

## Research Article

# Multidimensional Scaling-Based Data Dimension Reduction Method for Application in Short-Term Traffic Flow Prediction for Urban Road Network

Yi Zhao <sup>1</sup>, Satish V. Ukkusuri,<sup>2</sup> and Jian Lu <sup>3</sup>

<sup>1</sup>College of Automobile and Traffic Engineering, Nanjing Forestry University, Longpan Road #159, Nanjing, 210037, China

<sup>2</sup>Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN, 47907, USA

<sup>3</sup>School of Transportation, Southeast University, Si Pai Lou #2, Nanjing, 210096, China

Correspondence should be addressed to Yi Zhao; zhaoyi207@126.com

Received 2 July 2018; Revised 15 October 2018; Accepted 24 October 2018; Published 19 November 2018

Academic Editor: David F. Llorca

Copyright © 2018 Yi Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study develops a multidimensional scaling- (MDS-) based data dimension reduction method. The method is applied to short-term traffic flow prediction in urban road networks. The data dimension reduction method can be divided into three steps. The first is data selection based on qualitative analysis, the second is data grouping using the MDS method, and the last is data dimension reduction based on a correlation coefficient. Backpropagation neural network (BPNN) and multiple linear regression (MLR) models are employed in four kinds of urban traffic environments to test whether the proposed method improves the prediction accuracy of traffic flow. The results show that prediction models using traffic data after dimension reduction outperform the same prediction models using other datasets. The proposed method provides an alternative to existing models for urban traffic prediction.

## 1. Introduction

The success of ITS technology is heavily dependent on the availability of timely and accurate estimates of prevailing and emerging traffic conditions. Hence, there is a strong need for traffic prediction. The method used for prediction should be able to utilize advanced traffic models to analyze data, especially real-time traffic data, from different sources to estimate and predict traffic conditions. Accurate prediction of traffic conditions makes it possible to implement proactive ATMS and ATIS strategies in advance to meet various traffic control, management, and operation objectives.

Urban road systems have a greater need for traffic prediction than freeway and intercity highway systems. Increases in urban populations and vehicle ownership in the last few decades have resulted in severe traffic congestion, with serious economic and social consequences [1]. Therefore, an advanced short-term traffic flow prediction method is required to support the implementation of real-time traffic management measures, such as a dynamic signal control

schemes, traffic information release, and real-time traffic induction.

*1.1. Literature Review.* Short-term traffic flow prediction has been an important subject in the field of intelligent transportation research since the 1970s. Up to 200 kinds of short-term traffic flow prediction models have been developed in the past 50 years.

The first prediction model relies on statistical theory and assumes that traffic flow follows a linear system. Prediction models mainly include the time series model [2], historical average model [3], the Kalman filter [4], and linear regression model [5]. However, traffic flow is very uncertain and nonlinear [6]. Hence, nonlinear theoretical prediction models based on classification theory [7], chaos theory [8], wavelet theory [9], mutation theory [10], and the nonparametric regression method [11] have been developed to address this issue. Over the last decade, studies on traffic prediction models have emphasized the use of the artificial intelligence (AI) model rather than the traditional statistical model. The AI model has

no strict mathematical relationship between input and output data but tries to obtain knowledge through training. The most popular AI models are the support vector machine-based model [12, 13] and the neural network-based model [14, 15]. Certain traffic flow theory-based simulation models have also been used for traffic flow prediction. In the simulation models, a series of mathematical models, such as the motion wave model [16], cellular automata model [17], and three-phase traffic flow theory [18], have been used to capture the dynamic characteristics of traffic.

Among the aforementioned prediction models, each has its own advantages and disadvantages [19]. Therefore, researchers have tried to combine different prediction models to improve the stability and accuracy of a single model. Bates et al. [20] first combined two separate sets of forecasts of airline passenger data to form a composite set of forecasts, which showed the superiority and potential of a combined model. Fuzzy logic [21], Bayesian theory [22], and Grey relation analysis [23] are all commonly employed tools used to combine different models. The idea of combining different models has confirmed the need to keep improving the stability and accuracy of single prediction models.

*1.2. Challenges in Short-Term Traffic Prediction.* In 2014, Vlahogianni [24] pointed out that recent developments in technology and the widespread use of powerful computers and mathematical models give researchers an unprecedented opportunity to expand their horizons and direct work in 10 challenging, yet relatively underresearched, directions. The main motivations of this paper lie in overcoming two of these challenges.

The first challenge involves the fact that most short-term traffic forecasting algorithms were built to function at a freeway, arterial, or corridor level. Short-term traffic forecasting for urban arterials is more complex than that for freeways due to constraints such as signalization. This was pointed out by Kirby [25] 20 years ago, but there has not yet been any change. The limited research on urban traffic focuses on either arterial streets or small local areas. Yin et al. [26] used a fuzzy-neural approach to predict urban traffic flow, but only two adjacent junctions were adopted as examples. Stathopoulos and Karlaftis [27] presented a multivariate state space approach to predict traffic flow on urban arterial streets near a downtown. In almost all existing research on urban traffic prediction, the input data for prediction models only consist of historical data from one or several road sections. Moreover, only arterial streets or suburban roads, rather than central streets, are studied. Generally speaking, the characteristics of urban road traffic flow have not been suitably considered in existing prediction models. Hence, it is likely that the prediction results will be unstable and inaccurate if existing prediction models are directly applied to urban road networks.

