Electronic Research Archive

ERA, 31(1): 207–228.
DOI: 10.3934/era.2023011
Received: 06 September 2022
Revised: 06 October 2022
Accepted: 13 October 2022
Published: 28 October 2022

http://www.aimspress.com/journal/ERA

*Research article*

# Transfer learning for robust urban network-wide traffic volume estimation with uncertain detector deployment scheme

**Jiping Xing[1], Yunchi Wu[2], Di Huang[1] and Xin Liu[1,*]**

[1] Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, China
[2] School of Public Administration, Huazhong University of Science and Technology, Wuhan, China

* **Correspondence:** Email: 230208830@seu.edu.cn.

**Abstract:** Real-time and accurate network-wide traffic volume estimation/detection is an essential part of urban transport system planning and management. As it is impractical to install detectors on every road segment of the city network, methods on the network-wide flow estimation based on limited detector data are of considerable significance. However, when the plan of detector deployment is uncertain, existing methods are unsuitable to be directly used. In this study, a transfer component analysis (TCA)-based network-wide volume estimation model, considering the different traffic volume distributions of road segments and transforming traffic features into common data space, is proposed. Moreover, this study applied taxi GPS (global positioning system) data and cellular signaling data with the same spatio-temporal coverage to improve feature extraction. In numerical experiments, the robustness and stability of the proposed network-wide estimation method outperformed other baselines in the two subnetworks selected from the urban centers and suburbs.

**Keywords:** network-wide volume estimation; transfer component analysis; multi-source data fusion; taxi GPS data; cellular signaling data

## 1. Introduction

Traffic congestion is a severe problem that significantly affects the vibrant development of urban megacities, as congestion increases travel costs, noise, fuel consumption, and emissions. Timely and accurate network-wide traffic flow/volume estimation is vital for alleviating traffic congestion and

serves as a fundamental input to model congestion control strategies [1]. Currently, traffic volume data is collected mainly from fixed detectors, such as inductive loop detectors, radar detectors, and continuous counting stations. Nevertheless, the installation of detectors with full network-scale coverage is unrealistic and costly owing to the limited budget, particularly in suburban areas. As such, estimating network-wide traffic volume from limited detector deployment has intrigued great interest in research and practice.

Most of the existing studies on the estimation of network-wide traffic volumes can be divided into model-based approaches and data-driven approaches. The former includes traffic flow fundamental diagram-based models [2,3], multiple linear regression [4], compressive sensing [5], trajectory-based inference models [6], weighted mean models [7], and kriging-based models [8]. The latter employs cutting-edge artificial intelligence (AI) methods in the application of traffic volume estimation on large-scale road network [9–11]. The reported AI methods reported include semi-supervised learning-based model [12], mixture Gaussian graphical model [13], deep meta-learning-based model [14], generative adversarial network-based model, tensor decomposition [15], and dynamic Bayesian graph convolution network [16]. The main idea of both model-based and data-driven works was to model spatio-temporal correlations between the detector-installed and detector-uninstalled road segments, and they attempted to explicitly reveal the inherent relationships between network-wide volume estimation with geo-graph attributes and segment cascade characteristics.

However, an urban traffic network is a complex system, and thus estimating network-wide traffic volume is a challenging task. Existing studies rely on perfect knowledge and certainty from detector data and often ignore the inherent uncertainty and variability in the detector data. In particular, when the number of fixed detectors deployed is limited by their special function, they are mainly deployed on arterial segments or important intersections; secondary arterial and branch segments do not deploy detectors. Note that there are significant differences in the distribution of traffic volume between detector-installed and detector-uninstalled segments, and the correlation of geo-graph attributes (i.e., point of interest, environmental features) cannot be simply modeled as sufficient common features for traffic inference. Thus, the robustness and stability of traffic volume estimation in specific research scenarios are unsatisfactory.

To solve this problem, this study presents a network-wide transfer component analysis (TCA)-based model to estimate the traffic volume on a detector-uninstalled segment with different data distributions. We show that the TCA-based estimation model is able to explicitly consider network-wide variability in volume distribution used by different grades of road segments in the volume estimation, which significantly improves model robustness. We develop a stability estimation model that is scalable to improve the diversity of traffic features using two categories of probe vehicle data: cellular signaling (CS) and taxi global positioning system (GPS) data. Numerical experiments conducted on two practical urban networks show that our proposed model performs well in accurately estimating the traffic volume when the data distribution on detector-installed segments differs from that on detector-uninstalled segments. In particular, robust performance can still be maintained despite uncertainty in detector deployment.

The remainder of this paper is organized as follows. Section 2 reviews previous studies. Section 3 demonstrates the methodology, and Section 4 compares the performance of the proposed approach with actual case experiments. Finally, Section 5 provides conclusions and outlines future research.

## 2. Literature review

Existing studies on the estimation of missing traffic data mainly focus on estimating traffic speed [17,18], travel time [19–21], and traffic volume [22,23]. This study addresses the traffic volume estimation under uncertainty in detection data sources, and we present an overview of the existing literature on urban traffic volume estimation and the application of multi-source data fusion in traffic estimation.

### *2.1. The urban traffic volume estimation*

Traffic volume estimation in urban road networks can be classified into network-wide traffic flow estimation and city-wide traffic flow estimation based on the size of the research road network. For network-wide traffic volume estimation, existing studies can be divided into the model-based and data-driven approaches. The former is mainly based on the relationship between segments with and without detectors installed for multiple regression modeling [7]. In the latter studies, Zhang et al. [24] combined a bidirectional recurrent neural network and graph convolution to implement network-wide online traffic completion and prediction. In addition, based on a graph theory approach, Yi et al. [13] applied a tree ensemble-based model to extract network-scale features for hourly traffic volume inference. Luan et al. [16] proposed a dynamic Bayesian graph convolution network model to extrapolate the urban network congestion propagation. Applying the transfer learning theory, Li et al. [15] proposed a deep tensor adaptation network model to balance the impact of missing data on regional network traffic prediction.

For city-scale traffic state estimation, with the expansion of the network size, there will be a large number of road sections that are involved in traffic flow estimation. Compared to network-wide traffic flow estimation, city-wide studies generally require greater computational power and difficulty but do not have too high a requirement for estimation accuracy. It is mainly based on the GPS trajectory inference method of probe detector data to achieve city-wide traffic state information [25]. Liu et al. [26] developed a dynamic neural network model to perform a parallel estimation of the missing traffic speed on the spark platform. Cao et al. [27] proposed a semi-supervised route choice model to infer the traffic state by fusing probe detectors and automatic vehicle identification data. Using similar input data, Yu et al. [28] further added a simulation module to complement the data-driven approach for repairing sparse trajectories, thus enabling traffic volume estimation in city-scale networks.

However, despite the significant interest and a large number of studies in traffic flow estimation, some challenges remain to be addressed. The aforementioned two categories of approaches have mainly focused on modeling the inherent relationships between geo-graph attributes and data spatio-temporal correlation on both network-wide and city-scale traffic flow estimation, but such relationships in each segment are difficult to learn without any prior knowledge. In particular, urban traffic networks are complex systems. It is difficult to apply these methods to specific scenarios and obtain satisfactory performance. For example, compared with developed city areas, fixed detectors are usually more sparsely deployed in rural areas. At the same time, the variability between road segments in the city-scale network is large than in general road network-wide. As such, the estimation model that originated in areas with densely deployed detectors cannot be directly used in areas with sparely deployed detectors. Hence, there is scope to develop a kind of adaptively traffic estimation model that can handle different data distributions for improving the robustness of estimation performance.

### *2.2. The application of multi-source data fusion in traffic estimation*

The development of emerging information provides various data for urban traffic computing, such as mobile phone, taxi GPS, and license plate recognition data. In network-wide traffic volume estimation, the fusion of these data in modeling can well use the advantages of each type of data. Zhan et al. [29] integrated taxi GPS data with fixed detector data to estimate city-scale traffic flows using a Bayesian network approach. Wang et al. [30] fused three types of data to infer the travel routes of vehicles for network-wide traffic speed estimation. Seppecher et al. [31] applied CDR, GPS, and LBNS data to address the impact of sparse positioning data on the traffic state estimation for some segments. Saffari et al. [32] discussed the fusion of probe vehicles and loop detectors with the same spatiotemporal coverage, which were used to estimate the macroscopic fundamental diagram in large-scale urban networks. Moreover, Rodriguez-Vega et al. [33] proposed a dynamic traffic density estimation model using travel path reconstruction by fusing fixed detector turning ratios and aggregated floating vehicle data. In estimation of travel time distribution, Yun and Qin [34] applied the sources of taxi floating vehicle and radio frequency identification data were collected for determining minimum sampling size of floating cars. In traffic mobility analysis, with the fusion of mobile phone data and urban transportation data, Huang et al. [35] proposed a predictive model to evaluate crowd gatherings that cause traffic jams. The effects of built environment spatial variation [36] and topological properties of the urban network [37] were analyzed in bike-sharing mobility by fusion of four types of data. In summary, the large amount of fused data has led to multiple network-wide traffic state estimation methods, however, there is scope to further develop feature extraction methods based on data characteristics.

