Junzhuo LI, Doctoral student[1]
(Corresponding author)
E-mail: 19152202004@mails.guet.edu.cn
Wenyong LI, Ph.D.[1]
E-mail: traffic@guet.edu.cn
Guan LIAN, Ph.D.[2]
E-mail: lianguan@guet.edu.cn
[1] School of Information and Communication
Guilin University of Electronic Technology
No.1, Jinji Road, Qixing District, Guilin 541004, China
[2] School of Architecture and Transportation Engineering
Guilin University of Electronic Technology
No.1, Jinji Road, Qixing District, Guilin 541004, China

Transport Technology
Original Scientific Paper
Submitted: 9 Apr. 2022
Accepted: 13 July 2022

A NONLINEAR AUTOREGRESSIVE MODEL WITH EXOGENOUS VARIABLES FOR TRAFFIC FLOW FORECASTING IN SMALLER URBAN REGIONS

ABSTRACT

Data-driven forecasting methods have the problems of complex calculations, poor portability and need a large amount of training data, which limits the application of data-driven methods in small cities. This paper proposes a traffic flow forecasting method using a Nonlinear AutoRegressive model with eXogenous variables (NARX model), which uses a dynamic neural network Focused Time-Delay Neural Network (FTDNN) with a Tapped Delay Line (TDL) structure as a nonlinear function. The TDL structure enables the FTDNN to have short-term memory capabilities. At the same time, before the data is input into the FTDNN, the use of trend decomposition or differential calculation on the traffic data sequence can make the NARX model maintain long-term predictive capabilities. Compared with common nonlinear models, the FTDNN has structural advantages. It uses a simple TDL structure without the memory mechanism and the gated structure, which can reduce the parameters of the model and reduce the scale of data. Through the four-day data of Guilin City, the traffic volume forecast for five minutes is verified, and the performance of the NARX model is better than that of the SARIMA model and the Holt-Winters model.

KEYWORDS

intelligent transportation system; traffic flow forecasting; time series; NARX model; traffic data.

1. INTRODUCTION

Traffic congestion has generated economic, social and environmental issues in many cities around the world. Accurate traffic flow forecasting methods can assist traffic managers in formulating traffic policies to alleviate congestion. As urban traffic congestion becomes more and more serious, reliable short-term traffic forecasting becomes more and more important in Intelligent Transportation Systems (ITS) [1, 2].

The main objective of short-term traffic forecasting is the prediction of traffic measurements (e.g. travel time, traffic flow) in the near future, ranging from the next few minutes to up to several hours, based on historical traffic data [3]. Traffic forecasting methods fall into two main categories. One method is model-driven, which combines mathematical statistics, time series analysis and other methods with traffic theory, such as Autoregressive Integrated Moving Average (ARIMA) [4, 5]. Exponential Smoothing (ES) [6, 7]. Traditional time-domain analysis methods often require homogeneity of variance and linear correlation of data variance, while traffic flow data often fails to meet the assumptions of this model, resulting in low prediction accuracy [8–10]. With the development of time-domain analysis in the case of heteroscedasticity and nonlinearity [11, 12], more prediction models are applied to traffic flow prediction, such as the Autoregressive Conditional Heteroskedasticity (ARCH) model [13] and a series of extended models derived from it [14, 15]. The model-driven traffic forecasting method has a good explanatory ability and can describe the trend change characteristics of the traffic flow state better. However, for the complex urban traffic system, the model-driven

traffic prediction method needs to make a lot of assumptions about the model, and make elaborate and complex designs on the prediction model. This requires the forecaster to have rich traffic practice experience and statistical knowledge and it also makes the model too complex to solve [16].

As the variety and quantity of available traffic data provided by the ITS continues to increase, in recent years, data-driven traffic prediction methods have continuously developed [17-19]. The data-driven traffic forecasting method can get better forecasting results without making a lot of assumptions about the model. Forecasters do not need to have a deep understanding of urban transportation systems to achieve better forecasting accuracy. Many well-known traditional Machine Learning (ML) methods have been applied to traffic flow prediction, such as Support Vector Machines (SVM) [20], K-Nearest Neighbours (KNN) [21, 22], Random Forest [23, 24], etc. Machine learning methods can not only effectively capture the spatial and temporal relationships in the transportation network, but also better process high-dimensional data and catch nonlinear relationships, so that the rich traffic data can be more fully utilised by ML methods [17].

In recent years, Deep Neural Network (DNN) has attracted the attention of traffic researchers due to their excellent performance in accuracy and error. ML methods need to perform exploratory analysis such as dimensionality reduction for data, while the DNN can avoid complex feature engineering [25–27]. Recurrent Neural work (RNN) is a type of neural network with short-term memory capacity. RNN models form network structures with loops that accept information from upper neurons as well as their own error information. It is better suited for the prediction of urban traffic flow through back-propagation algorithms. However, when the time series is long, the RNN model experiences the gradient exploding problem and the gradient vanishing problem. Aiming at the problems in the prediction of long time series, the Long-Short Term Memory (LSTM) model introduces the gating mechanism to make up for the deficiency of the RNN model in long time series processing to some extent [28, 29].

The nonlinear method can predict the traffic volume very accurately under the condition of sufficient data, such as the scene of the highway. *Table 1* shows the performance and sources of SOTA methods for several highway traffic datasets.

The linear prediction model represented by ARIMA has simple parameter calculation and good interpretability. However, the limitation is that their tendency to concentrate on the mean values of the past series data seems unable to capture the rapid variational process underlying traffic flow [36]. The nonlinear structure of the ML and DNN can capture more subtle relationships in the transportation system, but

Table 1 – SOTA methods for highway traffic datasets

Dataset	Methodology	MAE	Literature
METR-LA	Traffic transformer	3.28	Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting [30]
PEMS-BAY	Traffic transformer	1.77	Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting [30]
PeMS07	ADN	21.62	Structured time series prediction without structural prior [31]
PeMS-M	ST-UNeT	3.38	ST-UNet: A Spatio-Temporal U-Network for graph-structured time series modeling [32]
PeMS04	SCINet	19.02	Time series is a special sequence: Forecasting with sample convolution and interaction [33]
PeMSD7	STG-NCDE	20.53	Graph neural controlled differential equations for traffic forecasting [34]
Q-Traffic	hybrid Seq2Seq	8.63	Deep sequence learning with auxiliary information for traffic prediction [35]
PeMSD4	STG-NCDE	19.21	Graph neural controlled differential equations for traffic forecasting [34]
PeMSD8	STG-NCDE	15.45	Graph neural controlled differential equations for traffic forecasting [34]
PeMSD3	STG-NCDE	15.57	Graph neural controlled differential equations for traffic forecasting [34]

the more complex relationships make the model difficult to explain. Meanwhile, more parameters in the nonlinear structure make the model difficult to calculate and require more data [37].

In order to avoid the insufficient data size in smaller urban region's traffic flow prediction, this paper adopts the traditional NARX model and uses Focused Time-Delay Neural Network (FTDNN) as a nonlinear function for prediction, which makes the model simple and easy to calculate and adapt to the nonlinear change of traffic flow sequence. Through the introduction of nonlinear functions, the NARX model is able to adapt to complex predictions better. It is a commonly used statistical time series model with good interpretability and operability, it is easily solved and is widely used in various forecasting scenarios [38, 39]. In this paper, the FTDNN is used to construct the nonlinear function of the NARX model, so that the neural network has short-term memory and can effectively learn the short-term correlation of urban traffic systems. The FTDNN had the Tapped Delay Line (TDL) in the network, which can be distributed at the input to the first layer of the static feedforward or throughout the network. FTDNN has no gate control and TDL distributed in the network replaces the memory mechanism for network parameter states in RNN, which greatly simplifies calculation. The traffic flow data of the North Zhongshan Road in Guilin, China from 6 to 10 April 2020 were used for 5-minute traffic flow forecast to verify the model. The results show that the prediction performance of the NARX model is better than that of the SARIMA model and Holt-Winters model. NARX can accurately predict urban traffic flow and it has the advantages of simple structure, easy calculation and needs less training data.