The second challenge is that identifying spatial and temporal flow patterns has been an important consideration in short-term traffic forecasting research. The most striking feature of urban traffic is that the source of road flow is very complex and the flow is affected by signal control. In order to obtain accurate prediction results, a large amount of traffic

flow data should be collected and the relationships between these data should be analyzed. Traffic flow from different road sections is relevant in terms of both time and space. In other words, a spatial-temporal correlativity analysis is required in urban traffic prediction. Some attempts at such an analysis have been made, such as Cheng's [28] study of spatiotemporal autocorrelation and Wang's [29] study of bipattern recognition. The principal methods for spatial-temporal correlativity analysis include the spatiotemporal random effects model [30], aggregation method [31], and state space approach [27]. The main purpose of this study is to combine spatial-temporal correlativity analysis and existing prediction models and to test whether the combined method can reach a better prediction result.

*1.3. Motivations and Contributions of the Paper.* Short-term traffic prediction models have been developed for more than half a century. However, their application to urban traffic is limited. With increasing vehicle ownership and network complexity, there are increasing challenges in short-term traffic prediction for urban traffic. In addition, developments in detection and communication techniques can provide more comprehensive information about traffic conditions for prediction. Hence, the aim of the study is urban traffic flow prediction.

Although large amount of data of urban road traffic can be collected, there are two factors that prevent these data from been directly used in urban road traffic prediction models. The first is computation speed. As the study focuses on short-term prediction, large amounts of data may reduce the computation speed of the models, decreasing the value of the short-term predictions. The second is the accuracy of the prediction results. Traffic data from adjacent or homogeneous road sections are highly similar. In other words, large amounts of data contain duplicate and redundant data, which affect the prediction results. Therefore, a method for reducing data dimensions should be applied to the prediction model. To the best of our knowledge, multidimensional scaling (MDS) has not been applied to urban traffic prediction. The advantage MDS-based data dimension reduction is its ability to visualize the level of similarity of a traffic flow data set. Hence, the proposed data dimension reduction method makes it possible to use less data to represent a whole data set, resulting in improved prediction accuracy.

To summarize, the research goals are as follows:

(a) To propose an MDS-based data dimension reduction method that can be used to conduct spatial-temporal correlativity analysis of urban traffic data and divide large amount of traffic flow data into smaller groups according to the level of similarity.

(b) To demonstrate that the proposed method can be combined with existing prediction models. The proposed method will be used to generate a smaller data set to represent all the traffic flow data. The accuracy of the prediction result using the small data set will then be evaluated against those obtained using other data sets.

(c) To illustrate that the proposed method can be combined with different kinds of prediction models and that

the combined methods can be adapted to different traffic environments in urban road networks.

## 2. Data Collection

As previously stated, the main characteristic of urban traffic is its complex direction and signal control management. Compared to that in a single freeway or intercity highway, the traffic direction in urban road networks is more fluid, and signal control results in poor continuity in traffic flow. To collect enough traffic flow data for the study, a virtual network was built in VISSIM to run a simulation, as shown in Figure 1. To reflect the complexity of an urban network, the virtual network is comprised of five main streets and six collector streets. The distance between adjacent streets is randomly set. It is assumed that the structure of the network has no effect on the prediction process; hence, other networks with different structures are not discussed. At each cross-section, lanes with the same driving direction are denoted as one road section. There are 142 road sections in total, and each section has a detector for collecting traffic flow information.

The traffic flow is input into the network through link origins, and the traffic flow subjects to Poisson distribution. In order to test the effectiveness of the proposed method under different environments, four kinds of OD input data are employed: OD<sub>1</sub>, OD<sub>2</sub>, OD<sub>3</sub>, and OD<sub>4</sub>. OD<sub>1</sub>, OD<sub>2</sub>, and OD<sub>3</sub> have the same flow pattern, but OD<sub>4</sub> has a different pattern. The same flow pattern means the same number of traffic origins and destinations, and the same main flow direction. The first three kinds of OD differ in the traffic loading level, with a low loading for OD<sub>1</sub>, medium loading for OD<sub>2</sub>, and high loading for OD<sub>3</sub>, while OD<sub>4</sub> has a medium traffic loading. A fixed signal timing scheme is adopted, and the scheme is optimized using built-in modules in VISSIM.

Based on existing research [32], common prediction intervals for short-term prediction include 5 min, 10 min, 15 min, and 20 min. Considering the signal cycle at intersections is ~2 min, using a prediction interval of 5 min may lead to instability of prediction result. Alternatively, a long prediction interval reduces the significance of the short-term traffic prediction. Hence, an interval of 10 min was used. Traffic data are collected on different days, and data should be collected in the same time period to obtain enough historical data. Hence, random seeds are renewed every hour while running the simulation. Thirty-nine groups of effective traffic flow data can be collected from each road section for simulation.

## 3. Application of MDS-Based Data Dimension Reduction Method

*3.1. Basic Idea of Proposed Data Dimension Reduction Method.* The proposed data dimension reduction method is composed of three steps, as follows.

*Step 1* (selection of historical data based on qualitative analysis). The road section for which predictions are going to be made is denoted as the target section. All the road sections of the network are connected, and the relationship

between them is either strong or weak. It is not necessary to consider the influence of all road sections on the target section. Two principles should be observed when selecting appropriate road sections that have an obvious influence on the target section. Firstly, road sections can be included if vehicles on the section can reach the target section in the upcoming prediction interval. Secondly, road sections where traffic is moving away from the target section are excluded, as there is a low possibility of traffic reaching the target section.

Only the historical data of sections that are supposed to exert an obvious influence on the target section can be used for prediction.

*Step 2* (data grouping using MDS method). There are many road sections that remain that must be considered after Step 1 is completed. Hence, the next step is to divide the sections into several groups, making it easier to reduce their dimensions. Similarity is the best criterion for grouping; however, as each section has 39 groups of traffic flow data, it is difficult to directly identify the level of similarity between the different sections. Hence, the MDS method is used here (a detailed explanation of it will be given in a later section). However, the MDS method only provides a visual representation of the patterns of similarities or differences. The final grouping result should involve grouping of road sections on the same street based on the MDS analysis results.