Furthermore, to solve the conundrum of modeling adaptability, transfer learning methods have gradually gained great development. And it has recently been applied in the field of transportation [16]. Among them, transfer component analysis (TCA) model acts as a classical transfer learning approach that is developed by [38]. It aims to learn some transfer feature components by a reproducing kernel Hilbert space and is suitable for handling different data distributions in different domains. In this study, we try to apply this TCA model to estimate network-wide traffic volume in detector-uninstalled segments with different data distributions, which could improve the performance robustness under the uncertainty of detector deployment in the road network.

## 3. The TCA-based estimation method

This section first introduces the notations used and then presents the simulation-based bi-level DNDP model with budget constraints. The notations used in the proposed framework are defined in Table 1.

### *3.1. Problem formulation*

When researching road networks from different regions, the deployment schemes of detectors are different because of the structural differences within the road network. For example, if the research road network is from a rural area, the road network is affected by fewer key segments and its detector deployment density is sparse compared to that of the urban road network. In this case, although the same estimation model is applied, it is difficult to find similar road segments to be used as model input

in rural areas with sparsely deployed detectors, the estimation performance is poorer than that of the urban road network. For this uncertain scenario of detector deployment, our issue is how to maintain stable performance in the network-wide traffic volume estimation. In this section, we present a TCA model derived from transfer learning, which takes account of the different distributions of detector data, in estimating network-wide traffic volume.

We consider an urban road network where some, but all of the road segments have installed detectors that provide traffic volume and speed data. The detector-installed road segments are set as source domains $\mathcal{D}_S = \left\{(\boldsymbol{x}_{S_1}, y_{S_1}), \ldots, (\boldsymbol{x}_{S_{n_1}}, y_{S_{n_1}})\right\}$, where $\boldsymbol{x}_{S_i} \in \mathbb{R}^m$ is the extracted probe data feature and $y_{S_i}$ is the actual volume data label. The road segments which do not have detectors-installed and therefore do not have corresponding data labels are treated as target domains $\mathcal{D}_T = \left\{\boldsymbol{x}_{T_1}, \ldots, \boldsymbol{x}_{T_{n_2}}\right\}$, where the input $\boldsymbol{x}_{T_i}$ is also assumed to be in $\mathbb{R}^m$, $n_1$, and $n_2$, which are denoted as the number of detector-installed segments and detector-uninstalled road segments, respectively.

Let $\mathcal{P}(\boldsymbol{X}_S)$ and $\mathcal{Q}(\boldsymbol{X}_T)$ denote the marginal distributions of $\boldsymbol{X}_S$ and $\boldsymbol{X}_T$, respectively, where $\boldsymbol{X}_S = (\boldsymbol{x}_{S_1}, \ldots, \boldsymbol{x}_{S_{n_1}})$, $\boldsymbol{X}_T = (\boldsymbol{x}_{T_1}, \ldots, \boldsymbol{x}_{T_{n_2}})$, and $Y_S = (y_{S_1}, \ldots, y_{S_{n_1}})$. Note that they can be different from $\mathcal{P}(\boldsymbol{X}_S) \neq \mathcal{Q}(\boldsymbol{X}_T)$, but there exists a weaker assumption of a transformation $\phi$, such as $P(\phi(\boldsymbol{X}_S)) \approx P(\phi(\boldsymbol{X}_T))$ and $P(Y_S|\phi(\boldsymbol{X}_S)) \approx P(Y_T|\phi(\boldsymbol{X}_T))$. The task is to learn this latent transformation space $\phi$ from the transfer components, $\mathcal{D}_S$ and $\mathcal{D}_T$, and then estimate the volume data labels, $y_T$, corresponding to the inputs, $\boldsymbol{x}_T$, in the modeled transformation space. Herein, a regressor $f$ trained on the transformed source domain $\phi(\boldsymbol{X}_S)$ and $Y_S$ was used to estimate the traffic volume $y_T$ on the corresponding target domain $\phi(\boldsymbol{X}_T)$.

The proposed TCA-based network-wide volume estimation model can be represented as follows:

$$f = argmin \frac{1}{N} \sum_{i=1}^{N_s} \ell(f(\boldsymbol{x}_i), y_i) + \lambda R(\phi(\mathcal{D}_S), \phi(\mathcal{D}_T)) \tag{1}$$

where $N_s$ is the number of input data in the source domain, $\phi$ is the transformation function applied to the source and target domains, and $\lambda$ is a tradeoff regularization parameter for preserving the important properties of $\boldsymbol{X}_S$ and $\boldsymbol{X}_T$.

### *3.2. The TCA-based estimation framework*

The proposed framework is shown in Figure 1. The main steps are as follows:

Step 1: Data preprocessing: The travel speeds of CS and taxi GPS data were calculated and the correlation between the two types of probe detector and license plate recognition (LPR) data was analyzed.

Step 2: Classification of road segments: To analyze the spatio-temporal variability of traffic volume in a subnetwork, we determined the distribution between each segment for the selection of proper detector-installed segments.

Step 3: TCA-based model construction: The speed features of the CS and taxi GPS data were extracted. With this feature transformation, segments with different speed distributions were applied

as inputs to construct a TCA-based network traffic volume estimation model.

Step 4: Evaluation of the model performance: The performance of network-wide volume estimation in two different sub-networks with different segment classifications and baselines was compared.

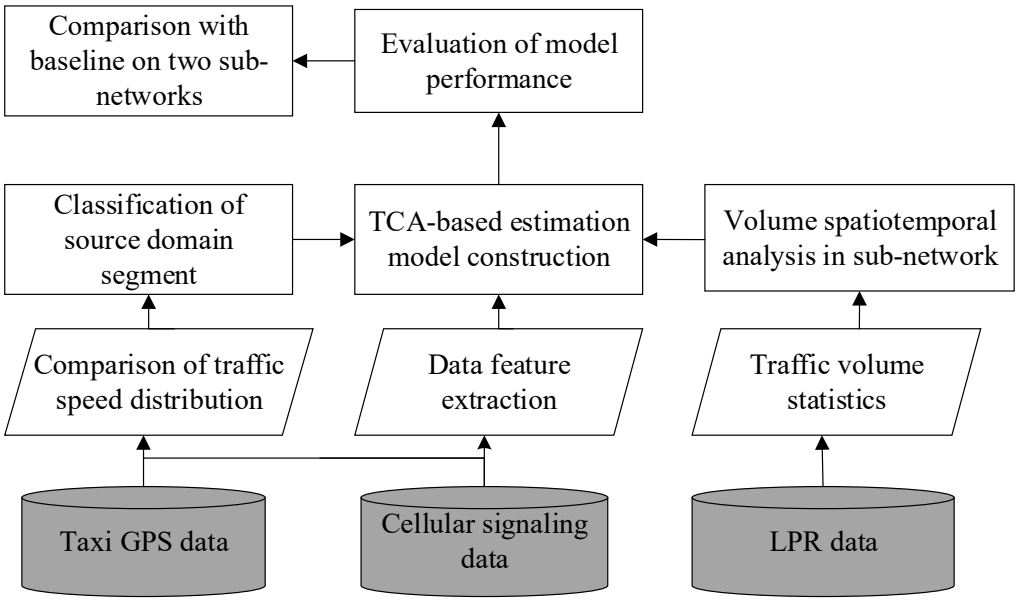



**Figure 1.** The TCA-based estimation framework.

*3.3. Mathematical models*

To calculate the distance between two different traffic volume distributions, an innovative nonparametric distance calculation method was proposed by embedding distributions in a reproducing kernel Hilbert space (RKHS), which is called the maximum mean discrepancy (MMD) method [37]. In our study, the calculation of MMD between the source domain dataset $\{\boldsymbol{x}_{S_1}, \ldots, \boldsymbol{x}_{S_{n1}}\}$ and the target domain dataset $\{\boldsymbol{x}_{T_1}, \ldots, \boldsymbol{x}_{T_{n2}}\}$ is denoted as

$$MMD(\boldsymbol{X}_S, \boldsymbol{X}_T) = \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(\boldsymbol{x}_{S_i}) - \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(\boldsymbol{x}_{T_i}) \right\|_{\mathcal{H}}^2 = \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} A^T \boldsymbol{x}_{S_i} - \frac{1}{n_2} \sum_{i=1}^{n_2} A^T \boldsymbol{x}_{T_i} \right\|_{\mathcal{H}}^2 \quad (2)$$

where $\|\cdot\|_{\mathcal{H}}$ is the RKHS norm. Let the kernel-induced feature map be an $\phi$. Herein, the distance between the two traffic data distributions is simplified as the distance between the two average components in an RKHS. When two distributions are the same, the MMD gradually approaches zero.