The rest of the paper is structured as follows. In Section 2, the structure of the NARX model is explained, including the series-parallel architecture used in training, the parallel architecture used in prediction and the TDL structure. In Section 3, the NARX model with FTDNN is verified and calculated by using traffic flow data of the city of Guili. Furthermore, trend decomposition methods and difference calculation methods are used to improve the performance of the model. In Section 4, the NARX model is compared with ARIMAS and Holt-Winters model. Finally, Section 5 presents the main conclusions of this paper.

2. MATERIALS AND METHODS

2.1 The ARX model

The AutoRegressive Model (AR) is a class of time series models commonly used in statistics, using the historical value of a variable *Y* to predict itself:

$$Y(t) = \omega_0 + \sum_{k=1}^{K} \omega_K Y(t-k) + \varepsilon_t \tag{1}$$

where *K* is a hyper-parameter, which has to do with the property of the time series itself and can be determined by drawing autocorrelation diagrams of time series. The value of *K* is generally equal to the number of periods in which the autocorrelation of the sequence decreases rapidly. When the time series has a long-term correlation, *K* can take a larger value, otherwise, it can take a smaller value. $\omega_0,...,\omega_k$ is the autocorrelation coefficient, $\varepsilon_t \sim N(0,\sigma^2)$ is the noise, and *t* is the period of time.

The AR model can only predict a time series itself and cannot be applied to the problem of multivariate prediction. Different from the AR model, the AutoRegressive eXogenous (ARX) model contains an input term called exogenous variable, which can solve multiple prediction problems. Its general structure is as follows:

$$A(q)\ \ Y(q) = B(q)\cdot(t-k) + \varepsilon(t) \tag{2}$$

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \tag{3}$$

$$B(q) = b_1 + b_2 q^{-1} + \ldots + b_m q^{m+1} \tag{4}$$

where *q* is the delay operator, *n* is the order of autocorrelation and *m* is the order of correlation in exogenous variables.

The ARX time series model is a linear representation of the dynamic system in discrete time. However, there are complex relationships among urban traffic systems, which cannot be simply described by linear relationships [36]. For example, the traffic flow at the downstream intersection will form an intermittent, high-density fleet due to signal control at the upstream intersection. Once upstream changes in flow, speed and density due to signal control or traffic jams, such traffic status is likely to be passed to downstream intersections.

2.2 The NARX model

The NARX can predict the nonlinear system, which can be represented as in *Equation 5*.

$$Y(t) = f(Y(t-1),\ldots,Y(t-m),x_1(t-1),\ldots x_n(t-1),\ldots,x_1(t-n),\ldots,x_n(t-n)) \tag{5}$$

where $f(\cdot)$ is a nonlinear function, which can adopt the general nonlinear relationship or feedforward neural network, support vector regression and other learning models.

In this paper, the FTDNN is used as a nonlinear function $f(\cdot)$. The FTDNN's basic unit is shown in *Figure 1* and the multi-layer basic unit forms a complete nonlinear function.

The FTDNN is a kind of simple dynamic network. As the training goes on, the value of d at the input terminal will change, thus affecting the structure of the neural network, so that FTDNN's dynamics is represented in the input layer of a static multi-layer network [40].

In the process of network training, the real value of time series is available, therefore the NARX model is trained with a series-parallel structure. In the long-term prediction process, the time series has no real values. The NARX model takes the previous output as the estimated value of the next time series and feeds it back to the input of the feedforward neural network to form a parallel structure. The structure of the NARX model is shown in *Figure 2*.

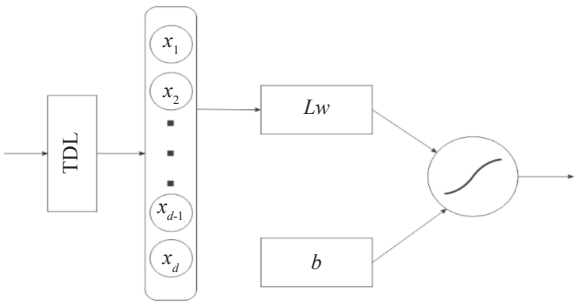
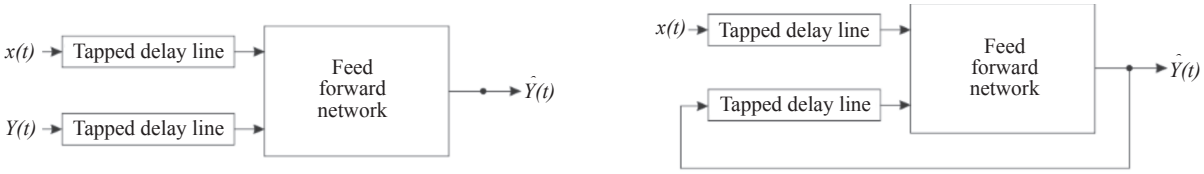


Figure 1 – Structure diagram of FTDNN's basic unit



a) Series-parallel architecture *b) Parallel architecture*

Figure 2 – Training and prediction structure of the FTDNN

The TDL is an important structure that enables the NARX models to have short-term memory. It has several delay units in series and the number of delay units is related to the autocorrelation and cross-correlation of time series. When the time series can maintain the correlation for a long time, more delay units are needed. When TDL has one input and multiple delayed outputs, the structure is shown in *Figure 3*.

2.3 Comparison of the FTDNN model and the RNN model

The FTDNN model has more advantages than the RNN model when dealing with inputs with short sequence lengths and feature sizes. For example, traffic data collected from small cities is collected by detectors with lower density and longer sampling intervals.

There is a sequence $\{(x_{1,1},x_{1,2},\ldots, x_{1,n}),\ldots,(x_{m,1},x_{m,2}, \ldots,x_{m,n})\}$ whose sequence length is m and the feature size is n. Input this sequence into the single-layer FTDNN and RNN models with U hidden units. The RNN model is shown in *Figure 4a* and the FTDNN model is shown in *Figure 4b*.

The RNN model requires m forward computations to complete a complete forward propagation. Each forward computation shares the parameters of the model and the number of parameters required to be computed is $(n+U+1)U$. The FTDNN model only needs one forward calculation to complete the forward propagation and the number of parameters to be calculated is $(mn+1)U$. When the data collection interval is five minutes, the traffic volume is predicted based on the traffic data of four

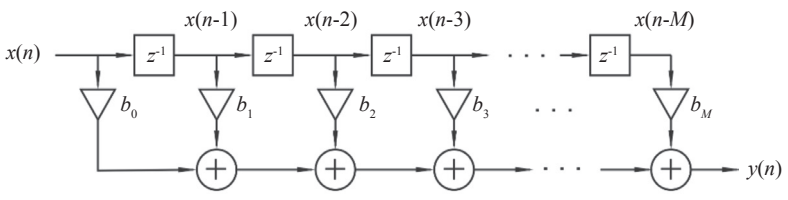
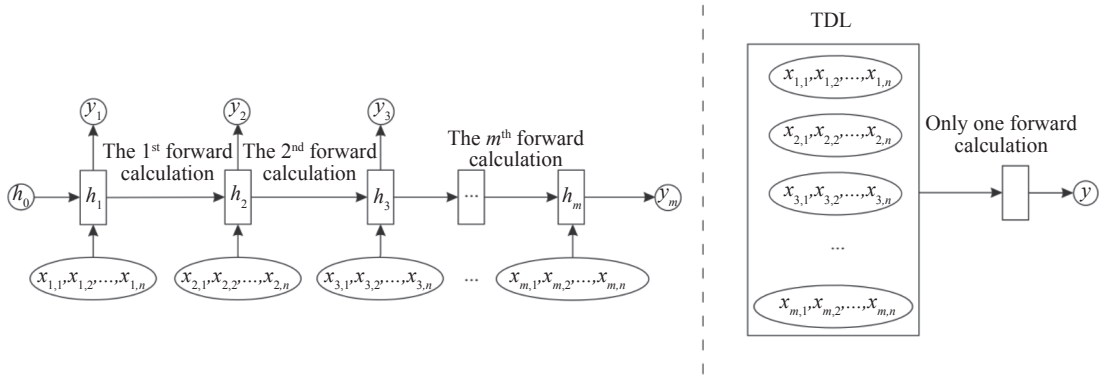