*Step 3* (data dimension reduction based on correlation coefficients). After the data grouping in Step 2, it can be assumed that road sections in the same group have affect the target section in similar ways. Hence, the data dimensions can be reduced by choosing one representative road section from each group. The data of the selected road sections will serve as the input data for the prediction section to which the BPNN model and MLR model will be applied. The BPNN model requires a low correlation of input data to enhance its robustness, while the MLR model requires a high correlation of input data. Considering that the historical data of the target section will also serve as the input data, the road sections with highest and lowest correlation with the target section in each group will be extracted. The input data for prediction can then be determined according to the requirements of the prediction model. The Pearson product-moment correlation coefficient, denoted as  $\rho_{ij}$ , will be employed to analyze the degree of correlation between the target section and the other road sections.

$$\rho_{ij} = \frac{E(\mathbf{ij}) - E(\mathbf{i})E(\mathbf{j})}{\sqrt{E(\mathbf{i}^2) - E(\mathbf{i})^2}\sqrt{E(\mathbf{j}^2) - E(\mathbf{j})^2}} \quad (1)$$

where  $\rho_{ij}$  is the Pearson product-moment correlation coefficient between the  $i^{\text{th}}$  and the  $j^{\text{th}}$  sections ( $i$  and  $j$  are section numbers) and  $E(\mathbf{i})$ ,  $E(\mathbf{j})$ ,  $E(\mathbf{ij})$ ,  $E(\mathbf{i}^2)$ , and  $E(\mathbf{j}^2)$  are the expectations for  $\mathbf{i}$ ,  $\mathbf{j}$ ,  $\mathbf{ij}$ ,  $\mathbf{i}^2$ ,  $\mathbf{j}^2$ , respectively ( $\mathbf{i}$  and  $\mathbf{j}$  are traffic data for different road sections).

*3.2. MDS Method.* MDS is a means of visualizing the level of similarity between individual cases in a dataset. It is a form

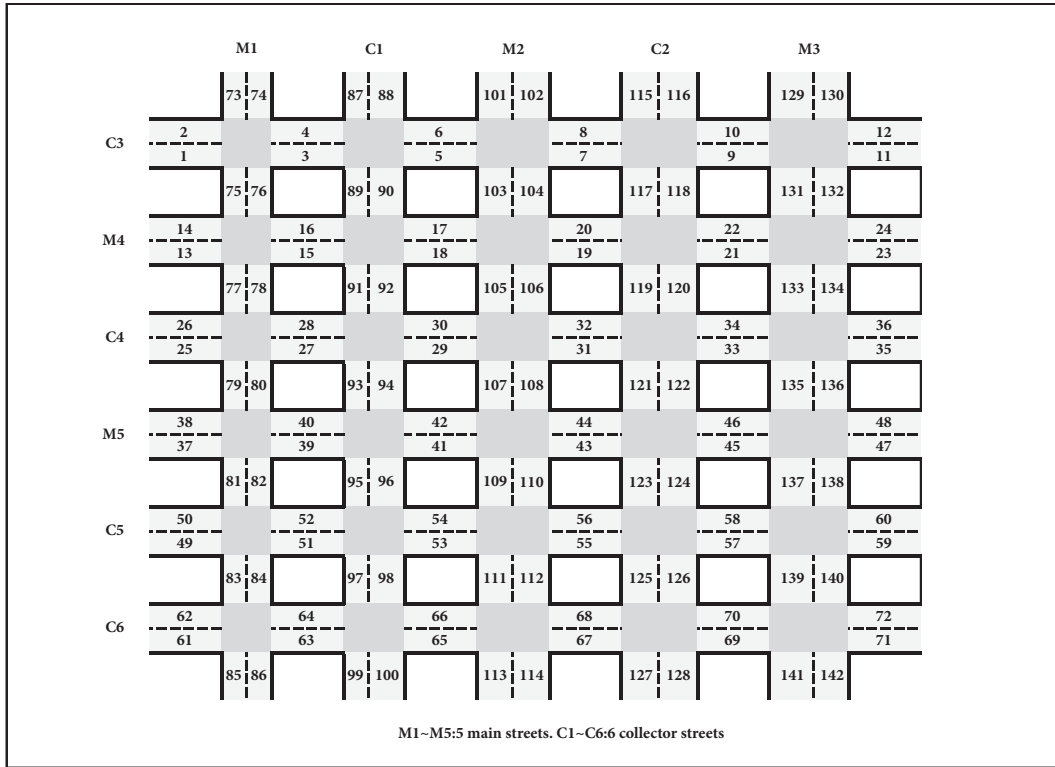


FIGURE 1: Structure of virtual network.

of nonlinear dimensionality reduction. The MDS algorithm places each object in an  $N$ -dimensional space, such that the between-object distances are preserved as well as possible.

If the traffic data set used for prediction is obtained from  $n$  road sections and the number of data values for each section is  $m$ , the data to be analyzed can be expressed in an  $n \times m$  matrix  $\mathbf{X}$ . As traffic flow data are of the nonmetric multidimensional type,  $\mathbf{X}$  should be normalized before the Euclidean distance is calculated. The Euclidean distance will be employed to describe the distance and dissimilarity between objects. The Euclidean distance  $\delta_{ij}$  from  $\mathbf{i}$  to  $\mathbf{j}$  is given by the Pythagorean formula. The Euclidean distances between each road section can be expressed using the distance matrix  $D(\mathbf{X})$ .

