

Estimation of traffic energy consumption based on macro-micro modelling with sparse data from Connected and Automated Vehicles

Wen-Long Shang^{a,d,**}, Mengxiao Zhang^{b,*}, Guoyuan Wu^{c,*}, Lan Yang^b, Shan Fang^b, Washington Ochieng^d

^a Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Beijing University of Technology, Beijing, China

^b School of Information Engineering, Chang'an University, Xi'an 710064, China

^c Center for Environmental Research and Technology, University of California at Riverside, Riverside 92507, USA

^d Centre for Transport Studies, Imperial College London, SW7 2AZ London, UK

HIGHLIGHTS

- Estimation of fuel consumption and emissions is closer to the real values than other methods.
- Reconstructing second-by-second vehicle trajectories based on macroscopic traffic data.
- Introduces CAVs trajectory data as a reference in the process of estimating the road spatio-temporal speed evolution.
- Proposes a spatio-temporal consistency validation framework by using macro and micro data.
- Experimental tests the effect of cell size and probe vehicle penetration on reconstruction accuracy.

ARTICLE INFO

Keywords:

Macro traffic
Vehicle trajectory reconstruction
CAVs data
Energy consumption estimation

ABSTRACT

Traffic energy consumption estimation is significant for the sustainable transportation. However, it is difficult to directly employ macro traffic flow data to accurately estimate the traffic energy consumption due to many traffic energy consumption models need second-by-second vehicle trajectory. To solve this problem, this paper proposes a traffic energy consumption model based on the macro-micro data, which the macro data derived from the fixed-location sensors and sparse micro data derived from the Connected and Automated Vehicles (CAVs). The completed vehicle trajectories are constructed by the nonparametric kernel smoothing algorithm and variational theory. To test the performance of the proposed method, the Next Generation Simulation micro (NGSIM) dataset and Caltrans Performance Measurement System macro dataset obtained from the same road and time are used. The results indicate that the proposed method not only can reflect the characteristics of traffic kinematic waves caused by traffic congestion, but also minimize the errors generated by the macro-micro transformation. In addition, it can significantly improve the accuracy of energy consumption estimation.

1. Introduction

Climate warming is a well-recognized global problem. With the rapid growth of the world economy, the massive use of fossil energy sources, such as coal, oil and natural gas, has resulted in excessive greenhouse gas emissions that is one of the main catalysts of global warming [1–4]. According to the U.S. Energy Information Administration (EIA), the transportation sector accounts for approximately 27% of total energy consumption and is thus one of the key contributors to greenhouse gas

emissions [5]. In recent years, the energy consumption and air quality degradation caused by road traffic have been become major issues of concern [6,7], and thus traffic management agencies are increasingly interested in the ability to estimate traffic energy consumption and greenhouse gas (GHG) emissions [8]. Typically, classical emissions models are used for such estimation, whether on a regional or road section basis. In this quantification process, the traffic/vehicle state data obtained from Connected and Automated Vehicles (CAVs), usually by means of inductive loop detectors [9,10] or vehicle global positioning

* Corresponding authors.

** Corresponding author at: Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Beijing University of Technology, Beijing, China.
E-mail addresses: shangwl@imperial.ac.uk (W.-L. Shang), mx_zmolly@163.com (M. Zhang).

<https://doi.org/10.1016/j.apenergy.2023.121916>

Received 6 March 2023; Received in revised form 8 August 2023; Accepted 5 September 2023

Available online 14 September 2023

0306-2619/© 2023 Elsevier Ltd. All rights reserved.

systems (GPS) [11,12], is fed into relevant macro/micro energy consumption and GHG emission models.

Misra evaluated the peak CO and NOx emissions from the downtown Toronto using the microscopic traffic simulation model PARAMICS and the CMEM microscopic emission model [13]. Zhao et al. used SUMO to obtain vehicle trajectory, and used the VT-Micro model to calculate the instantaneous fuel consumption and traffic emissions for mixed traffic flows [14]. While previous studies have been able to calculate accurate energy consumption and GHG emissions data by combining micro fuel consumption and emission models [15,16], differences in sensor cost and sampling frequency have made it difficult to obtain second-by-second speed trajectories on a large scale.

Inductive loop detectors are widely deployed in cities and highways, and are able to collect the macro traffic data that can reflect the traffic state within a certain period of time, such as the average speed, flow density, and traffic volume. This traffic data can be combined with energy consumption and GHG emission models to calculate traffic fuel consumption and emissions at a macro level [17,18]. Hao developed a traffic-related air pollution assessment framework with PeMS data and the macro-emissions model EMFAC to accurately assess the environmental impact of traffic congestion in real time [19]. With the MOBILE model, Choi and Frey estimated the pollutant emission inventories of motor vehicles at different speeds [20]. The average traffic speed does not take into account vehicular and traffic dynamics, however, and is therefore not suitable for accurately estimating energy consumption and GHG emissions in many scenarios.

In order to solve the problems of low accuracy of macroscopic energy consumption and emission models and the difficulty of obtaining microscopic traffic data, some scholars try to couple macroscopic traffic models with microscopic vehicle energy consumption and emission models to obtain microscopic traffic energy consumption and emission. Turkensteen used CMEM to calculate fuel consumption and carbon emissions of vehicles at a given speed and load, assuming that the vehicle is driven at a fixed speed [21]. Zegeye approximated macroscopic traffic as microscopic traffic variables by making The macroscopic traffic flow model METANET and the microscopic energy consumption and emission model VT-Micro are combined to achieve shorter simulation times and accurate estimation of fuel consumption [22]. Wang et al. reproduce traffic states and vehicle queue trajectories from macroscopic traffic data obtained from detection stations, derived accelerations from the reconstructed vehicle trajectories, and estimated accelerations as inputs to VT-Micro to calculate fuel consumption and emissions [23]. Chen et al. used the field speed data recorded by detectors to estimate second-by-second velocity and acceleration data of all passing vehicles and used CMEM to calculate the corresponding fuel consumption and emissions [24].

However, to couple macro and micro traffic data of different granularity with vehicle energy consumption and emission models to estimate energy consumption and emissions of roads, the problem that traffic data cannot be used as input to energy consumption and emission models must be solved. To couple macroscopic traffic data with microscopic energy consumption and emission models, microscopic traffic information is usually obtained from macroscopic traffic data using vehicle trajectory reconstruction methods. Therefore, the focus of this paper is to obtain second-by-second trajectory information of vehicle trips estimated from real-time traffic information (e.g., average traffic speed). Although a considerable number of studies have been conducted to reconstruct speed trajectories using traffic data collected from various surveillance systems, these can be divided into three main categories: in terms of methods, spatio-temporal interpolation, filtering, and multi-source data fusion.

1) *The spatio-temporal interpolation method*: This mainly relies on the spatio-temporal autocorrelation property of fixed detector data. It first calculates weights based on the distance between data points, and then fills in each point of the spatio-temporal region using the

weighted average of the measurements of each data point. Instantaneous [25], time-slice [26], dynamic time-slice [27], and linear models [28] can all be used to estimate the spatio-temporal speed distribution between fixed detectors within the observation range using the speed recursion method. Since all these models use an averaging approach for data collection and trajectory reconstruction, however, they are not very accurate in congested conditions, thus causing some high and low values to be discarded and resulting in reconstructed trajectories that do not reflect individual vehicles' different behaviors. Ni and Wang first offered a two-dimensional estimate of road speed based on fixed speed measurements obtained from an intelligent traffic system, and then reconstructed the vehicle trajectory based on its time of entry into the network [29]. Wu added random velocities to the velocities obtained using quadratic smoothing to generate a comprehensive speed trajectory for a trip, considering a priori knowledge of real-time traffic information on the travel route provided by intelligent transportation system (ITS) techniques [30]. Krol proposed the reconstruction of the vehicle trajectory using interpolation techniques to describe the density in the considered spatio-temporal framework. The density of each point was interpolated according to the density of the four corners of the spatio-temporal region, and then a triangular shaped fundamental diagram was used to construct the trajectory of each vehicle [31]. Since densities were collected in an aggregated manner, however, the density of each point in the block would be the same if the densities of the four corners were the same, implying that vehicles should travel at a constant speed. Coifman [32] proposed a method for estimating vehicle trajectories on highway sections using traffic data from a single dual-loop detector, using traffic flow theory to infer local traffic conditions. The drawback of the method, however, is that it cannot be applied during the transition period from uncongested to congested traffic conditions, thus limiting its practical application.

