

Research Article

Dynamic Spatiotemporal Causality Analysis for Network Traffic Flow Based on Transfer Entropy and Sliding Window Approach

Senyan Yang ^{1,2}, Lianju Ning ², Xilong Cai ¹ and Mingyu Liu ³

¹School of Modern Post (School of Automation), Beijing University of Posts and Telecommunications, Beijing 100876, China

²School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China

³Department of Civil Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Senyan Yang; yangsenyan@bupt.edu.cn and Lianju Ning; ninglj007@126.com

Received 15 October 2020; Revised 29 November 2020; Accepted 22 December 2020; Published 18 January 2021

Academic Editor: Ronghui Zhang

Copyright © 2021 Senyan Yang 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.

With the rapid development of sensor and communication technologies, a large amount of spatiotemporal traffic data has been accumulated, presenting the characteristics of big data. The potential information and regularity of traffic state evolution can be extracted from the huge traffic flow time series data and applied to intelligent transportation systems. This study proposes a dynamic spatiotemporal causality modeling approach to analyze traffic causal relationships for the large-scale road network. Transfer entropy algorithm is utilized to detect the spatiotemporal causality of network traffic states based on the extensive traffic time series data, which could measure the amount and direction of information transmission. A combination of Gaussian kernel density estimation and sliding window approach is proposed to calculate the transfer entropy and construct dynamic spatiotemporal causality graphs based on the causality significance test. The indexes of affected coefficient, influence coefficient, input degree, and output degree are defined to evaluate the causal interaction of traffic states among different road segments and identify the critical roads and potential bottlenecks of the existing road network. Experimental results based on real-world traffic sensor data indicate that the structures of traffic causality graphs are time-varying; the traffic cause-effect interaction among different road segments during the peak time is more significant than that during the nonpeak time; and the critical road segments can be identified, which are mainly located at the intersections of arterial roads, undertaking the convergence and dispersion of large traffic flows.

1. Introduction

The rapid development of sensing and communication technologies in transportation promotes the accumulation of huge multisource spatiotemporal traffic data, which is collected by loop detectors, vehicle GPS, and mobile phones [1], presenting the characteristics of traffic big data. The valuable knowledge can be extracted from the huge observational spatiotemporal traffic data, which could be applied in the data-driven intelligent transportation systems (ITS) [2].

The diversity, uncertainty, and huge volume of traffic big data bring greater challenges to ITS. Potential traffic evolution characteristics can be extracted from extensive historical data through data mining techniques, such as

correlation analysis and association rule mining. Some studies have integrated the extractive traffic correlations or association rules into the traffic flow prediction models to improve the prediction accuracy [3–5]. However, the directional causal interaction could not be captured by these data mining techniques. For the events A and B , association relationship can be extracted based on the statistical rule “ A ’s existing means B ’s existing”, but it is still not clear whether the occurrence of A leads to the occurrence of B , or other factors make A and B appear simultaneously [6]. Similarly, correlation analysis could determine whether A is related to B but could not verify the cause-effect relationship between them.

The causal relationships among different objects, events, or variables are widespread in the natural and social sciences

[7]. Causality can be detected from the observational nonstationary time series data, and the strengths also need to be qualified [8, 9]. Many causal models have been proposed based on the probability theory, graph theory, Bayesian networks, etc. [10]. Spatiotemporal traffic causality can be discovered from the complex network traffic state. For example, if a traffic jam event happens in a certain road segment r_a at time t , the traffic state of the upstream adjacent road segment r_b may be congested at the next time $t + 1$ due to the shock waves of traffic flow, and then r_a can be regarded as the cause segment of r_b .

In addition to the dissemination of traffic waves in the physical network, the dissemination of traffic information can also lead to spatiotemporal causality. For example, if a serious traffic accident happens on the road segment r'_a , the traffic management department would release the traffic accident information on the navigation platform and guide the drivers to avoid the congested segment r'_a . Then, more drivers prefer to choose another route, which leads to serious traffic congestion on the segment r'_b . Although r'_b is far away from r'_a in space, real-time traffic information sharing strengthens spatiotemporal causality between them.

However, few studies have focused on the spatiotemporal traffic causality modeling so far. Previous studies on traffic causality either define the causal relationship based on the prior knowledge such as the chronological order of traffic jam events, which could not quantify potential causal strength [11, 12], or use the data-driven Bayesian network method, which is not suitable for large-scale road networks because of sophisticated parameter estimations [13–15]. The traffic causal dependencies are usually combined with the network traffic evaluation, traffic outliers detection, or traffic state prediction [11, 16].

Motivated by the lack of research on traffic causality and the challenge of traffic big data analysis, this study develops a dynamics spatiotemporal traffic causality research framework to capture the underlying causal knowledge of network traffic state for the decision making in ITS. The main scientific contributions include the following: (i) transfer entropy algorithm is first utilized to extract the traffic cause-effect interactions from the extensive traffic time series data; (ii) spatiotemporal traffic causality could be calculated dynamically and efficiently through the sliding window technique; (iii) dynamic spatiotemporal causality graphs could reveal the traffic causal structures and identify the critical road segments and potential bottlenecks of the existing road network; (iv) the proposed approach can be applied in the real-time traffic management system and combined with practical applications, such as the network traffic station evaluation and prediction.

The remainder of the paper is organized as follows. Section 2 provides a summary of the literature regarding traffic causality analysis. Section 3 depicts the study framework and transfer entropy method. Section 4 describes spatiotemporal causality modeling for network traffic flow. Section 5 presents the computational experiments based on real-world traffic sensor data. The conclusions are summarized in Section 6.

2. Literature Review

Spatiotemporal data mining approaches have been widely applied to traffic congestion propagation and prediction. Inoue et al. [17] proposed a frequent pattern mining method to extract traffic congestion patterns from traffic sensor data and demonstrated the process of traffic congestion generation, diffusion, and dissipation from a data-driven perspective. Chawla et al. [18] proposed an optimized mining algorithm framework for inferring the root cause of anomalies from large taxis GPS data. Xiong et al. [19] developed a propagation graph approach to predict traffic congestion patterns in the near future based on the large real-world vehicle trajectory data.

Previous studies on spatiotemporal traffic causality mainly adopt simple prior knowledge methods and define the causal relationship according to the chronological order of traffic jams or abnormal conditions. Liu et al. [11] extracted the spatiotemporal causal interactions among the traffic outliers through constructing the outlier causality trees according to temporal order and spatial contiguity of detected outliers. Kapoor et al. [12] studied the causality of the traffic congestion at road intersections and how the congestion propagates from one point in the road network to all directions, and predicted possible propagation patterns.

In addition, dynamic Bayesian network method is utilized for spatiotemporal traffic causality modeling. Chu et al. [13] proposed a time-varying dynamic Bayesian network for traffic causality modeling, studied the region macro structure based on vehicle trajectory data, and extracted the road junction dependency structure from sensor data. Queen and Albers [14] proposed a multivariate dynamic Bayesian network model to capture the conditional independence and causality of traffic flow time series. The causality between variables in the Bayesian network and the lagging causality between time series in the dynamic Bayesian network could be identified by setting external interventions. Nguyen et al. [15] identified traffic congestion propagation patterns from spatiotemporal traffic data and estimated congestion propagation probabilities by dynamic Bayesian network. Potential causal relationship structure can be extracted based on the dynamic Bayesian network modeling, but a great number of computing resources would be consumed to estimate the parameters, especially for the large-scale road network.

The above spatiotemporal traffic causality modeling methods are either too simple to fully capture the potential spatiotemporal causality and nonlinearity characteristic or too computationally complex, which is not suitable for large-scale road network traffic state analysis.

Recently, the information-theoretic casual approaches have attracted much attention, which can measure and quantify the causality [7]. Granger causality test is an effective method to identify potential causality for time series data [20–22]. The principle of Granger causality is described as follows: if variable Y can be better predicted by using the historical values of both X and Y , rather than only the

historical values of Y , X can be regarded as the Granger cause variable for Y . Li et al. [16] developed a Granger causality-based causal dependence mining approach for traffic predictions and revealed the relationship between the road network structure and the correlation among traffic flow time series through causal dependence graph. A linear or nonlinear relationship between variables needs to be assumed for Granger causality analysis.

Transfer entropy is a relatively emerging method in information theory, which can evaluate the causality correlation because of its asymmetry. Compared to the Granger causality method, transfer entropy does not need to assume the form of the causal relationship between variables, which is suitable for the long time series analysis of nonlinear systems, and has been widely applied in neuroscience [23], chemistry [24], finance [25], industrial processes [26], and so on. Transfer entropy can measure both the direction and quantity of information transmission, which is suitable for the nonlinear spatiotemporal causality modeling of network traffic flow.

3. Methods

3.1. Study Framework. This study proposes a dynamic spatiotemporal causality modeling framework, as shown in Figure 1. Firstly, transfer entropy is adopted to detect the spatiotemporal traffic causality from the huge traffic time series data for the large-scale road network. The combined Gaussian kernel density estimation and sliding window approach is proposed to calculate the transfer entropy matrix, which can denote the dynamic nonlinear causal relationship of the traffic states among different road segments. Secondly, the causal correlation coefficient matrix is calculated based on the transfer entropy matrix. The affected coefficient and influence coefficient are defined to identify potential bottlenecks and critical road segments in the road network for different time periods. Finally, the dynamic spatiotemporal causality graphs are established based on the causality significance test. The input and output degrees are proposed to evaluate the spatiotemporal causality of network traffic states.

