

## Research Article

# A Progressive Extended Kalman Filter Method for Freeway Traffic State Estimation Integrating Multisource Data

Yingshun Liu,<sup>1</sup> Shanglu He ,<sup>1</sup> Bin Ran ,<sup>2</sup> and Yang Cheng <sup>3</sup>

<sup>1</sup>School of Automation, Nanjing University of Science & Technology, No. 200 Xiao Ling Wei Street, Nanjing, Jiangsu 210094, China

<sup>2</sup>School of Transportation, Southeast University, No. 2 Si Pai Lou, Nanjing, Jiangsu 210096, China

<sup>3</sup>Department of Civil & Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53705, USA

Correspondence should be addressed to Shanglu He; slhemickey@126.com

Received 1 December 2017; Accepted 26 March 2018; Published 6 May 2018

Academic Editor: Li Zhu

Copyright © 2018 Yingshun Liu 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.

Variable techniques have been used to collect traffic data and estimate traffic conditions. In most cases, more than one technology is available. A legitimate need for research and application is how to use the heterogeneous data from multiple sources and provide reliable and consistent results. This paper aims to integrate the traffic features extracted from the wireless communication records and the measurements from the microwave sensors for the state estimation. A state-space model and a Progressive Extended Kalman Filter (PEKF) method are proposed. The results from the field test exhibit that the proposed method efficiently fuses the heterogeneous multisource data and adaptively tracks the variation of traffic conditions. The proposed method is satisfactory and promising for future development and implementation.

## 1. Introduction

Various traffic detection methods have been used for the freeway traffic surveillance and ITS application, such as loop detectors, microwave sensors, video cameras, Bluetooth sensors, cellphone probes, and GPS probes. In most cases, more than one technology is available on a freeway segment. Thus, it is a challenge to fully unitize the multisource data for the construction of traffic state. The critical issues include the following:

- (i) The heterogeneity in the multisource measurements
- (ii) The difference in the spatial or temporal resolution
- (iii) The inconsistency in the measurements when they should represent the same traffic state

Although the multisource data provide plentiful traffic information, they still have the limitation in the overall spatial or temporal coverage. A traditional method to construct the network-wide picture of traffic state with limited measurements is the model-based estimation [1–3]. One of the extensively used techniques for the estimation, known as

online Bayesian method, is Kalman filter and its extended version, such as Extended Kalman Filter (EKF). Kalman filter [4] has been implemented to solve the traffic problems with linear relation [5]. Extended Kalman Filter fits the estimation with nonlinear traffic flow models [3, 6–9]. A series of works based on the EKF algorithm conducted by Wang and Papageorgiou [3, 6–8, 10] have been applied in freeway network with data from the fixed detector. The EKF algorithm also has efficient performance on the estimation using data from probe vehicles, such as GPS-based probe technique [9], as well as cellphone activity data-based probe technique [1] which uses the wireless communication records. However, these works used a single-source data.

It has been proved that integrating multisource data can increase the accuracy, reduce the ambiguous, and improve the robust [11]. Bachmann et al. [12] provided a comparative assessment of the different data fusion techniques for traffic speed estimation, including distribution fusion techniques, Kalman filter techniques, ordered weighted averaging, fuzzy integral, and artificial neural network. The majority of recent works on traffic data fusion are data-driven approaches [11, 13–17], which required a set of training data for the calibration

of method. Hence, a data-driven method might fail when there is no sufficient data for training and testing. Mannini et al. made a further application of fused results by using EKF to estimate the traffic speed [18]. On the other hand, some model-based works using the macroscopic traffic flow models were developed to fuse multisource data. Heilmann et al. applied the linear Kalman filter to combine data from local detector and electronic toll collection (ETC) system [19]. Deng et al. [20] applied Clark's approximation method and Newell's method to fuse data from three detectors, and their estimator is also Kalman filter-based for the linear measurement equations. Recently, Nantes et al. [2] devoted to the development of an incremental EKF estimator for the fusion of three heterogeneous data sources collected from the arterials, with the assumption that all measurements are independent. Comparatively, the model-based method is superior to the data-driven method by considering the relationship among the traffic variables.

In the application of a model-based method, it is important to define the relation between the measurements and the traffic state variables. The newly developed data collection technique usually requires an update or adjustment to this relation model. The multisource data applied in this study are the traffic features that are extracted from wireless communication records and traffic measurements from microwave sensors. The traffic features from wireless communication records are not direct traffic state variables, which is a challenge for the application in the traffic state estimation. The wireless communication records are also known as the cellular communication records, including the records of handoff, normal location update, and data transition. The traffic features extracted from these wireless communication records are different from the traffic information from the handoff-based collection method [21]. Our recent work shows that the application of these traffic features in the traffic estimation is feasible [1]. This study aims to solve the integration of these traffic features with the data from the traditional fixed sensors for the traffic state estimation. Accordingly, the above three critical issues of traffic data fusion will be investigated in this study.

In this paper, we develop a Progressive Extended Kalman Filter- (PEKF-) based estimator. There are several obvious difference between the work of Nantes et al. [2] and our work. First, we use different data sources, which leads to different measurement equations in state-space model. And it also results in different semantic, spatial, and temporal synchronization of data. Moreover, the assumption of their method is the dependence of each data source, while our proposed method is based on the foreknowledge of the precision of multiple collection approaches. Besides, our study focuses on the application on the freeway while they focus on the urban arterial roads.

The organization of the rest of the paper is as follows. Section 2 presents a brief analysis on these two-source heterogeneous data. The following Section 3 describes the link-based freeway traffic model. We built up a PEKF estimator in Section 4. Section 5 applied the field data to assess the performance of the proposed approach. Finally, several key conclusions were drawn in Section 6.

## 2. Analysis on the Characteristics of the Multisource Data

*2.1. The Multisource Data.* As mentioned above, the multisource data stem from the wireless communication records and the microwave sensors. Similar to other fixed sensors (i.e., inductive loop detectors), the microwave sensors collect the value of traffic flow and speed at the spot. However, the sparse placement of sensors on the freeway in China makes this technique only achieve the traffic information from a small part of the road network. To some extent, the probe-based techniques have the larger spatial coverage. The wireless communication records of cellphones provide a source to extract traffic state information. In some studies, they used parts of these records; for instance, some used the handoff records to obtain the traffic speeds [22–24]. To achieve more samples and to better mine the potential value of these wireless communication records, this study applied all the records of wireless communication activities from the cellular network to extract traffic information. Mainly, two types of activities are included in the records: (1) system signal, such as location area updates, handoffs, and billing records, and (2) user activities, such as phone calls, text messaging, and data services, such as web browsing, sending, and checking emails [1].