- 2) *The filtering method*: This is a spatio-temporal traffic state estimation method capable of considering the propagation of traffic congestion. It uses a nonlinear low-pass filter to interpolate the traffic state between fixed detectors. Examples of this method include Treiber and Helbing's [33] adaptive smoothing approach, where they used a nonlinear spatio-temporal low-pass filter for fixed detector data so that, in congested traffic, the perturbation moved upstream at a near-constant rate, whereas in free traffic, the information propagated downstream. Van Lint and Hoogendoorn [34], meanwhile, obtained reliable traffic state estimates by improving the adaptive smoothing method so that it could fuse data from individual traffic detectors. Wang proposed a generalized stochastic macroscopic traffic flow model and extended Kalman filtering based on a highway section or road network traffic state [23]. Chen established a trajectory estimation algorithm based on an extended vehicle tracking model for the undetected part of each trajectory [35]. Based on that, a particle filter-based trajectory fusion algorithm was proposed to fuse the estimated trajectories for minimization. Chen proposed a hybrid method to reconstruct the complete vehicle trajectories at signal intersections. This used variational networks and Kalman filtering to reproduce stochastic features of the reconstructed queue boundary curves [36]. Wei et al. used a particle filtering-based method to reconstruct the trajectories along the main missing trajectories between successive updates of detected vehicles at multiple intersections [37]. Feng et al. used a particle filtering approach that considered five spatio-temporal trajectory correction factors and reconstructed vehicle trajectories in a large-scale network based on data collected by automatic vehicle recognition and conventional detectors [38]. Xu et al. also estimated the trajectories of signal intersection arterials by integrating different data sources through particle filters [39].
- 3) *Multi-source data fusion method*: This approach aims to combine data from fixed and mobile detectors to compensate for the shortcomings

of fixed detector data and sparse data from mobile sensors. Mehran et al. [40] proposed a new data fusion framework based on the three-dimensional (3D) kinematic wave model and variational theory [41,42] to reconstruct vehicle trajectories on urban arterials by fusing cab and cross section data. Sun et al. [43] reconstructed vehicle trajectories by fusing probe vehicle and signal timing data based on variational theory, and validated the method using vehicle trajectory data measured at intersections. The results showed that while the method was robust to changes in probe vehicle penetration, the variational theory-based model ignored the random-wave properties of speed. Tsanakas [44] proposed a two-stage approach to reconstruct multi-vehicle trajectories by first calculating the speed of each cell through filtered smoothing, and then calculating the position at each time step using bicubic interpolation. Chen et al. [45] used the hybrid property to develop a macro-micro integration framework to reconstruct vehicle trajectories on highways, where data fusion provided a critical speed baseline for reconstructing the trajectory. Additionally, the car-following and inverse car-following models were used to produce candidate trajectories and determine the optimal trajectory with the minimum speed baseline difference based on dynamic programming.

These aforementioned studies, however, have certain limitations in regard to reconstructing vehicle trajectories. Given the low deployment rate of fixed detectors on highways and the long aggregation time, most studies attempted to replace fixed detector data by counting the speed aggregation values of microscopic vehicle trajectories at fixed positions in a short period of time (usually 30 s), rather than using real fixed detector data to reconstruct vehicle trajectories. Further, most of the studies were more readily applicable to free traffic flow conditions, and therefore do not reflect the stop-and-go behavior of vehicles under congested traffic flow conditions. Moreover, some methods were undertaken from the perspective of estimating travel time without considering the vehicle kinematic characteristics, resulting in abnormal vehicle speed and acceleration results.

This paper, therefore, proposes an energy consumption estimation method of regional traffic based on macro-micro modelling with sparse data from CAVs. There are three main contributions. First, the speed observed from fixed detectors is utilized to reconstruct second-by-second vehicle trajectories. Second, a small number of CAV trajectory data is introduced as a reference in the process of estimating cell speed, allowing a more accurate description of the evolution of road spatio-temporal speed and thus ultimately a more accurately reconstructed trajectory. Last, a spatio-temporal consistency validation framework is developed using Next Generation Simulation (NGSIM) data and Caltrans Performance Measurement System (PeMS) data. The framework compares the traffic energy consumption and emissions under the different aggregation approaches for macro and micro models. It addresses the lack of accuracy caused by the errors embedded when converting between macro and micro modelling during the process of estimating traffic energy emissions.

2. Energy consumption estimation based on the micro-trajectory reconstruction method

This section introduces how to utilize the micro-trajectory reconstruction method to estimate the energy consumption of regional road traffic. The vehicle speed trajectory is synthesized in the following three steps.

2.1. Spatio-temporal speed estimation of road sections based on nonparametric kernel smoothing

(1) Traffic fundamental diagram estimation

The macroscopic Fundamental Diagram (FD) is a model describing the relationship between traffic density and traffic flow on a road section. It can visually characterize road traffic flow characteristics and is important in both road area control and macroscopic modelling. The most common triangular shaped FD uses two straight lines to fit the non-congested and congested areas, where the slope of the non-congested area is the free-flow speed and the slope of the congested area is the congestion propagation speed.

In this study, we extracted the historical macro data (aggregated 5-min flow/density values) of fixed detectors in our case study for one month from June 1st to June 30th, 2005, from 7:50 to 8:35. Then, we fitted the triangular-shaped FD so as to estimate key parameters quickly. These parameters were: free-flow speed v_{free} , capacity q_{max} , and critical density k_{jam} , which are expressed as:

$$v_{free} = \min(130, \max(V_j^i)) \quad (1)$$

$$q_{max} = \max(Q_j^i) \quad (2)$$

$$k_{jam} = \max(\max(K_{ij}), 150) \quad (3)$$

where V_j^i , Q_j^i , and K_{ij} are the speed, traffic flow, and density, respectively, obtained from the output of loop detector j at moment i . The free-flow speed and capacity are the data of cross-sectional detector; the minimum value was set to 150 because the blockage phenomenon of the road section is difficult to detect, and the critical density is the maximum density of the entire network. The free-flow speed was limited to a maximum of 130 km/h owing to the speed limit of the case study road.

(2) Nonparametric kernel smoothing-based cell speed estimation

Because the data collected by loop detectors was discrete, it did not reflect the dynamic changes of traffic in time and space. The approach reported in this section, therefore, improves the roadway speed estimation accuracy by fusing the fixed loop detector data and probe vehicle data, thus compensating for the deficiencies of the individual data sources on their own.

Spatio-temporal speed contours (where the x-axis is time, the y-axis is space, and the inner color (or z-axis) represents speed) are the basis of various traffic studies and applications. They are typically used to divide a spatio-temporal network into different cells, before applying one of various methods to estimate the traffic speed of all cells. He [46] demonstrated that using non-rectangular parallelogram cells to slice the spatio-temporal network further was better able to take into account the backward waves compared to the traditional rectangular parallelogram cells, reducing the errors in microscopic vehicle trajectory and travel time estimation in the presence of traffic congestion. For this purpose, we sliced the spatio-temporal network using the congestion wave speeds estimated in Section 3.1, sliced the spatio-temporal network into multiple non-rectangular parallelogram cells based on the time and spatial intervals, and used the speed at the center of the cell as the average speed of the whole cell.

Treiber-Helbing [33] proposed a nonparametric method called the Generalized Treiber-Helbing Filter (GTF) to achieve spatio-temporal interpolation of fixed detector velocities. Van Lint and Hoogendoorn [34] extended the GTF by fusing data from multiple sources (EGTF). As shown in Fig. 1, for each parallelogram cell, the range of data points (t , x) that have an effect on the speed of the cell center (t , x) is customized. According to the description in the literature [44], this range is defined in this paper as a circular region with the cell center(t , x) as the center and a radius r (shown as a red circle in the Fig. 1), which is related to the distance between the two fixed detectors. The speed of the cell center is estimated from the distances of the different source data points contained in this range from the cell center point (t , x), and from the speed

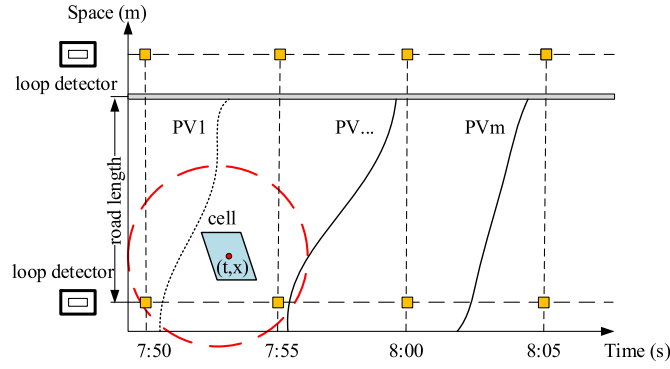


Fig. 1. Multi-source data fusion for cell speed estimation.

measurements of these data points.

To capture traffic disturbances, two auxiliary speed surfaces were established, one for free-flow and the other for congestion conditions. For each data source, the two auxiliary speed surfaces were estimated as follows:

$$v^{\text{free}}(t, x) = \frac{\sum_{s \in S} \beta_s^{\text{free}}(t, x) V_s}{\sum_{s \in S} \beta_s^{\text{free}}(t, x)} \quad (4)$$

$$v^{\text{con}}(t, x) = \frac{\sum_{s \in S} \beta_s^{\text{con}}(t, x) V_s}{\sum_{s \in S} \beta_s^{\text{con}}(t, x)} \quad (5)$$

where V_s is the speed measurement at data point s ; S is the set of speed counts in the range $A(t, x)$; β_s^{free} and β_s^{con} are given by the following equations:

$$\beta_s^{\text{free}}(t, x) = \varphi_O \left(t - t_s - \frac{x - x_s}{c_{\text{free}}}, x - x_s \right) \quad (6)$$

$$\beta_s^{\text{con}}(t, x) = \varphi_O \left(t - t_s - \frac{x - x_s}{c_{\text{con}}}, x - x_s \right) \quad (7)$$

where φ_O is a kernel function that assigns a weight to each speed count V_s , defined as follows:

$$\varphi_O = \exp \left(- \left(\frac{|t|}{\tau} + \frac{|x|}{\sigma} \right) \right) \quad (8)$$

and where τ and σ are the temporal and spatial smoothing widths, respectively.