The advantages of this research framework are as follows: (i) the nonlinear cause-effect interaction relationship could be extracted from huge traffic flow time series data, contributing to the deeper insights into complex network traffic state; (ii) the causal orientations and strength can be determined based on the asymmetry of transfer entropy; (iii) the sliding window approach can guarantee the computational efficiency of transfer entropy for the large-scale road network; (iv) the dynamic spatiotemporal traffic causality graphs can reveal the time-varying traffic causal structures.

3.2. Basic Concepts for Information Entropy. The basic concepts for information entropy proposed by Shannon [27] are briefly explained. x_i ($i = 1, 2, \dots, n$) is set as the states of discrete variable X . The information $I(x_i)$ for variable x_i is defined as (1), and $I(x_i) \geq 0$. The larger the probability of x_i , the smaller the information $I(x_i)$, and the smaller the

uncertainty of x_i . Conversely, the smaller the probability of x_i , the larger the information $I(x_i)$, and the larger uncertainty of x_i . When the probability of x_i is 1, the information $I(x_i)$ is 0.

$$I(x_i) = -\log p(x_i). \quad (1)$$

Information entropy $H(x)$ is defined as the mathematical expectation of information $I(x_i)$ for X , which is calculated as (2). Information entropy $H(x)$ can reflect the average uncertainty and information amount of X . The larger the information entropy $H(x)$, the larger the amount of information for the variable X ; the smaller the information entropy $H(x)$, the smaller the amount of information for the variable X .

$$H(x) = E[I(x_i)] = -\sum_{i=1}^n p(x_i) \log p(x_i). \quad (2)$$

Mutual information $M(X, Y)$ is proposed to quantify the common information between two correlative variables, as shown in (3). The larger the mutual information $M(X, Y)$, the stronger the correlation between the variables X and Y . However, mutual information is symmetrical and cannot represent the directionality of information transfer.

$$M(X, Y) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \quad (3)$$

3.3. Transfer Entropy. In 2000, Schreiber [28] proposed transfer entropy to measure the amount of information transfer and asymmetric interaction between systems based on information theory. Transfer entropy $TE_{Y \rightarrow X}$ for two discrete systems X, Y is calculated as (4), where x_i and y_i represent the state value of X and Y at the time i , respectively. $x_i^{(k)} = [x_i, x_{i-1}, \dots, x_{i-k+1}]$ denotes the values of X during the time period $[i, i-k+1]$, representing the state of the past k moments. $y_i^{(l)} = [y_i, y_{i-1}, \dots, y_{i-l+1}]$ denotes the values of Y during the time period $[i, i-l+1]$, representing the state of the past l moments. $p(x_{i+1}, x_i^{(k)}, y_i^{(l)})$ is the joint probability. $p(x_{i+1}|x_i^{(k)}, y_i^{(l)})$ and $p(x_{i+1}|x_i^{(k)})$ are the conditional probabilities. Transfer entropy has been regarded as an indicator of causality due to its asymmetric nature.

$$TE_{Y \rightarrow X} = \sum_{x_{i+1}, x_i^{(k)}, y_i^{(l)}} p(x_{i+1}, x_i^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+1}|x_i^{(k)}, y_i^{(l)})}{p(x_{i+1}|x_i^{(k)})}. \quad (4)$$

Transfer entropy represents the difference of the information entropy for x_{i+1} between the situations with both $x_i^{(k)}$ and $y_i^{(l)}$ known and with only $x_i^{(k)}$ known. If the state of X at a certain time is completely determined by its historical state, not connected with Y , the transfer entropy is 0. The parameters k, l are the sampling period of X, Y for the calculation of transfer entropy. With the increasing of k, l , more computational resources and data are required to estimate the joint probability density. Considering the time delay impacts of information propagation, Bauer et al. [29]

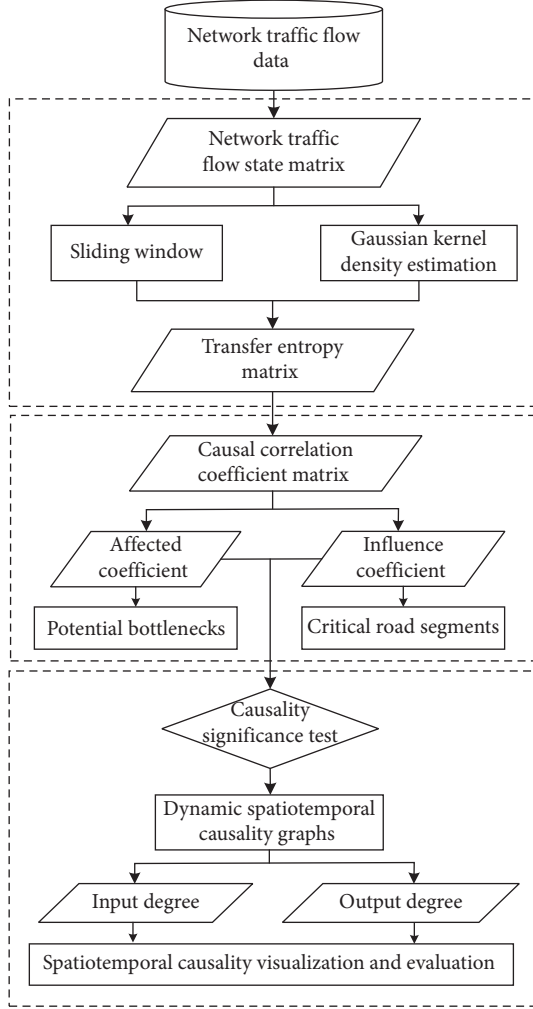


FIGURE 1: Research framework.

modified the calculation of transfer entropy by embedding the prediction horizon h , as shown in

$$TE_{Y \rightarrow X} = \sum_{x_{i+1}, x_{i+h-1}^k, x_i^l} p(x_{i+h}, x_{i+h-1}^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+h} | x_i^{(k)}, y_i^{(l)})}{p(x_{i+h} | x_i^{(k)})}. \quad (5)$$

Schreiber's definition for transfer entropy is based on the assumption that the system should be approximated by the stationary Markov process, and the current system state only depends on the past states within the limited time periods. If the assumption of Markov process is not satisfied, transfer entropy may not be suitable to measure the causal relationship for this system [24]. The evolution of traffic flow has been regarded as being eligible for the nature of "Markov process" [30]. Therefore, transfer entropy is suitable for spatiotemporal causality modeling of the network traffic flow.

As the prediction horizon varies, the reference $x_i^{(k)}$ would change. This study adopts the modified transfer entropy proposed by Shu and Zhao [24] as (6). $x_i^{(k)}$ is

replaced by $x_{i+h-1}^{(k)}$, which is more suitable for estimating the transfer entropy considering the time delay.

$$TE_{Y \rightarrow X} = \sum_{x_{i+1}, x_{i+h-1}^k, x_i^l} p(x_{i+h}, x_{i+h-1}^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+h} | x_{i+h-1}^{(k)}, y_i^{(l)})}{p(x_{i+h} | x_{i+h-1}^{(k)})}. \quad (6)$$

3.4. Calculation Method for Transfer Entropy. The joint probability density in (5) is estimated by the kernel density estimation function. The probability density is estimated in (7). $K(x - x_i)$ is the value of kernel function at x_i . The probability density $\hat{p}(x)$ is the average of the kernel function value over a certain range. Kernel density estimation method does not depend on the prior distribution of the data and is also applicable for the non-Gaussian distribution data. Gaussian kernel function is used to estimate the probability density of traffic state for each road segment as (8). The parameter θ denotes the width of the window for the calculation of the kernel function values.

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N K(x - x_i), \quad (7)$$

$$K(x - x_i) = \frac{1}{\sqrt{2\pi}\theta} e^{-((x - x_i)^2 / 2\theta^2)}. \quad (8)$$

The joint probability density $\hat{p}(x, y)$ for x, y is shown in (9), and the corresponding joint Gaussian kernel function is calculated in (10).

$$\hat{p}(x, y) = \frac{1}{N} \sum_{i=1}^N K(x - x_i, y - y_i), \quad (9)$$

$$\begin{aligned} K(x - x_i, y - y_i) &= K(x - x_i)K(y - y_i) \\ &= \frac{1}{2\pi\theta^2} e^{-((x - x_i)^2 + (y - y_i)^2 / 2\theta^2)}. \end{aligned} \quad (10)$$

The interactions among different variables vary with time. Sliding window technique is utilized to dynamically calculate the transfer entropy between variables along the timeline, which can reduce the sampling data size and improve the efficiency of causal relationship analysis. The sliding window is described by the window width w and the moving step length l . The original state space is divided into n continuous subspaces S_i . Each window consists of w time intervals. The moving step length l is smaller than w . The window width w should not be too small; otherwise, the small sampling data within the window would affect the accuracy of kernel density estimation. The moving step length l should not be too large; otherwise, it could not reflect the variation of the information transmission process timely. As shown in Figure 2, for the time series data with a length of L time intervals, the sliding window starts to move with a fixed step length l . For each window, the probability density is calculated, and then the transfer entropy vector can be obtained with the dimensions of $p = L - w + 1$ in

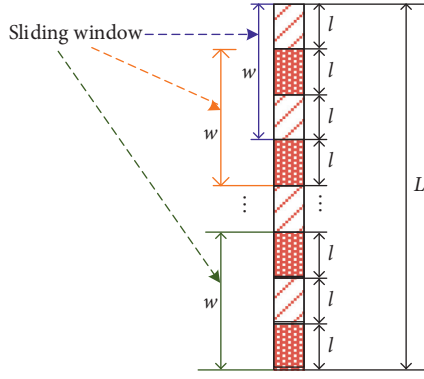


FIGURE 2: Schematic diagram of sliding window.

time, which can reflect the time-varying transferred information.