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (2)$$

$$\delta_{ij} = \sqrt{\sum_{h=1}^m (x_{ih} - x_{jh})^2} \quad (3)$$

$$D(\mathbf{X}) := \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nn} \end{bmatrix} \quad (4)$$

where  $x_{ij}$  is the  $j^{\text{th}}$  traffic flow data of road section  $i$  ( $i \in [1, n]$  and  $j \in [1, m]$ ) and  $\delta_{ij}$  is the Euclidean distance from  $\mathbf{i}$  to  $\mathbf{j}$  (which are data from the  $i^{\text{th}}$  and  $j^{\text{th}}$  sections, respectively).

The goal of the MDS method is to find an embedding from  $n$  section vectors into  $\mathbb{R}^N$ , an  $N$ -dimensional vector space, such that the distances are preserved. If  $N$  is 2 or 3,  $n$  section vectors should be plotted to visualize the similarities. Supposing  $n$  vectors can be found in  $\mathbb{R}^k$ , whose Euclidean distance matrix  $\widehat{D}(\mathbf{Y})$  is close to  $D(\mathbf{X})$ , then matrix  $\mathbf{Y}$ , which is composed of the  $n$  vectors, is called the fitting chart of  $D(\mathbf{X})$ .

$$\mathbf{Y} = (\mathbf{y}_1; \mathbf{y}_2; \cdots; \mathbf{y}_n) = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \quad (5)$$

$$\widehat{\delta}_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\| = \sqrt{\sum_{s=1}^k (y_{is} - y_{js})^2} \quad (6)$$

$$\widehat{D}(\mathbf{Y}) = \begin{bmatrix} \widehat{\delta}_{11} & \widehat{\delta}_{12} & \cdots & \widehat{\delta}_{1n} \\ \widehat{\delta}_{21} & \widehat{\delta}_{22} & \cdots & \widehat{\delta}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\delta}_{n1} & \widehat{\delta}_{n2} & \cdots & \widehat{\delta}_{nn} \end{bmatrix} \quad (7)$$



where  $\mathbf{Y}$  is a matrix composed of  $n$  vectors in  $\mathbb{R}^k$ ;  $\mathbf{y}_i$  is the  $i^{\text{th}}$  vector in  $\mathbb{R}^k$ , denoted as  $[y_{i1}, y_{i2}, \dots, y_{ik}]$ ; and  $\delta_{ij}$  is the Euclidean distance between  $\mathbf{y}_i$  and  $\mathbf{y}_j$ .

There are various approaches to determine  $\mathbf{Y}$ . Usually, MDS is formulated as an optimization problem, where  $\mathbf{Y}$  is a minimizer of the cost function  $\sigma(\mathbf{Y})$ .

$$\sigma(\mathbf{Y}) = \min \left( \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - \widehat{\delta}_{ij})^2 \right) \quad (8)$$

The most common approach used to determine  $\mathbf{Y}$  is an iterative process commonly referred to as the Shepard–Kruskal algorithm. However, the algorithm is not a major focus of this study. SPASS can be employed to solve the algorithm problem. The goodness-of-fit statistic *Stress*, which is based on the differences between the actual distances and their predicted values, is adopted to express how well  $\mathbf{X}$  is represented by  $\mathbf{Y}$ . The smaller the *Stress* value, the better the fit. The goodness-of-fit is excellent when the *Stress* is less than 0.025, good when it is 0.025–0.050, fair when it is 0.050–0.100, and poor when it is 0.100–0.200.

$$\text{Stress} = \sqrt{\frac{\sum_{i=1}^{I-1} \sum_{j=i+1}^I (\delta_{ij} - \widehat{\delta}_{ij})^2}{\sum_{i=1}^{I-1} \sum_{j=i+1}^I \delta_{ij}^2}} \quad (9)$$

**3.3. Dimension Reduction of Test Data Using Proposed Method.** Road section 104 in the virtual network in Figure 1 is used as a case study. The proposed data dimension reduction method is applied to reduce the dimensions of the collected data. According to the principles in Step 1, only 56 road sections should be considered, including the target section. All 56 road sections are renumbered as shown in Figure 2, where the target section is No. 42.

The historical data for 56 road sections can be expressed in a  $56 \times 39$   $\mathbf{X}$  matrix. According to Step 2 of the method,  $\mathbf{X}$  matrices under four kinds of ODs are analyzed using the MDS method, and the results are shown in Figure 3. The *Stress* values for all the OD types are under 0.050, which indicates a good goodness-of-fit. Based on the similarity analysis, road sections on the same street are grouped with a priority. The 56 road sections can be divided into four groups for each OD. For each kind of OD, there are several road sections that are not within the group, such as section 6 for OD<sub>2</sub> and OD<sub>3</sub>, and section 18 for OD<sub>4</sub>. Apart from these exceptions, the first three ODs have the same grouping result, but OD<sub>4</sub> has a different grouping result. Figure 4 illustrates the grouping results more clearly, where each color represents a group.

Based on the grouping result, the sections with the lowest and highest correlation with the target section are selected by employing the Pearson product-moment correlation coefficient. The results are shown in Table 1.

## 4. Prediction Model Construction and Prediction Result Analysis

The BPNN and MLR models were employed to test the effectiveness of the proposed method for data dimension

reduction. The BPNN model is a widely accepted knowledge discovery model, and its main feature is its ability to learn and adapt to any uncertain system with complex non-linear relationships. The MLR model relies on linear relationships. Ideally, the input data should have a low correlation for the BPNN model and a high correlation for the MLR model. This difference makes it impossible to test the effectiveness of proposed data dimension reduction method on different prediction models.

**4.1. Evaluation Index Selection.** To test the prediction performance of the two prediction methods, two measures, the mean absolute percentage error (MAPE) and the root-mean-square deviation (RMSE), are employed. They are calculated as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^S ((Y(t) - \widehat{Y}(t)) / Y(t))}{S} \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{S} \sum_{t=1}^S \left( \frac{Y(t) - \widehat{Y}(t)}{Y(t)} \right)^2} \quad (11)$$

where  $Y(t)$  is the observed value,  $\widehat{Y}(t)$  is the predicted value, and  $S$  is the number of observed values.