In this process, nonlinear mapping $\phi$ was applied to embed the source and target domain data into a shared low-dimensional latent transforming space and then solve a semi-definite program (SDP) to obtain the corresponding kernel matrix $A$. However, the high cost of SDP solvers limits their applications in the solution of our kernel problem, which would decrease the computational efficiency. Hence, based on kernel characteristics, the MMD distance can be converted as $tr(A^T XLX^T A)$, the details of which can be found in Wang et al. [40], where $tr(\cdot)$ denotes the trace of the matrix, $A$ is the correspondence matrix of the feature transformation function $\phi$, $X$ is the matrix for joining the

source and target domains, and

$$L_{ij} = \begin{cases} \frac{1}{n_1^2} & \boldsymbol{x}_i, \boldsymbol{x}_j \in \boldsymbol{X}_s \\ \frac{1}{n_2^2} & \boldsymbol{x}_i, \boldsymbol{x}_j \in \boldsymbol{X}_t \\ -\frac{1}{n_1 n_2} & otherwise \end{cases} \tag{3}$$

Thus, the objective function of MMD can then be written as

$$\min_{A>0} \mathrm{tr}(A^T X L X^T A) \tag{4}$$

where the objective is to minimize the distance between different distributions.

To ensure that the variance between $\phi(\boldsymbol{X}_S)$ and $\phi(\boldsymbol{X}_T)$ preserves improved properties to a feasible extent, the distance between the marginal distributions $P(\phi(\boldsymbol{X}_S))$ and $P(\phi(\boldsymbol{X}_T))$ can be simultaneously reduced, which is a new dimensionality reduction method in the feature space spanned by the learned latent components. Based on the data information divergence matrix conversion, the variance maximization can be formalized as:

$$max\ A^T\ XH(A^T X)^T \tag{5}$$

Integrating Eq (4) with Eq (5), the objective is denoted as

$$\min_A \frac{tr(A^T XLXA)}{A^T XHXA} \tag{6}$$

Herein, the final optimization function in the TCA-based estimation model becomes

$$f = arg\min_A tr(A^T XLXA) - \lambda tr(A^T A), \tag{7}$$

$$s.t.\ A^T KHKA = I_m \tag{8}$$

where $\lambda > 0$ is a tradeoff regularization parameter, $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix, and $m \leq n_1 + n_2 - 1$. To simplify the notation, we removed the subscript $m$ from $I_m$ in Eq (8).

Although this optimization problem involves a non-convex norm constraint $A^T KHKA = I_m$, it can still be efficiently solved using the trace optimization problem. With the help of Lagrangian transformations, it can be reformulated as

$$f = \underset{A}{argmax} tr(A^T(XLX + \lambda I)A^{-1}A^T XHXA) \tag{9}$$

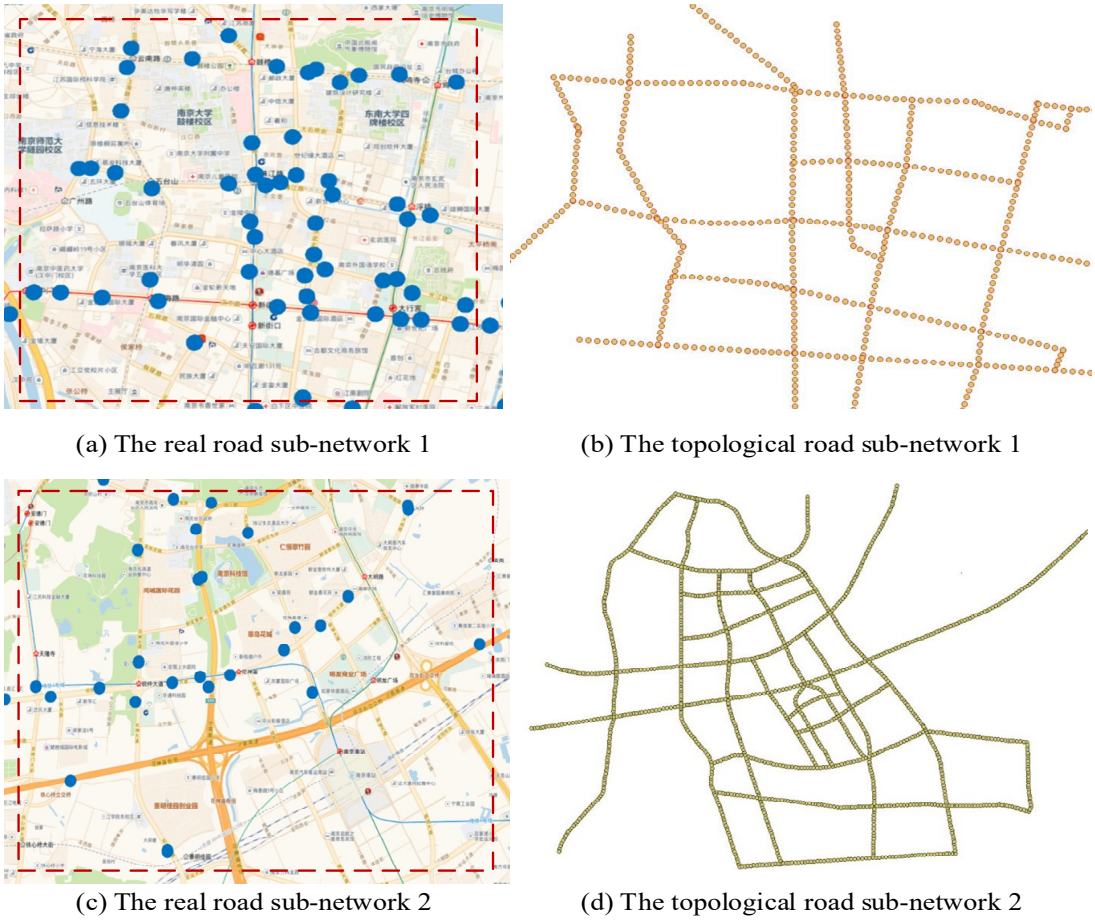
## 4. Experiments

In this section, two representative real road networks from different urban functional areas and two different types of probe detector data were applied to the proposed problem. Different features and data sizes were set to evaluate the performance of the proposed model.

*4.1. Research road network selection*

Previous traffic volume estimation studies mainly focused on the effective utilization of the relationship between the detector-installed and detector-uninstalled segment sets. Our study is aimed at estimating network traffic volume when the deployment of a fixed detector is uncertain. Hence, it is crucial to consider a particular scenario in which the detector-installed segment set differs significantly from the detector-uninstalled segment set. To this end, two subnetworks from different urban functionals were selected for our study. Both were extracted from the actual road networks in Nanjing, China. Subnetworks 1 and 2 were taken from the urban center and suburbs, respectively. The actual road subnetwork and its topology are shown in Figure 2. The layout of the LPR detector is indicated in blue.

Subnetwork 1 included 76 single-directional segments (as shown in Figure 2(a)). The number of LPR detector-installed segments was 35 and the coverage rate was approximately 46%. Subnetwork 2 included 99 single-directional traffic segments (shown in Figure 2(b)). It had 22 LPR detector-installed segments, with a coverage rate of approximately 22%. Note that the density of the LPR detector deployment in subnetwork 2 was significantly sparser than that in subnetwork 1.



(a) The real road sub-network 1 (b) The topological road sub-network 1

(c) The real road sub-network 2 (d) The topological road sub-network 2

**Figure 2.** The real and topological subnetworks with deployment of LPR detectors marked as filled blue circles.

### *4.2. Dataset*

In this section, we first introduce the dataset, analyze the spatiotemporal distribution diversity of road segment in road network, and then preprocess the data.

#### 4.2.1. Dataset description

To reduce the impact of size reduction in the feature transformation process of the TCA model, we applied two types of probe vehicle data to enhance feature diversity and market penetration. The selected data were taxi GPS, CS, and LPR. The three types of data labels are listed in Table 1.

1) Taxi GPS data

The taxi GPS data can be viewed as classic probe vehicle detector data collected by the local government. The GPS location information could be updated at an interval of 30 s. Subnetwork 1 has 312 million trajectory records, with each taxi driver in subnetwork 1 making approximately 13 trips per day. Subnetwork 2 has nearly 109 million taxi trajectory records, with 426 taxis per day and an annual mean of four trips for each driver.