By combining data from different data sources, the EGTF speed field is expressed as:

$$v^{\text{EGTF}}(t, x) = \frac{\sum_{m \in M} \alpha_m \sum_{s \in A_m(t, x)} \left[w_m(x, t) \beta_{s, m}^{\text{con}}(x, t) + (1 - w_m(x, t)) \beta_{s, m}^{\text{free}}(x, t) \right] V_{s, m}}{\sum_{m \in M} \alpha_m \sum_{s \in A_m(t, x)} \left[w_m(x, t) \beta_{s, m}^{\text{con}}(x, t) + (1 - w_m(x, t)) \beta_{s, m}^{\text{free}}(x, t) \right]} \quad (9)$$

where $V_{s, m}$ is the speed value measured by the data source m at point (t_s, x_s) , M is the set of data sources, $m \in M$, α_m is the data source-specific weight for each source m reliability, $A_m(t, x)$ is the set of speed counts located within the user predefined region (t, x) and $A(t, x)$, and $w_m(x, t)$ is the weight of the influence of the two auxiliary speed surfaces of data source m on this metric. This is therefore an adaptive weighted s-type function that depends on the congestion level at point (t, x) expressed as:

$$w_m(x, t) = \frac{1}{2} \left[1 + \tanh \left(\frac{\hat{v} - \min(v_m^{\text{free}}(x, t), v_m^{\text{con}}(x, t))}{\Delta v} \right) \right] \quad (10)$$

where $\hat{v}$ is the speed threshold between the free-flow and congestion conditions, which indicates the transition width in the surroundings.

2.2. Microscopic vehicle trajectory reconstruction based on 3D kinematic wave theory

2.2.1. Discretization of cell lattices

3D kinematic wave theory is a traffic flow model combining kinematic wave theory and the accumulation curve principle to provide an accurate estimate of the traffic shock wave caused by traffic congestion in the opposite direction of vehicle travel, assuming that vehicles follow the first-in-first-out principle in a closed area. This is illustrated in Fig. 2a, where a third coordinate (z -axis) is added to the spatio-temporal axis of the two-dimensional kinematic wave to represent the cumulative vehicle numbers. The red curve in the figure, meanwhile, is the reference probe vehicle trajectory of CAVs, the green dashed line represents the forward wave, and the black dots on the time axis are the entry and exit times of the vehicles. Daganzo [41] proposed a traffic flow calculation model for road networks based on 3D kinematic wave theory with variational theory and comparative capacity constraints. As shown in Fig. 2b, their approach divides the spatio-temporal network according to the forward and backward wave speed. The i -coordinate of the coordinate system of the divided variational network is aligned with the backward wave, and the j -coordinate is aligned with the forward wave; each grid node has a value representing the cumulative number of vehicles at that spatio-temporal location. The cumulative number of vehicles passing through an unknown node is obtained in a discrete spatio-temporal space based on the boundary state. The application of variational theory based on 3D kinematic waves is therefore here based on a discretized spatio-temporal network.

Because the forward and backward arc slopes of the spatio-temporal network are deterministic, the dynamic changes in the traffic state are not fully considered. Thus, the free-flow speed of the small grid point is set as the average speed of the cell where it is located. The basic parameters of the initial variational network—i.e., time, and space steps of the backward wave—are determined as follows:

$$w = \frac{q_{\text{max}}}{k_{\text{jam}} - \frac{q_{\text{max}}}{u}} \quad (11)$$

$$t_{\text{step}} = 1s \quad (12)$$

$$s_{\text{step}} = \frac{u \times w \times t_{\text{step}}}{u + w} \quad (13)$$

where w is the backward arc slope, and u is the forward arc slope (i.e., the average speed of the traffic in the cell where point (x, t) is located); t_{step} is the time step, which is a predetermined value, and s_{step} is the space step.

2.2.2. Multi-vehicle trajectory reconstruction based on 3D kinematic wave theory

After completing the network division as above, the initial cumulative vehicle number of the first column nodes in the initial variational network is set to one. The reference probe vehicle trajectory can be regarded as a path connected by nodes with the same cumulative vehicle number on the variational network whose starting and ending points correspond, respectively, to the cumulative vehicle numbers of the upstream and downstream nodes, as recorded by the fixed detector. Considering that it is difficult to match the probe vehicle trajectory exactly with the spatio-temporal network, however, the probe vehicle trajectory must be gridded as a constraint to calculate the cumulative number of vehicles at subsequent nodes. For the cumulative number of vehicles at each time point at the upper and lower boundaries of the initial variational network, it is necessary to determine when vehicles enter and leave the network based on other external conditions.

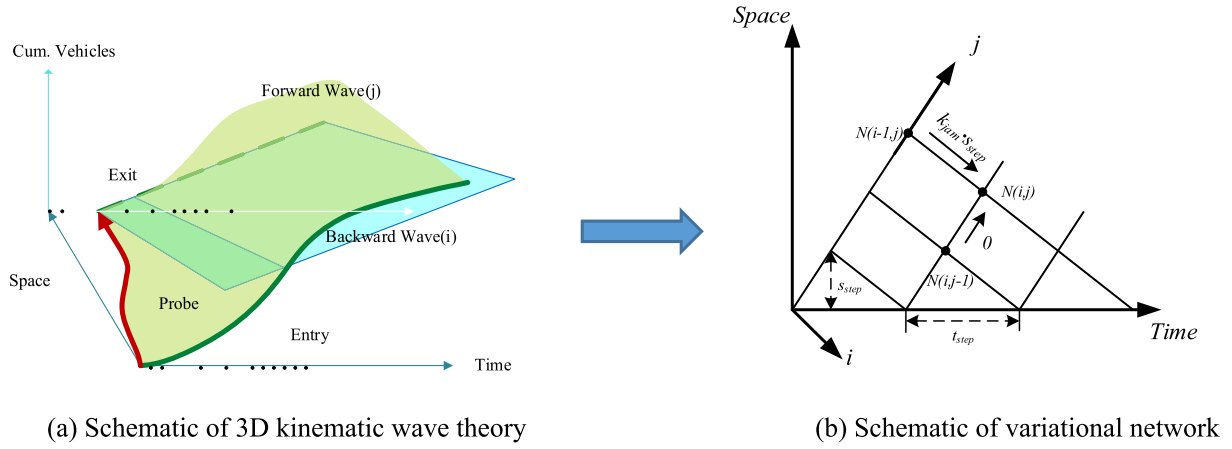


Fig. 2. Schematic diagram of variational theory based on 3D kinematic waves.

In the triangular-shaped FD, the increase in the number of vehicles at the unknown nodes (i, j) in the network takes only two values, i.e., the increment of the forward wave is zero, which means that there is no change in the cumulative number of vehicles along the forward wave, and the increment of the backward wave is $k_{jam} \cdot s_{step}$, which is the maximum change allowed in the cumulative number of vehicles. According to the literature [41,42], the solution of the kinematic wave problem based on variational theory is a set of spatio-temporally continuous shortest paths. Since the traffic flow cannot be reversed in space and time, it can only be searched from the node with a low cumulative number of vehicles to the node with a high or equal number of vehicles. Therefore, for the acyclic variational network in this paper, it is simple and efficient to use Dijkstra shortest path algorithm for searching. Under the constraint of detecting the vehicle trajectory as well as the spatio-temporal network boundary, the cumulative number of vehicles for each unknown number of nodes in the network can be determined according to the shortest path algorithm, expressed as follows:

$$N(i, j) = \min\{N(i, j-1), N(i-1, j) + k_{jam} \cdot s_{step}\} \quad (14)$$

where $N(i, j)$ denotes the cumulative number of vehicles at node (i, j) , and $k_{jam} \cdot s_{step}$ represents the amount of variation in the cumulative number of vehicles that may occur at node (i, j) . Finally, nodes with the same cumulative number of vehicles are connected in order to output the vehicle trajectory.