**4.2. Construction of BPNN Model.** The BPNN model employed contains three layers: input layer, hidden layer, and output layer. The other components are set as follows.

**Input data:** Three groups of input data for each OD are used for prediction: group AX, group SX, and group TY. Group AX contains all 56 road sections selected via qualitative analysis. Group SX contains the four sections with the lowest correlation and the target section, i.e., five sections in total. Group TY contains only the target section.

**Temporal correlation:** Based on the concept of the time series prediction model, for all sections, the prediction input data are composed of the historical data of the last three intervals before the prediction period. In other words, if the goal is to predict the traffic flow of the target section in time period  $t$ , the input data for section X include  $X(t-1)$ ,  $X(t-2)$ , and  $X(t-3)$ .

**Data assignment for each section:** There are 39 data values for each section. Eighteen are used for training, eight are used for testing, and ten are used for prediction.

**Number of neurons:** The range of the number of neurons is  $[\sqrt{N(\text{in}) + N(\text{out})}, \sqrt{N(\text{in}) + N(\text{out})} + 10]$ , where  $N(\text{in})$  and  $N(\text{out})$  are the number of nodes of the input and output layer, respectively. The final number of neurons is determined via training and testing.

**Function setting:** The functions that prove to be most effective based on several attempts of testing are selected and assigned to the BPNN model. All the BPNN models have the same functions. The transmission function for the hidden layer is 'tansig' and that for the output layer is 'purelin'. The training function is 'trainlm' and the weight learning function is 'learnrnm'.

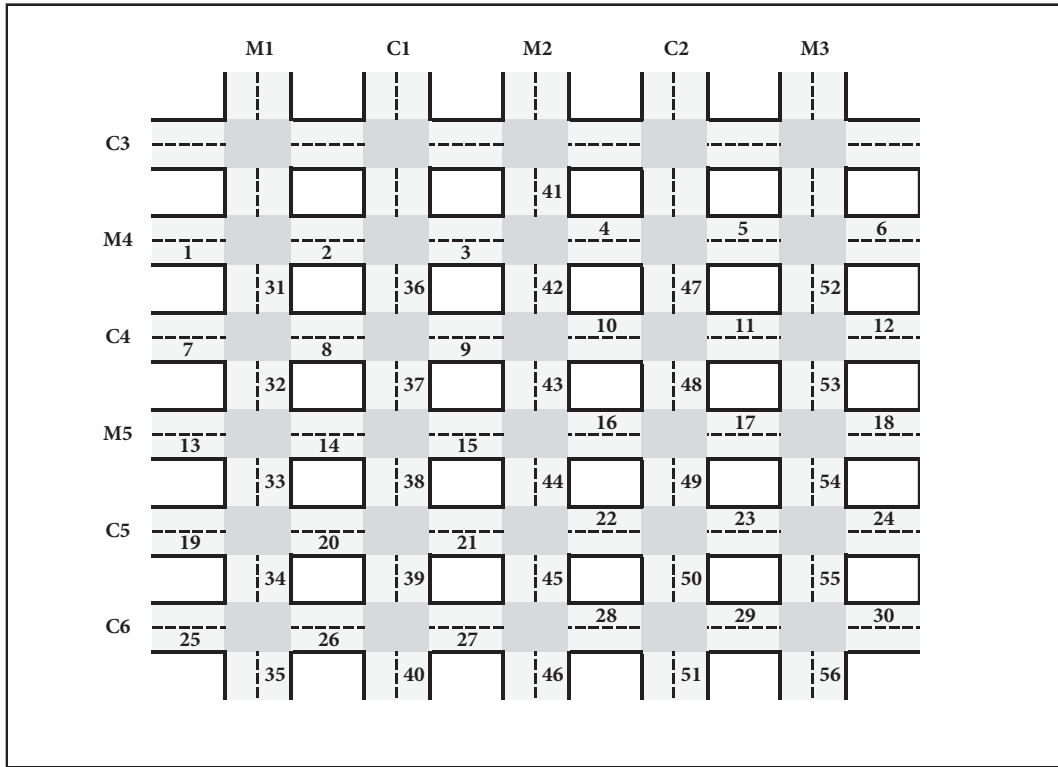


FIGURE 2: Road section selection based on qualitative analysis.