The road network consisting of m road segments is utilized to illustrate the calculation of the transfer entropy matrix. The traffic state of each road segment can be treated as a variable, and thus the total number of variables for this traffic system is m . Transfer entropy between any two variables is calculated, and then two-dimensional transfer entropy matrix $T_{m \times m}$ is obtained for each sliding window, as shown in (11). Considering the directionality of transferring entropy, $T_{m \times m}$ is not a symmetric matrix, and for each pair of road segments, $t_{ij} \neq t_{ji}$. The elements on the diagonal are 0.

$$T_{m \times m} = \begin{bmatrix} 0 & t_{12} & \cdots & t_{1m} \\ t_{21} & 0 & \cdots & t_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & 0 \end{bmatrix}. \quad (11)$$

After eliminating diagonal zero elements, the transfer entropy matrix $T_{m \times m}$ for each sliding window is transformed into a row vector $te_i = [t_{12}^i, t_{13}^i, \dots, t_{1m}^i, t_{21}^i, t_{23}^i, \dots, t_{m(m-1)}^i]$. After the window slides from the beginning to the end of the time series, p transfer entropy vectors can be obtained. Then, all the transfer entropy vectors are integrated together to form a transfer entropy matrix with the dimensions of $p \times (m^2 - m)$ for road network traffic state, which can represent the transferred information among different road segments. The sliding window can improve the computational efficiency of transfer entropy by using limited data within each window, which makes it suitable for the real-time traffic management system.

$$te = \begin{bmatrix} t_{12}^1, t_{13}^1, \dots, t_{1m}^1, t_{21}^1, t_{23}^1, \dots, t_{m(m-1)}^1 \\ t_{12}^2, t_{13}^2, \dots, t_{1m}^2, t_{21}^2, t_{23}^2, \dots, t_{m(m-1)}^2 \\ \vdots \\ t_{12}^p, t_{13}^p, \dots, t_{1m}^p, t_{21}^p, t_{23}^p, \dots, t_{m(m-1)}^p \end{bmatrix}. \quad (12)$$

4. Spatiotemporal Causality Modeling for Network Traffic Flow

4.1. Causality Significance Test Method. For the causal inference, the causal relationship is assumed to exist between any two different traffic state variables X and Y , and then the causality needs to be verified based on the observed data. The cause variables and effect variables can be discriminated through the causality analysis. Transfer entropy is asymmetry, because the amount of information transferred in opposite directions is different. To characterize the direction and strength of the causality, the causal correlation coefficient $\rho_{X,Y}$ is defined to model the causal strength [29]. The causal orientation and strength are measured by the difference between $TE_{Y \rightarrow X}$ and $TE_{X \rightarrow Y}$, as shown in

$$\rho_{X,Y} = \|TE_{Y \rightarrow X} - TE_{X \rightarrow Y}\|. \quad (13)$$

When the transfer entropy $TE_{Y \rightarrow X}$ in the direction of $Y \rightarrow X$ is larger than $TE_{X \rightarrow Y}$ in the direction of $X \rightarrow Y$, Y is the cause variable of X , and the direction of information transfer is $Y \rightarrow X$. Conversely, when $TE_{Y \rightarrow X}$ is smaller than $TE_{X \rightarrow Y}$, X is the cause variable of Y , and the direction of information transfer is $X \rightarrow Y$. When $TE_{Y \rightarrow X}$ is equal to $TE_{X \rightarrow Y}$, $\rho_{X,Y} = 0$, and there is no causality between X and Y . Due to data noise or interference, the causal correlation coefficient $\rho_{X,Y}$ is generally not equal to 0. If $\rho_{X,Y}$ is too small, the causal correlation is not significant. Then, it is necessary to set a causal correlation coefficient threshold to define the significant causality, namely, the causality significance test. If $\rho_{X,Y}$ exceeds the threshold, the causality between X and Y is significant.

Causality significance test can be regarded as a hypothesis testing problem to determine the causal relationships. The null hypothesis is that if $\rho_{X,Y}$ is small, there is no causality between X and Y . If $\rho_{X,Y}$ is large enough, the null hypothesis is rejected, and there exists causal relationship between X and Y . Bauer et al. [29] used the Monte Carlo method to reconstruct a new alternative time series for causality significance test, which should satisfy the following assumptions: the causality between X and Y is completely destroyed, and the statistical distribution of X and Y remain unchanged. This study utilizes the method proposed by Duan et al. [31] to disrupt the original time series for X and Y with the L time intervals. The new time series X' and Y' are constructed, as shown in (14). The statistical distribution of the reconstructed time series X' and Y' is consistent with the original time series X and Y .

$$\begin{cases} X' = [x_i, x_{i+1}, \dots, x_{i+M-1}], \\ Y' = [y_j, y_{j+1}, \dots, y_{j+M-1}], \end{cases} \quad (14)$$

where M is the length of X' and Y' ; i, j are randomly selected from $\{1, 2, \dots, L - M + 1\}$; and $\|i - j\| \geq e$, where e is much larger than the prediction horizon h to make sure that there

is almost no causality correlation between X' and Y' . Then, the causal correlation coefficient $\rho = [\rho_1, \rho_2, \rho_3, \dots, \rho_N]$ are calculated for X' and Y' . The causality significance test is carried out according to (15), where the μ_ρ and σ_ρ are the mean and standard deviation of $\rho_1, \rho_2, \rho_3, \dots, \rho_N$. The significance threshold ε is set as $\mu_\rho + 3\sigma_\rho$. When the causality coefficient $\rho_{X,Y}$ is smaller than ε , there is no causal relationship between X and Y ; when the causality coefficient is larger than ε , there is significant causality between X and Y .

$$\rho_{X,Y} - \mu_\rho \geq 3\sigma_\rho. \quad (15)$$

4.2. Network Traffic State Evaluation. To evaluate the network traffic state, the influence coefficient and affected coefficient are defined for each road segment. For road segment i , the influence coefficient $R_{\text{out}}(i)$ denotes the sum of the transfer entropy from road segment i to the other road segments in the network as (16), which can describe the impacts of road segment i on the other road segments. In the same way, the affected coefficient $R_{\text{in}}(i)$ denotes the sum of the transfer entropy from the other road segments to road segment i as (17), which can describe the impacts of other road segments on the target road segment i .

A data-driven method for identifying the potential bottlenecks and critical road segments is proposed from the perspective of spatiotemporal causality analysis. The road segments with large $R_{\text{in}}(i)$ can be regarded as the potential bottleneck segments, which are most likely to be affected by the traffic state of other road segments in the network. The road segments with large $R_{\text{out}}(i)$ can be regarded as the critical road segments, which are most likely to affect the traffic state of other road segments.

$$R_{\text{out}}(i) = \sum_j \text{TE}_{i \rightarrow j}, \quad (16)$$

$$R_{\text{in}}(i) = \sum_j \text{TE}_{j \rightarrow i}. \quad (17)$$

4.3. Dynamic Spatiotemporal Traffic Causality Graphs. The time-varying network traffic state leads to dynamic spatiotemporal causality graphs. Due to the asymmetry of transfer entropy, the spatiotemporal traffic causality graphs are directed graphs, representing the dynamic causal structure for traffic state variables, as shown in Figure 3. The road network consists of n road segments that are denoted by the nodes r_1, r_2, \dots, r_n . The directed edges demonstrate the significant causal relationship between the traffic states of two road segments. The structures of spatiotemporal causality graphs at different time slices are quite different. For example, r_i is the cause segment of r_j at time $t - 1$, while there is no link between them at t and $t + 1$. The causal strength between any two road segments is defined as the weight of directed edges. The causality coefficient ρ_{r_i, r_j} for road segment r_i and r_j during $[t - 1, t]$ is calculated based on the transfer entropy $\text{TE}_{r_i \rightarrow r_j}$ and $\text{TE}_{r_j \rightarrow r_i}$ according to (13). The causality matrix W^t for transfer entropy at time t is

calculated as (18). The weight of the directed edges with strong causality correlation is set to 1, and the weight is set to 0 with no obvious causality correlation. Then, the redundant connections can be removed for the construction of causality graphs. The calculation process for the causality matrix is shown in Figure 4.

$$W^t = \begin{cases} w_{i,j}^t = 1, & \text{TE}_{i \rightarrow j}^t > \text{TE}_{j \rightarrow i}^t, \rho_{i,j}^t \geq \varepsilon, \\ w_{i,j}^t = 0, & \text{others.} \end{cases} \quad (18)$$

Based on the dynamic spatiotemporal causality graphs, four indicators are proposed to evaluate the impacts of any road segment in the road network from the perspective of causal dependence. The input degree $D_{\text{in}}^t(i)$ is defined as (19), denoting the impacts of the traffic states of the other $n - 1$ road segments on that of r_i at time t . The output degree $D_{\text{out}}^t(i)$ is defined as (20), denoting the influence of traffic state for road segment r_i on the other $n - 1$ road segments. The sum of input degrees $\text{Sum}D_{\text{in}}(i)$ and the sum of output degrees $\text{Sum}D_{\text{out}}(i)$ are defined to quantify the cause-effect relationship between the road segment r_i and the other road segments during the time period T , as shown in (21) and (22).

$$D_{\text{in}}^t(i) = \sum_j^{n-1} w_{ji}^t, \quad (19)$$

$$D_{\text{out}}^t(i) = \sum_j^{n-1} w_{ij}^t, \quad (20)$$

$$\text{Sum}D_{\text{in}}(i) = \sum_t^T \sum_j^{n-1} w_{ji}^t, \quad (21)$$

$$\text{Sum}D_{\text{out}}(i) = \sum_t^T \sum_j^{n-1} w_{ij}^t. \quad (22)$$

5. Experiments and Discussion

5.1. Data Description. The expressway network of Shanghai in China is utilized to test the proposed causality analysis method. Traffic flow data is collected by the loop detectors distributed on the network, as shown in Figure 5. The detailed data pre-processing process has been illustrated in our previous study, including data aggregation, missing data estimation, and data noise reduction [32]. Traffic speed data on May 6, 2014, for 432 road segments in the expressway network is used to verify the proposed spatiotemporal traffic causality approach. The time interval of traffic flow data is 10 min.