The extraction procedure is named as the cellphone activity (CA) data-based method [1]. This method ambiguously locates the on-board cellphones on the freeway according to the geographic information of nearby cell towers. Two traffic features are extracted which are nominated as unique cellphone counts (UCC) and pseudo speed (PS). Compactly, the method includes the following steps. First, find out the cell towers whose signal covers the freeway. Second, project these cell towers on the freeway as the virtual sensor sites. Third, determine the moving direction of a on-board cellphone on the freeway by tracking the change of virtual sensor sites, that is, the cell towers that continuously record the activities of this cellphone. Fourth, the UCC of a freeway is calculated by counting the cellphones in the same direction and recorded by the cell towers on the same freeway link. Fifth, the PS of a cellphone ambiguously equals the distance of two adjacent virtual sensor sites divided by the time between two sequential records. Because of the ambiguous positioning of the on-board cellphones, UCC and PS could not directly equal the traffic state variables. It was found that these traffic features had a link-based characteristic, which meant that some freeway links could generate good data that closely related to the traffic state variables [1]. Besides, it was proven that the physical meaning of UCC changes depended on the traffic state. Accordingly, the relation models were set up as follows [1].

(1) The relation model of PS and traffic speed is as follows:

$$m_i^{vc}(k) = v_i(k) + \eta_i^{vc}(k), \quad (1)$$

where  $v_i(k)$  (in km/h) denotes the traffic speed at Link  $i$  at the time interval  $k$ ,  $m_i^{vc}(k)$  (in km/h) denotes the measurements of PS at Link  $i$  at the time interval  $k$ , and  $\eta_i^{vc}(k)$  denotes the corresponding PS noise.

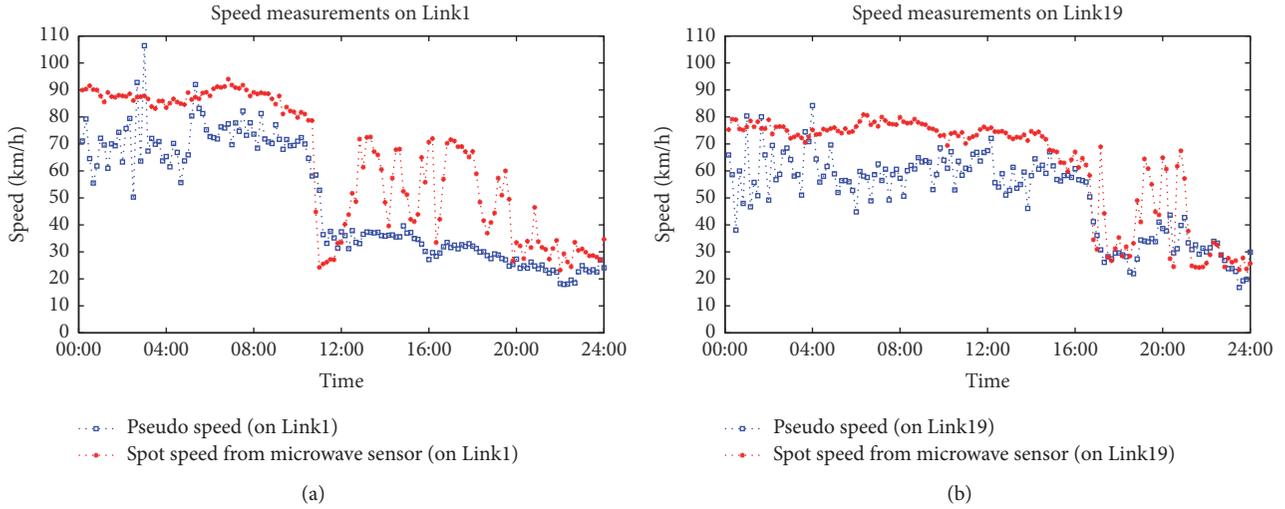


FIGURE 1: Speed measurements from multisensors: (a) on Link 1; (b) on Link 19.

(2) The relation model of UUC and the traffic state variable (density or flow) is as follows:

Since the relationship depended on the traffic state, a conditional function was developed [1]:

$$m_i^c(k) = \begin{cases} \theta(k) \cdot q_i(k) + \eta_i^{cq}(k) & \text{when } m_i^{vc}(k) \geq v_{cr} \\ \varphi \cdot \rho_i(k) + \eta_i^{cp}(k) & \text{when } m_i^{vc}(k) < v_{cr} \end{cases} \quad (2)$$

where  $q_i(k)$  (in veh/h) denotes the traffic flow at Link  $i$  during the time interval  $k$ .

$\rho_i(k)$  (in veh/km/h) denotes the average traffic density of Link  $i$  during the time interval  $k$ ,  $m_i^c(k)$  denotes the measurements of UCC during the time interval  $k$ ,  $\eta_i^{cq}(k)$  and  $\eta_i^{cp}(k)$  denote the corresponding UCC noise,  $\theta(k)$  is a linear coefficient which is a time-dependent variate,  $\varphi$  is a linear coefficient which is a constant, and  $v_{cr}$  (in km/h) is the critical speed to identify the traffic state.

The later six parameters will be discussed in Section 5.

**2.2. Characteristics of the Heterogeneous Data.** In this study, it is essential to identify the differences between the heterogeneous multisource data to guide the fusion research. Obviously, the multisource data in this study differ in semantic, temporal, and spatial coverage. The microwave sensors provide the spot measurements while the traffic features of CA data-based method are the link-average values. Besides, they exhibit different accuracy. To explore their different in the precision, we use the field data collected from the Jiangsu freeway. The comparison between the speed measurements from these two collection methods is shown in Figure 1. It can be inferred from the figure that under free-flow condition (the speed is over 50 km/h), the PSs from the wireless communication records are not stable which fluctuate over a large range. Figure 1(a) shows that the PSs are all below

50 km/h from 11:00 to 24:00 on Link 1 and from 16:40 to 24:00 on Link 19, but the spot speeds from the microwave sensors show that the traffic switches between the free-flow state and the low-speed state. The incident/accident reports from the freeway operation center indicated that the traffic speeds were slow during that time period. It proves that the PSs are more accurate under the low-speed state compared with the speeds from microwave sensors. The possible cause for the instability of the microwave measurements is that the speed is collected on the spot which could not reflect the average speed of a link. There will be stop-and-go phenomenon especially under the low-speed or congested state.

Based on the comparison, we make the following inferences to promote the construction of the estimator. First, measurements from the microwave sensors are more stable and accurate than the traffic features from wireless communication records under the free-flow condition. Second, under the low-speed or congested condition, PS is regarded as the accurate speed measurements. Obviously, it is important to identify the traffic condition when both data sources are available but conflict. According to the inferences, PS is more reliable under the low-speed or congested condition. To avoid the impact of the biased PS under the free-flow condition, we set the following rules to identify the traffic condition when speeds from two-source conflict.