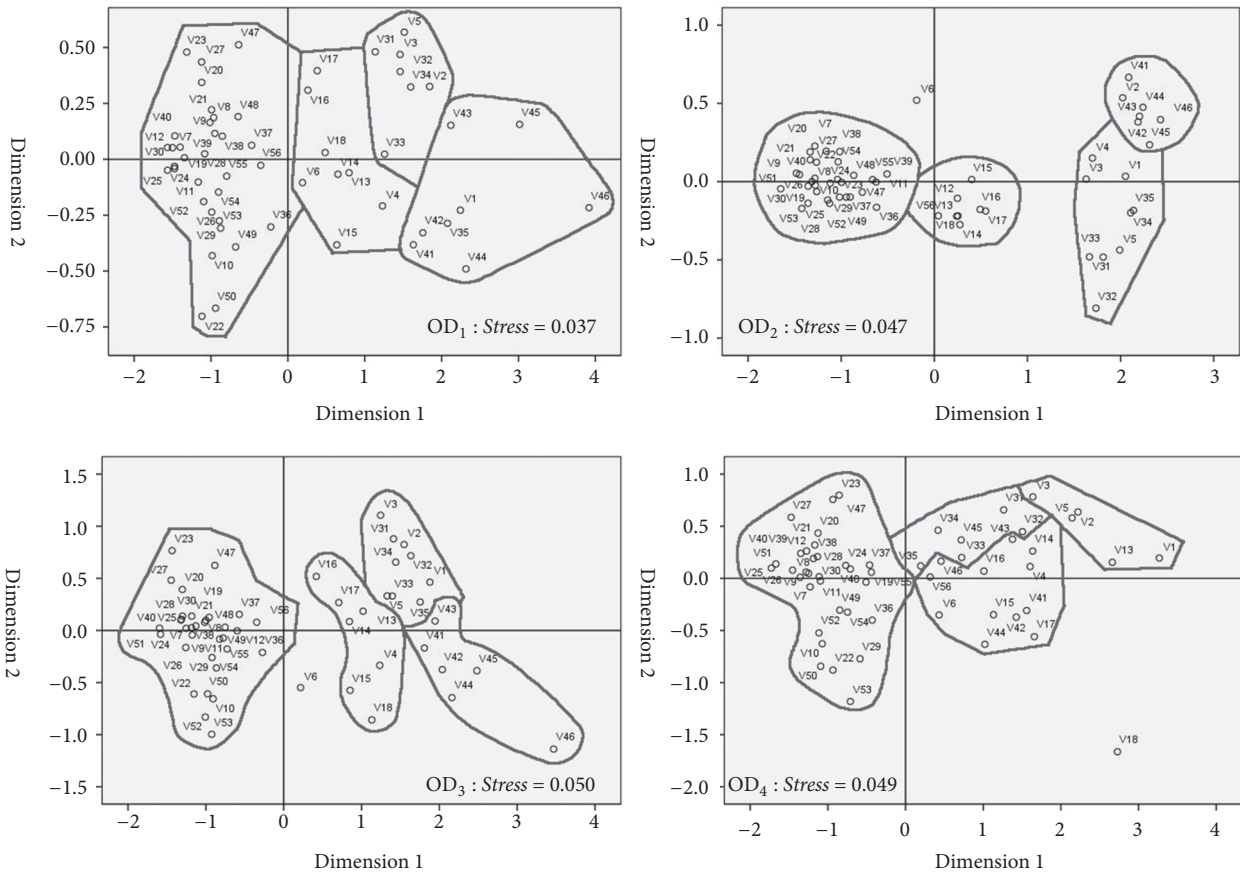


FIGURE 3: Road section selection based on qualitative analysis.

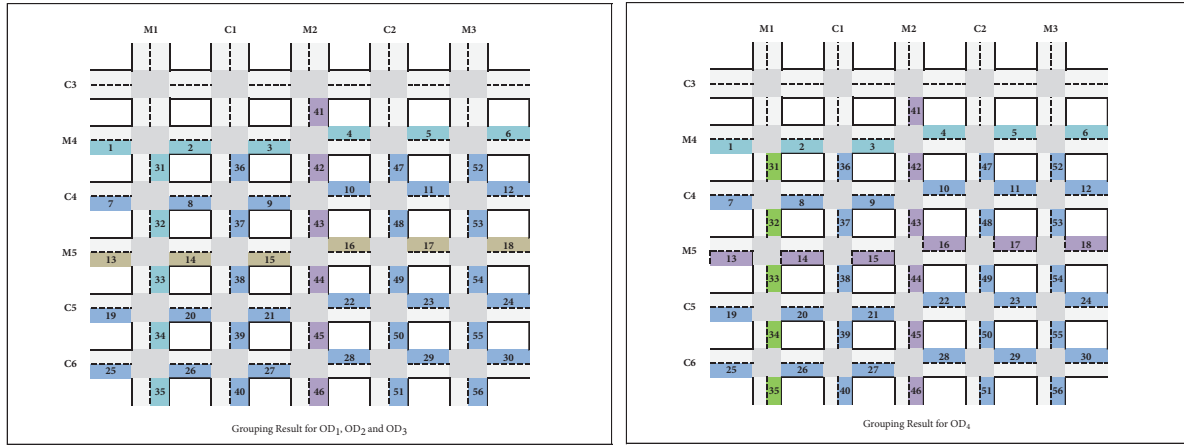


FIGURE 4: Grouping results for different kinds of ODs.

TABLE 1: Correlation analysis based on grouping result.

Input data type	Sections with lowest correlation in each group	Sections with highest correlation in each group
OD <sub>1</sub>	27\32\12\45	39\2\15\42
OD <sub>2</sub>	46\32\13\53	43\39\15\4
OD <sub>3</sub>	48\35\17\45	26\4\15\42
OD <sub>4</sub>	5\31\14\56	4\32\39\42

**4.3. Construction of MLR Model.** The dependent variable in the analysis is the traffic of the target section. The independent variables are divided into three groups for each OD, i.e., group AX, group SX, and group TY. ANOVA is used to select sections with a significant difference, owing to the large error that occurs when 56 road sections are used for MLR analysis. The ANOVA results are shown in Table 2. Group AX contains all the sections after implementing ANOVA. Group SX contains the four sections with the highest correlation and the target section. Group TY contains only the target section.

As with the BPNN model, temporal correlation is also considered in the MLR model. For each section that is used as an independent variable, the input data are composed of the historical data of the last three time intervals before the prediction period. Only the data from the last 10 intervals of the target section are used for prediction, while the others are used for fitting.

**4.4. Assessment of Prediction Results.** Because the initial weights in the BPNN model are random, the prediction result varies when the model is rerun. The prediction process is repeated 30 times using the BPNN model. The 10 best results are selected, and the mean values of the evaluation indices are calculated, as shown in Table 3. The prediction result obtained using the MLR model is shown in Table 4.