2) CS data

This type of probe detector data was provided by a telecommunications service company. The tag information is listed in Table 1. Among them, approximately 6.8 million data records are available daily in subnetwork 2. These data are used to identify the location of each mobile phone signal transmission using the time difference of the arrival location technique, where the positioning accuracy can be within 100 m.

3) LPR data

LPR systems use cameras to capture vehicle data in real time with high precision, are located along road segments, and can recognize the license number plate of each passing vehicle using image recognition technology. This equipment can detect vehicle types by identifying the different colors of license plates, which is useful for converting the volumes to passenger car units.

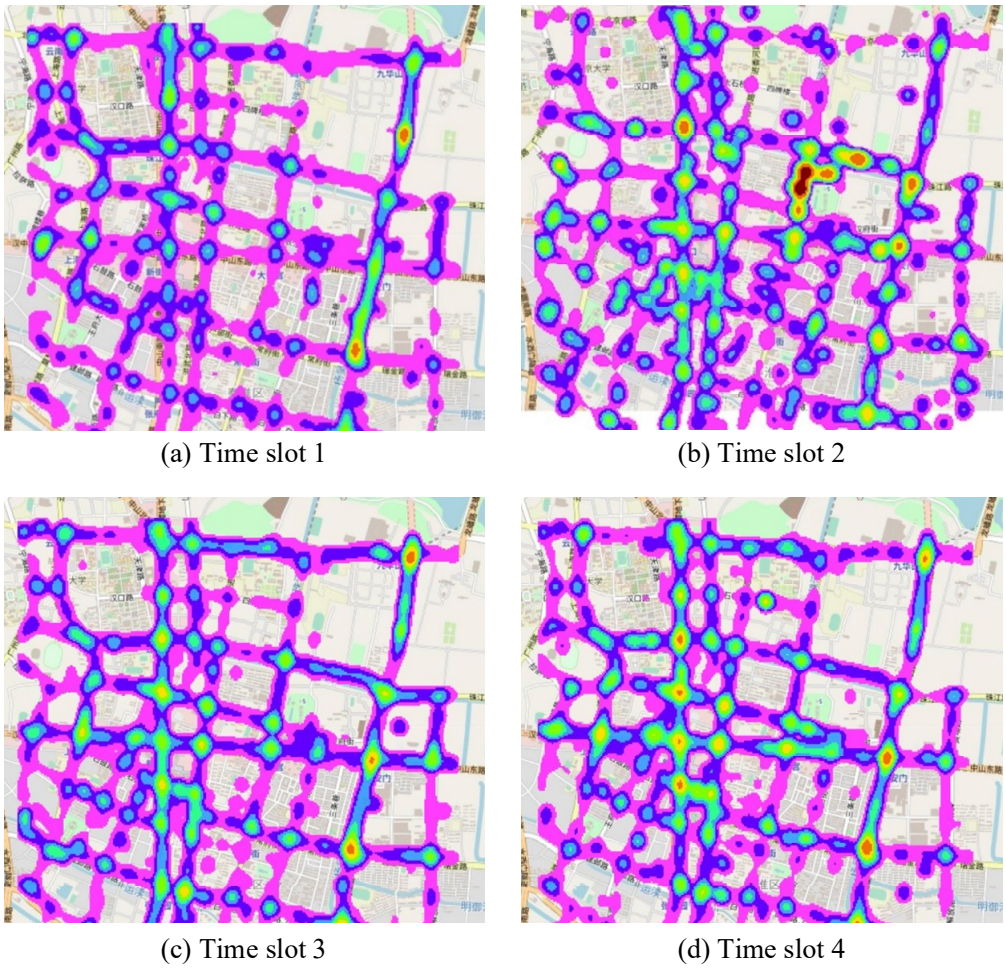
**Table 1.** The information of three data types.

| Type of data | Data label | Amount of data | Time period |
|---|---|---|---|
| Taxi GPS data | Taxi License ID, Get on time, Get off time, Longitude, Latitude, Timestamp | 307 million data | 26/09/2016–31/10/2016 |
| CS data | Mobile phone ID, Longitude, Latitude, Timestamp | 6.8 million records each day | 1/10/2016–07/10/2016 & 15/10/2016–21/10/2016 |
| LPR data | The encrypted vehicle plate ID, Timestamp, Vehicle type, The segment ID with the LPR detector installed | 9 million data | 01/10/2016–31/10/2016 |

### 4.2.2. Spatiotemporal analysis

The spatiotemporal distribution of the traffic volume in different road segments is usually affected by the functional variability of urban areas [41–43]. For example, owing to the difference in the grade of road segments and the number of points of interest around road segments, there is a difference in traffic volume in different road segments and time slots [44]. To visualize this difference, we mapped the taxi GPS data in subnetwork 1 to observe the variation of traffic at spatiotemporal scales.

As Figure 3 illustrates, the spatio-temporal distribution of taxi GPS data, as a partial traffic volume, can represent the traffic state of the road network. The spatiotemporal distribution varies in different road segments and time slots (0.00–2:00 a.m. as time slot 1; 7:00–9:00 a.m. as time slot 2; 1:00–3:00 p.m. as time slot 3; 5:00–7:00 p.m. as time slot 4). Thus, the spatio-temporal features in the detector-installed segment set were significantly different from those in the detector-uninstalled segment set, and they cannot be transferred directly.



(a) Time slot 1 (b) Time slot 2

(c) Time slot 3 (d) Time slot 4

**Figure 3.** The spatiotemporal distribution of taxi GPS data.

### 4.2.3. Dataset preprocessing of probe detector

The aforementioned label information of CS and taxi GPS data, as a medium for recording travel

trajectory information, can both be viewed as probe detector data. Owing to the advantages of the full spatial and temporal coverage of these two types of probe data, the traffic flow parameter features extracted by these two types of data can be used to analyze the relationship between detector-installed and detector-uninstalled segments and then implement traffic volume estimation on detector-uninstalled segments. Information on partial traffic volume and traffic speed can usually be acquired. Owing to the impact of traffic demand from points of interest around the road network, these partial traffic volumes are not always representative of some segments. However, the traffic speed extracted by these two types of probe data is a more reliable representation of the actual network traffic state than the traffic volume [43,45–47]. Travel speed has a direct relationship with traffic volume in network volume estimation. Road segments with similar average speed patterns have a high probability of similar volume patterns. In this subsection, we calculate the probe traveling speed as the basis for constructing the estimation model.

In the preprocessing of the two types of probe data, these data must first be matched into the map based on the grid scanning method, and some CS data need to be excluded for non-motorized trips and signaling drift data. The details can be found in Xing et al. [48]. We must calculate the instantaneous travel speed of each traveler and the average travel speed from each road segment. The travel speeds were extracted by calculating the positioning intervals.

The instantaneous speed $v_{i,j,k}$ for the probe vehicle $i$ $(i = 1,\ldots,I)$, locating point $j$ $(j = 1,..,J)$, and road segment $k$ $(k = 1,\ldots,K)$ at time slot $t$ $(t = 1,\ldots,T)$ is denoted as

$$v_{i,j,k} = \frac{Dist(x_i \cdot P_{j+1,k}, x_i \cdot P_{j,k})}{|x_i \cdot P_{i,t+1} - x_i \cdot P_{i,t}|} \tag{10}$$

where $Dist$ denotes the distance between two successive positioning points, $P_j$ and $P_{j+1}$. The average speed of probe vehicle $i$ in the segment $k$ is denoted by $\bar{v}_{i,k}$, the average speed of road segment $k$ is set as $\bar{v}_k$, and they can be acquired as follows:

$$\bar{v}_{i,k} = \frac{1}{L}\sum_{l=1}^{L} v_{i,j,k} \tag{11}$$

$$\bar{v}_k = \frac{1}{I}\sum_{i=1}^{I} \bar{v}_{i,k} \tag{12}$$

### *4.3. Research segment matching*

To filter the road segment set for evaluating the performance of our proposed model in a complex and diverse urban road network, experiments for data correlation analysis and speed distribution classification in each segment were conducted.

#### 4.3.1. The correlation analysis of probe and fixed detector data

Our research aims to implement a feature transformation called $\phi$ to solve the problem of uncertain detector deployment in network traffic volume estimation, which needs to be satisfied as $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$. Before that, the correlation between features $X_S$ and $X_T$ and labels $Y_S$ and $Y_T$ should first be evaluated. In our experiments, we needed to correspondingly analyze the correlation between the two types of probe and LPR detector data in our research network. Because of

the lack of LPR detector data in detector-uninstalled segments, we assumed consistency in this correlation between the detector-installed and detector-uninstalled segments. The Pearson correlation coefficient (PCC) method was applied as follows:

$$PCC = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{13}$$

where $x_i$ and $y_i$ denote the partial probe and LPR volumes in $i$ th, $i = 1,2,\ldots,n$, $n$ is the data size, $\bar{x}$ and $\bar{y}$ is the mean of $x_i$ and $y_i$, respectively.