Figure 3 – TDL structure schematic

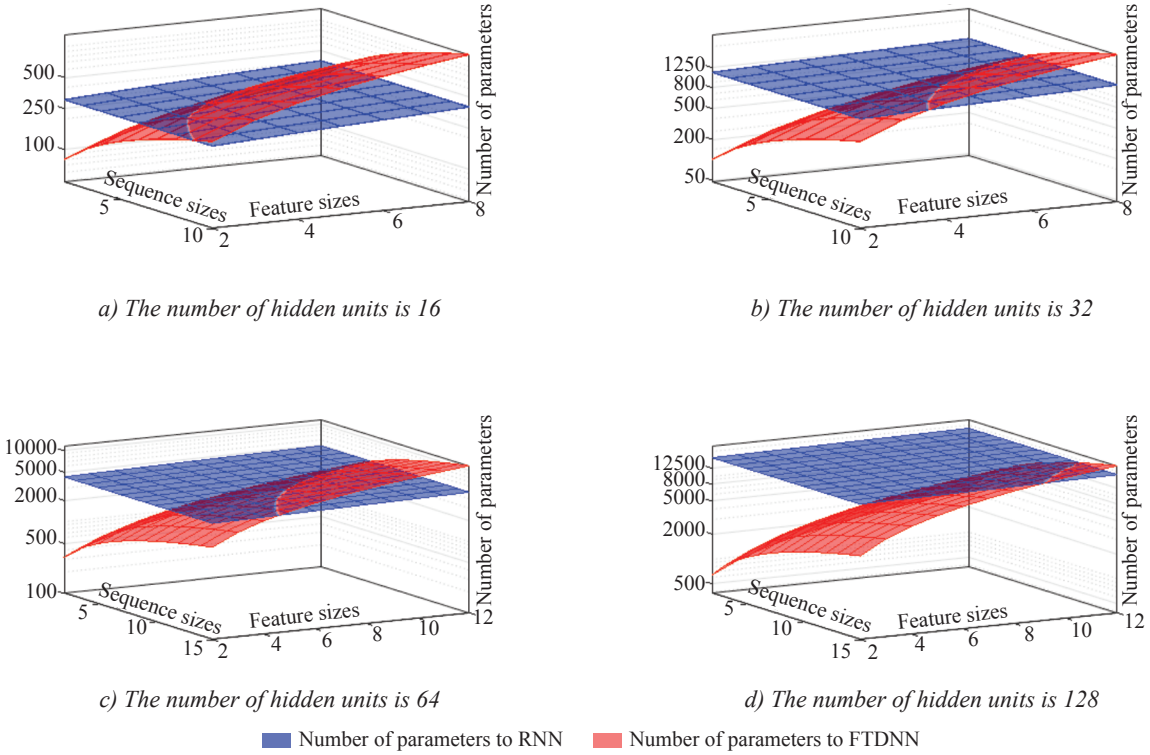


a) Forward propagation structure of RNN model *b) Forward propagation structure of FTDNN model*

Figure 4 – Forward propagation structure of the RNN model and the FTDNN model

characteristics (speed, density, occupancy rate, traffic flow) within half an hour. The RNN model with 64 hidden units needs to perform six forward computations for 5440 parameters; the FTDNN with 64 hidden units needs only one forward computation for 1600 parameters.

By comparing the parameter number calculation formulas of the two models, it is found that the FTDNN model has computational advantages when the data sampling interval is large and the data feature size is small. As the data sampling time decreases, the sequence length of traffic data will increase. At the same time, when the density of data collectors increases, the number of related features will increase, and the number of parameters of the FTDNN model will increase rapidly, surpassing the RNN model and no longer having computational advantages. *Figure 5* shows the number of parameters of the FTDNN model and the RNN model when the number of hidden nodes is 16, 32, 64 and 128. Due to the low quality of data collection in small cities, large data collection intervals and low detector density, which lead to short sequence lengths and small feature sizes of traffic data, the FTDNN model is more suitable for small city prediction calculations.



a) The number of hidden units is 16 *b) The number of hidden units is 32*

c) The number of hidden units is 64 *d) The number of hidden units is 128*

Figure 5 – Comparison of the number of parameters between RNN and FTDNN

2.4 Long-term correlation of the NARX model

Although the TDL structure enables the NARX model to have short-term memory capacity, the daily change trend of traffic flow has a fixed cycle and has a significant influence on the traffic flow, which requires the model to have long-term memory. Trend decomposition methods and difference calculation methods are commonly used to improve data.

According to the Cramer decomposition theorem (the formula is as follows), any time series can be decomposed into two parts, one of which is the steady trend component and the other is the residuum component.

$$x_t = \mu_t + \varepsilon_t = \sum_{j=0}^{d} \beta_j t^j + \psi(B) a_t \tag{6}$$

The Cramer decomposition theorem demonstrates that the fluctuations of any sequence can be regarded as both deterministic and stochastic effects. Therefore, the daily change trend of traffic flow is regarded as a deterministic influence and dummy variables D are constructed according to the time period of the day in which the output variables to be predicted are located. When training the model, dummy variables are taken as dependent variables and the previous variables are trained together, so the NARX model with dummy variables will not only consider the traffic flow data of the associated intersection but also comprehensively analyse the daily change trend of the intersection in the regression. Its functional relationship can be expressed in *Equation 7*.

The difference operation is also a simple and effective way to catch the deterministic trend of the time series. For the traffic flow data with a fixed period, the difference calculation with cycle length as step length can better catch the daily change trend of the traffic flow.

The first order and C steps difference calculation is carried out on the traffic flow data and then the NARX model after difference operation training can effectively catch the daily change trend of traffic flow in the data, where c represents the cycle length of traffic sequences, ∇_k^p represents the p order k step difference operation on a time series [41]. Its functional relationship can be expressed in *Equations 8-10*.

3. EXPERIMENT

3.1 Data preparation

This paper uses the data of the city of Guilin to test the performance of the NARX model. The data includes the time, speed, license plate number and other information of vehicles passing through fixed monitoring points in the Guilin urban area from Monday, 6 April 2020 to Sunday, 12 April 2020. In this paper, the monitoring data of some sections of the North Zhongshan road in Guilin was used for the experiment and the test area is shown in *Figure 6*. Taking the step length of five minutes as the time series, the traffic flow information of four detection points was counted and the traffic flow time series of 1.2.3.4 detection points were established respectively. The monitoring points are the Beichen Interchange, the Intersection of the Zhongshan North Road and the Qunzhong Road, the Intersection of the Zhongshan North Road and the North First Ring Road, the Intersection of Yushan Road and the Zhongshan North Road, in that order.

