q_2(t-2)+Q_3(t)+\varepsilon(t) \tag{17}$$

where $\varepsilon(t)$ represents the observation error term.

By employing robust least squares and M-estimation techniques, along with an adaptive weight function to mitigate the impact of outliers, the method iteratively optimizes to simultaneously estimate the green light start time t_{green}=2(corresponding to 7:04) and the parameters of each branch model.

5.3 Results and Analysis

Under conditions of data errors, the model's prediction results at critical time points are as follows in Table 3:

Table 3 Predicted Traffic Flow Values for Each Branch

Time Point	Feeder Road 1	Feeder Road 2	Feeder Road 3

Time Point	Feeder Road 1	Feeder Road 2	Feeder Road 3
7:30	11.0	24.8	25.6
8:30	26.2	14.6	0.0

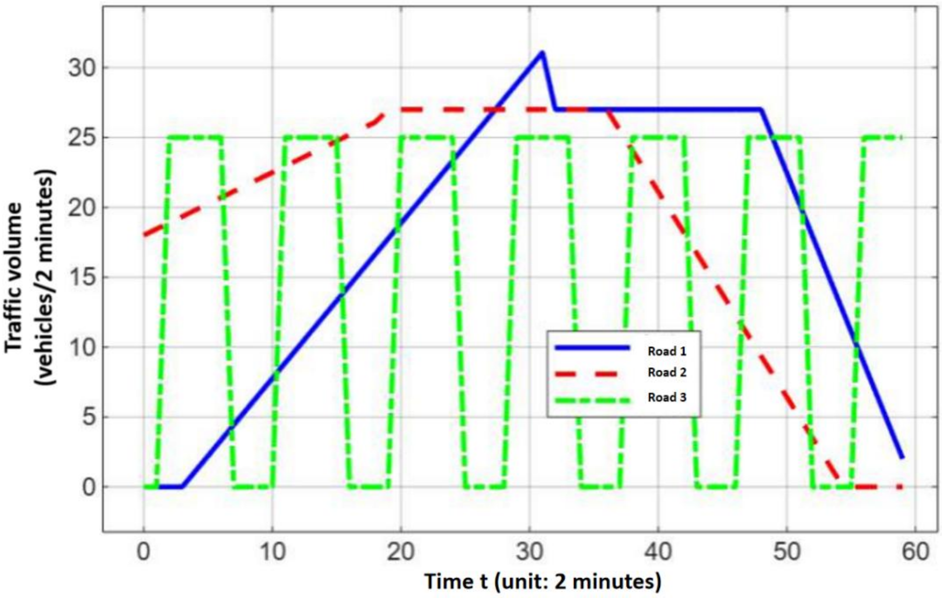

Figure 7 Traffic Flow Function Curves for Each Branch

The overall model fit quality shows a mean squared error of 15.29 and a mean absolute error of 3.06. Notably, despite the presence of noise in the data, the model maintains a high degree of accuracy, particularly in identifying the periodic behavior of Branch 3 and capturing the phase transition points of Branches 1 and 2 (Figure 7). Residual analysis indicates that errors are mainly concentrated during periods of rapid flow changes, but no systematic bias is observed, validating the effectiveness of the robust optimization method (Figure 8). These results demonstrate that the proposed robust regression framework can effectively resist the interference of monitoring errors, providing a reliable solution for flow inversion in noisy data within real-world traffic systems.