One specific segment was considered as an example to illustrate the temporal variation in the probe and LPR volumes. As shown in Figure 4, the PCC was calculated at 5, 15, 30, and 60 min intervals in sequence. Each time slot showed a good correlation (as a strong correlation is generally set at 0.6 to 0.8) with PCC values of 0.76, 0.78, 0.79, and 0.8, respectively.

To satisfy the correlation requirement, correlation analyses in other remaining LPR detector-installed segments must also be implemented. Combining the calculation of the empirical cumulative distribution of regression residuals, we set 0.65 as the threshold to determine whether the probe and LPR data were correlated. If the segment does not reach the threshold, it cannot be used for our proposed model.

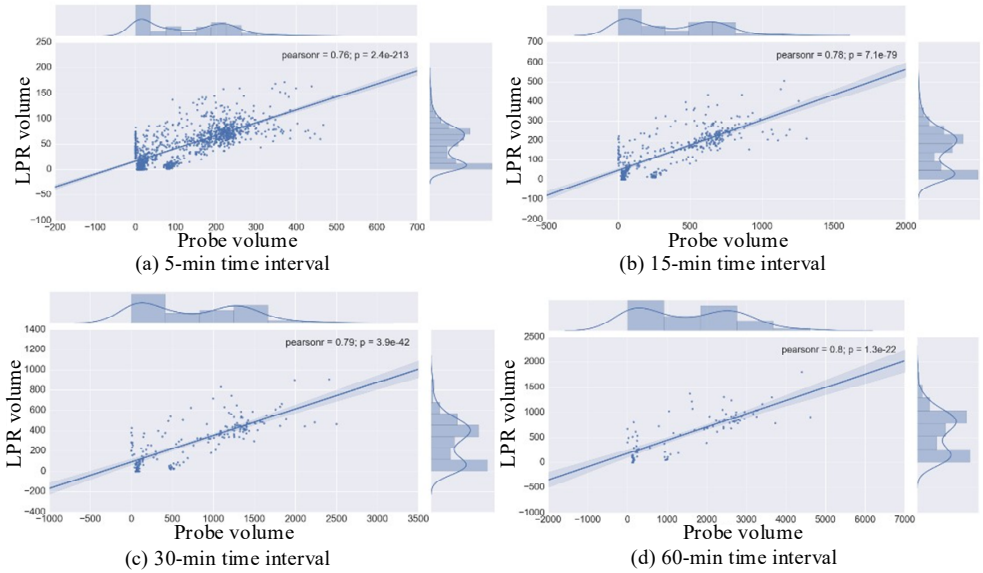



**Figure 4.** The correlation analysis of the probe and actual volumes.

### 4.3.2. Classification of research segment with speed distribution

Our proposed model transforms the traffic features of the detector-installed segment with these of the detector-uninstalled object segment to implement traffic volume estimation. Before this, we needed to reveal the relationships between detector-installed and detector-uninstalled segments in urban networks using the full spatio-temporal coverage of probe detector data. With the aforementioned correlation analysis between the probe and fixed LPR detector data, it can be found that the probe detector data is a decent representation of the actual traffic state of the road segment.

Furthermore, compared to the traffic volume information extracted by the probe detector data, the traffic speed from the probe detector data could well reflect the actual traffic state. In this study, we selected the speed distribution as our measure to determine the similarity between detector-installed and detector-uninstalled segments. The selected detector-installed segment could then be used as a training dataset for the estimation model of the detector-uninstalled segment.

In the process of distinguishing and selecting segments in road networks based on traffic speed distribution, we first applied the Kolmogorov-Smirnov test to judge whether the distribution form was uniform between each road segment. Furthermore, with a uniform distribution form, we continue to measure the degree of variation in the speed distribution between road segments. In present studies of quantifying the distribution difference in estimation of probe traffic parameters, the Kullback-Leibler (KL) divergence and the Hellinger distance are generally employed to describe this differences [29,49]. We select KL divergence to calculate the distance between different speed distributions in this study, which evolves from the relative entropy in the information system [50]. The distribution of probe speed can be shown in Figure 5. Segments similar to the objective segment to be estimated were selected.

We presented an instance of subnetwork 2 with a sparse deployment of the LPR detector, which was used to illustrate the selection of the road segment in our proposed TCA-based volume estimation model. In the road network topology shown in Figure 6, we labeled all the road segments in the network. For instance, the 4-digit number "6611" denotes the LPR detector-installed segment and Nos. 1–99 represent the segment numbers. Assuming that we estimate the traffic volume in LPR detector-uninstalled segment No. 52, we need to find the corresponding similar segment No. 53 or 54 first, and if they are LPR detector-installed segments, they can be directly used for modeling. Otherwise, we need to find other LPR detector-installed segments that are as similar as possible to replace. Suppose that the similar segment Nos. 91 and 92 of road segment No. 90 are not LPR detector-installed segments; then, segment No. 3 or 9, which have different road grades from road segment No. 90, are selected for modeling. This is exactly the application scenario of our proposed TCA-based model, which was used to increase the robustness of network volume estimation.

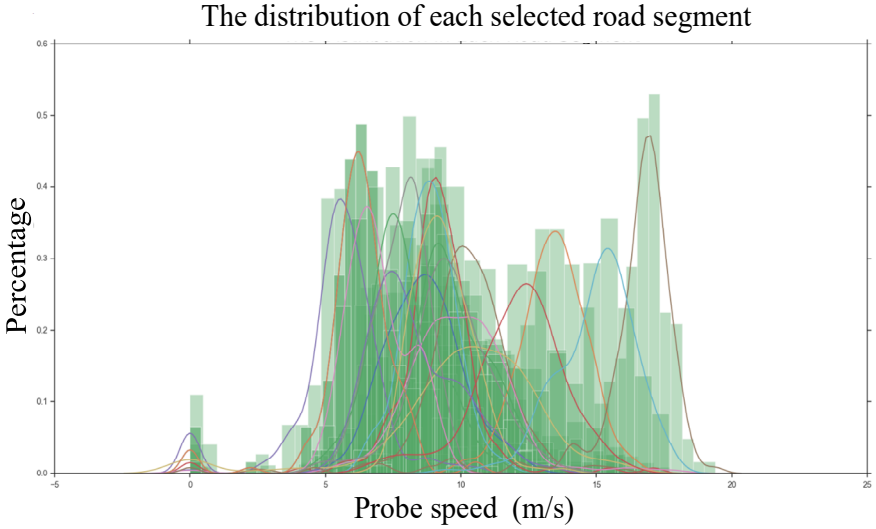



**Figure 5.** The taxi speed distribution in each segment.

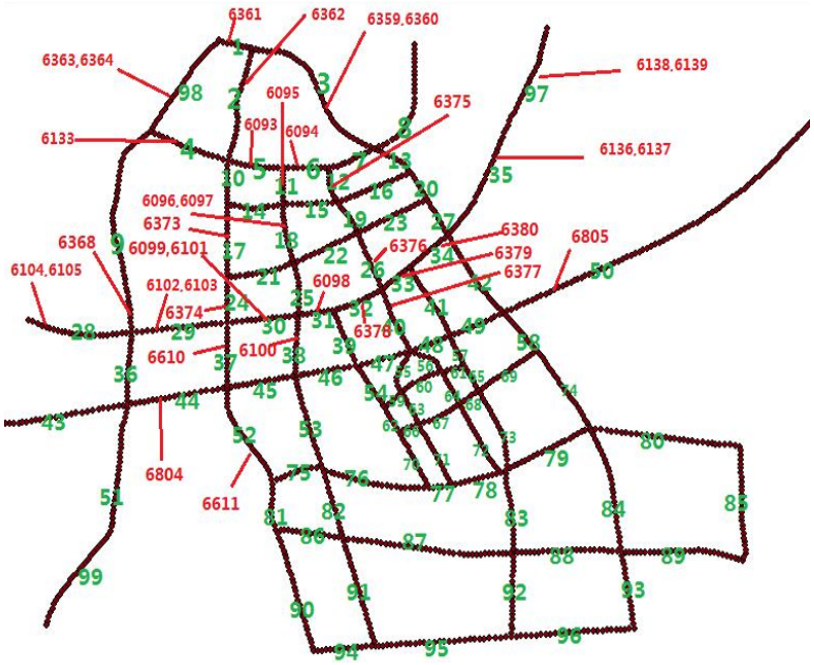



**Figure 6.** The notation of research segment in sub-network 2.

*4.4. Experimental settings*

In this study, the k-nearest neighbors (KNN) algorithm was chosen as a baseline regressor, which is computationally efficient since the hyperparameter training process is skipped. It can be viewed as a lazy machine learning method. In the training of the source domain dataset, two types of probe detector data were used for feature input. The traffic volume extracted by the LPR detector was viewed as the output. For each probe, the average speed information is evolved from traffic volume using fundamental diagrams (FD) of traffic flow [51,52]. However, limited the fact that the intrinsic relationship between traffic volume, density and speed in FD curves are different for each individual segment, and lies in the need for calibration using sufficient amount of traffic data. Hence, we delicately translate the probe speed information into the number of probe vehicles in different speed intervals as input for our data-driven model.