The condition is low-speed or congested if one of the following conditions is met:

- (i)  $(m_i^{vc}(k) < v_{cr} \text{ and } m_i^{vc}(k-1) < v_{cr} \text{ but } m_i^{vm}(k) \geq v_{cr})$ .
- (ii)  $(m_i^{vc}(k) < v_{cr} \text{ and } m_i^{vc}(k+1) < v_{cr} \text{ but } m_i^{vm}(k) \geq v_{cr})$ .
- (iii)  $(m_i^{vm}(k) < v_{cr} \text{ but } m_i^{vc}(k) \geq v_{cr})$ .

Otherwise, it is the free-flow condition, where  $m_i^{vm}(k)$  (in km/h) denotes the speed measurement (the time mean speed) from the microwave sensors on Link  $i$  at the time interval  $k$ .

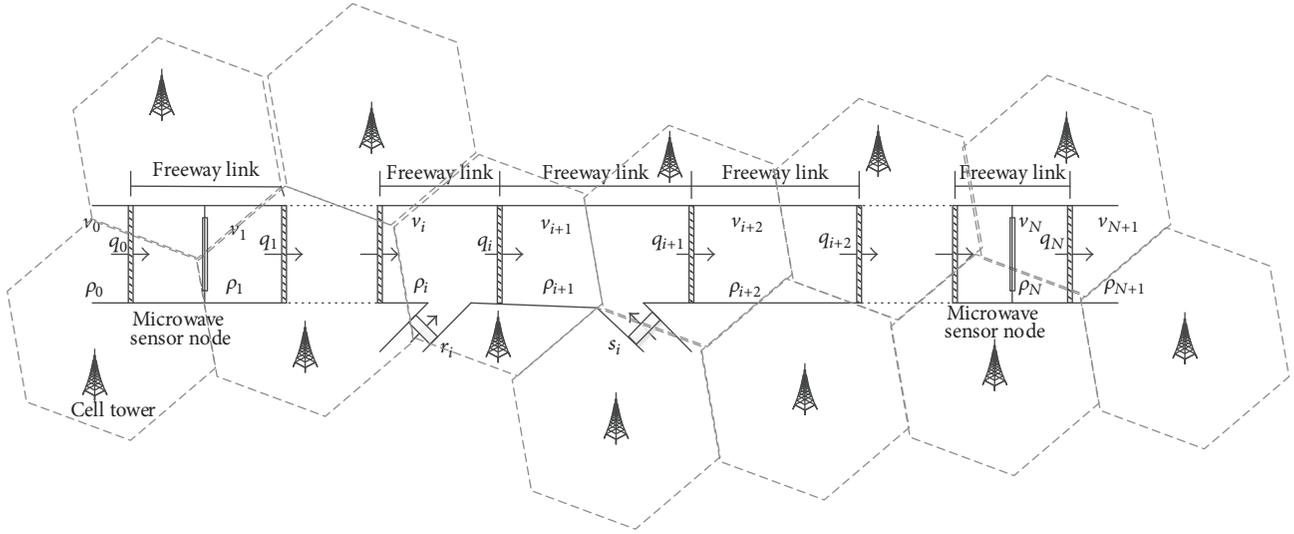


FIGURE 2: A freeway stretch, link division, sensor location, and cellular signal coverage [1].

### 3. Freeway Traffic Flow Modelling with Multisource Data

The models include a macroscopic traffic flow model and a measurement model. The macroscopic traffic flow model describes the relationship between traffic variables including speed, density, and flow. And the measurement model shows the relationship between the heterogeneous measurements and traffic variables.

**3.1. Macroscopic Freeway Traffic Flow Model.** This study utilized one of the validated macroscopic traffic flow models, that is, the METANET model [3]. The macroscopic model is discretized in space and time. As shown in Figure 2, a freeway stretch in a travel direction is divided into  $N$  links with length  $L_i$ ,  $i = 1, \dots, N$ . Time is discretized in an equal time step  $T$ .

For the spatial synchronization of multisource measurements, the division of the freeway stretch follows some rules. First, if there is a microwave sensor on a link, it is better to locate in the middle of the link (or within the link), which is different from some works that put the fixed detector at the boundary of a link [3, 6]. Second, as an advantage of CA data-based method, any length can be used as long as one or more cell towers vertically projected to a link and the signal from this or these cell towers could cover this link. Other rules including the geometric difference (e.g., lane drops, on/off-ramps) could be applied, and it is optimal to have at most one on- or off- ramp within a link [3]. In this study, we applied a uniform length of 1 kilometer as a link length shown in the following test section.

Based on the divided freeway link and time step, the applied second-order macroscopic traffic flow model for Link  $i$  ( $i$  is the number of the link) includes the following four equations.

(1) Conservation equation is

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + i_i(k) - o_i(k)]. \quad (3)$$

(2) Dynamic speed equation is

$$v_i(k+1) = v_i(k) + \frac{T}{\tau} [v_e(\rho_i(k)) - v_i(k)] + \frac{T}{L_i} v_i(k) [v_{i-1}(k) - v_i(k)] - \frac{\gamma T}{\tau L_i} \frac{[\rho_{i+1}(k) - \rho_i(k)]}{\rho_i(k) + \kappa} + \xi_i^v(k). \quad (4)$$

(3) Stationary speed equation is

$$v_e(\rho_i(k)) = v_f \exp \left[ \frac{1}{2} \left( \frac{\rho_i(k)}{\rho_{cr}} \right)^2 \right]. \quad (5)$$

(4) Flow equation is

$$q_i(k) = \rho_i(k) \cdot v_i(k) + \xi_i^q(k), \quad (6)$$

where  $T$  (in hour) is the discretized time step,  $L_i$  (in km) is the length of Link  $i$ ,  $\rho_i(k)$  (in veh/km/lane) is the traffic density of the Link  $i$  at the time interval  $k$ ,  $q_i(k)$  (in veh/h) is the traffic flow of the Link  $i$  at the time interval  $k$ ,  $i_i(k)$  (in veh/h) is the ramp inflow of the Link  $i$  at the time interval  $k$ ,  $o_i(k)$  (in veh/h) is the ramp outflow of the Link  $i$  at the time interval  $k$ ,  $v_i(k)$  (in km/h) is the space mean speed of the Link  $i$  at the time interval  $k$ ,  $\tau$ ,  $\nu$ ,  $\delta$ , and  $\kappa$  are model parameters, and all links share the same value of these parameters,  $\xi_i^v(k)$  is zero-mean Gaussian white noise in the speed equation,  $v_e(\rho_i(k))$  (in km/h) is the average corresponding speed according to

density,  $v_f$  (in km/h) is the free low speed,  $\rho_{cr}$  (in veh/km/h) is the critical density, and  $\xi_i^q(k)$  denotes zero-mean Gaussian white noise in the flow equation.