Figure 6 – Diagrammatic drawing of the experiment road section

$$Y(t) = f(Y(t-1), \ldots, Y(t-d), x_1(t-1), \ldots, x_n(t-1), \ldots, x_1(t-d), \ldots, x_n(t-d), D) \tag{7}$$

$$\nabla_c^1 Y(t) = f(\nabla_c^1 Y(t-1), \ldots, \nabla_c^1 Y(t-d), \nabla_c^1 x_1(t-1), \ldots, \nabla_c^1 x_n Y(t-1), \ldots, \nabla_c^1 x_1 Y(t-d)) + \nabla_c^1 Y(t-c) \tag{8}$$

$$\nabla^p Y(t) = (1-B)^p Y(t) = \sum_{i=0}^{p} (-1)^i C_p^i Y(t-i) \tag{9}$$

$$\nabla_k Y(t) = \nabla Y(t) - \nabla Y(t-k) = (1-B^k) Y(t) \tag{10}$$

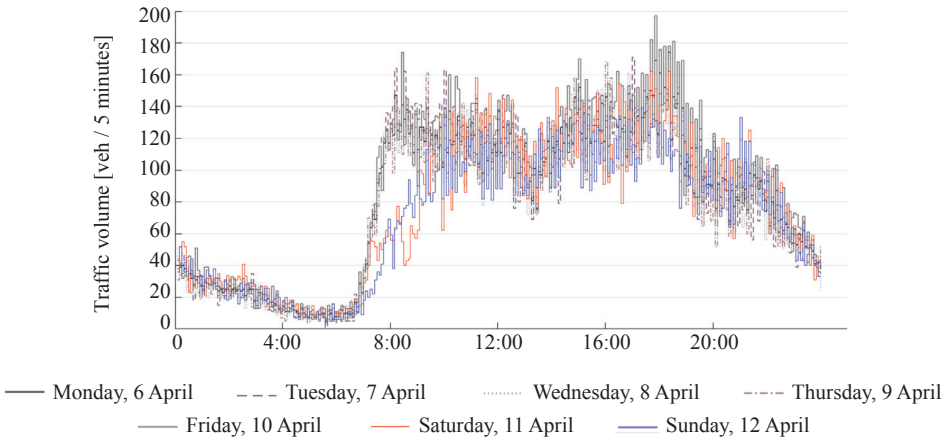


Figure 7 – Daily trends in traffic volume

The variation trend of daily traffic volume in a week was counted and the graph was drawn in *Figure 7*. The weekday morning peak hour occurs at 8 o'clock, earlier than on rest days, and the traffic increase rate in the morning peak on weekdays is higher than that on rest days. In addition, the evening peak hours of working days are basically the same as those of rest days, but the traffic is higher than that of rest days. Considering the difference in travel behaviour between working days and rest days, it is not suitable to use the same model to predict the traffic volume, so the subsequent predictions all use the traffic data on working days.

3.2 The NARX model with FTDNN

The traffic flow of detection point 4 was predicted by using the traffic flow of detection point 4 and detection 1,2,3 as exogenous variables. The data from 6 to 9 April were used for the training of the model, and the time series was divided into the training set, validation set and test set according to a scale of 0.7:0.15:0.15 for training. The training process is shown in *Figure 8*.

As the training progresses, the Mean Square Error (MSE) of the model decreases continuously. When the iteration process is carried out to the 10^{th} operation, the MSE of the validation set and test set will not be reduced, and then the training of the model has been completed. The fitting residuum of the training set is smaller than that of the validation set and the test set, so there is some overfitting phenomenon. The training model in step 10 is taken for drawing *Figure 9* and the analysis. For the three data sets, the predicted results are evenly distributed around the actual data and the predicted trend is consistent with the actual trend. Therefore, the training model can better fit the original distribution of traffic flow.

The time series diagram and residual diagram of detection point 4 from 6 April to 9 April is drawn, as shown in *Figure 10*. The model better fitted the 4 days traffic trend.

Then, the traffic flow on 10 April is predicted according to the training model to test the prediction accuracy. The time series diagram and residual diagram of observation point 4 on 10 April is drawn, as shown in *Figure 11*.

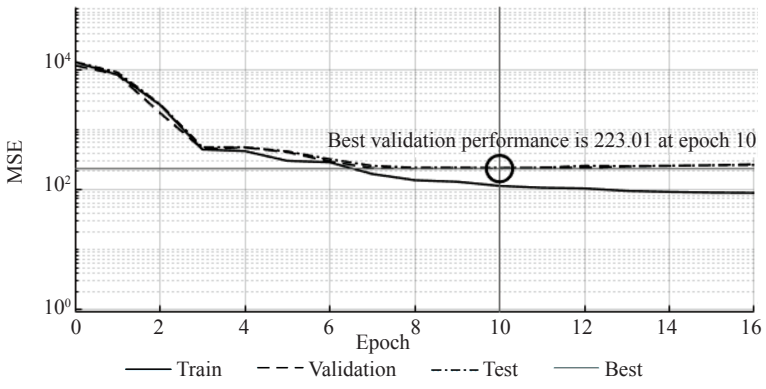


Figure 8 – The performance of the NARX model with FTDNN in different epochs

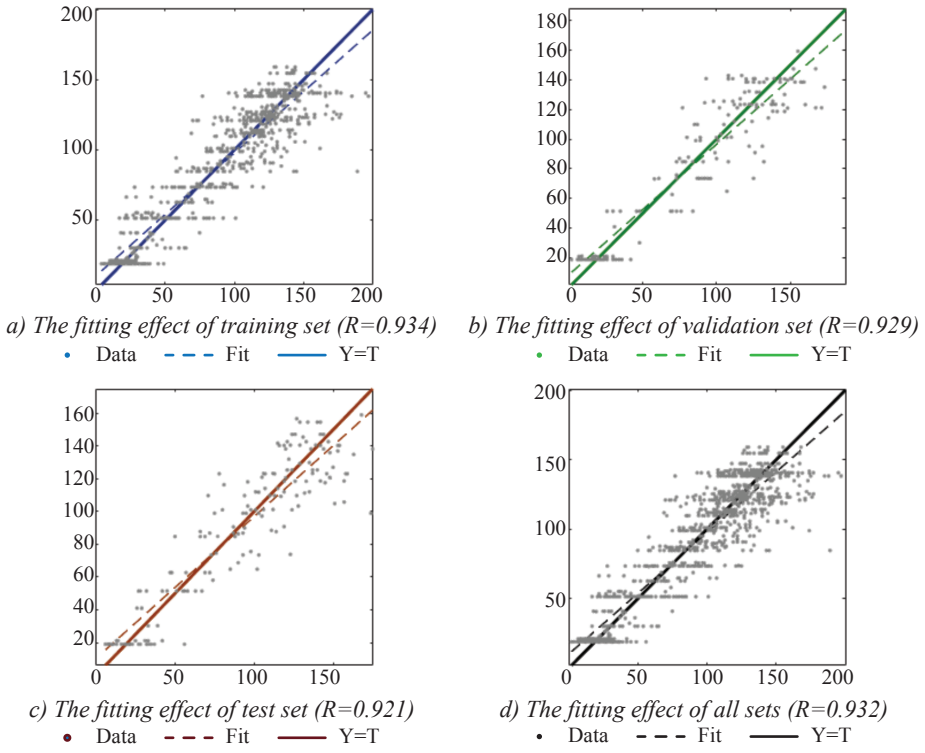


Figure 9 – The fitting performance of the NARX model with FTDNN in different data sets