In this study, we divided the intervals by speed thresholds, and each interval represented one input feature unit. To incorporate sufficient common features between the source and target domains in the Hilbert space transformation process, the feature information from both probe detector datasets should be reasonably mined. Note that we set 5 km/h as an interval unit, and the feasible field, combined with the actual maximum speed limit, was set from 0 to 50 km/h, and the two types of data were sequentially divided into a total of 20 intervals.

In the parameter settings, the TCA-based estimation model was applied to both domain data as a dimensionality-reduction procedure. The number of dimensions indicated the number of transformed features. Adequate features contributed to achieving good performance, but this also decreased the computational efficiency and overfitted the distance between two domain distributions. Thus, we set the grid search method to evaluate the number of dimensions by empirically searching the threshold space for optimal performance settings and reporting the best results.

In our experimental setup, the dimensionality of the latent space varied from 5 to 50, we balanced

the training time and estimation performance, and it was fixed at 10. Herein, we experimented with the radial basis function (RBF), Laplacian, and linear kernels for feature extraction or reweighting in TCA. As for the time-series properties of our data, we finally set the RBF kernel in our experiment. Furthermore, our proposed TCA-based estimation model involved other two parameters, kernel width $\sigma$ and regularization parameter $\lambda$ .We first set $\lambda = 1$ and searched for the best $\sigma$ value in the range $\sigma \in \{0.01, 0.1, 1, 10, 100\}$, and then we fixed $\sigma$ and searched for the best $\lambda$ value also at $[10^{-2},\ 10^{2}]$. As traffic speed data is gradual sequence data, which is insensitive to the setting of $\sigma$ in TCA-based estimation model, it was set at 10. In addition, the wide range of $\lambda$ for the estimation performance was stable.

### *4.5. Evaluation*

To evaluate the performance of our proposed TCA-based volume estimation model, we performed a comparative analysis of scenarios with different road subnetworks, different combinations of road segments, and different amounts of probe detector data, with setting other baseline scenarios.

To prove the performance of the proposed method compared to other models in the intelligence-based models, multiple linear regression (MLR), weighted mean approach (WMA) [53,54], and artificial neural networks (ANN) are viewed as baselines for comparing the performance. The mean absolute percentage error (MAPE) is generally applied to measure the approximation between the estimated and actual traffic volumes by calculating the average of absolute differences [55,56]. The definition is as

$$MAPE_j = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\left|q_{ij}-\widehat{q}_{ij}\right|}{q_{ij}}\right]\cdot 100 \tag{14}$$

where $q_{ij}$ is the actual volume in $i$ th ($i = 1, \cdots, N$) time slot from $j$ th ($j = 1, \cdots, M$) segment, $\hat{q}_{ij}$ is the estimated volume in $i$ th time slot from $j$ th segment.

In subnetwork 1, the source domain dataset was set to 35 road segments with LPR detectors installed, and the target domain dataset was the remaining 41 road segments without LPR detectors installed. Meanwhile, in subnetwork 2, the source and target domain datasets were 22 and 77 segment sets, respectively. Our proposed framework selected a detector-installed segment in the source domain set to be viewed as a training dataset and a model with features from an objective detector-uninstalled segment in the target domain. The number of input features was 20, as determined in the aforementioned section, and the time slot was set to 15 min. Hence, in each modeling process, the source domain dataset was $D_s = \{20 \times 96, 1 \times 96\}$and the target domain dataset was $D_t = \{20 \times 96\}$. Because the objective detector-uninstalled segment does not have actual traffic volume data, to validate the estimated model performance, we artificially removed the labels of the source domain, treated it as a detector-uninstalled segment in the target domain, and compared the estimated traffic volume with the actual value. Furthermore, to make the results representative, we selected the average of three measurements for comparison.

In our experiments, we set up scenarios with different subnetworks and road segments with different levels of traffic volume distribution differences and evaluated the model performance by inputting different amounts of probe detector data. In setting up different levels of traffic distribution differences of road segments, we divided the set of source domain road segments into two categories

based on the KL divergence values: 1) category 1: road segments with smaller distribution differences; and 2) category 2: road segments with larger distribution differences. Furthermore, because of the lack of a CS dataset in subnetwork 1, we only applied the taxi GPS dataset to model the traffic volume estimation.

**Table 2.** MAPE performance comparison with other baselines.

| Method | Sub-network 1 (%) | | | Sub-network 2 (%) | | | | |
|---|---|---|---|---|---|---|---|---|
| | Category 1 | Category 2 | Difference | Category 1 | | Category 2 | | Average difference |
| | | | | 1 data | 2 data | 1 data | 2 data | |
| MLR | 20.58 | 25.86 | 5.28 | 24.58 | 21.17 | 28.09 | 23.26 | 2.80 |
| WMA | 25.35 | 31.24 | 5.89 | 26.13 | 24.84 | 32.82 | 30.57 | 6.20 |
| ANN | 19.32 | 23.36 | 4.04 | 21.26 | 19.75 | 25.14 | 20.27 | 2.20 |
| TCA | 18.13 | 20.14 | 2.02 | 19.16 | 18.31 | 18.74 | 17.21 | 0.75 |

Table 2 shows a comparison of the MAPE results for different research scenarios. The results are shown in Figure 7 for a better interpretation. In general, we observed that the TCA-based model performs significantly better than the three baseline methods, with statistical significance. In modeling two categories of segments with different levels of traffic distribution differences, we found that category 1 outperformed category 2. In particular, the difference in the performance of the proposed TCA-based estimation model between the two categories is smaller, only about 1%, and much lower than the difference between the two categories by the three baselines, which is almost in the range of 4–5%. This verifies that the TCA-based model can construct a more effective and robust representation of urban network-wide traffic volume estimation, which is particularly suitable in the case of the sparse deployment of fixed detectors in road networks.

Second, in the modeling of the two sub-networks, we noticed that the performance in subnetwork 1 generally performed better than that in subnetwork 2 in all baseline models, which is attributed to the higher LPR detector deployment density in subnetwork 1 than in subnetwork 2. Thus, it is relatively easy to find similarly distributed segments in the source-domain set. Furthermore, it should be noted that the performance of the TCA-based model in subnetwork 1 is not significantly different than that in subnetwork 2, which also indicates that it is preferred for networks with high uncertainty in the deployment of LPR detectors.

Third, in the sensitivity analysis of different types of probe detector data, we detected a discrepancy between the performances using two-probe and one-probe detector data with different sub-networks. In the performance of the baselines, the use of one more floating vehicle data in category 2 improved the performance. However, in category 1, the performance decreased. Furthermore, the estimation performance of the TCA-based model improved with the increase in the data type, which is mainly because the increase in the input data type enriches the information of the features.