2.3. Estimation of energy emission based on the MOVES model

In this study, the vehicle specific power (VSP) method is adopted to estimate the energy consumption of road traffic on the basis that VSP can better characterize the behavior of motor vehicles on real-world roads [47]. The MOVES model uses the VSP method to characterize the driving characteristics of a motor vehicle in addition to its speed, with VSP representing the tractive power of the vehicle expressed as follows:

$$VSP_i = \frac{Av_i + Bv_i^2 + Cv_i^3 + mv_i a_i}{m} \quad (15)$$

where v is the speed (m/s), a is the acceleration (m/s^2), m is the weight of the vehicle (t), and A , B , C are the vehicle road load factors. The MOVES emission model clusters the VSP values into different VSP bins at 1 kW/t intervals and combines them with instantaneous speed intervals to obtain the vehicle operating mode. The emissions of a vehicle operating on the road are basically estimated by multiplying the emission rate with the VSP distribution, as shown in Eq. (16):

$$emission = running\ time \times \sum_{VSP\ bin} emission\ rate \times VSP\ distribution \quad (16)$$

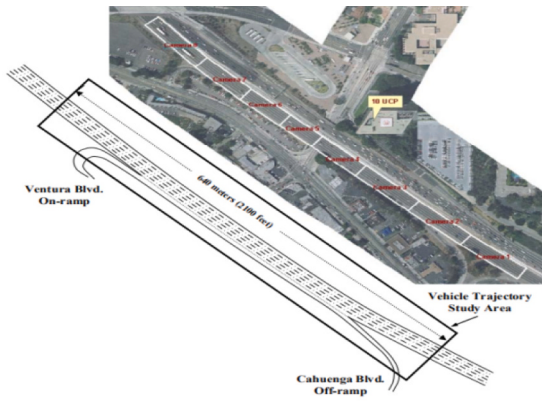
The emission rates for different VSP bins represent vehicle emissions at different power demands. The VSP distribution (or operating mode distribution in the MOVES model) is the fraction of time spent in each VSP bin. The emission rate is fixed for a specific type of vehicle; so the distribution of VSP in a fleet can determine the emissions in a traffic network.

3. Case study

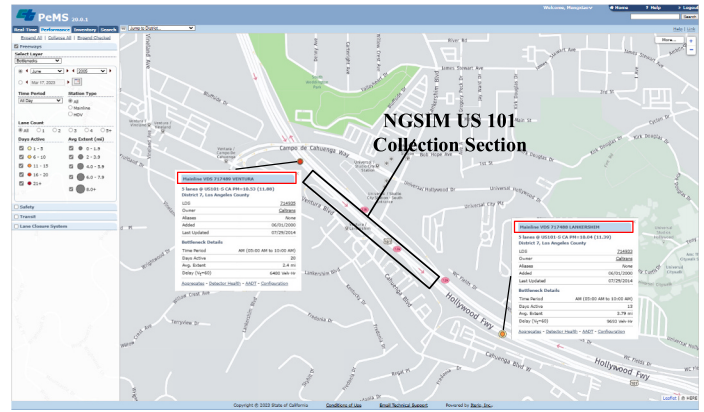
To investigate the impact of macro and micro data conversion methods on regional traffic fuel consumption emission results, this study constructs a data validation framework with spatiotemporal consistency. As shown in Fig. 3, the Performance Measurement System (PeMS) macro data (low-density scatter data) and Next Generation Simulation (NGSIM) micro data (vehicle trajectory data) in the US101 freeway in the US State of California State are selected for comparison and analysis. The PeMS data are obtained from the inspection stations 717,488 and 717,489 of the California Highway 101, which are located on the same roadway as the NGSIM US 101 trajectory data (shown in Fig. 3). The PeMS data are extracted from June 1st to June 30th, 2005; while the NGSIM US 101 trajectory data corresponded to the time interval between 7:50 and 8:35 on June 15th, 2005. The NGSIM data set has become the microscopic real data standard, underlying the vast majority of empirically based advances of the past decade. However, the data of CAV is unknown in the scenario of this article, so a certain proportion of CAVs trajectories can be used as observed CAVs data.

As shown in Fig. 4, we first estimate the spatio-temporal speed evolution of road segments by fusing the observed speed values from different fixed detectors and probe vehicles, using a nonparametric kernel smoothing method to compensate for the shortcomings of using a single data source to characterize the basic state of road segments. Second, to consider the complex and variable road traffic environment and better reflect the congestion transmission of the actual road, a selection of the vehicle trajectories in the NGSIM dataset are chosen to be the virtual probe vehicles, which are used to constrain and reconstruct the microscopic vehicle trajectory based on the variational theory of 3D kinematic waves. Finally, the energy emission consumption of the reconstructed trajectory is estimated based on the MOVES model.

To verify the effectiveness of the proposed model, we conduct a comparative analysis using a four-step experiment. First, the NGSIM data are filtered to reduce errors in the baseline data. Second, the reconstruction accuracy of the proposed under different parameter combinations is compared using root mean square error (RMSE) and mean absolute error (MAE) metrics. Again, the trajectory conversion



(a) NGSIM US101 data map



(b) PeMS data map

Fig. 3. Micro and macro data with spatiotemporal consistency.

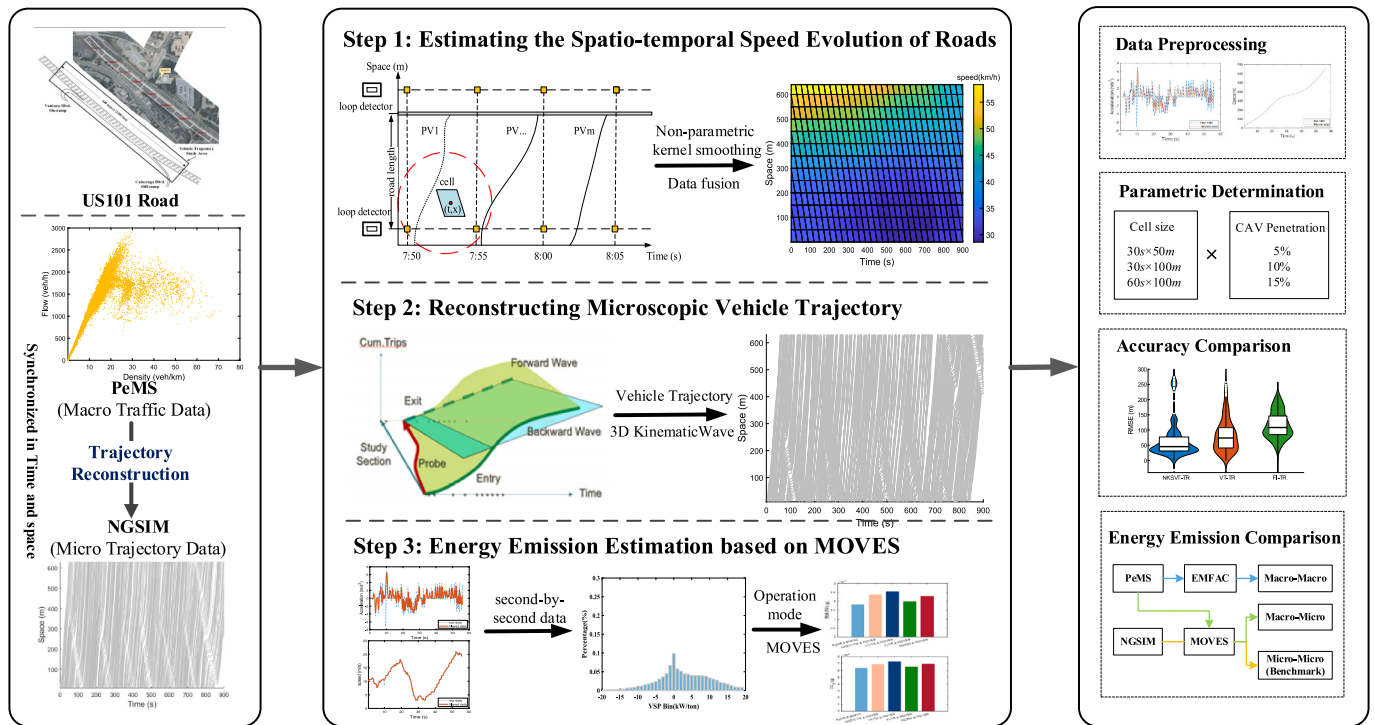


Fig. 4. Flow Chart of the Proposed Method.

accuracies of the three methods are compared using the RMSE and MAE metrics, and the trajectory reconstruction accuracy of the three methods are compared under free flow and congested flow conditions. Finally, to verify the effects of macro data on the micro-trajectory method on traffic energy consumption and emissions, the California motor vehicle emission models—i.e., Emissions Factors (EMFAC) and MOVES—are selected, and traffic data of different granularities are input into the corresponding energy consumption and emission models. We compared the fuel consumption and emission of macro data with the macro energy consumption model, macro data after micro disaggregation and micro energy consumption model, and verified the environmental characterization effect under the combination of traffic and energy consumption models at different levels.

4. Results and discussion

We conducted experiments to compare the performance of the proposed 3D kinematic wave trajectory reconstruction method based on Nonparametric Kernel Smoothing and Variational Theory (NKSVT-TR) with trajectory reconstruction based on four-corner interpolation (FI-TR), and trajectory reconstruction based on variational theory (VT-TR). The pre-processed NGSIM US101 Lane 1 7:50–8:05 vehicle trajectory data are used as the observation data, while the overtaking and lane-changing behaviors of vehicles are not considered. For these three trajectory reconstruction methods, it is assumed that the total number of trajectories and the entry and exit times of the reconstructed vehicles are known and could be measured using road loop detectors.