**3.2. Measurement Model and Date Synchronization.** According to the characteristics of the multisource measurements, a set of measurement equations are established as follows.

**3.2.1. Cellphone Density Measurement Equation.** The conditional function (2) shows the relationship between UCC and the traffic state variables (flow and density) under different traffic conditions. Based on the relationship among flow, density, and speed, this relationship model could be transformed as follows:

$$m_i^{pc}(k) = \begin{cases} \frac{m_i^c(k)}{\theta(k) \cdot m_i^{vc}(k)} & m_i^{vc}(k) \geq v_{cr} \\ \frac{m_i^c(k)}{\varphi} & m_i^{vc}(k) < v_{cr}, \end{cases} \quad (7)$$

where  $m_i^{pc}(k)$  (in veh/km/h) denotes the transferred density from UCC at Link  $i$  at the time interval  $k$ .

Afterwards, the cellphone density measurement model is set up as

$$m_i^{pc}(k) = \rho_i(k) + \eta_i^{pc}(k), \quad (8)$$

where  $\eta_i^{pc}(k)$  is the measurement error of UCC.

**3.2.2. Cellphone Speed Measurement Equation.** The speed measurement equation could directly apply (1) in Section 2.

**3.2.3. Speed Measurement Equation of Microwave Sensor.** Speed measurements from the microwave sensors are average spot speeds, that is, the time mean speed. It is known that there is the difference between the time mean speed and space mean speed. In the model-based estimation, it is the space mean speed required. Rakha and Zhang [25] formulated the conversion between these two kinds of speeds. This formulation works when the variance of time mean speeds is known. However, it is difficult to achieve this variance some times. According to the analysis in Section 2, it is found that the speed measurements from the microwave sensors are more stable and reliable than the PS under the free-flow condition, because the PS are oscillating under the low-speed condition. Therefore, this study proposed the following conditional function using the two-source speeds as follows.

If  $m_i^{vm}(k)$  conflicts with  $m_i^{vc}(k)$  under the low-speed or congested condition, then

$$m_i^{vm}(k) = m_i^{vc}(k) = v_i(k) + \eta_i^{vc}(k). \quad (9)$$

Otherwise,

$$m_i^{vm}(k) = v_i(k) + \eta_i^{vm}(k), \quad (10)$$

where  $\eta_i^{vm}(k)$  is the noise of speed measurements from the microwave sensor.

**3.2.4. Flow Measurement Equation of Microwave Sensor.** Consider a microwave sensor installed in the middle Link  $i$ , as illustrated in Figure 2. For the flow measurement, we have

$$m_i^{qm}(k) \times \left(\frac{60}{TI}\right) = q_i(k) + \eta_i^{qm}(k), \quad (11)$$

where TI (in min) is the time interval to collect the flow measurements.

$m_i^{qm}(k)$  (in veh/TI) is the flow measurement from the microwave sensor located at Link  $i$  during the time interval  $k$ .

$\eta_i^{qm}(k)$  denotes the corresponding flow measurement noise.

Considering the relation in flow equation (6), the traffic density can be calculated from the measurements of flow and speed. Accordingly, (11) can be transferred as follows, except for the measurement  $q_0$ .

$$\frac{m_i^{qm}(k) \times (60/TI)}{m_i^{vm}(k)} = \rho_i(k) + \frac{\xi_i^q(k) + \zeta_i^{qm}(k)}{m_i^{vm}(k)}. \quad (12)$$

## 4. Design of PEKF-Based Estimator

**4.1. State-Space Model.** Based on the traffic flow model and measurement model mentioned above, a state-space model is set up with the state vectors  $\mathbf{x}$ ,  $\mathbf{u}$ ,  $\xi$ , and  $\boldsymbol{\eta}^c, \boldsymbol{\eta}^m$  defined as follows:

$\mathbf{x} = [\rho_1 \ v_1 \ \rho_2 \ v_2 \ \cdots \ \rho_N \ v_N]^T$  denotes the density-speed state vector.

$\mathbf{u} = [q_0 \ v_0 \ \rho_{N+1} \ i_1 \ \cdots \ i_N \ o_1 \ \cdots \ o_N]^T$  denotes the boundary state vector.

$\xi = [\xi_1^q \ \xi_1^v \ \cdots \ \xi_N^q \ \xi_N^v]^T$  denotes the vector of the process noise.

$\boldsymbol{\eta}^c = [\eta_1^{pc} \ \eta_1^{vc} \ \cdots \ \eta_N^{pc} \ \eta_N^{vc}]^T$  denotes the vector of the measurement noise from CA data.

$\boldsymbol{\eta}^m = [\eta_1^{qm} \ \eta_1^{vm} \ \cdots \ \eta_N^{qm} \ \eta_N^{vm}]^T$  denotes the vector of the measurement noise from microwave sensors.

Then the macroscopic flow model and measurement model are rewritten to a compact state-space form including a process function (13) and two measurement functions (14) and (15) as follows:

$$\mathbf{x}(k) = \mathbf{f}[\mathbf{x}(k-1), \mathbf{u}(k), \xi(k-1)] \quad (13)$$

$$\mathbf{z}^c(k) = \mathbf{h}^c[\mathbf{x}(k), \boldsymbol{\eta}^c(k)] \quad (14)$$

$$\mathbf{z}^m(k) = \mathbf{h}^m[\mathbf{x}(k), \boldsymbol{\eta}^m(k)], \quad (15)$$

where the process function relates to (3), (4), (5), and (6). The measurement functions relate to (8), (1), (10), and (12).  $\mathbf{f}$ ,  $\mathbf{h}^c$ , and  $\mathbf{h}^m$  are nonlinear differential vector functions. The vector  $\mathbf{z}^c$  consists of all available measurements from CA data and  $\mathbf{z}^m$  contains the measurements from the microwave sensors. In the application, the measurements will be transferred via function (7) and (10). The random variables  $\xi(k)$ ,  $\boldsymbol{\eta}^c(k)$ , and  $\boldsymbol{\eta}^m(k)$  represent the process and measurement noise, respectively.  $k$  is the number of time steps.