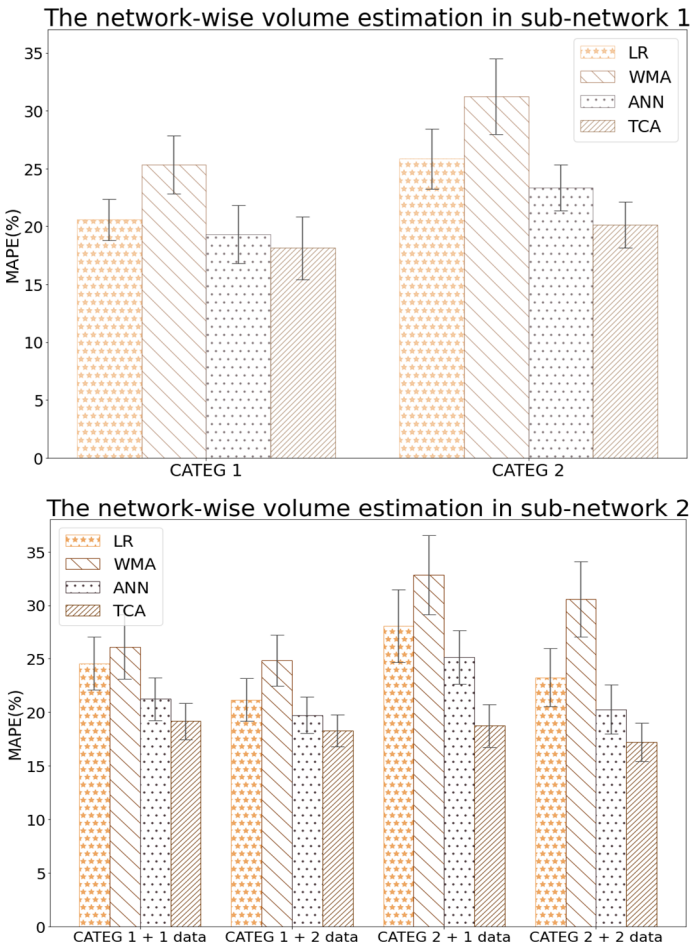



**Figure 7.** The result of MAPE from different baselines of the volume estimation.

Lastly, in the performance of different scenarios, MLR and WMA, as a type of weight regression model, can achieve fast estimation results, but lack the process of model training and perform poorly in volume estimation. The ANN model, as a classical machine learning model, performs well when the detector deployment density is high and the volume distribution difference is small, even better than the proposed TCA-based model. However, its excessive hyperparameter-setting process increases the risk of overfitting.

## 5. Conclusions

This study addresses the network-wide volume estimation model under different volume distributions with the fusion of different types of probe detector data. Real-world CS and taxi GPS data were used to improve the traffic input features of our proposed approach in two subnetworks. To solve the sparse fixed detector deployment problem, we developed a TCA-based network-wide volume-estimation model. The numerical results demonstrate that the proposed method outperforms MLR, WMA, and ANN in different research scenarios. Specifically, when no similar detector-installed segments can be found for undetected segments to be estimated, the MAPE by the TCA-based model maintains a stable estimated result compared to the other baselines. This agrees with our hypothesis that the TCA-based model can achieve stable performance under different volume distributions.

In this study, only the feature-based transfer learning (i.e., the proposed TCA-based) model was implemented as a network-wide traffic volume estimation model. It is worth studying the performance of models using different modeling paradigms, such as a combination of deep reinforcement and feature transfer learning models. Moreover, only the traffic volume estimation was considered in our method. Detector deployment policies based on volume estimation should be considered in the future.

**Acknowledgments**

The Authors declare that they did not receive any specific funding for this work.

**Conflict of interest**

The authors declare there is no conflict of interest.

**References**

1. Z. He, W. Zhang, N. Jia, Estimating carbon dioxide emissions of freeway traffic: a spatiotemporal cell-based model, *IEEE Trans. Intell. Transp. Syst.*, **21** (2020), 1976–1986. https://doi.org/10.1109/tits.2019.2909316
2. Q. Cheng, Z. Liu, Y. Lin, X. Zhou, An s-shaped three-parameter (S3) traffic stream model with consistent car following relationship, *Transp. Res. Part B Methodol.*, **153** (2021), 246–271. https://doi.org/10.1016/j.trb.2021.09.004
3. Q. Cheng, Z. Liu, J. Guo, X. Wu, R. Pendyala, B. Belezamo, et al., Estimating key traffic state parameters through parsimonious spatial queue models， *Transp. Res. Part C Emerging Technol.*, **137** (2022), 103596. https://doi.org/10.1016/j.trc.2022.103596
4. Z. Shan, D. Zhao, Y. Xia, Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model, in *16th International IEEE Conference on Intelligent Transportation Systems*, **1** (2013), 118–123. https://doi.org/10.1109/ITSC.2013.6728220
5. Z. Liu, Z. Li, M. Li, W. Xing, D. Lu, Mining road network correlation for traffic estimation via compressive sensing, *IEEE Trans. Intell. Transp. Syst.*, **17** (2016), 1880–1893. https://doi.org/10.1109/tits.2016.2514519
6. Z. He, G. Qi, L. Lu, Y. Chen, Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data, *Transp. Res. Part C Emerging Technol.*, **108** (2019), 320–339. https://doi.org/10.1016/j.trc.2019.10.001
7. J. Aslam, S. Lim, X. Pan, D. Rus, City-scale traffic estimation from a roving sensor network, in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, **2012** (2012), 141–154. https://doi.org/10.1145/2426656.2426671
8. Y. Song, X. Wang, G. Wright, D. Thatcher, P. Wu, P. Felix, Traffic volume prediction with segment-based regression kriging and its implementation in assessing the impact of heavy vehicles, *IEEE Trans. Intell. Transp. Syst.*, **20** (2019), 232–243. https://doi.org/10.1109/tits.2018.2805817
9. Z. Liu, Z. Wang, Q. Cheng, R. Yin, M. Wang, Estimation of urban network capacity with second-best constraints for multimodal transport systems, *Transp. Res. Part B Methodol.*, **152** (2021), 276–294. https://doi.org/10.1016/j.trb.2021.08.011

10. Z. Cheng, J. Lu, H. Zhou, Y. Zhang, L. Zhang, Short-Term traffic flow prediction: an integrated method of econometrics and hybrid deep learning, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 5231–5244. https://doi.org/10.1109/TITS.2021.3052796
11. L. Li, J. Zhang, Y. Wang, B. Ran, Missing value imputation for traffic-related time series data based on a multi-view learning method, *IEEE Trans. Intell. Transp. Syst.*, **20** (2019), 2933–2943. https://doi.org/10.1109/tits.2018.2869768
12. C. Meng, X. Yi, L. Su, J. Gao, Y. Zheng, City-wide traffic volume inference with loop detector data and taxi trajectories, in *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, **2** (2017), 1–10. https://doi.org/10.1145/3139958.3139984
13. Z. Yi, X. C. Liu, N. Markovic, J. Phillips, Inferencing hourly traffic volume using data-driven machine learning and graph theory, *Comput. Environ. Urban Syst.*, **85** (2021), 101548. https://doi.org/10.1016/j.compenvurbsys.2020.101548
14. Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, J. Zhang, Urban traffic prediction from spatio-temporal data using deep meta learning, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, **2019** (2019), 1720–1730. https://doi.org/10.1145/3292500.3330884
15. J. Li, N. Xie, K. Zhang, F. Guo, S. Hu, X. Chen, Network-scale traffic prediction via knowledge transfer and regional MFD analysis, *Transp. Res. Part C Emerging Technol.*, **141** (2022), 12–34. https://doi.org/10.1016/j.trc.2022.103719
16. S. Luan, R. Ke, Z. Huang, X. Ma, Traffic congestion propagation inference using dynamic Bayesian graph convolution network, *Transp. Res. Part C Emerging Technol.*, **135** (2022), 103526. https://doi.org/10.1016/j.trc.2021.103526
17. Z. Liu, Y. Liu, Q. Meng, Q. Cheng, A tailored machine learning approach for urban transport network flow estimation, *Transp. Res. Part C Emerging Technol.*, **108** (2019), 130–150. https://doi.org/10.1016/j.trc.2019.09.006
18. Q. Cheng, Z. Liu, W. Y. Szeto, A cell-based dynamic congestion pricing scheme considering travel distance and time delay, *Transportmetrica B Transport Dyn.*, **7** (2019), 1286–1304 https://doi.org/10.1080/21680566.2019.1602487
19. Q. Cheng, S. Wang, Z. Liu, Y. Yuan, Surrogate-based simulation optimization approach for day-to-day dynamics model calibration with real data, *Transp. Res. Part C Emerging Technol.*, **105** (2019), 422–438. https://doi.org/10.1016/j.trc.2019.06.009
20. D. Huang, J. Xing, Z. Liu, Q. An, A multi-stage stochastic optimization approach to the stop-skipping and bus lane reservation schemes, *Transportmetrica A Transport Sci.*, **17** (2021), 1272–1304. https://doi.org/10.1080/23249935.2020.1858206
21. D. Huang, Y. Wang, S. Jia, Z. Liu, S. Wang, A Lagrangian relaxation approach for the electric bus charging scheduling optimisation problem, *Transportmetrica A Transport Sci.*, **202232** (2022). https://doi.org/10.1080/23249935.2021.2023690
22. Z. Liu, X. Chen, Q. Meng, I. Kim, Remote park-and-ride network equilibrium model and its applications, *Transp. Res. Part B Methodol.*, **117** (2018), 37–62. https://doi.org/https://doi.org/10.1016/j.trb.2018.08.004
23. J. Xing, W. Wu, Q. Cheng, R. Liu, Traffic state estimation of urban road networks by multi-source data fusion: review and new insights, *Physica A Stat. Mech. Appl.*, **595** (2022), 127079. https://doi.org/https://doi.org/10.1016/j.physa.2022.127079