5.2. Sensitivity Analysis for Transfer Entropy. In this study, considering the limited computational resource, the parameters k and l are set as 1. The maximum time delay is set as 40 minutes, and the parameter h of the prediction horizon is set as [1, 4]. The transfer entropy between any two segments can be calculated for different directions. For

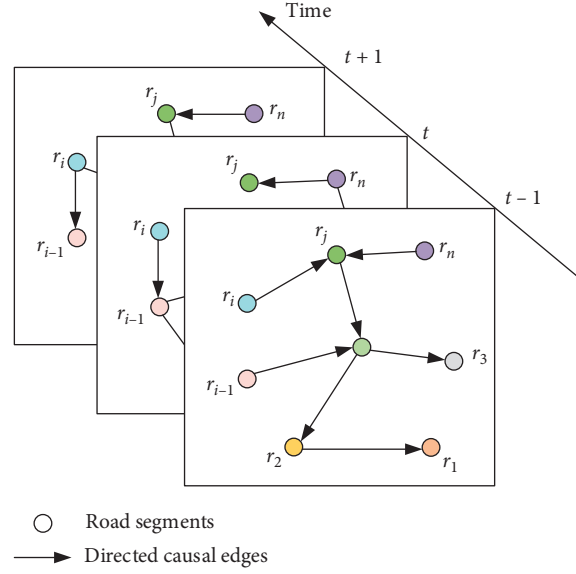


FIGURE 3: Dynamic spatiotemporal traffic causality graphs.

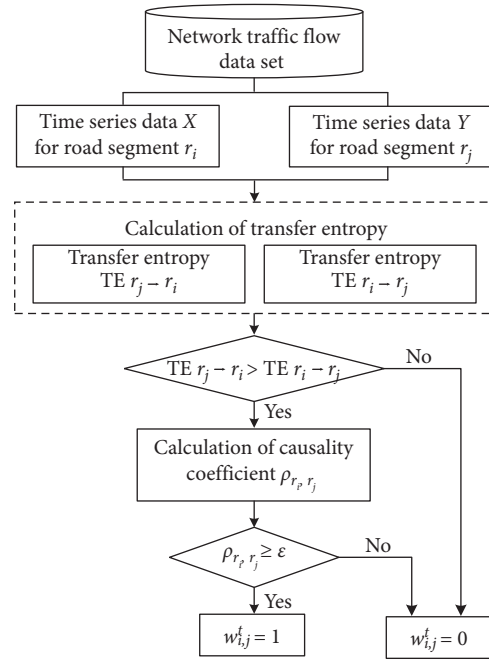


FIGURE 4: Calculation process for causality matrix.

example, r_{220} is the upstream road segment of r_{223} , and the variation of $TE_{r_{220} \rightarrow r_{223}}$ and $TE_{r_{223} \rightarrow r_{220}}$ is shown in Figure 6.

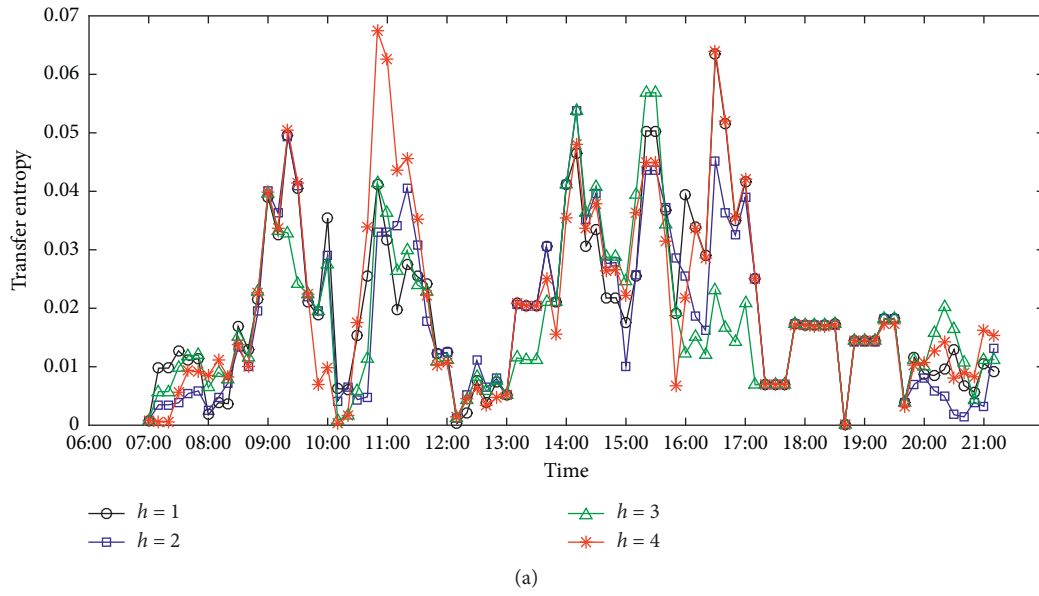
Transfer entropy in different directions for r_{220} and r_{223} is shown in Table 1. For example, $TE_{r_{223} \rightarrow r_{220}}$ is larger than $TE_{r_{220} \rightarrow r_{223}}$ at the evening peak time 18:30, while $TE_{r_{223} \rightarrow r_{220}}$ is larger than $TE_{r_{220} \rightarrow r_{223}}$ at the nonpeak time 13:30. Therefore, the downstream segment r_{223} has a more obvious impact on the upstream segment r_{220} in the evening peak congestion periods, while the upstream segment r_{220} has a more obvious impact on the downstream segment r_{223} in the nonpeak period.

In addition to the parameters k, l, h , the causality coefficient threshold ϵ also needs to be set. When constructing the spatiotemporal causal graph, as the causality threshold ϵ increases, the number of remaining directed edges with significant causality is reduced. The transfer entropy and causality coefficients between any two segments are calculated based on the traffic speed data set for the entire road network on May 6, 2014. The mean value μ_ρ and standard deviation σ_ρ of transfer entropy for the disturbed sequence X' and Y' are 0.0151 and 0.0116, respectively, and then the threshold ϵ is set according to $\epsilon = \mu_\rho + 3\sigma_\rho = 0.05$. Different settings for ϵ would affect the structures of the

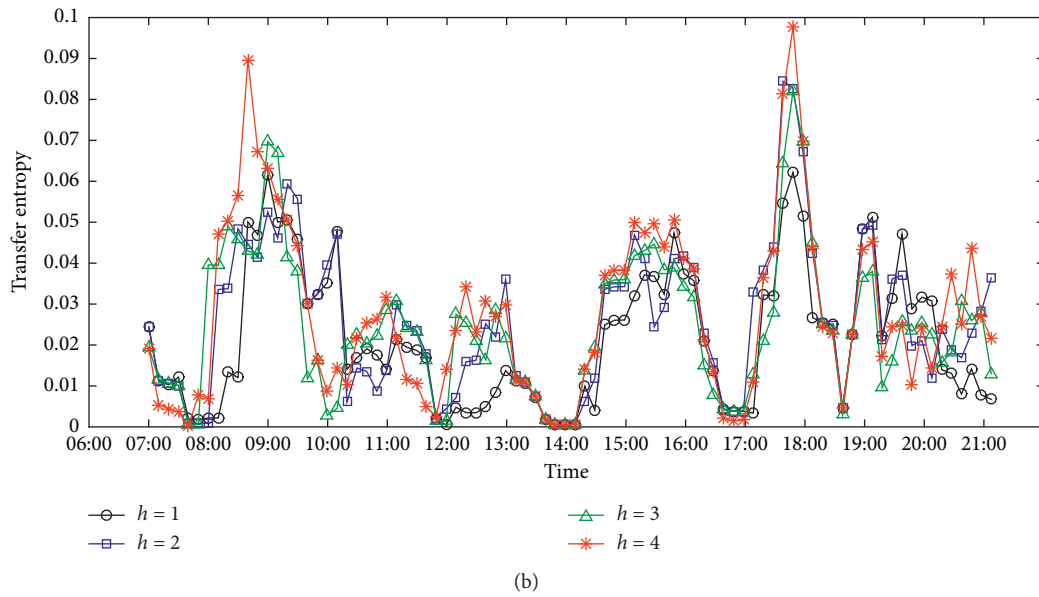


• Loop detectors
— Expressway segments

FIGURE 5: Expressway network of Shanghai [32].



(a)



(b)

FIGURE 6: Variation of transfer entropy in different directions: (a) $TE_{r_{220} \rightarrow r_{223}}$; (b) $TE_{r_{223} \rightarrow r_{220}}$.

TABLE 1: Transfer entropy values in different time periods.

	Morning peak time 8:30		Nonpeak time 13:30		Evening peak time 18:30	
	$TE_{r_{220} \rightarrow r_{223}}$	$TE_{r_{223} \rightarrow r_{220}}$	$TE_{r_{220} \rightarrow r_{223}}$	$TE_{r_{223} \rightarrow r_{220}}$	$TE_{r_{220} \rightarrow r_{223}}$	$TE_{r_{223} \rightarrow r_{220}}$
$h = 1$	0.0169	0.0120	0.0205	0.00722	0.0173	0.0250
$h = 2$	0.0133	0.0485	0.0205	0.00749	0.0172	0.0247
$h = 3$	0.0152	0.0459	0.0112	0.00749	0.0173	0.0234
$h = 4$	0.0137	0.0567	0.0205	0.00751	0.0172	0.0228

spatiotemporal causal graph. The number of directed edges for traffic causal graphs with different ε is shown in Table 2.