4.1. Data pre-processing

In reality, the NGSIM data have a number of outliers and observation errors that can have a big impact when used directly in the calculation of microscopic vehicle behavior. Furthermore, when the observed trajectory data are used to calculate vehicle speed and acceleration by first-order and second-order derivatives, such observation errors are multiplied. Pre-processing of the data is therefore necessary before the verification experiment.

We used Montanino and Punzo's [48] four-step method to correct the NGSIM trajectory data, which was divided into the following four steps: (1) the very large and very small outliers of acceleration were removed; (2) trajectory data corresponding to the speed values of high and medium frequencies were filtered using a low-pass filter (frequency of 1.25 Hz); (3) trajectory data corresponding to acceleration values that did not conform to physical characteristics were eliminated; and (4) the low-pass filter in step (2) was used to filter and denoise the trajectory data. The vehicle acceleration and speed of NGSIM US101 after step-by-step processing are shown in Fig. 5; the vehicle acceleration, speed, and position profiles before and after processing are shown in Fig. 6.

4.2. Sensitivity analysis of model parameters

The experiments used randomly sampled NGSIM data as the virtual probe detector vehicle trajectory as well as VT-TR, including the vehicle position at a fixed time interval (0.1 s) and the speed information. Fixed detector data was obtained from the loop detector data from the PeMS US101 section and included the average speed data at a fixed position. Since the cell size and penetration rate of the probe vehicle are very important to the results of this study the values of their parameters had to be determined in advance. In this regard, limitations in computational power meant that, when calculating the spatio-temporal speed contour of the road section, the road has to be discretized into a cell, with the central speed of that cell calculated separately to represent the entire

speed field. Tsanakas [44] and He [49] applied parallelogram cells to estimate the spatio-temporal speed network, considering cell sizes of (10×50) , (30×50) , (30×100) , (60×50) , and (60×100) (in sec $\times$ m). In this study, therefore, the cell sizes of (30×50) , (30×100) , and (60×100) spatio-temporal grids are considered to verify the accuracy of the cell time and spatial intervals for spatio-temporal speed network estimation. Meanwhile, the number of probe vehicles is an important factor affecting the spatio-temporal speed estimation; thus, for each cell size, three probe vehicle penetration rates are set: 5%, 10%, and 15% [44], respectively. As listed in Table 1, nine sets of experimental combinations are used.

The RMSE and MAE metrics have been widely used to test reconstruction accuracy [44,45]. Our experiments therefore compare the errors between the instantaneous position, speed, and acceleration of the reconstructed trajectories and reference (actual) values. The RMSE and MAE of the vehicles are calculated as follows:

$$\text{RMSE} = \frac{\sum_{n=1}^N \sqrt{\frac{\sum_{i=1}^{I_n} (z_n(t_i) - \bar{z}_n(t_i))^2}{I_n}}}{N} \quad (17)$$

$$\text{MAE} = \frac{\sum_{n=1}^N \sqrt{\frac{\sum_{i=1}^{I_n} (z_n(t_i) - \bar{z}_n(t_i))^2}{I_n}}}{N} \quad (18)$$

where $z_n(t_i)$ is the estimated value of the variable (position, speed, acceleration) at time step i , $i = 1, 2, \dots, I_n$, $\bar{z}_n(t_i)$ that corresponds to the reference value; N is the total number of vehicles.

Table 1 lists the average RMSE and MAE of the vehicles for the proposed method at the different combinations of cell size and probe vehicle penetration. It can be seen that the RMSE is generally larger than the MAE, indicating that the errors are not uniformly distributed in the reconstructed area. The estimated trajectories are more accurate at the entrance and exit of the road, with errors typically increasing near the middle of the road. This happens because fixed sensors were deployed at

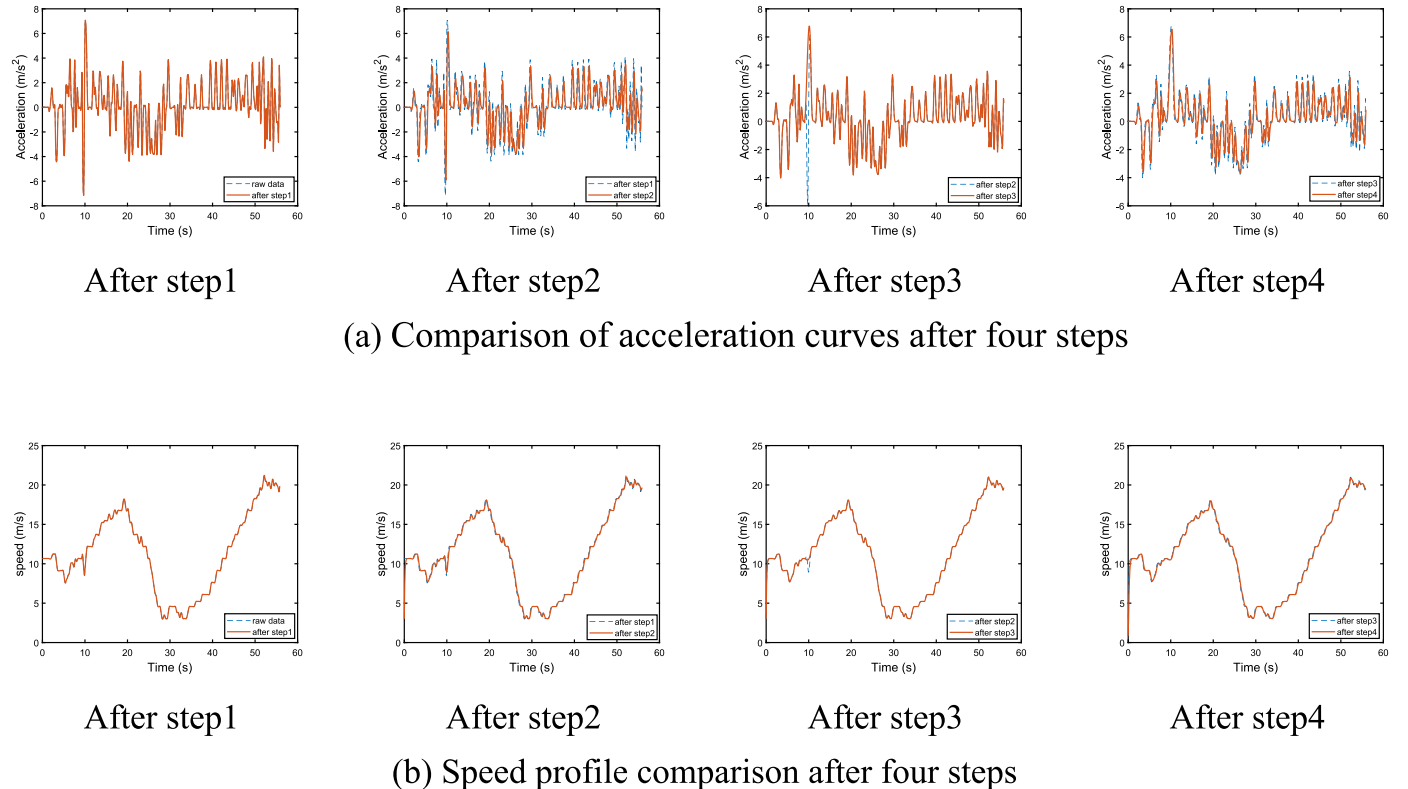


Fig. 5. Comparison of acceleration and speed profiles of vehicle No. 356 after step-by-step processing.

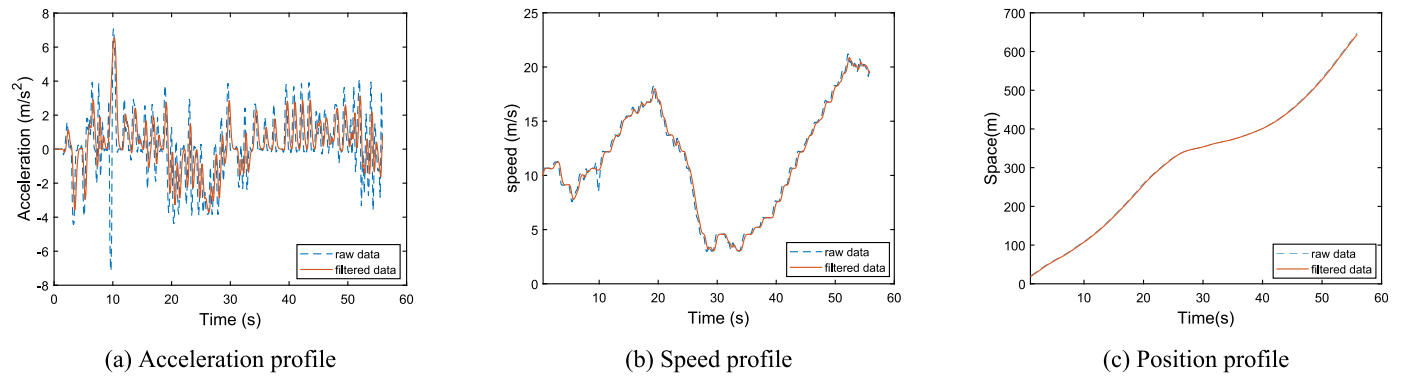


Fig. 6. Comparison of acceleration, speed, and position profiles before and after treatment of vehicle No. 356.