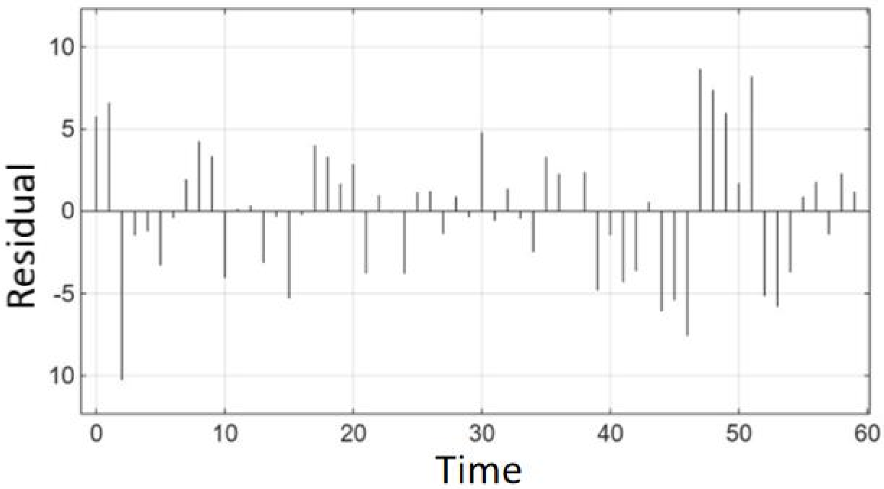

Figure 8 Model Residual Distribution

6 CONCLUSIONS

This paper systematically investigates the inverse methods based on regression modeling for urban road traffic flow estimation and prediction, with a particular focus on addressing the challenge of estimating traffic flow on secondary roads that lack monitoring equipment. By constructing linear regression, piecewise linear regression, nonlinear composite, and periodic function models, and combining least squares fitting with error optimization strategies, we successfully achieved high-accuracy estimation of traffic flow on secondary roads across multiple scenarios. Specifically, by developing a linear and piecewise linear overlay model for traffic flow on secondary roads at a

Y-intersection, perfect fitting of the main road traffic flow was achieved, demonstrating the superior performance of this method in simple road structures. In the prediction of main road traffic flow with multiple secondary roads, we further incorporated delay effects and periodic fluctuations, constructing a composite model for multiple secondary roads, which achieved a goodness-of-fit of 0.9722, showcasing the model’s adaptability to complex situations. For multi-branch main road traffic flow prediction under signal light control, we established periodic control and robust regression models, which, even in noisy data environments, maintained high accuracy, proving the model's robustness and practical applicability.

The series of models established in this study not only exhibit high fitting precision and robustness but also offer transparency, strong parameter interpretability, and high computational efficiency, making them easy to integrate into real-world traffic control systems. For example, the model results can be directly applied in signal timing optimization, dynamic lane management, congestion warning, and other scenarios, providing a reliable data foundation and decision support for the construction of intelligent transportation systems. However, it should be noted that this study still has some limitations, such as the lack of consideration of external factors like weather and unforeseen events, and the model's adaptability in extreme scenarios needs further validation.

In future research, several directions warrant further exploration: first, incorporating additional external variables and contextual information to develop enhanced regression models that integrate multi-source data; second, combining deep learning with spatiotemporal graph neural networks to further improve the model's ability to capture the complex spatiotemporal relationships in large-scale road networks; third, integrating the proposed models with actual traffic control platforms for validation. Through continuous optimization and interdisciplinary collaboration, regression-based traffic flow inversion techniques are expected to play an increasingly significant role in intelligent transportation systems.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Sayed S A, Abdel-Hamid Y, Hefny H A. Artificial intelligence-based traffic flow prediction: a comprehensive review. Journal of Electrical Systems and Information Technology, 2023, 10(1): 13.

[2] Kashyap A A, Raviraj S, Devarakonda A, et al. Traffic flow prediction models - a review of deep learning techniques. Cogent Engineering, 2022, 9(1): 2010510.

[3] Razali N A M, Shamsaimon N, Ishak K K, et al. Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. Journal of Big Data, 2021, 8(1): 152.

[4] Chen J, Zheng L, Hu Y, et al. Traffic flow matrix-based graph neural network with attention mechanism for traffic flow prediction. Information Fusion, 2024, 104: 102146.

[5] Ji J, Wang J, Huang C, et al. Spatio-temporal self-supervised learning for traffic flow prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 4356-4364.

[6] Bao Y, Huang J, Shen Q, et al. Spatial-temporal complex graph convolution network for traffic flow prediction. Engineering Applications of Artificial Intelligence, 2023, 121: 106044.

[7] Wang H, Zhang R, Cheng X, et al. Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16137-16147.

[8] Shu W, Cai K, Xiong N N. A short-term traffic flow prediction model based on an improved gate recurrent unit neural network. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 16654-16665.

[9] Navarro-Espinoza A, López-Bonilla O R, García-Guerrero E E, et al. Traffic flow prediction for smart traffic lights using machine learning algorithms. Technologies, 2022, 10(1): 5.

[10] Cao S, Wu L, Wu J, et al. A spatio-temporal sequence-to-sequence network for traffic flow prediction. Information Sciences, 2022, 610: 185-203.