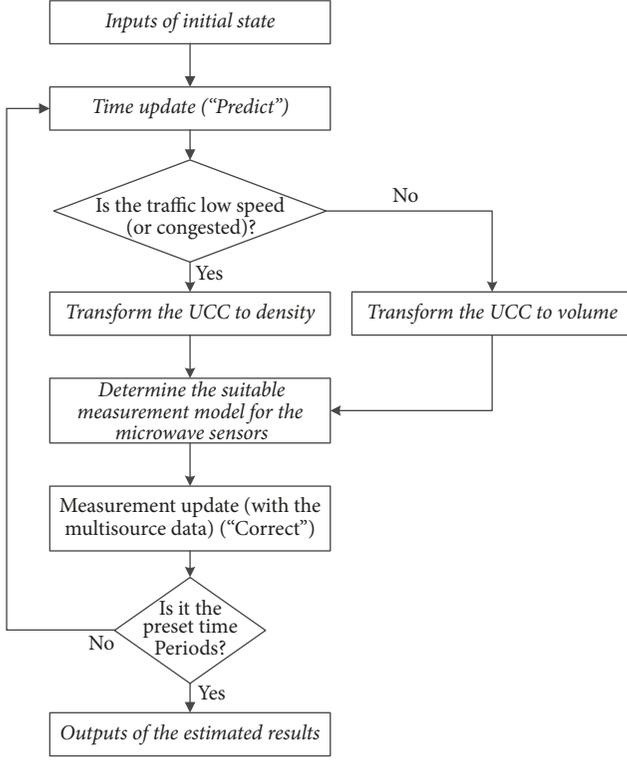


FIGURE 3: The operation of the Progressive Extended Kalman Filter estimator.

**4.2. The PEKF Estimator for Multisource Data.** The state-space model contains the nonlinear equations, and, thus, this study set up the estimator based on the Extended Kalman Filter (EKF) technique. As the exploration in the previous sections, the relationship between the multisource measurements and the traffic state variables highly relies on the judgment of the traffic condition. Therefore, we make some adjustments to the traditional EKF estimator, and the updated estimator is nominated as the progressive EKF (PEKF) estimator. The operation diagram of PEKF estimator is shown in Figure 3. The adjustments are mainly shown in the measurement update step. The purposes of these adjustments are to integrate the multisource data from the wireless communication records and microwave sensors in this study. As shown in Figure 3, the adjustments include the following contents. First, speeds from the microwave sensors (named as  $m$ -speed in the following description) and PSs are used to identify the traffic condition as shown in Section 2. Second, based on the identified traffic condition, the UCC are transformed to the corresponding traffic state variables via (7). Consequently, the speed measurement model can be decided via (10). Similar to the traditional EKF estimator, the partial derivatives of the process function and measurement function, transformed using the Jacobian matrix, are used to linearize the model [23, 26]. The proposed estimator includes the following steps.

**Step 0 (initialization).** Set time interval  $k = 1$ , and let  $E[\mathbf{x}(0)] = \widehat{\mathbf{x}}(0)$ , and  $E[(\mathbf{x}(0) - \widehat{\mathbf{x}}(0))^2] = \mathbf{p}(0)$ , where the

initial state  $\mathbf{x}(0)$  and covariance  $\mathbf{p}(0)$  are determined by the historical data.

**Step 1 (time update).** State estimate extrapolation is as follows:  $\widehat{\mathbf{x}}(k)^- = \mathbf{f}[\widehat{\mathbf{x}}(k-1), \mathbf{u}(k), 0]$

Error covariance extrapolation is

$$\mathbf{P}(k)^- = \mathbf{A}(k) \cdot \mathbf{P}(k-1) \cdot \mathbf{A}(k)^T + \mathbf{W}(k) \cdot \mathbf{Q}(k-1) \cdot \mathbf{W}(k)^T, \quad (16)$$

where  $\mathbf{P}(k)$  is the error covariance at the time step  $k$ .

**Step 2 (measurement update)**

**Step 2.1.** Identify the traffic state based on the rules in Section 2.2.

**Step 2.2.** Transform UCC to traffic density according to the traffic state using the conditional (7).

**Step 2.3.** Make a decision on the speed measurement model for the microwave sensor according to the traffic state by the conditional (10).

**Step 2.4.** Update the input measurement matrix  $\mathbf{h}$ .

**Step 2.5.** Measurement update using (14) and (15) is as follows. Kalman gain calculation is

$$\mathbf{K}(k) = \mathbf{P}(k)^- \cdot \mathbf{H}(k)^T \cdot \left( \mathbf{H}(k) \cdot \mathbf{P}(k)^- \cdot \mathbf{H}(k)^T + \mathbf{V}(k) \cdot \mathbf{R}(k) \cdot \mathbf{V}(k)^T \right)^{-1} \quad (17)$$

State estimate update is as follows:  $\widehat{\mathbf{x}}(k) = \widehat{\mathbf{x}}(k)^- + \mathbf{K}(k) \cdot (\mathbf{z}(k) - \mathbf{h}(\widehat{\mathbf{x}}(k)^-), 0)$ , where  $\mathbf{z}$  could be  $\mathbf{z}^c$  or  $\mathbf{z}^m$  and  $\mathbf{h}$  could be  $\mathbf{h}^c$  or  $\mathbf{h}^m$  depending on which source data is using.

Error covariance update is as follows:  $\mathbf{P}(k) = (\mathbf{I} - \mathbf{K}(k) \cdot \mathbf{H}(k)) \cdot \mathbf{P}(k)^-$ .

**Step 3.** Let  $k = k + 1$  and go back to Step 1 until the preset time periods end, where  $\mathbf{I}$  is the identity matrix.

$\mathbf{A}$  and  $\mathbf{W}$  are the Jacobian matrix of partial derivatives of  $\mathbf{f}$  with respect to  $\mathbf{x}$  and  $\boldsymbol{\xi}$ , respectively:

$$\begin{aligned} \mathbf{A}(k) &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \left( \widehat{\mathbf{x}}(k-1), \mathbf{u}(k), 0 \right) \\ \mathbf{W}(k) &= \frac{\partial \mathbf{f}}{\partial \boldsymbol{\xi}} \left( \widehat{\mathbf{x}}(k-1), \mathbf{u}(k), 0 \right). \end{aligned} \quad (18)$$

$\mathbf{H}$  and  $\mathbf{V}$  are the Jacobian matrix of partial derivatives of  $\mathbf{h}$  with respect to  $\mathbf{x}$  and  $\boldsymbol{\eta}$ .  $\mathbf{h}$  is the updated measurement matrix stemming from the multiple sources.  $\boldsymbol{\eta}$  is the corresponding noise matrix compounding  $\boldsymbol{\eta}^c$  and  $\boldsymbol{\eta}^m$ :

$$\begin{aligned} \mathbf{H}(k) &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \left( \widehat{\mathbf{x}}(k), 0 \right) \\ \mathbf{V}(k) &= \frac{\partial \mathbf{h}}{\partial \boldsymbol{\eta}} \left( \widehat{\mathbf{x}}(k), 0 \right). \end{aligned} \quad (19)$$