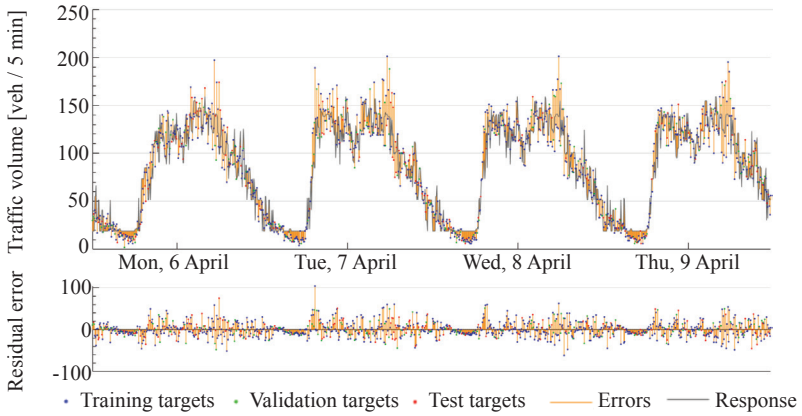


Figure 10 – The fitting performance of the NARX model with FTDNN on time series

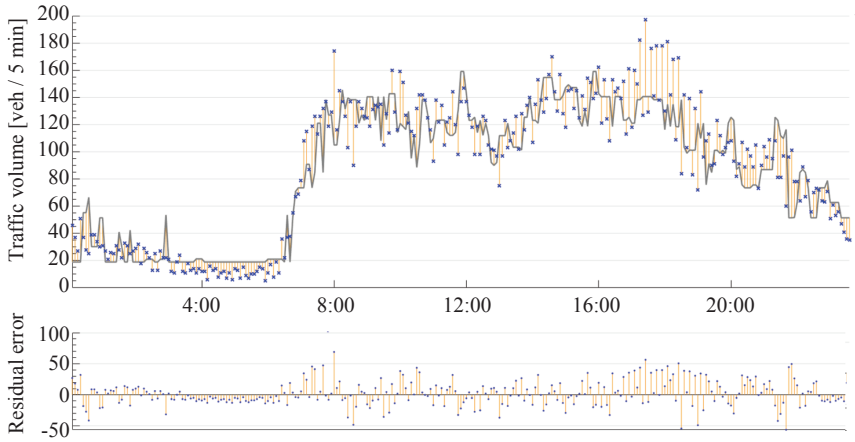


Figure 11 – The prediction performance of the NARX model with the FTDNN on time series

It is observed that the NARX model with the FTDNN can predict the traffic flow information well. But at 3 to 4 o'clock and 16 to 20 o'clock, the prediction model continuously gives larger or smaller predictions. Because the FTDNN could not accurately grasp the long-term daily change trend of the traffic flow time series, it gave a trend of wrong prediction in these two periods, which reduced the accuracy of the prediction.

3.3 The NARX model after trend decomposition

The traffic flow sequence after trend decomposition is trained by the FTDNN model according to *Equation 7*. The training data are consistent with the data set of Subsection 3.1. The training process is shown in *Figure 12*. Due to the increase in the number of input variables, the training converges at step 13, which is slower than the ten steps of Section 1. On the other hand, the performance curves of training set, validation set and test set are very close, so the overfitting phenomenon has been significantly improved.

The time series diagram and residual diagram of detection point 4 from 6 April to 9 April is drawn, as shown in *Figure 13*. The NARX model after trend decomposition also maintains a high fitting accuracy.

According to the NARX model after trend decomposition, the traffic flow on 10 April is predicted to test the accuracy of the model. The time series diagram and residual diagram of detection point 4 on 10 April are drawn as shown in *Figure 14*. The prediction model maintains a high prediction accuracy, and the deviation in the prediction is significantly improved at 3–4 and 4–7 hours.

3.4 The NARX model after difference operation

Before the training, the time series was calculated by difference according to its cycle, namely 288 steps per day, and the calculation was carried out according to *Equation 8*. Due to the difference calculation, the four-day data from 6 April to 9 April were reduced by 288 phases, only producing a total of 864 data phases over three days.

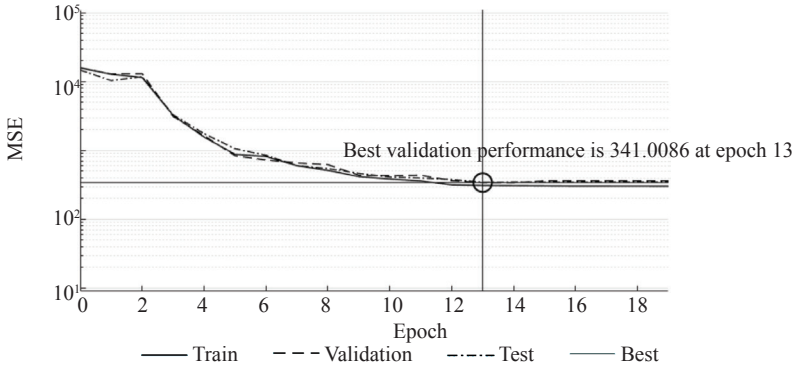


Figure 12 – The performance of the NARX model after trend decomposition in different epochs

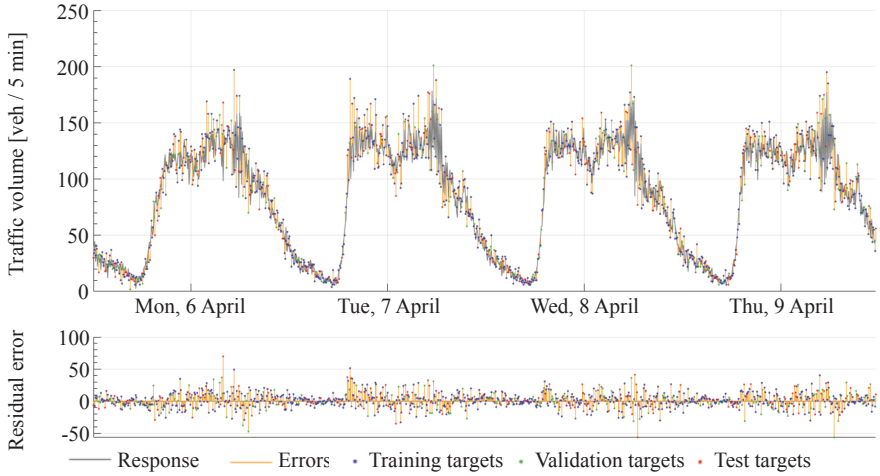


Figure 13 – The fitting performance of the NARX model after trend decomposition on time series

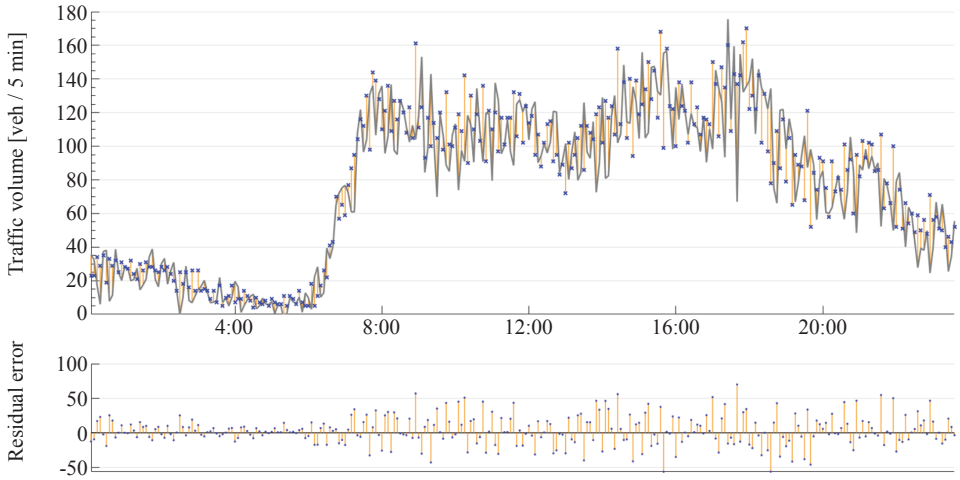


Figure 14 – The prediction performance of the NARX model with dummy variables on time series

In order to confirm that the changing trend of daily time is caught fully by the difference operation, the autocorrelation function after the difference operation of the time series is drawn for checkout, as shown in *Figure 15*. With the increase of the lag phases, the autocorrelation coefficient decreases rapidly, which indicates that the time series after the difference operation is a stationary series, and the difference operation can effectively catch the daily change trend of the time series.