In most existing research, only the historical data of the target section are used in the prediction model. Therefore, group TY is treated as the representative of regular prediction method. Group AX uses the data of all the related sections for

prediction, without considering the spatial-temporal correlativity of the sections. Group SX uses the selected sections by employing the proposed data dimension reduction method, which takes the spatial-temporal correlativity of the sections into consideration. The improvements in the results for methods using group SX and group AX are compared against the method that uses group TY. The degree of improvement is denoted as P in Tables 3 and 4.

The following conclusions can be made from the comparison of the prediction results:

- According to the MAPE index, the prediction accuracy of the BPNN model is more than 85%, while that of the MLR model is ~80%. The BPNN model is better than the MLR model when a single prediction method is employed.
- For the two kinds of OD flow patterns, both the method using group AX and that using group SX result in greater improvements in the prediction accuracy than the method using group TY.
- For the three different levels of traffic loading, both the method using group AX and that using group SX result in greater improvements in the prediction accuracy than the method using group TY.
- When using the BPNN model for prediction, the prediction accuracy of the method using group AX and that of the method using group SX are very similar. However, because of the difference in the size

TABLE 2: ANOVA results for 56 road sections.

OD type	ANOVA result	Amount
OD <sub>1</sub>	6, 15, 20, 24, 35, 37, 41, 45, 46	9
OD <sub>2</sub>	1, 4, 5, 11, 16, 18, 22, 41, 51	9
OD <sub>3</sub>	2, 6, 12, 15, 40, 41, 45, 46	8
OD <sub>4</sub>	1, 5, 13, 16, 35, 40, 41, 44, 47	9

TABLE 3: Prediction result using BPNN model for different input data.

Input Type		OD <sub>1</sub>		OD <sub>2</sub>		OD <sub>3</sub>		OD <sub>4</sub>	
		Mean	P	Mean	P	Mean	P	Mean	P
Group TY	MAPE	0.118		0.113		0.130		0.109	
	RMSE	0.142		0.145		0.162		0.143	
Group SX	MAPE	0.100	15.4%	0.092	18.9%	0.110	15.1%	0.093	14.6%
	RMSE	0.122	14.1%	0.116	19.7%	0.135	16.6%	0.131	8.9%
Group AX	MAPE	0.102	13.8%	0.089	22.8%	0.115	11.2%	0.092	15.6%
	RMSE	0.132	7.1%	0.114	22.6%	0.150	7.3%	0.127	11.6%

TABLE 4: Prediction result using MLR model for different input data.

Input Type		OD <sub>1</sub>		OD <sub>2</sub>		OD <sub>3</sub>		OD <sub>4</sub>	
		Result	P	Result	P	Result	P	Result	P
Group TY	MAPE	0.213		0.089		0.194		0.161	
	RMSE	0.244		0.106		0.252		0.191	
Group SX	MAPE	0.138	35.3%	0.077	14.4%	0.152	22.0%	0.108	33.3%
	RMSE	0.151	38.0%	0.098	7.1%	0.173	31.6%	0.134	29.7%
Group AX	MAPE	0.175	17.8%	0.110	-22.8%	0.179	7.7%	0.151	6.5%
	RMSE	0.203	16.9%	0.142	-34.2%	0.204	19.3%	0.173	9.3%

of the input data, the method using group AX requires more time for prediction than that using group SX.

- (e) When using the MLR model for prediction, the prediction accuracy of the method using group SX is better than that of the method using group AX.
- (f) Based on the above, it can be concluded that the proposed data dimension reduction method can improve the prediction accuracy for different conditions, using different prediction models and conditions under different traffic environments.

**4.5. Application in Urban Traffic Management.** The most common problem in urban traffic management is traffic jams. When a traffic jam occurs, administrators have a limited amount of time to stop it from spreading and causing other problems. More time for management can be obtained if traffic jams can be prejudged. In addition, if the traffic flow on roads that lead to an intersection can be predicted, the occurrence of congestion at the intersection can be determined in advance. The proposed short-term traffic flow prediction method can play an important role in predicting traffic flow towards intersections and identifying possible traffic jams. Then, based on the prediction results, the management measures can be appropriately adjusted.

The proposed method enables the application of existing prediction models to the urban traffic environment, as it overcomes the complexity and instability limitations of existing prediction models. Large amounts of historical traffic data from different road sections can be collected for each road leading to an intersection. Then, the data dimension can be reduced using the proposed MDS-based data dimension reduction method and a smaller dataset can be obtained to represent the larger dataset. By using a small dataset as the input data in existing prediction models, such as the BPNN and MLR models, a better prediction result is likely to be reached. The result is the basis for predicting traffic jams. The proposed MDS-based data dimension reduction method can also be combined with other prediction models.

## 5. Concluding Remarks

This study presents an MDS-based data dimension reduction method and applies it to short-term traffic flow prediction in urban road networks. A virtual network is built to generate traffic flow data, and the proposed method is used to reduce the data dimensions. The test results show significant advantages of applying the proposed method to short-term traffic flow prediction over the regular prediction method, which only considers the target section's historical data.



Additionally, the proposed prediction method is better at considering all the related sections, in terms of both efficiency and accuracy.

The key contributions of the study are as follows:

First, an MDS-based data dimension reduction method was developed and successfully applied to short-term traffic flow prediction in an urban road network.

Second, this research has demonstrated that the proposed data dimension reduction method can be combined with existing prediction models and adapted to different traffic environments, regardless of the flow patterns or traffic loading levels. Furthermore, the proposed method is compatible with different types of prediction models, both nonlinear and linear.

Third, an alternative method to apply various prediction methods to urban traffic prediction is presented. In urban traffic systems, the network is more complex and analysis of the spatial-temporal correlativity is necessary.