Table 1

Errors under different combinations of parameters.

Serial number	Cell size	Probe Vehicle Penetration (%)	Position		Speed		Acceleration	
			RMSE	MAE	RMSE	MAE	RMSE	MAE
1	30s × 50 m	5	77.5968	65.0811	4.7823	3.8522	1.4874	1.0299
2		10	72.1628	62.9308	4.5860	3.7160	1.4783	1.0359
3		15	66.7624	57.8796	4.582	3.7862	1.4289	1.0011
4	30s × 100 m	5	78.3946	65.9687	4.8361	3.8865	1.5038	1.0294
5		10	70.9197	61.8651	4.5667	3.6965	1.4941	1.0470
6		15	68.7614	59.8540	4.7245	3.8973	1.4577	1.0208
7	60s × 100 m	5	79.0716	66.5784	4.7947	3.8825	1.4612	1.0170
8		10	72.0659	62.9452	4.5320	3.6734	1.4786	1.0421
9		15	68.4411	59.3672	4.5489	3.7447	1.4396	1.0069

both ends of the road, thus can't capture the vehicle condition between the detector. As expected, higher probe vehicle penetration and smaller cells significantly improved the overall performance of the method. This happens because a higher probe vehicle penetration provides more observations and thus more accurate input. In respect to the different cell sizes (e.g., scenes 1, 2, and 3; scenes 4, 5, and 6; and scenes 7, 8, and 9), it can be seen both that more speed information is obtained as the probe vehicle penetration increases, and that the speed estimates for the spatio-temporal grid are more accurate with the lowest RMSE and MAE for the position. Higher penetration rates, however, produce slightly larger speed and dithering errors compared to medium penetration rates (e.g., speed errors for scenes 2 and 3, scenes 5 and 6, and scenes 8 and 9) because dense data points interfere more frequently with speed continuity. It is also notable that for the same probe vehicle penetration, a larger cell improves acceleration and dithering performance compared to a smaller cell since the short spatio-temporal interpolation width serves to establish fast local speed transitions. The combined position, speed, and acceleration error analysis results showed that the reconstruction accuracy of the third cell combination is higher; thus, the third cell combination (i.e., cell size of 30×50 and probe vehicle penetration of 15%) is chosen as the initial model parameter for all subsequent experiments.

4.3. Model accuracy comparison

4.3.1. Multi-vehicle trajectory reconfiguration

Vehicles on a road can be divided by state into driving, stopping, and stop-and-go, which correspond, respectively, to free flow, stopping waves, and congestion sections in the vehicle trajectory spatio-temporal diagram. To better simulate the multiple states of vehicles in the driving process, therefore, as well as to better reflect the influence of surrounding vehicles on the state of self-vehicles, it is necessary to analyze the spatio-temporal consistency of the reconstructed multi-vehicle and reference trajectories in a certain time period. Fig. 7 shows a comparison of the reconstructed vehicle trajectories obtained using the three

methods with the corrected NGSIM data in the first 15 min. For all three reconstruction methods, the entry time of the reconstructed vehicles and average speed observed by the fixed detector are used as model inputs; the NKSVT-TR and VT-TR methods added the probe vehicle trajectories with a penetration rate of 15% (48 vehicles) as the reference trajectory. The spatio-temporal network of multi-vehicle trajectories for the three methods in free-flow and congested flow states are presented in Figs. 8 and 11 in order to analyze the trajectory reconstruction effects in more detail.

As shown in Fig. 7, in comparison with the vehicle trajectory reconstructed using the FI-TR method shown in Fig. 7d, the two 3D kinematic wave methods based on variational theory (Fig. 7b and Fig. 7c) better reflect the stop-and-go behavior of vehicles in a congested environment. This is due to the fact that the variational theory based on 3D kinematic waves considers the trajectory curve of the probe vehicle and therefore generates shape constraints on the curve of the reconstructed trajectory. Meanwhile, the vehicle trajectories reconstructed by the FI-TR method are close to a straight line, although the combination of the vehicle travel period adds a random value to the vehicle speed. This happens because the FI-TR method assumes that the vehicle is in a closed and empty roadway during the reconstruction process, and not affected by the surrounding environment and vehicles. In addition, the nonparametric kernel smoothing method in Fig. 7b improves the estimation performance because the NKSVT-TR method results in fewer trajectory gaps and crossings than the VT-TR method. Thus, NKSVT-TR captures the speed disturbances of cells more effectively, independently of the variational theory approach used in the second stage.

Figs. 8 and 9 show the performance of the three reconstruction methods under different traffic states, where the orange curve represents the trajectories of the probe vehicles and the gray curve represents the reconstructed trajectories of each vehicle. Figs. 8a~8c show the performances of the three methods under free-flow conditions. As shown in the figures, the vehicle is closer to a free-flow state downstream of the road. The reconstructed multi-vehicle trajectories in Fig. 8a are closer to the trajectories of the probe vehicles at both ends compared with Fig. 8b

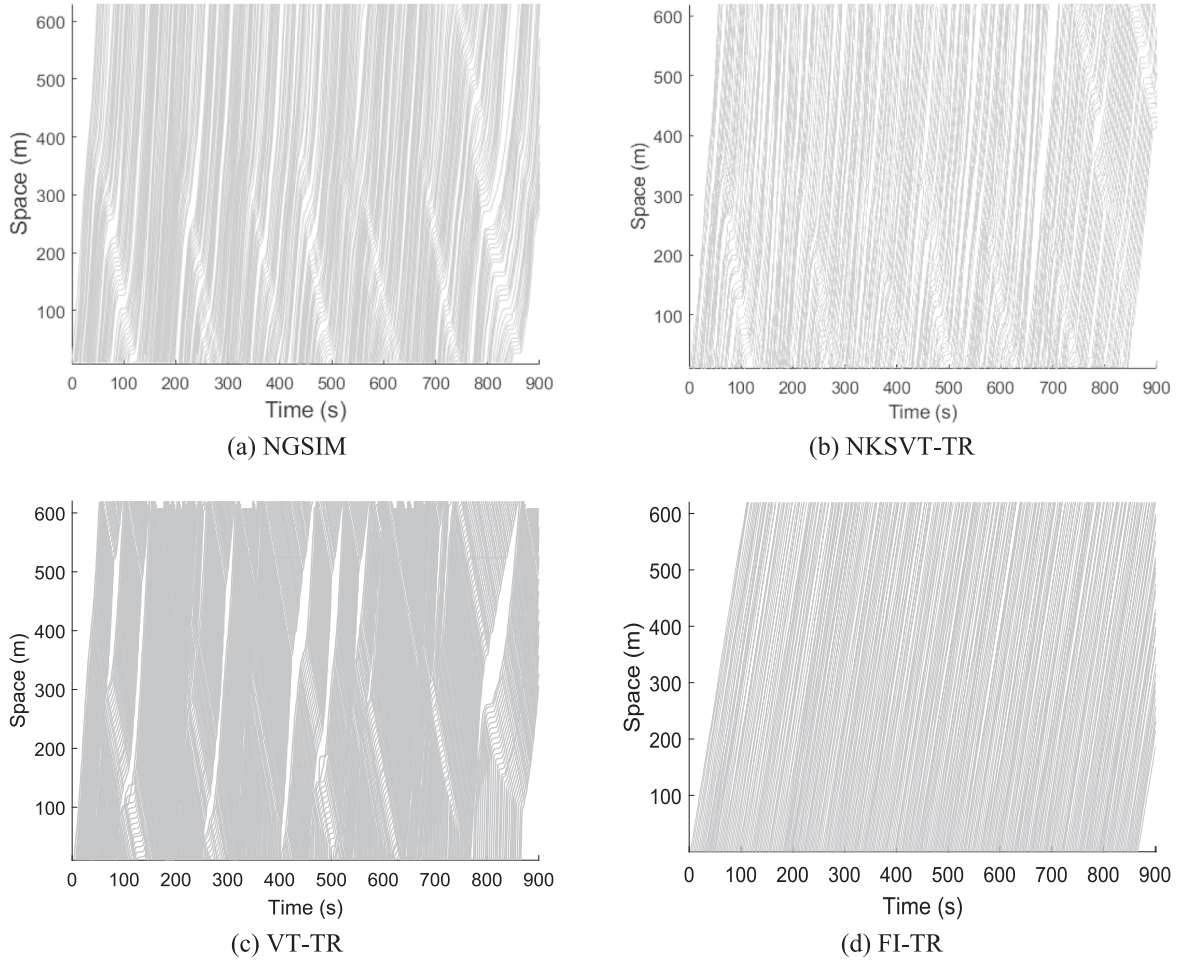


Fig. 7. Comparison of reconstructed trajectories.