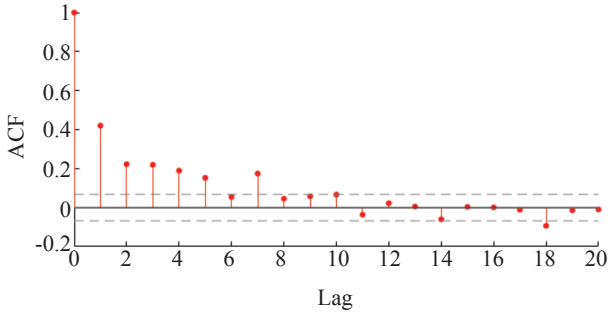


Figure 15 – Autocorrelation coefficient graph of time series after difference operation

The time series diagram and residual diagram of detection point 4 from 7 April to 9 April are drawn as *Figure 16a*, then forecast of the traffic flow on 10 April and its time series diagram and residual diagram are drawn as *Figure 16b*. The NARX model after difference operation caught the temporal correlation of the traffic flow sequence. By observing the residual plot on the lower side of *Figure 16b*, there is no continuous prediction deviation like the residual plot in *Figure 11*. The NARX model after the difference operation has better fitting accuracy and prediction accuracy and avoids the prediction deviation caused by the daily change trend of traffic flow.

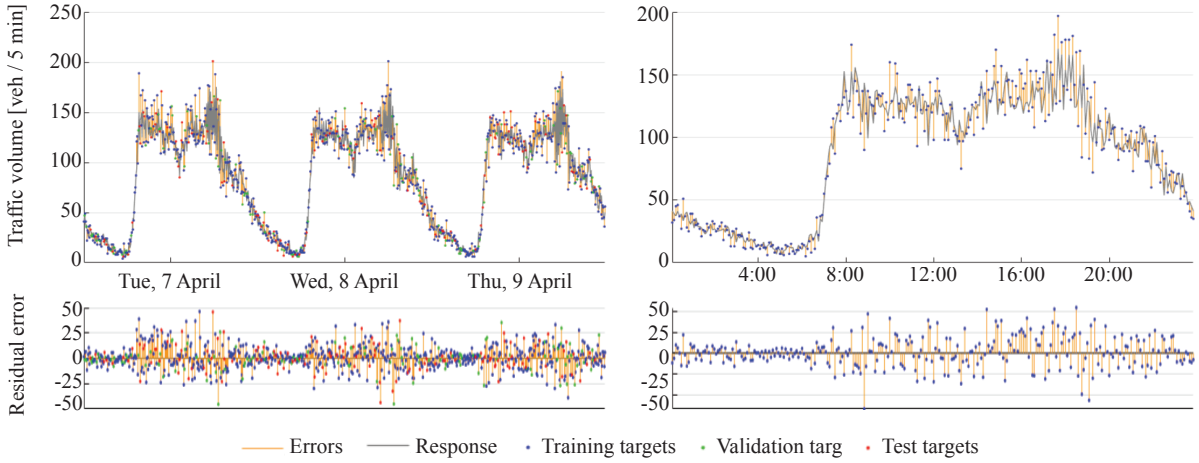


Figure 16 – a) The fitting performance of the NARX model after difference operation from 7 to 9 April b) The prediction performance of the NARX model after difference operation on 10 April

4. DISCUSSION

4.1 Comparison of the model fitting performance

By the MSE and coefficient of determination of the NARX model, the fitting performance was compared. *Table 2* shows the comparison of the fitting performance. The fitting residuum of the NARX model after trend decomposition is significantly larger than the NARX model after difference operation and NARX model. All three methods have good fitting performance. Their coefficient of determination can be stable above 0.9, indicating that the fitting performance is significant.

4.2 Establishing the SARIMA model

To determine that the NRAX model has a better prediction performance, the prediction results are compared with the SARIMA and Holt-Winters methods.

Table 2 – Comparison of fitted performance

Method	Fitting performance		
	Data set	MSE	R
NARX model	Training set	114.99	0.975
	Validation set	223.01	0.952
	Testing set	229.34	0.957
NARX model after trend decomposition	Training set	309.84	0.934
	Validation set	341.01	0.929
	Testing set	346.37	0.921
NARX model after difference operation	Training set	137	0.972
	Validation set	180.11	0.961
	Testing set	163.24	0.965

First of all, the data is pre-processed, the difference operation is carried out on the model to make the sequence stable and pass the white noise test. An autocorrelation graph of the time series is drawn as *Figure 17*, and a partial autocorrelation graph is drawn as *Figure 17b*, to judge the order of the model.

By observing the figure, it is found that the autocorrelation function of the time series shows the second-order censored characteristic, and the partial autocorrelation function shows the trailing characteristic. Therefore, the SARIMA (0,0,2)×(0,1,0) 288 model is adopted for prediction. The maximum likelihood method was used to estimate the parameters of the model and the prediction model was obtained:

$$\nabla_{288}^{1}\hat{x}_t = (1 + 0.81715B - 0.0265B^2)\varepsilon_t, \varepsilon_t \sim N(0, 209) \quad (11)$$

where B is the delay operator, $B = \frac{x_{t-1}}{x_t}, B^2 = \frac{x_{t-2}}{x_t}$.

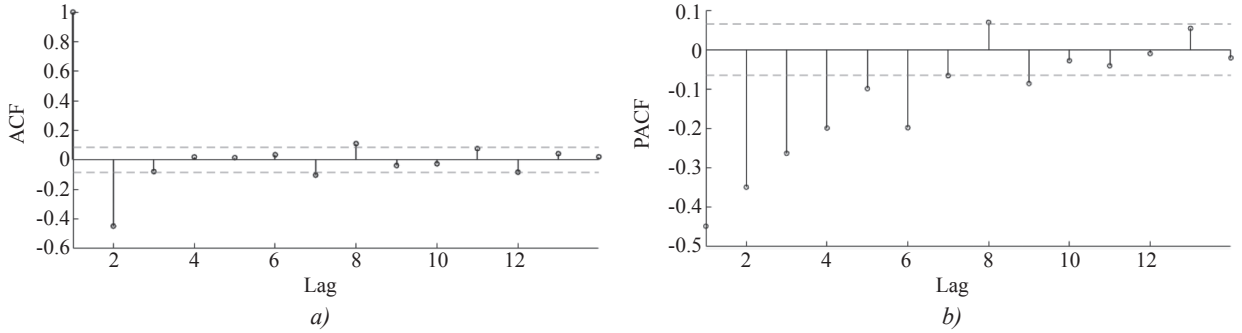
4.3 Establishing the Holt-Winters model

The solution of the Holt-Winters model is to calculate the three smoothing coefficients α, β and γ, which satisfy the following constraint: $0<\alpha,\beta$ and $\gamma<1$, and respectively represent the smoothing of horizontal series, trend series and seasonal series. According to the judgment of the traffic flow time series, the linear growth trend of the traffic flow has little influence on the traffic flow prediction in short term, so parameter β is specified as 1. Then, other parameters are calculated, α=0.0613, γ=0.548. The following model is constructed:

$$a_t = 0.0613(\hat{x}_t - s_{t-288}) + 0.9387a_{t-1} \quad (12)$$

$$s_t = 0.548(\hat{x}_t - a_t) + 0.452s_{t-288} \quad (13)$$

where a_t is the horizontal trend of the time series; s_t is the daily change trend of the time series, $\{s_t\}$ is a sequence of 288 values, depending on the value



a) *b)*

Figure 17 – a) Autocorrelation diagram of time series after difference operation
b) Partial autocorrelation diagram of time series after difference operation