24. Z. Zhang, X. Lin, M. Li, Y. Wang, A customized deep learning approach to integrate network-scale online traffic data imputation and prediction, *Transp. Res. Part C Emerging Technol.*, **132** (2021), 103372. https://doi.org/10.1016/j.trc.2021.103372
25. L. Li, R. Jiang, Z. He, X. Chen, X. Zhou, Trajectory data-based traffic flow studies: a revisit, *Transp. Res. Part C Emerging Technol.*, **114** (2020), 225–240. https://doi.org/10.1016/j.trc.2020.02.016
26. Z. Liu, P. Zhou, Z. Li, M. Li, Think like a graph: real-time traffic estimation at city-scale, *IEEE Trans. Mobile Comput.*, **18** (2019), 2446–2459. https://doi.org/10.1109/tmc.2018.2873642
27. Q. Cao, G. Ren, D. Li, J. Ma, H. Li, Semi-supervised route choice modeling with sparse automatic vehicle identification data, *Transp. Res. Part C Emerging Technol.*, **121** (2020). https://doi.org/10.1016/j.trc.2020.102857
28. Y. Yu, X. Tang, H. Yao, X. Yi, Z. Li, Citywide traffic volume inference with surveillance camera records, *IEEE Trans. Big Data*, **7** (2021), 900–912. https://doi.org/10.1109/tbdata.2019.2935057
29. X. Zhan, Z. Yu, X. Yi, S. V. Ukkusuri, Citywide traffic volume estimation using trajectory data, *IEEE Trans. Knowl. Data Eng.*, **29** (2017), 272–285. https://doi.org/10.1109/TKDE.2016.2621104
30. P. Wang, Z. Huang, J. Lai, Z. Zheng, Y. Liu, T. Lin, Traffic speed estimation based on multi-source GPS data and mixture model, *IEEE Trans. Intell. Transp. Syst.*, **23** (2021), 10708–10720. https://doi.org/10.1109/tits.2021.3095408
31. M. Seppecher, L. Leclercq, A. Furno, D. Lejri, T. V. da Rocha, Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths, *Transp. Res. Part C Emerging Technol.*, **129** (2021), 103183. https://doi.org/10.1016/j.trc.2021.103183
32. E. Saffari, M. Yildirimoglu, M. Hickman, Data fusion for estimating Macroscopic Fundamental Diagram in large-scale urban networks, *Transp. Res. Part C Emerging Technol.*, **137** (2022), 103555. https://doi.org/10.1016/j.trc.2022.103555
33. M. Rodriguez-Vega, C. Canudas-de-Wit, H. Fourati, Dynamic density and flow reconstruction in large-scale urban networks using heterogeneous data sources, *Transp. Res. Part C Emerging Technol.*, **137** (2022), 103569. https://doi.org/10.1016/j.trc.2022.103569
34. M. Yun, W. Qin, Minimum sampling size of floating cars for urban link travel time distribution estimation, *Transp. Res. Rec. J. Transp. Res. Board*, **2673** (2019), 24–43. https://doi.org/10.1177/0361198119834297
35. Z. Huang, X. Ling, P. Wang, F. Zhang, Y. Mao, T. Lin, et al., Modeling real-time human mobility based on mobile phone and transportation data fusion, *Transp. Res. Part C Emerging Technol.*, **96** (2018), 251–269. https://doi.org/10.1016/j.trc.2018.09.016
36. C. Wu, I. Kim, H. Chung, The effects of built environment spatial variation on bike-sharing usage: a case study of Suzhou, China, *Cities*, **110** (2021), 103063. https://doi.org/10.1016/j.cities.2020.103063
37. C. Wu, H. Chung, Z. Liu, I. Kim, Examining the effects of the built environment on topological properties of the bike-sharing network in Suzhou, China, *Int. J. Sustainable Transp.*, **15** (2021), 338–350. https://doi.org/10.1080/15568318.2020.1780652
38. S. J. Pan, I. W. Tsang, J. T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, *IEEE Trans. Neural Networks*, **22** (2011), 199–210. https://doi.org/10.1109/TNN.2010.2091281

39. S. J. Pan, J. T. Kwok, Q. Yang, Transfer learning via dimensionality reduction, in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, **2** (2008), 677–682. https://dl.acm.org/doi/abs/10.5555/1620163.1620177
40. J. Wang, Y. Chen, W. Feng, H. Yu, Q. Yang, Transfer learning with dynamic distribution adaptation, *ACM Trans. Intell. Syst. Technol.*, **11** (2020), 1–25. https://doi.org/10.1145/3360309
41. Z. Cheng, L. Zhang, Y. Zhang, S. Wang, W. Huang, A systematic approach for evaluating spatiotemporal characteristics of traffic violations and crashes at road intersections: an empirical study, *Transportmetrica A Transp. Sci.*, **2022** (2022), 1–23. https://doi.org/10.1080/23249935.2022.2060368
42. Y. Gu, A. Chen, S. Kitthamkesorn, Accessibility-based vulnerability analysis of multi-modal transportation networks with weibit choice models， *Multimodal Transp*., **1** (2022), 100029. https://doi.org/10.1016/j.multra.2022.100029
43. Y. Zheng, W. Li, F. Qiu, A slack arrival strategy to promote flex-route transit services, *Transp. Res. Part C Emerging Technol.*, **92** (2018), 442–455. https://doi.org/10.1016/j.trc.2018.05.015
44. S. Wang, D. Yu, M. P. Kwan, L. Zheng, H. Miao, Y. Li, The impacts of road network density on motor vehicle travel: an empirical study of Chinese cities based on network theory, *Transp. Res. Part A Policy Pract.*, **132** (2020), 144–156. https://doi.org/10.1016/j.tra.2019.11.012
45. Z. Zhang, M. Li, X. Lin, Y. Wang, Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data, *Transp. Res. Part C Emerging Technol.*, **121** (2020), 102870. https://doi.org/10.1016/j.trc.2020.102870
46. D. Huang, S. Wang, A two-stage stochastic programming model of coordinated electric bus charging scheduling for a hybrid charging scheme, *Multimodal Transp.*, **1** (2022), 100006. https://doi.org/10.1016/j.multra.2022.100006
47. D. Xiao, I. Kim, N. Zheng, Recent advances in understanding the impact of built environment on traffic performance, *Multimodal Transp*., **1** (2022), 100034. https://doi.org/10.1016/j.multra.2022.100034
48. J. Xing, Z. Liu, C. Wu, S. Chen, Traffic volume estimation in multimodal urban networks using cellphone location data, *IEEE Intell. Transp. Syst. Mag.*, **11** (2019), 93–104. https://doi.org/10.1109/mits.2019.2919593
49. A. H. F. Chow, Z. C. Su, E. M. Liang, R. X. Zhong, Adaptive signal control for bus service reliability with connected vehicle technology via reinforcement learning, *Transp. Res. Part C Emerging Technol.*, **129** (2021), 103264. https://doi.org/10.1016/j.trc.2021.103264
50. S. Wang, D. Yu, X. Ma, X. Xing, Analyzing urban traffic demand distribution and the correlation between traffic flow and the built environment based on detector data and POIs, *Eur. Transport Res. Rev.*, **10** (2018). https://doi.org/10.1186/s12544-018-0325-5
51. R. Zhong, J. Luo, H. Cai, A. Sumalee, F. Yuan, A. Chow, Forecasting journey time distribution with consideration to abnormal traffic conditions, *Transp. Res. Part C Emerging Technol.*, **85** (2017), 292–311. https://doi.org/https://doi.org/10.1016/j.trc.2017.08.021
52. W. Qin, X. Ji, F. Liang, Estimation of urban arterial travel time distribution considering link correlations, *Transportmetrica A Transport Sci.*, **16** (2020), 1429–1458. https://doi.org/10.1080/23249935.2020.1751341
53. S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.*, **22** (1951), 79–86. https://doi.org/10.1214/aoms/1177729694

54. Y. Jiang, O. A. Nielsen, Urban multimodal traffic assignment, *Multimodal Transp*., **1** (2022), 100027. https://doi.org/10.1016/j.multra.2022.100027
55. R. Yan, S. Wang, Integrating prediction with optimization: models and applications in transportation management, *Multimodal Transp.*, **1** (2022), 100018. https://doi.org/10.1016/j.multra.2022.100018
56. Y. Zheng, W. Li, F. Qiu, H. Wei, The benefits of introducing meeting points into flex-route transit services, *Transp Res. Part C Emerging Technol.*, **106** (2019), 98–112. https://doi.org/10.1016/j.trc.2019.07.012


AIMS Press