Considering computational complexity, the three key parameters are set as $k = 1$, $l = 1$, $h = 1$. The time-varying transfer entropy in different directions, $TE_{r_{220} \rightarrow r_{223}}$ and $TE_{r_{223} \rightarrow r_{220}}$, is shown in Figure 7. The variation of the traffic causality coefficient $\rho_{r_{220}, r_{223}}$ is shown in Figure 8. The transfer entropy between the adjacent road segments changes greatly over time. The direction and amount of information transmission in different time periods are quite diverse. For example, the direction of information transmission between 10:00 and 18:00 is mainly $r_{223} \rightarrow r_{220}$, while the direction of information transmission between 18:00 and 21:00 is mainly $r_{220} \rightarrow r_{223}$. The distribution of causality coefficients for network traffic flow is concentrated, as shown in Figure 9.

5.3. Spatiotemporal Traffic Causality Analysis. Transfer entropy values among different road segments fluctuate greatly with time and space, reflecting the variation of information transfer. Each road segment may be a potential cause or effect segment. The influence coefficient $R_{out}(I)$ and affected coefficient $R_{in}(I)$ for all the road segments in the morning peak time, evening peak time, and nonpeak time are shown in Figure 10. Each road segment in the network is represented by one bubble. The bubble size denotes the average speed of the road segments. The distribution of the bubbles is determined by both $R_{out}(I)$ and $R_{in}(I)$, which can quantitatively describe the casual interaction of network traffic flow state. The bubbles for the morning peak time are the most scattered. Table 3 lists the potential bottleneck segments with the largest $R_{in}(I)$ and the critical road segments with the largest $R_{out}(I)$. The spatial locations of potential bottleneck segments and critical road segments in the road network are shown in Figure 11. The critical road segments are mainly distributed in the central and western regions of Shanghai City.

Not all traffic causality correlations are significant. In this section, the threshold ε is set as 0.05 for the causality significance test. The sum of input degrees $SumD_{in}(i)$ and output degrees $SumD_{out}(i)$ are calculated for the morning peak period (7:00–10:00), nonpeak period (13:00–16:00), and evening peak period (17:00–20:00). The distributions of $SumD_{in}(i)$ and $SumD_{out}(i)$ are shown in Figures 12 and 13, respectively. The output degree distribution is more concentrated than the input degree distribution. On the whole, $SumD_{in}(i)$ for the morning peak hours is larger than that of the evening peak hours. The distribution of $SumD_{in}(i)$ for the nonpeak hours is scattered with smaller

values, and thus the road segments are more likely to be affected by the traffic state of other road segments for traffic congestion. Similarly, $SumD_{out}(i)$ for the morning peak hours is larger than that for the evening peak and nonpeak hours. The road segments are more likely to affect the traffic state of other road segments. Generally, the causal interaction among different road segments during the peak time periods is more significant than that of nonpeak time periods.

5.4. Spatiotemporal Traffic Causality Visualization and Evaluation. The spatiotemporal traffic causality graphs for the Shanghai expressway network at 8:30 in the morning peak time, 13:30 in the nonpeak time, and 18:30 in the evening peak time are visualized in Figure 14, which can represent the spatial distribution characteristics of the input degree and output degree. Moreover, the circles represent the output degrees of the expressway segments. The larger the circle, the larger the output degree, demonstrating more significant impacts on the traffic states of other road segments in the network. The directed causal edges describe the causal-effect relationship between two segments. The head arrow for each directed edge connects to the affected road segment, while the end of each directed edge connects to the cause road segment. The density of arrows around each circle can present the impacts of other road segments on the target segment.

The distributions of circles and directed causal edges are diverse for different periods. The output degrees of road segments in the peak time are generally larger than those of nonpeak time. In the morning peak time, the road segments with larger output degrees are mainly located at the north-south expressway, east-west expressway, inner ring road, and southern middle ring road. In the evening peak hours, the road segments with a larger output degree are mainly located at the eastern inner ring expressway. Generally, these critical road segments are mainly located at the intersections of arterial roads, which undertake the convergence and dispersion of large traffic flows, and have more significant impacts on the traffic state of other roads.

The circles distributed in the eastern outer ring area are smaller than other areas, and the directed causal edges are also sparser than other areas, especially during the nonpeak period, indicating that the road segments in this area have no significant cause-effect interaction with other road segments. The main reason for this phenomenon is that the eastern region develops relatively late with weaker network accessibility, which is less likely to be affected by the traffic states

TABLE 2: Total number of directed edges for traffic causality graph with different ε .

ε	Total number of directed edges
0.01	5081635
0.05	184323
0.06	39318
0.09	33

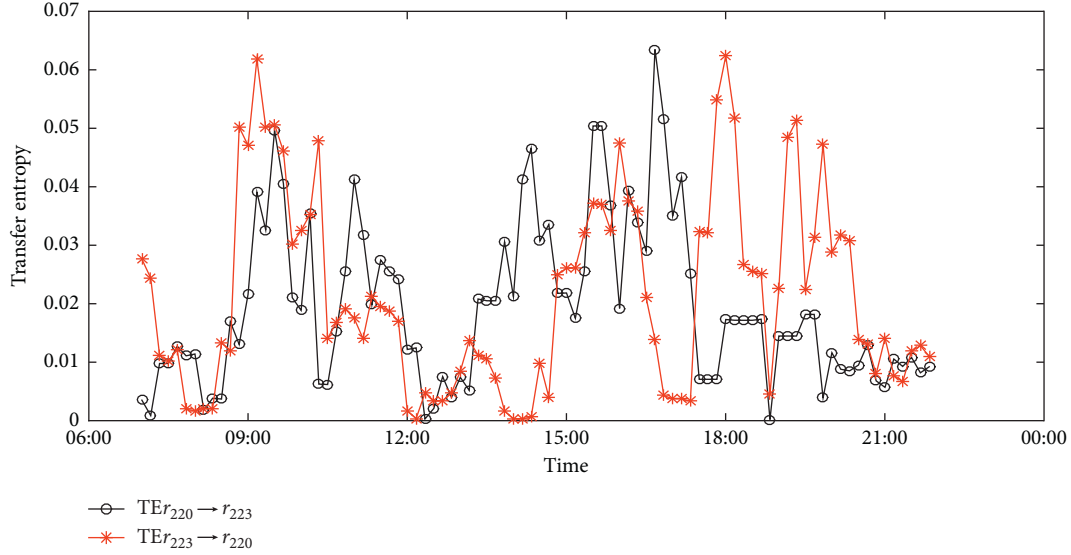


FIGURE 7: Variation of transfer entropy in different directions.

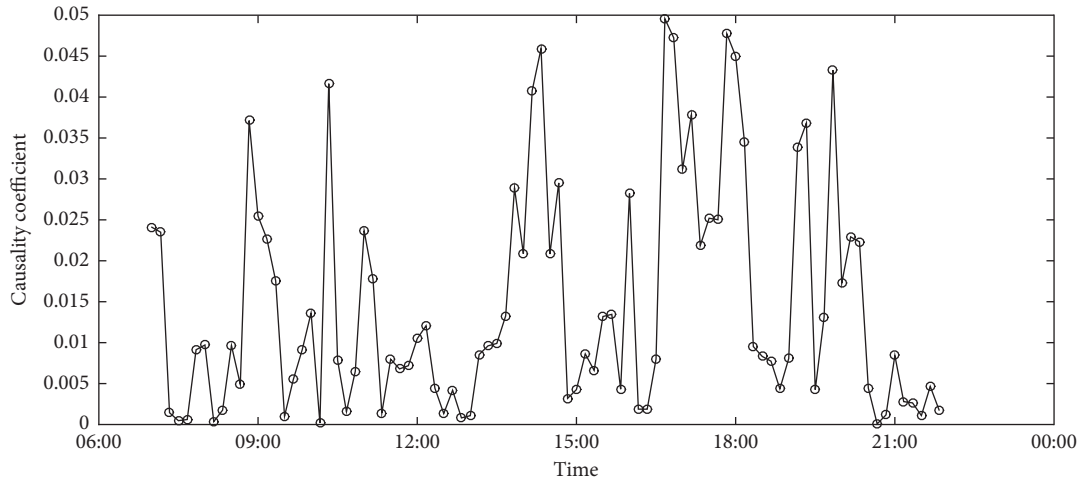


FIGURE 8: Variation of traffic causality coefficient.

in other regions. In addition, the traffic condition in the eastern region is normally smooth, having less impact on the traffic flow in other regions.

The spatial locations of critical road segments $r_{302}, r_{391}, r_{286}, r_{181}$ with large output degree at 8:30 am are visualized in Figure 15. The specific spatial structures of $r_{302}, r_{391}, r_{286}, r_{181}$ are shown in Figure 16. The yellow segments are the critical road segments, and the green segments are the entrance and exit ramps or the

interchange ramps. The road segments $r_{302}, r_{391}, r_{286}$ are located at the intersections and near the import and export of expressways with intricate traffic flow. r_{181} is located in the middle segment of the east-west expressway, which is the main corridor in Shanghai and bears the largest traffic volume in the east-west direction. These critical road segments are normally congested, which may affect the traffic states of other segments in the road network.