During the course of this study, several elements for future research were identified, which are as follows. (1) The results reported here are from a hypothetical network. Hence, results from real-world implementation would make the conclusions stronger. (2) Although different ODs are tested in the research, the same network is used. Networks with different structures should be analyzed to prove the effectiveness of the method. (3) The proposed method can also be applied to other prediction models, such as combined prediction models or models that can predict more sections at one time.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

An earlier version of this paper has been presented at "Transportation Research Board 97th Annual Meeting" as a presentation.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] I. Thomson and A. Bull, "Urban traffic congestion: its economic and social causes and consequences," *Cepal Review*, pp. 105–116, 2002.
- [2] M. Zare and M. Shirazi, "Arima Model for Network Traffic Prediction and Anomaly Detection," in *Proceedings of the International Symposium on Information Technology*, vol. 4, pp. 1–6, 2008.
- [3] M. Billy and P. K. D. D. Williams, "Urban Freeway Traffic Flow Prediction," *Transportation Research Record*, vol. 1644, pp. 132–141, 2014.
- [4] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.
- [5] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of Local Linear Regression Model for Short-Term Traffic Forecasting," *Transportation Research Record*, vol. 1836, pp. 143–150, 2014.
- [6] W. Hui, "Vehicle speed prediction based on clustering and adaptive neuro-fuzzy inference," *Computer Engineering and Applications*, vol. 44, no. 6, pp. 240–242, 2008.
- [7] A. S. Nair, J. C. Liu, L. Rilett, and S. Gupta, "Non-Linear Analysis of Traffic Flow," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pp. 681–685, Oakland, CA, USA, 2001.
- [8] S. Y. Chen and W. Wang, "Chaos Forecasting for Traffic Flow based on Lyapunov Exponents," *China Civil Engineering Journal*, vol. 37, no. 9, pp. 96–99, 2004.
- [9] G. He, S. F. Ma, and Y. Li, "Study on the Short-term Forecasting for Traffic Flow Based on Wavelet Analysis," *Systems Engineering—Theory & Practice*, vol. 22, no. 9, pp. 101–106, 2002.
- [10] X.-J. Cheng, J. Liu, and M.-S. Ma, "Algorithm of short-term traffic flow forecasting using fractal theory," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 4, pp. 106–110, 2010.
- [11] G. A. Davis and N. L. Nihan, "Nonparametric Regression and Short-Term Freeway Traffic Forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, NY, USA, 1995.
- [13] Q. H. Xu and R. Yang, "Traffic Flow Prediction Using Support Vector Machine Based Method," *Journal of Highway and Transportation Research and Development*, vol. 22, no. 12, pp. 131–134, 2005.
- [14] J. Hua and A. Faghri, "Applications of Artificial Neural Networks to Intelligent Vehicle-Highway Systems," *Transportation Research Record*, vol. 1453, pp. 83–90, 1994.
- [15] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction models—a comparison of neural network and non-parametric regression approaches," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 1706–1709, San Antonio, TX, USA, 1994.
- [16] G. F. Newell, "A simplified theory of kinematic waves in highway traffic, Part I: general theory," *Transportation Research Part B: Methodological*, vol. 27, no. 4, pp. 281–287, 1993.
- [17] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [18] B. S. Kerner, *The physics of traffic*, Springer-Verlag, 2004.
- [19] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: overview of objectives and methods," *Transport Reviews*, vol. 24, no. 5, pp. 533–557, 2004.
- [20] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Operational Research Quarterly*, vol. 20, no. 4, pp. 451–468, 1969.
- [21] J. E. Jarrett and J. S. Plouffe, "The fuzzy logic method for simpler forecasting," *International Journal of Engineering Business Management*, vol. 3, no. 3, pp. 18–52, 2011.
- [22] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 79–94, 2014.
- [23] H. Y. Chen, J. B. Zhao, and C. L. Liu, "Properties of combination forecasting model based on degree of grey incidence," *Journal of*

- Southeast University (Natural Science Edition)*, vol. 34, no. 1, pp. 130–134, 2004.
- [24] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Short-term traffic forecasting: where we are and where we’re going,” *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.
- [25] H. R. Kirby, S. M. Watson, and M. S. Dougherty, “Should we use neural networks or statistical models for short-term motorway traffic forecasting?” *International Journal of Forecasting*, vol. 13, no. 1, pp. 43–50, 1997.
- [26] H. B. Yin, S. C. Wong, J. Xu, and C. K. Wong, “Urban traffic flow prediction using a fuzzy-neural approach,” *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 2, pp. 85–98, 2002.
- [27] A. Stathopoulos and M. G. Karlaftis, “A multivariate state space approach for urban traffic flow modeling and prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 2, pp. 121–135, 2003.
- [28] T. Cheng, J. Haworth, and J. Wang, “Spatio-temporal autocorrelation of road network data,” *Journal of Geographical Systems*, vol. 14, no. 4, pp. 389–413, 2012.
- [29] J. Wang, P. Shang, and X. Zhao, “A new traffic speed forecasting method based on bi-pattern recognition,” *Fluctuation and Noise Letters*, vol. 10, no. 1, pp. 59–75, 2011.
- [30] Y. Wu, F. Chen, C. Lu, and S. Yang, “Urban Traffic Flow Prediction Using a Spatio- Temporal Random Effects Model,” *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 282–293, 2015.
- [31] R. Giacomini and C. W. Granger, “Aggregation of space-time processes,” *Journal of Econometrics*, vol. 118, no. 1-2, pp. 7–26, 2004.
- [32] S. Oh, Y.-J. Byon, K. Jang, and H. Yeo, “Short-term travel-time prediction on highway: a review of the data-driven approach,” *Transport Reviews*, vol. 35, no. 1, pp. 4–32, 2015.