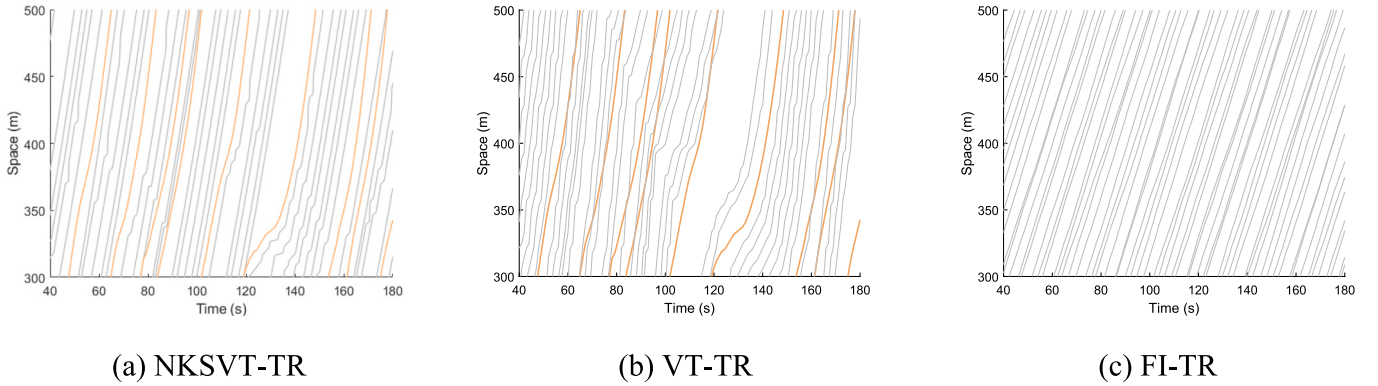


Fig. 8. Comparison of reconstructed trajectories of the three methods under free-flow conditions.

and Fig. 8c, and the reconstructed effect is better. The reconstructed vehicle trajectories in Fig. 8b cause some vehicle trajectories to cross due to not having a brother accurate estimate of the basic road conditions. The reconstructed vehicle trajectories in Fig. 8c at the free flow state is smoother and neater, but the trajectory similarity is higher and more desirable without affecting the vehicle trajectories downstream because of the stop-and-go behavior of the vehicle in the middle of the road.

Figs. 9a~9c show the performances of the three methods under congested flow conditions. Among the two methods based on 3D

kinematic wave theory, Fig. 9a better reflects the stop-and-go behavior under congested traffic conditions compared with Fig. 9b. And because the local speed under congested conditions is not clearly reflected, the trajectory estimated in Fig. 9b conflicts with the trajectory of the probe vehicle. As shown in Fig. 9c, meanwhile, the FI-TR method of reconstruction shows the same ideal trajectory in congested flow conditions as in free-flow conditions; an outcome that can be attributed mainly to the fact that this approach does not consider the influence of other vehicles in the road.

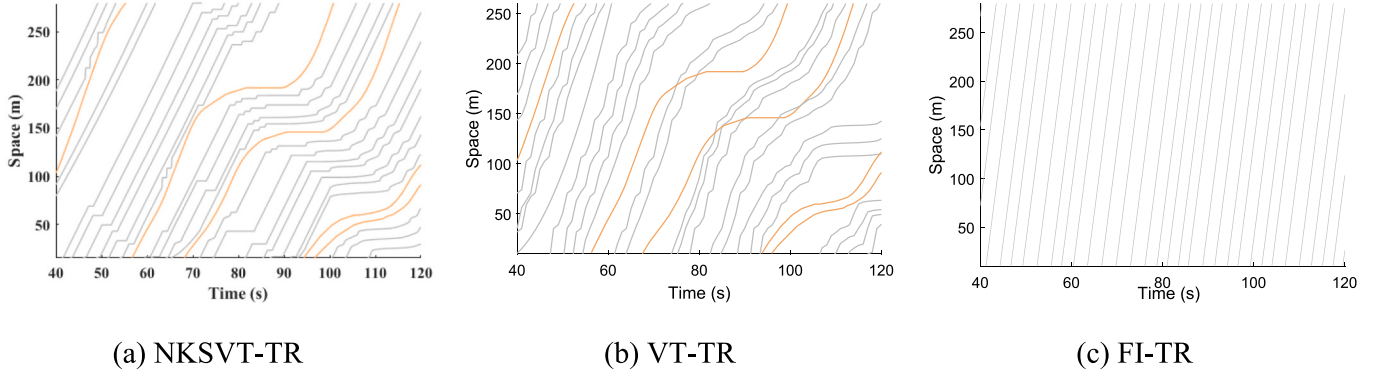


Fig. 9. Comparison of reconstructed trajectories of the three methods under congested flow conditions.

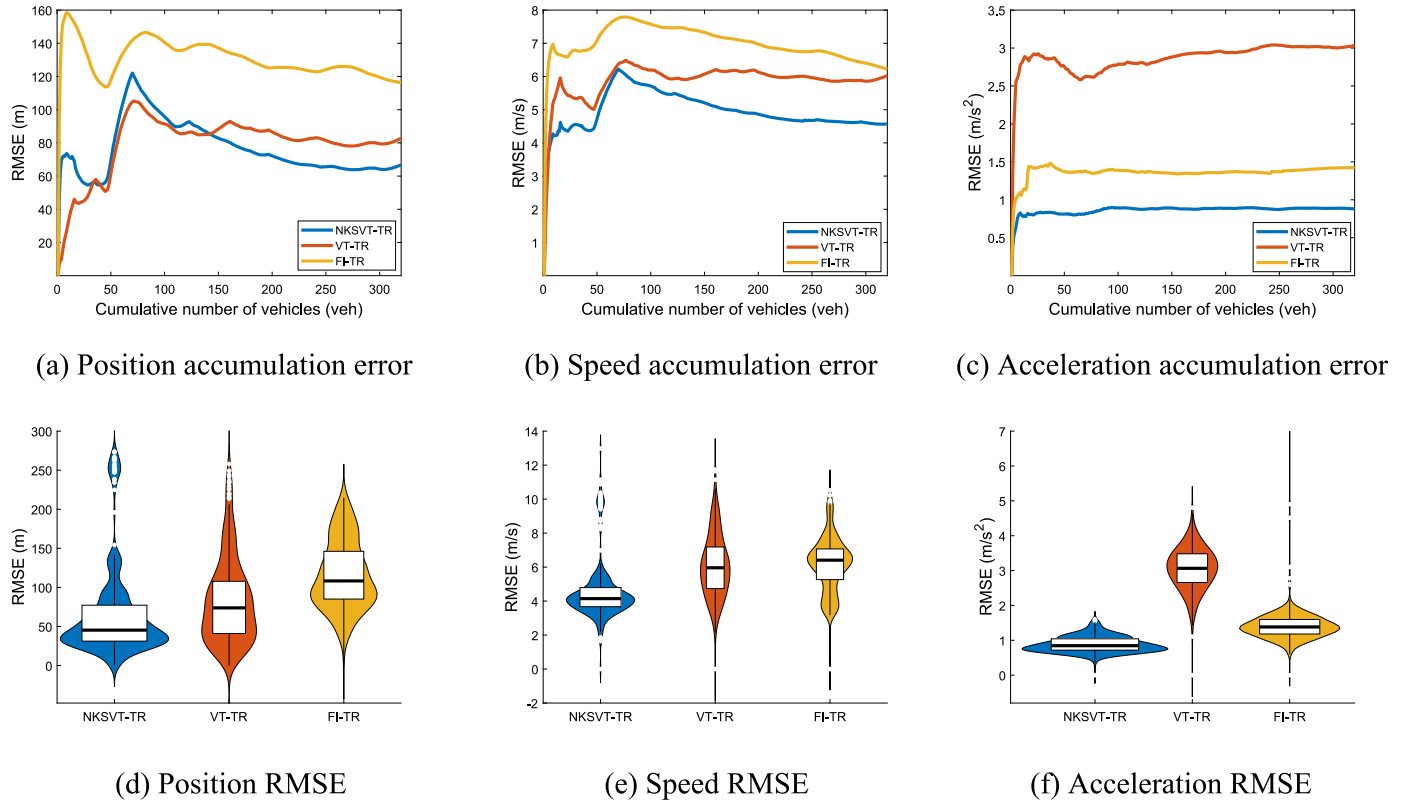


Fig. 10. Comparison of the accuracy of the three reconstruction methods.

4.3.2. Comparison of the accuracy of the various reconstructed trajectories

To verify the effectiveness of the method in this study, the errors under the microscopic and reference variables of the reconstructed trajectories were calculated for all three methods. The convergence profiles of the average RMSE of the vehicles as the cumulative number of vehicles increase are shown in Figs. 10a – 10c. In comparison with the other two methods, the method proposed in this study has a lower overall position error. In contrast, the vehicle trajectories reconstructed by VT-TR exhibit a large error owing to the neglect of the random wave error in speed, while the FI-TR method ignores the constraints of other vehicles in the network on the road, thus making the resulting trajectory too idealized; a feature that again causes a larger trajectory error. The speed error in Fig. 10b and acceleration error in Fig. 10c exhibit a similar pattern to the position error in Fig. 10a. In comparison with the other two reconstruction methods, therefore, the NKSVT-TR method is superior for estimating the local speed surface.