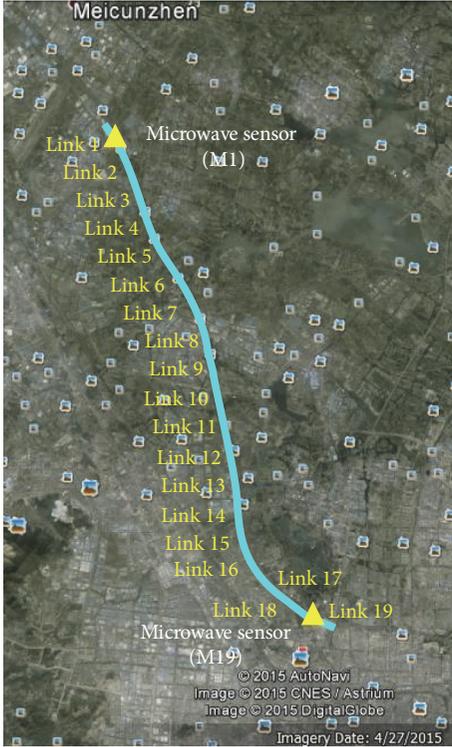


FIGURE 4: The test bed, part of Ning-Hu freeway, and a northbound segment. *Note.* The background photo is from Google Earth.

## 5. Test and Results

### 5.1. A Test Freeway Segment and Data Collection

**5.1.1. Freeway Test Bed.** The test bed is a 19 km, four-lane northbound segment of Ning-Hu freeway connecting Wuxi and Suzhou, China, which is part of the busiest freeway in China, as shown in Figure 4. Two microwave sensors (M1 and M19) are located near the ends of the segment. According to the freeway-division rules, the segment is divided into 19 links with uniform lengths of 1 km each. The links are labelled Link 1 to Link 19 from north to south opposite to the traveling direction. Microwave sensors are within Link 1 and Link 19, respectively. The segment is almost geometrically homogeneous without any lane drops.

**5.1.2. Data Collection.** The multisource data for the field test was provided by the Jiangsu Freeway Operation Center and was collected on September 30, 2014. Both collection methods (i.e., the microwave sensors and the CA data-based method) integrated their measurements by a time interval of ten minutes starting from 00:00, September 30, 2014. The spatial coverage of the microwave sensor is limited to Link 1 and Link 19, while the cellular signal covers all links.

In addition to these two data sources, there is a traffic incident/accident report which could be used in the following evaluation of the estimation results. The report indicates that an incident occurred on Link 4 at about 11:12 AM. The traffic was also quite heavy. These two traffic phenomena together

TABLE 1: Parameter settings.

Parameter	Value
$\tau$	0.02 h
$\kappa$	40 veh/km
$v_f$	100 km/h
$\nu$	24 km <sup>2</sup> /h
$\rho_{cr}$	120 veh/km

lead to the traffic moving slowly from Link 1 to Link 10 as well as the nearby affected links. The report also shows that the traffic moved slowly from 10:00 to the late evening of that day due to the heavy traffic. The cause is the China National Day Holiday which begins from October 1 and the free tolling policy on freeway that works during this holiday, and, thus, it attracts tremendous traffic. Therefore, the traffic flow climbs high when it is closing to the midnight of the test day. Consequently, the high traffic flow triggered the speed drop and the large density.

### 5.2. Model Parameters

**5.2.1. Parameters about Time.** The time step  $T$  of the estimator is 10 seconds to meet the requirement that  $T < L/v_f = 1 \text{ km}/100 \text{ (km/h)} = 36 \text{ seconds}$ . The time interval for measurement update is 10 minutes. The estimator works when the measurements are available. It means that the time update and measurement update will work 60 steps (10 min/10 sec) using the same measurements.

**5.2.2. Parameters about Links and Boundaries.** Although there are 19 links, the measurements from the microwave sensors are only available on two links (Link 1 and Link 19). Our recent work found that the traffic features (UCC and PS) extracted from the wireless communication records on Link 6, Link 10, Link 14, and Link 18 had the closest relation with the traffic variables [1]. This study uses the UCC and PS from these links for the sake of high accuracy. Besides, the PSs on Link 1 and Link 19 are also used under the low-speed or congested condition, because these PSs are more reliable than the microwave speeds under this condition as discussed in Section 2. To avoid the impact of inaccurate boundary inputs, Link 1 and Link 19 are regarded as the boundary links, and, thus, the measurements on these links are used as the boundary variables.

**5.2.3. Parameters in the Macroscopic Traffic Flow Model.** The empirical parameters are set as Table 1. And the models of all links shared the same values. This study uses the value suggested in the work of Wang et al. [6] as the process noise.

**5.2.4. Parameters in the Measurement Model.** The coefficients in the measurement model for the UCC were calibrated using the historical data as follows [1]. The constant parameter  $\varphi$  approximately equals 8.5 for all links (Link 6, Link 10, Link 14, and Link 18). The time-varying parameter  $\theta(k)$  is valued as follows: from the time 00:00 to 05:00, it approximately equals

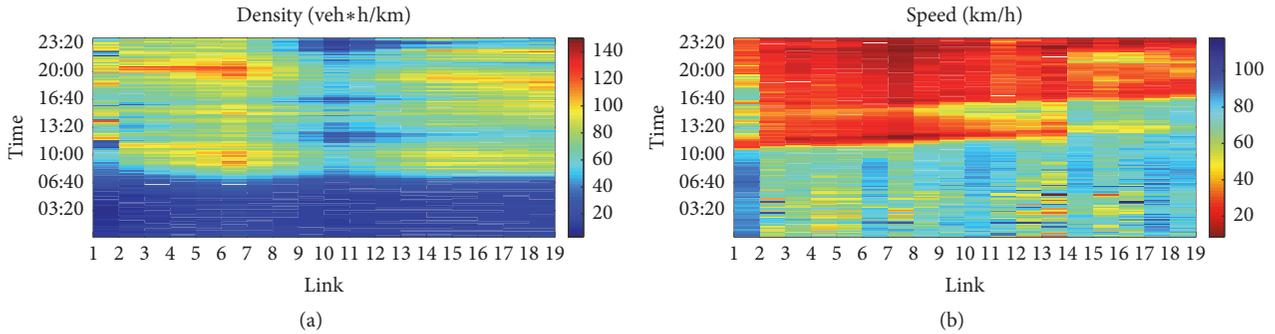


FIGURE 5: The estimation results. (a) The density estimates. (b) The speed estimates.