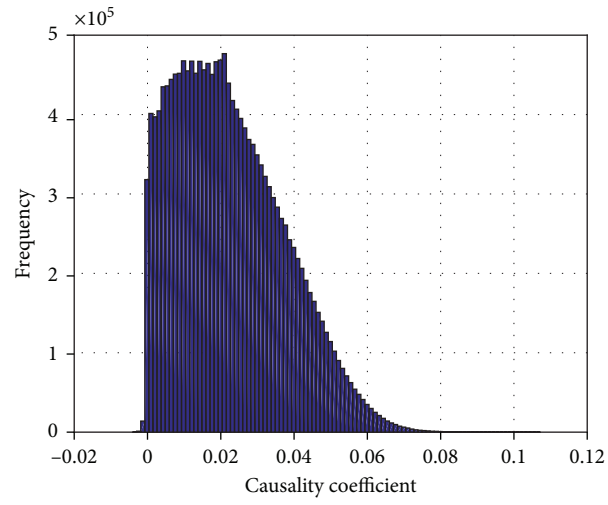


FIGURE 9: Distribution of causality coefficients.

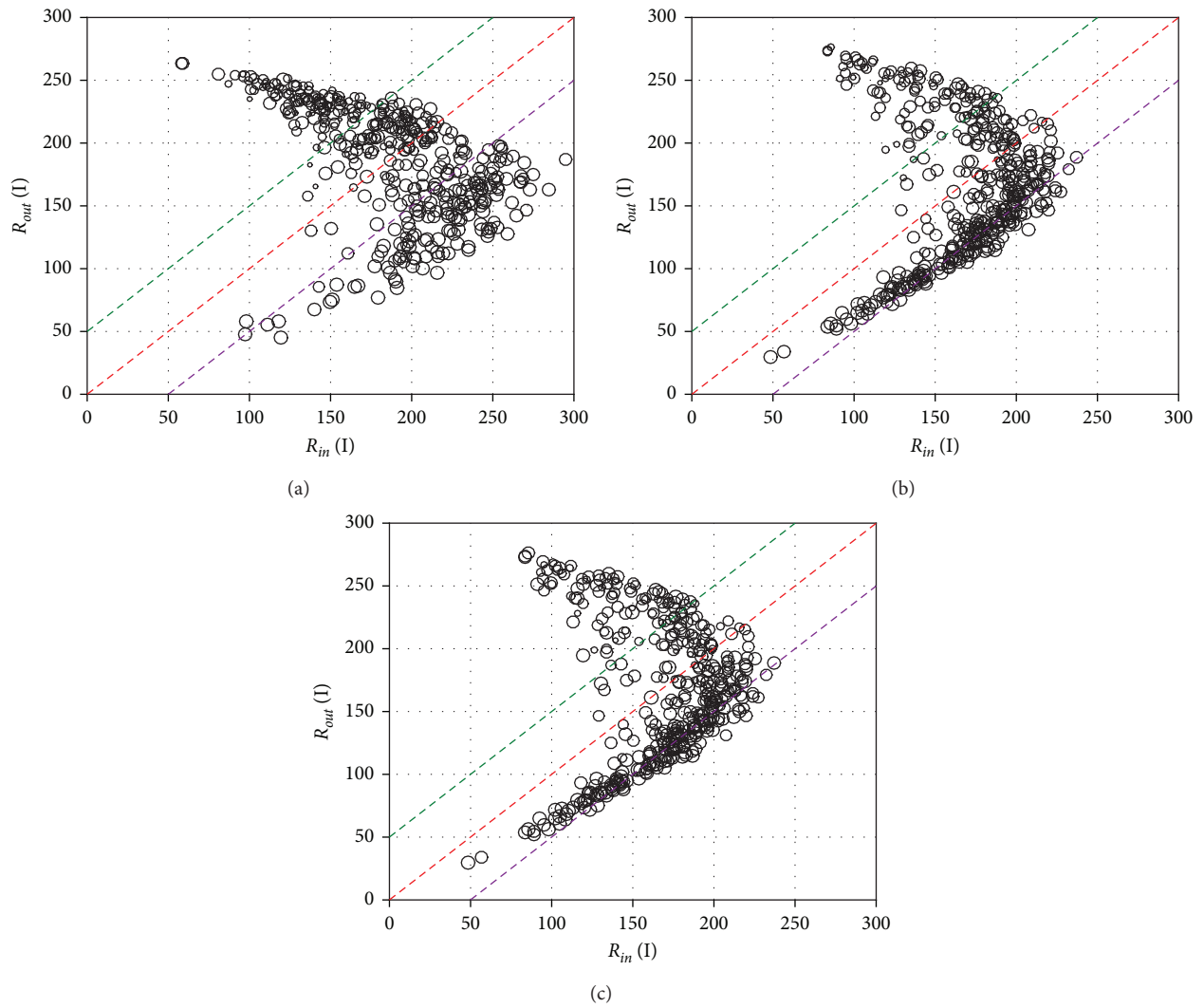
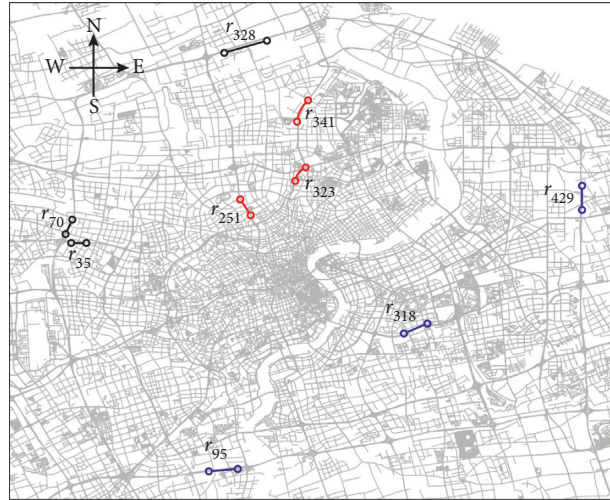


FIGURE 10: Distribution of influence coefficients and affected coefficients for road network: (a) morning peak time; (b) nonpeak time; (c) evening peak time.

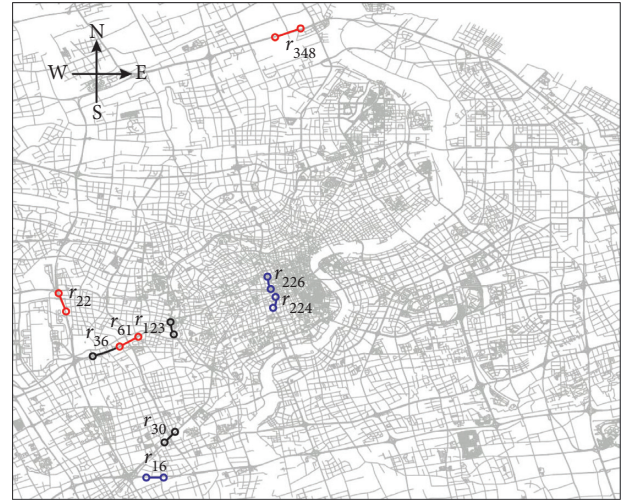
TABLE 3: Potential bottleneck segments and critical road segments.

	Potential bottleneck segments	Critical road segments
Morning peak time	r_{35}, r_{328}, r_{70}	r_{123}, r_{36}, r_{30}
Nonpeak time	$r_{341}, r_{323}, r_{251}$	r_{22}, r_{348}, r_{61}
Evening peak time	r_{429}, r_{318}, r_{95}	r_{16}, r_{226}, r_{224}



—○— Morning peak time
 —○— Nonpeak time
 —○— Evening peak time

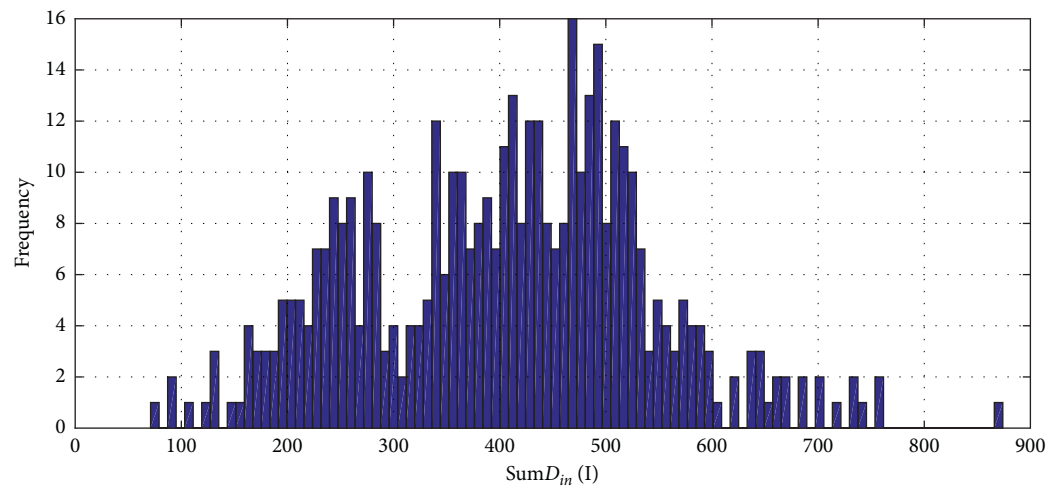
(a)



—○— Morning peak time
 —○— Nonpeak time
 —○— Evening peak time

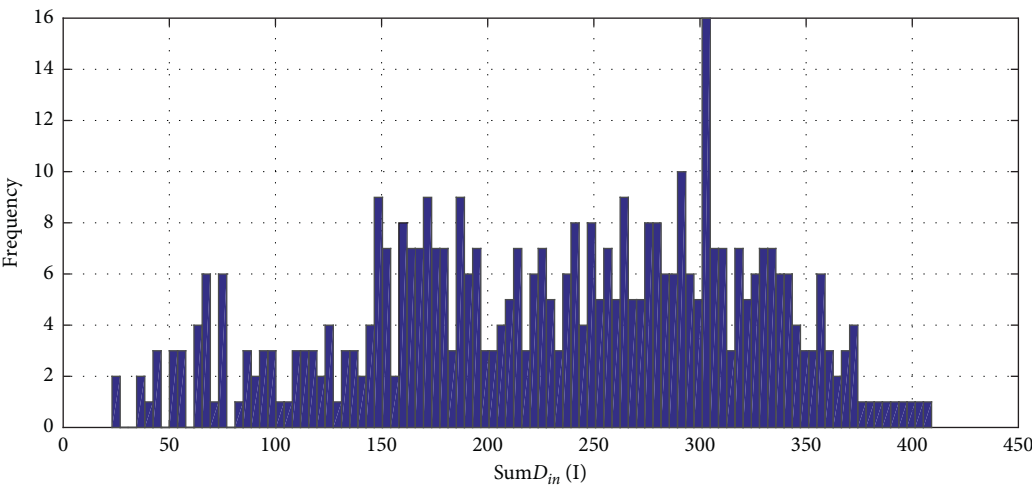
(b)

FIGURE 11: Spatial distribution of critical road segments: (a) potential bottleneck segments; (b) critical road segments.