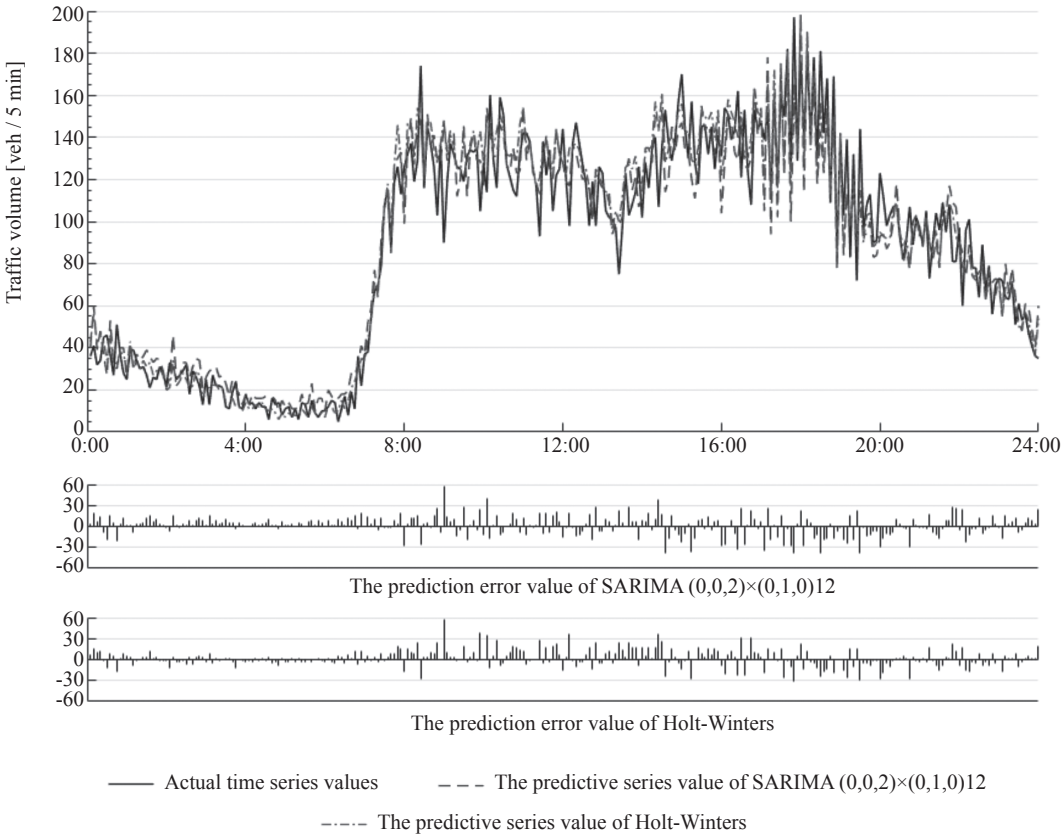


Figure 18 – The fitting performance and prediction performance of the time series analysis method

of t, there are different values of s_t. The prediction result figure of the SARIMA (0,0,2)×(0,1,0) model and Holt-Winters model is drawn and the residual diagram between the predicted value and the forecast target are calculated, as shown *Figure 18*.

4.4 Comparison of the model prediction performance

The forecasting results of the SARIMA (0,0,2)×(0,1,0) model and Holt-Winters model were compared with the forecasting results of the NARX model, as shown in the *Table 3*.

Table 3 – Comparison of predicted performance

Method	Prediction performance	
	MSE	R^2
NARX model with FTDNN	306.87	0.918
NARX model after trend decomposition	211.48	0.938
NARX model after difference operation	218.97	0.957
SARIMA	270.50	0.939
Holt—Winters	360.64	0.930

The forecasting performance of the NARX model with the FTDNN is between the SARIMA of the Holt-Winters. Through trend decomposition or difference operation, the NARX model with FTDNN can better learn the long-term daily change trend of traffic flow, which has a minimum MSE and a maximum determination coefficient.

In addition, time series methods require a series of tests on the model, such as the stable and white noise test, which needs to judge the order of the model based on the experience of the analyst [42]. However, the NARX model with the FTDNN is not influenced by the experience of the analysts, so its forecasting stability is better than that of the time series methods.

5. CONCLUSIONS

The traffic prediction model based on the DNN has problems of complex calculation and large data size demand. This paper uses the NARX model with the FTDNN to predict the traffic flow sequence. It uses the estimation of the number of delay periods in the TDL instead of the estimation of the memory mechanism for the network parameter states in the RNN, which reduces the demand for data size and

avoids the difficulty of calculation. In addition, without complicated gate control, the FTDNN can have long-term predictive ability through simple trend decomposition or difference calculation.

The model was verified by using the monitoring data of the North Zhongshan Road in Guilin City from 6 April to 10 April 2020. The results show that the model can quickly predict the state of the urban traffic flow with a smaller data size. Compared with the time series analysis method, this method does not require the analysts to have rich analysis experience, it is easy to implement and has higher prediction stability and accuracy.

The proposed model has well general applicability and scalability and can be suitable for various data sets and multivariate scenarios. By taking serialised multi-source traffic data, such as traffic flow speed, head-to-head time interval and road occupancy as additional exogenous inputs, the model is able to make more reliable predictions about traffic volume.

ACKNOWLEDGMENTS

This research was funded by the National Natural Science Foundation of China No. 61963011, Guangxi Science and Technology Base and Talent Special Project No. AD19245021, Project of Natural Science Youth Foundation of Guangxi Province No. 2020JJB170049.

李俊卓 博士研究生[1]
电子邮箱：19152202004@mails.guet.edu.cn
（通讯作者）
李文勇 博士[1]
电子邮箱：traffic@guet.edu.cn
廉冠 博士[2]
电子邮箱：lianguan@guet.edu.cn
[1] 信息与通信学院
桂林电子科技大学
中国桂林市七星区金鸡路1号，541004
[2] 建筑与交通工程学院
桂林电子科技大学
中国桂林市七星区金鸡路1号，541004

用于小城市区域交通流量预测的具有外生变量的非线性自回归模型

摘要

数据驱动的预测方法存在计算复杂、可移植性差、需要大量训练数据等问题，限制了数据驱动的方法在小城市的应用。本文提出了一种带有外生变量的非线性自回归模型*(NARX*模型*)*的交通流预测方法，该方法将具有抽头延迟线*(TDL)*结构的聚焦时滞神经网络*(FTDNN)*作为非线性函数。*TDL*结构使*FTDNN*具有短期记忆能力。同时，在数据输入*FTDNN*之前，对交通数据序列进行趋势分解或差分计算，可以使*NARX*模型保持长期预测能力。与常见的非线性模型相比，*FTDNN*具有结构上的优势。它使用了一个简单的*TDL*结构，没有记忆机制和门控结构，可以减少模型的参数，降低数据的规模。通过桂林市数据，对5分钟交通流量进行预测验证，所提出预测方法性能优于*SARIMA*模型和*Holt-Winters*模型。

关键字

智慧交通系统；交通流量预测；时间序列；*NARX*模型；交通数据

REFERENCES

[1] Crivellari A, Beinat E. Forecasting spatially-distributed urban traffic volumes via multi-target LSTM-based neural network regressor. *Mathematics*. 2020;8(12): 2233. doi: 10.3390/math8122233.

[2] Qu W, et al. Short-term intersection traffic flow forecasting. *Sustainability*. 2020;12(19): 8158. doi: 10.3390/su12198158.

[3] Vlahogianni EI, Karlaftis MG, Golias JC. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*. 2014;43: 3-19. doi: 10.1016/j.trc.2014.01.005.

[4] Hamed MM, Almasaeid HR, Said ZMB. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*. 1995;121: 249-254. doi: 10.1061/(Asce)0733-947x(1995)121:3(249).

[5] Van der Voort M, Dougherty M, Watson S. Combining kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies.* 1996;4(5): 307-318. doi: 10.1016/S0968-090x(97)82903-8.

[6] Jomnonkwao S, Uttra S, Ratanavaraha V. Forecasting road traffic deaths in Thailand: Applications of time-series, curve estimation, multiple linear regression, and path analysis models. *Sustainability.* 2020;12(1): 395. doi: 10.3390/su12010395.