0.3; from 05:00 to 07:30, it gradually increases from 0.3 to 1.0; from 07:30 to the rest of a day, it approximately equals a constant 1.0 under the free-flow condition. Another critical parameter in the measurement update step is the critical speed for the identification of traffic condition. It is valued empirically as 50 km/h.

The measurement noise should be the standard deviation of measurements from cellphone activity data and microwave sensors. However, it is difficult to value the noise when there is no sufficient ground-truth data. Some existing studies indicate that the Extended Kalman Filter was found to have a slight sensitivity to the standard deviation-based measurement noise, which guaranteed the estimator's performance even with poor foreknowledge of model noise [3, 7]. Therefore, the deviation was set as a small value; that is, the density deviation was 1 veh/km and the speed deviation was 1 km/h for both sources.

**5.3. Estimation Results.** Figure 5 shows the estimation results via the proposed PEKF-based method. The gradual changes in the color represent the variations in the traffic state. The blue represents the low-density and high-speed condition, while the red indicates the high-density and low-speed condition. As shown in Figure 5, both the estimated speeds and densities are changing among the free-flow, dense, and congested conditions. For instance, the speeds of the entire freeway segment during 00:00 to 10:00 AM are rarely red as shown in Figure 5(b). With the color getting red, it means that the traffic slows down. When the color is dark red, the traffic speed is quite low which indicates a congestion. In Figure 5(a), the density is lower than 40 veh/km from 00:00 to 7:00 showing in the blue. It is reasonable that there is little traffic on the freeway at the midnight of a working day and the traffic increases in the day time.

On the other hand, the incident/accident report data are used to qualitatively evaluate the accuracy of results for lacking the "ground-truth" speeds and density. Figure 5(b) shows that the speed drops firstly on Link 1 and Link 2 at about 11:00 AM. Consequently, this phenomenon happens on Link 3 and Link 4 quickly. And speeds on Link 5 and Link 6 also decrease around that time. The variation of speed indicates that the estimation results exhibit the influence of traffic incidents. However, the reported incident happened

on Link 4 at about 11:00, but the estimation results on Link 4 have a small delay. Since the measurements are available only on Link 1 and Link 6 and the distance between two links is 6 km long, it is understandable for such a small delay. In Figure 5(a), the density on Link 1, Link 2, and Link 3 decreases when the incident happens. Since Link 1, Link 2, and Link 3 are downstream links of Link 4, this decline phenomenon reflects the propagation of a shock wave. Another incident record is about the heavy traffic later that day which induces the low speed. From Figure 5(b), we can see the large area of red later that day which validates the correctness of the estimation results. And the changing color in Figure 5(a) shows the variation of density which reflects the oscillation of traffic under the low-speed traffic condition. Roughly, the estimation results efficiently integrated the measurements from the multiple sources and exhibit the dynamic and real-time traffic state.

**5.4. Evaluation.** An evaluation of the proposed PEKF-based estimator is conducted by making a comparison between the results from multisource data and single-source data. We applied the estimation results from the previous work [1], which estimated the traffic state using the wireless communication data and EKF-based data assimilation method. Since the distance between the microwave sensors is extremely sparse, the data generated from these sensors is not qualified for the application of a model-based approach. Hence, the direct measurements from microwave sensors are used for the comparison. To make a precise comparison, the absolute difference (AD) is chosen as the performance measure. It can be calculated by the following equation:

$$AD(k) = |x_s(k) - x_f(k)|, \quad (20)$$

where  $AD(k)$  is the absolute difference at the time interval  $k$ ,  $x_s(k)$  is the state estimates (density or speed) from single data source at the time interval  $k$ ,  $x_f(k)$  is the state estimates (density or speed) from multiple sources at the time interval  $k$ .

The iteration process of the progressive EKF-based estimator shows that the measurements from microwave sensors are usually taken as the more precise value compared with the traffic features from the wireless communication records.

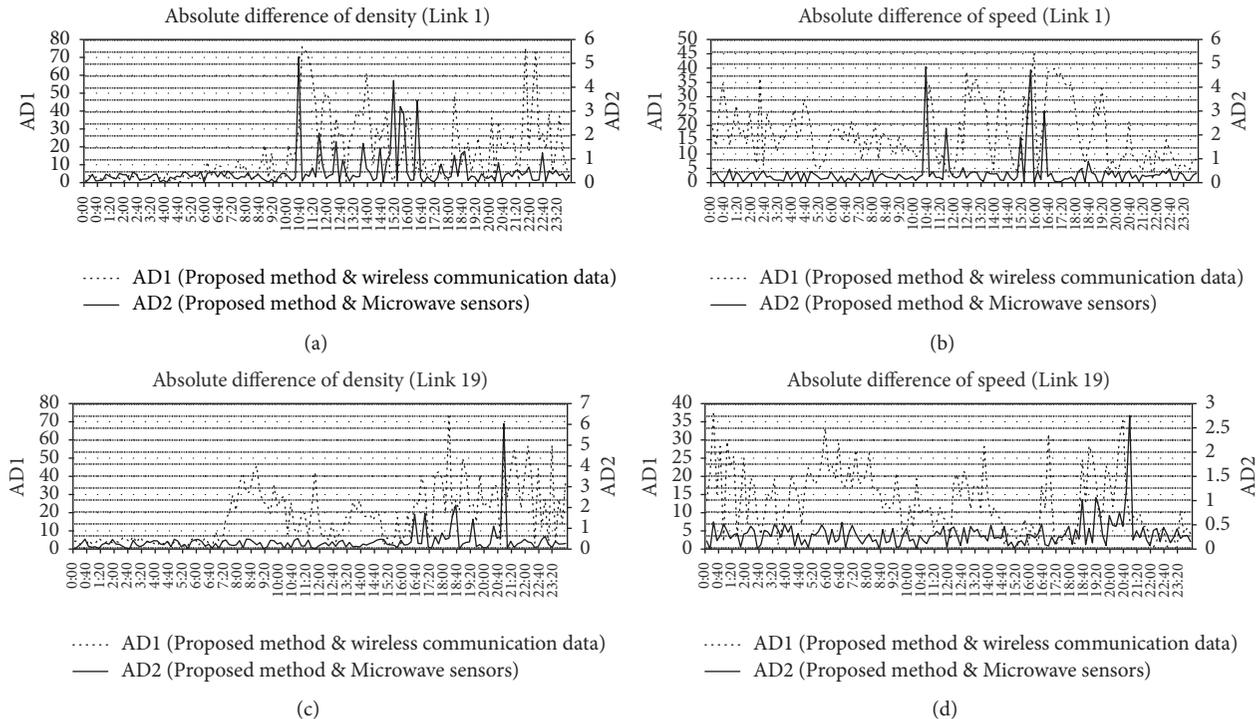


FIGURE 6: The AD results. (a) The ADs of density on Link 1. (b) The ADs of speed on Link 1. (c) The ADs of density on Link 19. (d) The ADs of speed on Link 19.