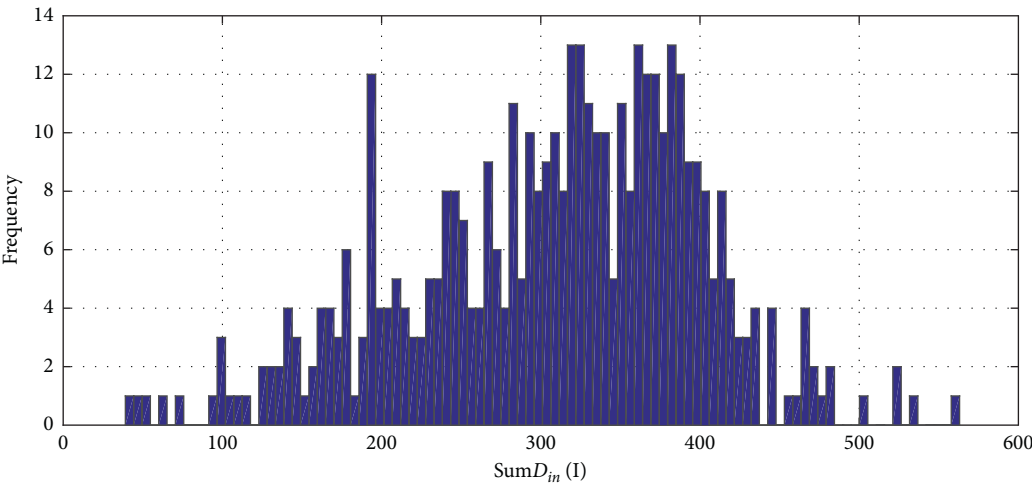


(a)

FIGURE 12: Continued.

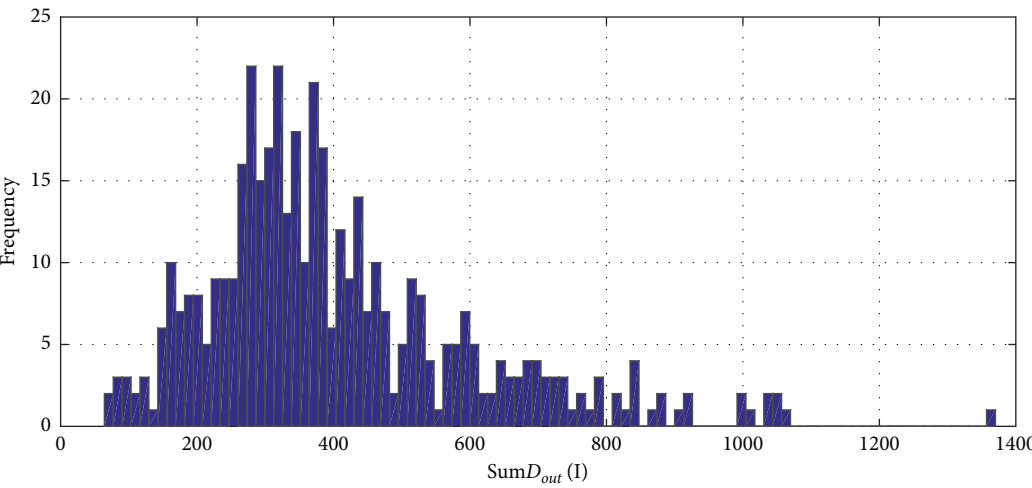


(b)



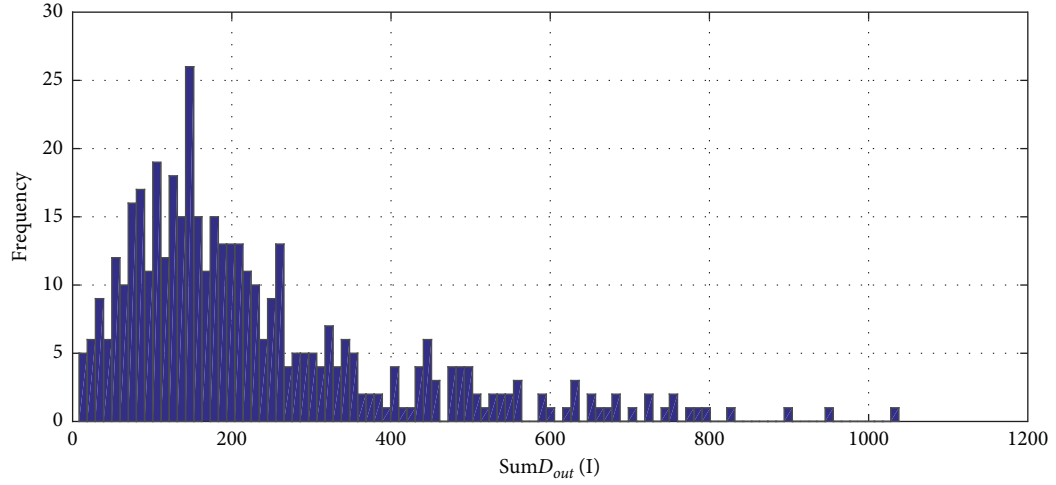
(c)

FIGURE 12: Distributions of $\text{Sum}D_{in}(i)$ for different time periods: (a) morning peak time; (b) nonpeak time; (c) evening peak time.

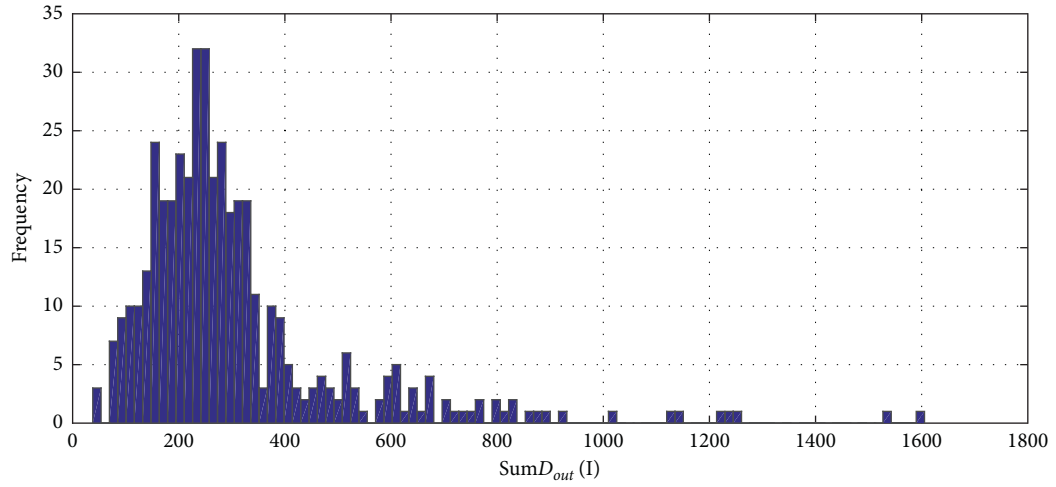


(a)

FIGURE 13: Continued.



(b)



(c)

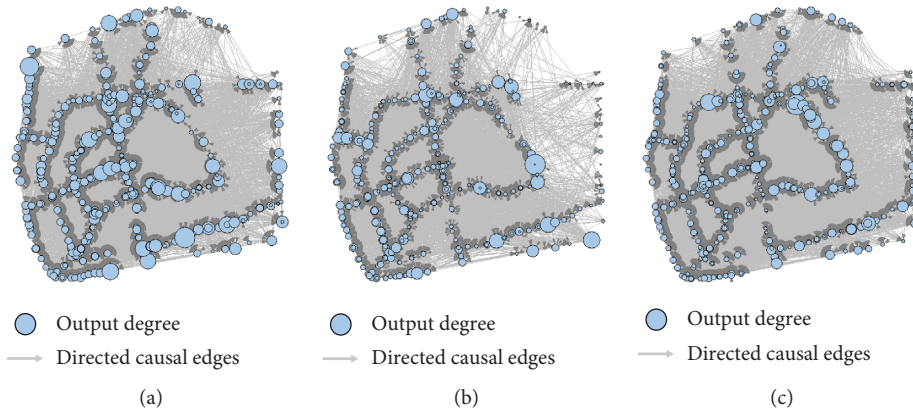
FIGURE 13: Distributions of $\text{Sum}D_{\text{out}}(i)$ for different time periods: (a) morning peak time; (b) nonpeak time; (c) evening peak time.