[7] Andrysiak T, Saganowski L, Maszewski M. Time series forecasting using Holt-Winters model applied to anomaly detection in network traffic. *Advances in Intelligent Systems and Computing.* 2018;649: 567-576. doi:10.1007/978-3-319-67180-2_55.

[8] Zhang H, et al. A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics. *Applied Intelligence*. 2018;48: 2429-2440. doi: 10.1007/s10489-017-1095-9.

[9] Yao RH, Zhang WS, Zhang LH. Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network. *Journal of Transportation Engineering, Part A: Systems.* 2020;146(8): 04020086. doi: 10.1061/JTEPBS.0000388.

[10] Huang W, et al. Real-time prediction of seasonal heteroscedasticity in vehicular traffic flow series. *IEEE Transactions on Intelligent Transportation Systems*. 2018;19(10): 3170-3180. doi: 10.1109/Tits.2017.2774289.

[11] Hafner CM, Kyriakopoulou D. Exponential-type GARCH models with linear-in-variance risk premium. *Journal of Business and Economic Statistics*. 2019;39(2): 589-603. doi: 10.1080/07350015.2019.1691564.

[12] Bentes SR. Forecasting volatility in gold returns under the GARCH, IGARCH and FIGARCH frameworks: New evidence. *Physica A: Statistical Mechanics and its Applications*. 2015;438: 355-364. doi: 10.1016/j.physa.2015.07.011.

[13] Krithikaivasan B, Zeng Y, Deka K, Medhi D. ARCH-based traffic forecasting and dynamic bandwidth provisioning for periodically measured nonstationary traffic. *IEEE/ACM Transactions on Networking*. 2007;15: 683-696. doi: 10.1109/Tnet.2007.893217.

[14] Anand NC, Scoglio C, Natarajan B. GARCH - Non-linear time series model for traffic modeling and prediction. *NOMS 2008 IEEE Network Operations and Management Symposium, 7-11 Apr. 2008, Salvador, Brazil*. New Jersey: IEEE; 2008. p. 694-697. doi: 10.1109/Noms.2008.4575191.

[15] Ding C, et al. Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility. *IEEE Transactions on Intelligent Transportation Systems*. 2018;19: 1054-1064. doi: 10.1109/Tits.2017.2711046.

[16] Smith BL, Williams BM, Oswald RK. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*. 2002;10(4): 303-321. doi: 10.1016/S0968-090X(02)00009-8.

[17] Bogaerts T, et al. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies*. 2020;112: 62-77. doi: 10.1016/j.trc.2020.01.010.

[18] Mahdy B, et al. A clustering-driven approach to predict the traffic load of mobile networks for the analysis of base stations deployment. *Journal of Sensor and Actuator Networks*. 2020;9(4): 53. doi: 10.3390/jsan9040053.

[19] Mamera M, van Tol JJ, Aghoghovwia MP, Kotze E. Sensitivity and calibration of the FT-IR spectroscopy on concentration of heavy metal ions in river and borehole water sources. *Applied Sciences*. 2020;10(21): 7785. doi: 10.3390/app10217785.

[20] Zhang SQ, Lin KP. Short-term traffic flow forecasting based on data-driven model. *Mathematics*. 2020;8(2): 152. doi: ARTN 15210.3390/math8020152.

[21] Cai PL, et al. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*. 2016;62: 21-34. doi: 10.1016/j.trc.2015.11.002.

[22] Wang ZY, Ji SW, Yu BW. Short-term traffic volume forecasting with asymmetric loss based on enhanced KNN method. *Mathematical Problems in Engineering*. 2019;2019: 4589437. doi: Artn 458943710.1155/2019/4589437.

[23] Evans J, Waterson B, Hamilton A. Forecasting road traffic conditions using a context-based random forest algorithm. *Transportation Planning and Technology*. 2019;42: 554-572. doi: 10.1080/03081060.2019.1622250.

[24] Ou JS, Xia JX, Wu YJ, Rao WM. Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and bias-corrected random forests. *Transport Res Rec*. 2017;2645: 157-167. doi: 10.3141/2645-17.

[25] Szegedy C, et al. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition, 7-12 June 2015, Boston, Massachusetts*. New Jersey: IEEE; 2015. p. 1-9. doi: 10.1109/cvpr.2015.7298594.

[26] Sun T, et al. Bidirectional spatial-temporal network for traffic prediction with multisource data. *Transport Res Rec*. 2020;2674: 78-89. doi: 10.1177/03611981209273930.

[27] Wang JW, Chen RX, He ZC. Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transportation Research Part C: Emerging Technologies*. 2019;100: 372-385. doi: 10.1016/j.trc.2019.02.002.

[28] Cho K, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *2014 Conference on Empirical Methods in Natural Language Processing, 25-29 Oct. 2014, Doha, Qatar*. Strousburg: Association for Computational Linguistics; 2014. p. 1724-1734. doi: 10.3115/v1/D14-1179.

[29] Guo YP, Zhang DL, Ling YX, Chen HH. A joint neural network for session-aware recommendation. *IEEE Access*. 2020;8: 74205-74215. doi: 10.1109/ACCESS.2020.2984287.

[30] Cai L, et al. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Transactions in GIS*. 2020;24(3): 736-755. doi: 10.1111/tgis.12644.

[31] Drakulic D, Andreoli J-M. Structured time series prediction without structural prior. *arXiv*. 2022;02: 03539. doi: 10.48550/arXiv.2202.03539.

[32] Yu B, Yin HT, Zhu ZX. ST-UNet: A spatio-temporal U-network for graph-structured time series modeling. *arXiv*. 2019;03: 05631. doi: 10.48550/arXiv.1903.05631.

[33] Liu MH, et al. Time series is a special sequence: Forecasting with sample convolution and interaction. *arXiv*. 2021;06: 09305. doi: 10.48550/arXiv.2106.09305.

[34] Choi J, Choi H, Hwang J, Park N. Graph neural controlled differential equations for traffic forecasting. *arXiv*. 2021;12: 03558. doi: 10.48550/arXiv.2112.03558.

[35] Liao BB, et al. Deep sequence learning with auxiliary information for traffic prediction. *24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 19-23 Aug. 2018, London, England*. New York: Association for Computing Machinery; 2019. p. 537-546. doi: 10.1145/3219819.3219895.

[36] Hong W-C. Application of seasonal SVR with chaotic immune algorithm in traffic flow forecasting. *Neural Computing and Applications*. 2012;21: 583-593. doi: 10.1007/s00521-010-0456-7.

[37] Cai L, et al. A noise-immune LSTM network for short-term traffic flow forecasting. *Chaos*. 2020;30: 023135. doi: 10.1063/1.5120502.

[38] Duran-Hernandez C, Ledesma-Alonso R, Etcheverry

G. Using autoregressive with exogenous input models to study pulsatile flows. *Applied Sciences*. 2020;10(22): 8228. doi: 10.3390/app10228228.

[39] Tian Y, et al. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water.* 2018;10(11): 1655. doi: 10.3390/w10111655.

[40] Jang B, et al. PIMD signal modeling based on FTDNN. *2019 2nd IEEE International Conference on Information Communication and Signal Processing. 28-30 Sep. 2019, Weihai, China.* New Jersey: IEEE; 2019. p. 1-4. doi: 10.1109/ICICSP48821.2019.8958484.

[41] Pwasong A, Sathasivam S. A new hybrid quadratic regression and cascade forward backpropagation neural network. *Neurocomputing*. 2016;182: 197-209. doi: 10.1016/j.neucom.2015.12.034.

[42] Chen HF, Zhao WX. New method of order estimation for Arma/Armax processes. *SIAM Journal on Control and Optimization.* 2010;48(6): 4157-4176. doi: 10.1137/090768680.