ADs between estimates from multisources and wireless communication data and ADs between estimates from multisources and measurements from microwave sensors in different links are present in Figure 6. Obviously, the progressive EKF-based method generated estimates which are more close to the more accurate measurements from microwave sensors. Generally, it will improve the total precision of estimation by fusing multisource data.

## 6. Conclusions

This paper proposed a Progressive Extended Kalman Filter (PEKF) method to estimate the freeway traffic state and integrate the heterogeneous data from the wireless communication data and the microwave sensors. The challenges in the application of these two data sources are the heterogeneities in the spatial coverage and the semantic of measurements. Via the characteristic analysis of the multisource data, we set up the relations between the measurements and traffic state variables and propose some rules to solve the conflicts between two data sources. The state-space model is constructed accordingly. The EKF technique is applied to establish the estimator, and it is improved to fit the multisource data and is nominated as the PEKF estimator. The test with the field data indicates that the proposed method successfully integrates the heterogeneous data, especially the new combination of measurements from the wireless communication records and microwave sensors. Moreover, the estimation results indicated that the PEKF is able to track the dynamic traffic conditions. The qualitative analysis between the estimates and

traffic incident reports validates the accuracy of the proposed method roughly. The final comparisons indicates the advantages of the proposed methods; that is, it has larger coverage compared with the microwave sensors, and it is more accurate than the estimates from wireless communication data.

This study makes an effort to use the limited, heterogeneous, and multisource data for the traffic state estimation, but there are still plenty of works before the field application. The most urgent task is to validate the method with the ground-truth data. Second, the test on other freeway links is required due to the spatial characteristic of the traffic features from the wireless communication data.

## Conflicts of Interest

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

## Acknowledgments

The authors gratefully acknowledge the data support from the Jiangsu Freeway Operation Center and the Traffic Operations and Safety Laboratory (TOPS), University of Wisconsin-Madison.

## References

- [1] S. He, Y. Cheng, F. Ding, J. Zhang, and B. Ran, "Extended Kalman Filter-based Freeway Traffic State Estimation Using Cellphone Activity Data," in *Transportation Research Board 95th Annual Meeting*, 2016.

- [2] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 99–118, 2016.
- [3] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.
- [4] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [5] Z. Lu, W. Rao, Y.-J. Wu, L. Guo, and J. Xia, "A Kalman filter approach to dynamic OD flow estimation for urban road networks using multi-sensor data," *Journal of Advanced Transportation*, vol. 49, no. 2, pp. 210–227, 2015.
- [6] Y. Wang, M. Papageorgiou, and A. Messmer, "a unified macroscopic model-based approach to real-time freeway network traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 3, pp. 190–212, 2006.
- [7] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: A case study," *Transportation Science*, vol. 41, no. 2, pp. 167–181, 2007.
- [8] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: adaptive capabilities and real data testing," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1340–1358, 2008.
- [9] Y. Yuan, H. Van Lint, F. Van Wageningen-Kessels, and S. Hoogendoorn, "Network-wide traffic state estimation using loop detector and floating car data," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 18, no. 1, pp. 41–50, 2014.
- [10] Y. Wang, M. Papageorgiou, A. Messmer, P. Coppola, A. Tzimitsi, and A. Nuzzolo, "An adaptive freeway traffic state estimator," *Automatica*, vol. 45, no. 1, pp. 10–24, 2009.
- [11] S. He, J. Zhang, Y. Cheng, X. Wan, and B. Ran, "Neural network based freeway traffic state data fusion system using cellular handoff probe system and microwave sensors," in *Transportation Research Board 94th Annual Meeting*, 2015.
- [12] C. Bachmann, B. Abdulhai, M. J. Roorda, and B. Moshiri, "A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 33–48, 2013.
- [13] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 89–107, 2013.
- [14] N.-E. El Faouzi, "Data-driven aggregative schemes for multi-source estimation Fusion: A road travel time application," in *Proceedings of the Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2004*, pp. 351–359, USA, April 2004.
- [15] M. Treiber, A. Kesting, and R. E. Wilson, "Reconstructing the traffic state by fusion of heterogeneous data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 6, pp. 408–419, 2011.
- [16] K. Choi and Y. Chung, "A data fusion algorithm for estimating link travel time," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 7, no. 3-4, pp. 235–260, 2002.
- [17] J. W. C. V. Lint and S. P. Hoogendoorn, "A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 8, pp. 596–612, 2010.
- [18] L. Mannini, S. Carrese, E. Cipriani, and U. Crisalli, "On the short-term prediction of traffic state: An application on urban freeways in ROME," pp. 176–185.
- [19] B. Heilmann, N.-E. El Faouzi, O. de Mouzon et al., "Predicting Motorway Traffic Performance by Data Fusion of Local Sensor Data and Electronic Toll Collection Data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 6, pp. 451–463, 2011.
- [20] W. Deng, H. Lei, and X. Zhou, "Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach," *Transportation Research Part B: Methodological*, vol. 57, pp. 132–157, 2013.
- [21] S. He, Y. Cheng, G. Zhong, and B. Ran, "A data-driven study on the sample size of cellular handoff probe system," *Advances in Mechanical Engineering*, vol. 9, no. 4, pp. 1–9, 2017.
- [22] S. L. He, J. Zhang, Y. Cheng, X. Wan, and B. Ran, "Freeway Multisensor Data Fusion Approach Integrating Data from Cellphone Probes and Fixed Sensors," *Journal of Sensors*, vol. 2016, Article ID 7269382, 2016.
- [23] Z. T. Qiu, *Macroscopic traffic state estimation for large scale freeway network using wireless network data*, Doctor of Philosophy, University of Wisconsin-Madison, Madison, 2007.
- [24] J. Zhang, S. He, W. Wang, and F. Zhan, "Accuracy analysis of freeway traffic speed estimation based on the integration of cellular probe system and loop detectors," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 19, no. 4, pp. 411–426, 2015.
- [25] H. Rakha and W. Zhang, "Estimating traffic stream space-mean speed and reliability from dual and single loop detectors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1925, pp. 38–47, 2005.
- [26] G. Welch and G. Bishop, *An Introduction to the Kalman filter*, University of North Carolina at Chapel Hill, 1995.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