FIGURE 14: Visualization of spatiotemporal traffic causality graphs: (a) morning peak time; (b) nonpeak time; (c) evening peak time.

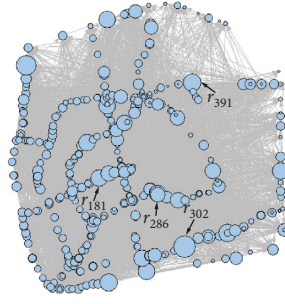
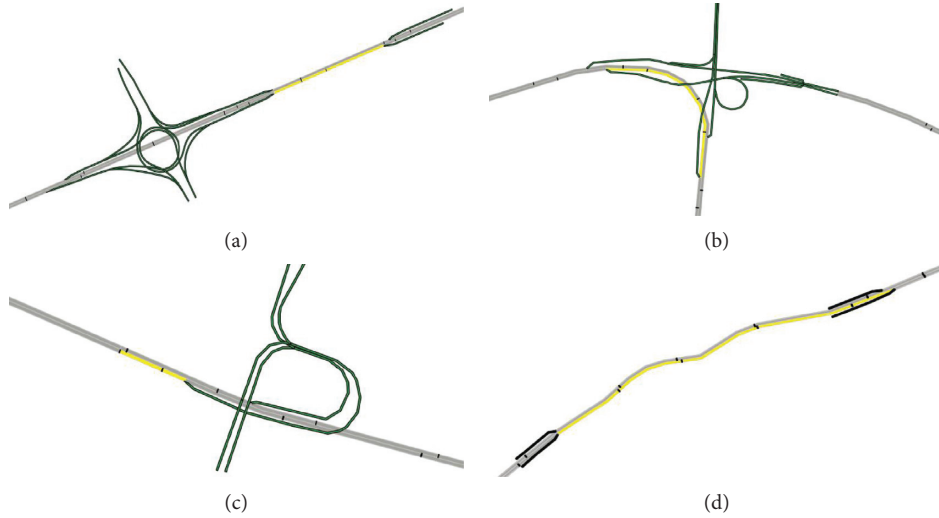


FIGURE 15: Spatial locations of critical road segments.

FIGURE 16: Spatial structures of critical road segments: (a) r_{302} ; (b) r_{391} ; (c) r_{286} ; (d) r_{181} .

Generally, the identified critical road segments are consistent with the spatial structure and traffic condition of the road network, which can prove that the transfer entropy is suitable to evaluate the causal interaction of network traffic flow. Real-time traffic control measures can be taken for the time-varying critical road segments or potential bottlenecks to prevent traffic jams and improve traffic operation efficiency. Furthermore, the potential flaws of road network structure can be optimized in the future.

6. Conclusions

This study proposes a novel dynamic spatiotemporal causality modeling framework, which can represent information transmission of network traffic flow and identify the potential bottlenecks and critical road segments of the existing road network. Gaussian kernel density estimation method is used to calculate the transfer entropy among different road segments. To reveal the dynamic variation of traffic causality, the sliding window technique is utilized for the calculation of the transfer entropy. Causality significance test is performed to construct spatiotemporal causality graphs. This study can effectively extract the potential nonlinear causal relationships from massive traffic data and provide a data-driven research framework to identify the

critical road segments and potential bottlenecks in the road network. The detected dynamic spatiotemporal traffic causality can be combined with the traffic prediction in the real-time traffic management system.

The experimental results based on the traffic sensor data for the Shanghai expressway network indicate that transfer entropy for network traffic flow is asymmetrical, which fluctuates significantly with space and time. The output and input degrees in the peak time are generally larger than those in nonpeak hours with more information transfer and stronger causal interaction for the network traffic flow. The critical road segments with larger output degrees are mainly located in the intersections, bearing the convergence and dispersion of large traffic flows and having significant impacts on the traffic state of other segments in the road network. The causal correlation of the road segments with smooth traffic condition at the nonpeak time is weaker than that of peak time.

This study does not consider the connectivity of road network. In the future, we would integrate the network topology structure into spatiotemporal traffic causality analysis and then develop a traffic congestion propagation pattern identification model. In addition, the traffic causality analysis can be further combined with traffic congestion prediction.

Data Availability

The data used to support the findings of this study are not publicly available. Please contact the corresponding author for details.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the China Postdoctoral Science Foundation (No. 2019M660566), National Natural Science Foundation of China (No. 72001029), and Beijing Municipal Natural Science Foundation (No. 9182012).

References

- [1] G. Leduc, "Road traffic data: collection methods and applications, working papers on energy," *Transport and Climate Change*, vol. 1, pp. 1–55, 2008.
- [2] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [3] J. Tang, L. Li, Z. Hu, and F. Liu, "Short-term traffic flow prediction considering spatio-temporal correlation: a hybrid model combining type-2 fuzzy C-means and artificial neural network," *IEEE Access*, vol. 7, pp. 101009–101018, 2019.
- [4] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.
- [5] F. Wen, G. Zhang, L. Sun, X. Wang, and X. Xu, "A hybrid temporal association rules mining method for traffic congestion prediction," *Computers & Industrial Engineering*, vol. 130, pp. 779–787, 2019.
- [6] C. Silverstein, S. Brin, R. Motwani, and J. Ullman, "Scalable techniques for mining causal structures," *Data Mining and Knowledge Discovery*, vol. 4, pp. 163–192, 2000.
- [7] K. Hlavackovaschindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.
- [8] B. Huang, K. Zhang, M. Gong, and C. Glymour, "Causal discovery and forecasting in nonstationary environments with state-space models," *Proceedings of Machine Learning Research*, vol. 97, p. 2901, 2019.
- [9] B. Huang, K. Zhang, J. Zhang et al., "Causal discovery from heterogeneous/nonstationary data," *Journal of Machine Learning Research*, vol. 21, pp. 1–53, 2020.
- [10] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge, UK, 2000.
- [11] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1010–1018, San Diego, CA, USA, August 2011.
- [12] V. Kapoor, D. Saxena, V. Raychoudhury, and S. Kumar, "Real time building and maintaining causal congestion graph for intelligent traffic management," in *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 770–775, IEEE, Athens, Greece, March 2018.
- [13] V. W. Chu, R. K. Wong, W. Liu, and F. Chen, "Causal structure discovery for spatio-temporal data," *Database Systems for Advanced Applications*, pp. 236–250, Springer, Berlin, Germany, 2014.
- [14] C. M. Queen and C. J. Albers, "Intervention and causality: forecasting traffic flows using a dynamic Bayesian network," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 669–681, 2009.
- [15] H. Nguyen, W. Liu, and F. Chen, "Discovering congestion propagation patterns in spatio-temporal traffic data," *IEEE Transactions on Big Data*, vol. 3, pp. 169–180, 2016.
- [16] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 292–307, 2015.
- [17] R. Inoue, A. Miyashita, and M. Sugita, "Mining spatio-temporal patterns of congested traffic in urban areas from traffic sensor data," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 731–736, IEEE, Rio de Janeiro, Brazil, November 2016.
- [18] S. Chawla, Y. Zheng, and J. Hu, "Inferring the root cause in road traffic anomalies," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, pp. 141–150, IEEE, Brussels, Belgium, December 2012.
- [19] H. Xiong, A. Vahedian, X. Zhou, Y. Li, and J. Luo, "Predicting traffic congestion propagation patterns: a propagation graph approach," in *Proceedings of the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pp. 60–69, Seattle, WA, USA, November 2018.
- [20] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Journal of the Econometric Society*, vol. 37, no. 3, pp. 424–438, 1969.
- [21] C. W. J. Granger, "Testing for causality," *Journal of Economic Dynamics and Control*, vol. 2, pp. 329–352, 1980.
- [22] S. L. Bressler and A. K. Seth, "Wiener-Granger Causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.
- [23] M. Wibral, R. Vicente, and M. Lindner, "Transfer entropy in neuroscience," in *Directed Information Measures in Neuroscience*, pp. 3–36, Springer, Berlin, Germany, 2014.
- [24] Y. Shu and J. Zhao, "Data-driven causal inference based on a modified transfer entropy," *Computers & Chemical Engineering*, vol. 57, pp. 173–180, 2013.
- [25] Y. Teng and P. Shang, "Transfer entropy coefficient: quantifying level of information flow between financial time series," *Physica A: Statistical Mechanics and Its Applications*, vol. 469, pp. 60–70, 2017.
- [26] F. Yang, L. S. Sirish, and D. Xiao, "Signed directed graph modeling of industrial processes and their validation by data-based methods," in *Proceedings of the 2010 Conference on Control and Fault-Tolerant Systems (SysTol)*, pp. 387–392, IEEE, Nice, France, October 2010.
- [27] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [28] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, p. 461, 2000.
- [29] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy,"

- IEEE Transactions on Control Systems Technology*, vol. 15, pp. 12–21, 2006.
- [30] G. Yu, J. Hu, C. Zhang, L. Zhuang, and J. Song, “Short-term traffic flow forecasting based on Markov chain model,” in *Proceedings of the IEEE IV2003 Intelligent Vehicles Symposium*, pp. 208–212, IEEE, Columbus, OH, USA, June 2003.
 - [31] P. Duan, F. Yang, S. L. Shah, and T. Chen, “Transfer zero-entropy and its application for capturing cause and effect relationship between variables,” *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 855–867, 2014.
 - [32] S. Yang, J. Wu, Y. Xu, and T. Yang, “Revealing heterogeneous spatiotemporal traffic flow patterns of urban road network via tensor decomposition-based clustering approach,” *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 120688, 2019.