Figs. 10d – 10(f) show the RMSE violin plots of the three micro-variables of the reconstructed vehicle trajectories compared with the reference NGSIM data, where the thick black line in the rectangular box is the median of the data; the range of the black box is from the lower to the upper quartile, and the width of the graph indicates the probability density falling at the corresponding value. The results show that all the proposed methods in this paper have better estimation effect than the other two methods. Specifically, the median of the RMSE of the proposed method is 57.68%, 25.92% and 75.73% lower in position, speed, and acceleration, respectively, than that derived from the VT-TR method. Further, the median RMSE of the proposed method is, respectively, 71.16%, 31.10% and 69.63% lower for the same three micro-variables than that of the FI-TR method. In addition, by comparing the graph widths of the three methods for position, speed, and acceleration, it can be seen that NKSVT-TR is more robust.

4.3.3. Energy emission analysis of road segments

To further analyze the effects of errors in the micro traffic data, macro traffic data, and different macro and micro conversion methods on traffic energy emission calculations, this section compares the results of four different combinations of traffic models and energy consumption and emission models. We used MOVES, i.e., a microscopic energy consumption and emission model from the United States, and EMFAC, i.e., a macroscopic energy consumption and emission model from California, USA, for the energy consumption and emission calculations. The MOVES model calculates the microscopic emissions of vehicles by meticulously portraying the activity characteristics of vehicles at the microscopic level. The EMFAC model, meanwhile, as a macroscopic fuel consumption emission model, can calculate rich motor vehicle pollutants, such as HC, CO, NO_x, PM, CO₂, N₂O, and CH₄. Because microscopic data have higher vehicle driving detail characterization and thus relatively high accuracy in this experiment, the combined results of the microscopic NGSIM data and microscopic energy consumption emission MOVES model are used as the baseline data for this experiment.

The experimental framework is illustrated in Fig. 11. First, the combination of macro PeMS data and the macro fuel consumption emission model EMFAC (PeMS & EMFAC) is used as a reference result representative of the macro calculation results. Second, the macro PeMS data are disaggregated into micro data using three reconstruction methods (NKSVT-TR, VT-TR, and FI-TR) and then input to the MOVES model to calculate the energy consumption emission results, i.e., the combinations of NKSVT-TR & MOVES, VT-TR & MOVES, and FI-TR & MOVES. Finally, NGSIM & MOVES are used as the benchmark for the calculation of energy consumption emissions in the experimental section.

In general, this study analyzes the results of different emission estimation methods by comparing Mean Absolute Percentage Error (MAPE) values in the interval of $[-20, 20]$ kW/t. When performing emission calculations, the MOVES model requires an input of 1 Hz trajectory data, and thus the experiments downsampled the NGSIM data to meet the model input requirements.

The VSP distributions of the three reconstruction methods with the reference data are shown in Fig. 12. In comparison with Fig. 12a, the MAPEs of the VSP distributions of the three methods with reference data were 10.57%, 16.18%, and 23.74%, respectively. This result indicates that the VKS-VT method in this study can more accurately reflect the microscopic characteristics of vehicles for better calculation of vehicle emissions. Frey [47] indicates that VSP and speed can jointly determine the vehicle operation mode to reflect the instantaneous speed and acceleration level of the vehicle. From Fig. 12b, it can be seen that the VSP generated by the method in this paper is distributed in the interval of $[-20, 20]$ kW/t, and the value of VSP = 0 has the largest proportion, which is consistent with the performance of Fig. 12a. Meanwhile,

Fig. 12c has VSP < 0, implying that the vehicle state is mostly in the deceleration and braking states. This is because the VT-TR method only connects the grid point connections as vehicle trajectories without proper smoothing of the trajectories, resulting in vehicles prone to acceleration outliers at the connection points and overlapping reconstructed and detected vehicle trajectories owing to the lack of accurate estimation of the road traffic state (as shown in Fig. 9b). Fig. 12d has the largest proportion in the VSP > 0 part, indicating that the vehicle is mostly in coasting as well as acceleration states. This is because the method determines the driving speed of the vehicle and thus the influence of the surrounding vehicles based only on the speed values of the upstream and downstream fixed detectors, while the speed values of the downstream detectors are generally larger than the values of the upstream detectors. The reconstruction process therefore assumes that the vehicle is in a process of slow acceleration.

Figs. 13a and Fig. 13b show the calculated total vehicle fuel consumption and pollutant emissions (CO₂, CO, and NO_x) for each of the five combinations.

As shown in Fig. 13a, the accuracy of the fuel consumption calculation results for the micro-combination NGSIM & MOVES is improved by 26.64% compared with the macro-combination PeMS & EMFAC. The accuracy of the microscopic fuel consumption results reconstructed by the three reconstruction methods NKSVT-TR, VT-TR, and FI-TR is improved by 31.55%, 41.09%, and 10.02%, respectively. By comparing the fuel consumption results of the reconstructed trajectories of the three reconstruction methods with those of NGSIM, it is found that the relative error between the fuel consumption calculated by NKSVT-TR, VT-TR, and FI-TR and the fuel consumption emissions calculated by the micro-combination NGSIM & MOVES is 3.88%, 10.46% and 11.95%, respectively. This shows that the proposed method is able to calculate a relatively more accurate level of fuel consumption. As shown in Figs. 13b - 13d, by comparing the emission results of PeMS & EMFAC and NGSIM & MOVES, it is found that the results of the micro-combination NGSIM & MOVES are 9.17%, 33.68%, and 45.26% more accurate, respectively, while NKSVT-TR & MOVES, VT-TR & MOVES, and FI-TR & MOVES are more accurate than the PeMS & EMFAC macro combination by, respectively, 8.75%, 14.83%, and 3.03% for CO₂, 53.03%, 54.46%, and 8.41% for CO, and 83.75%, 111.20%, and 19.97% for NO_x. After comparing the pollutant emission results between the reconstructed and reference NGSIM trajectories obtained using the three reconstruction methods, the relative errors between the NKSVT-TR, VT-TR, and FI-TR method and the reference trajectories by, respectively, 0.39%, 5.18%, and 5.62% for CO₂, 14.48%, 15.55%, and 18.90% for CO, and 26.50%, 45.40%, and 17.41% for NO_x. This shows that method proposed in this study is able to pollutant emissions markedly more accurately than other current methods. The fuel consumption and emission results are summarized in Table 2.

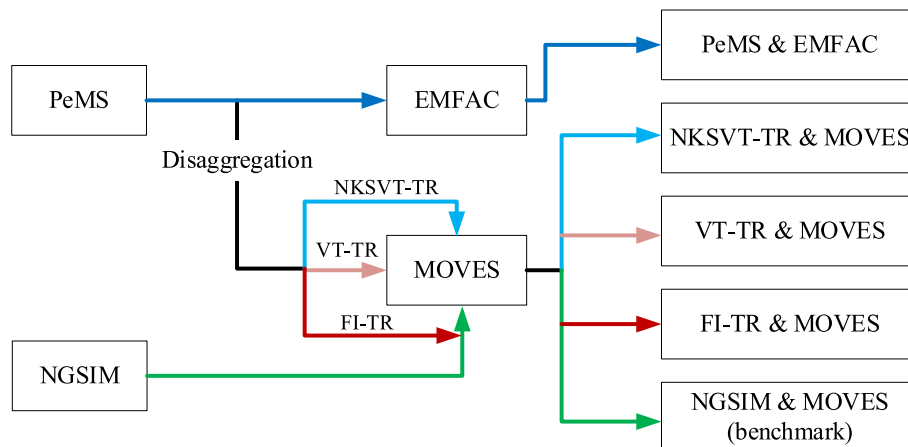


Fig. 11. Cross-validation of traffic data with energy consumption models.

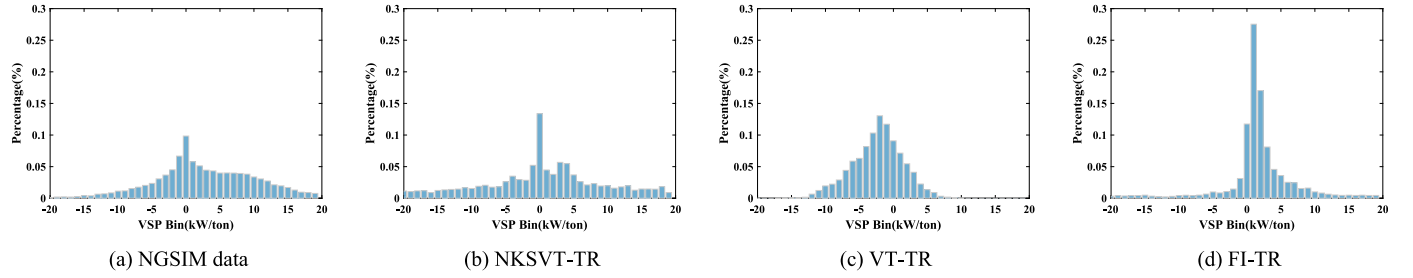


Fig. 12. Comparison of VSP distribution of reconstructed trajectories.

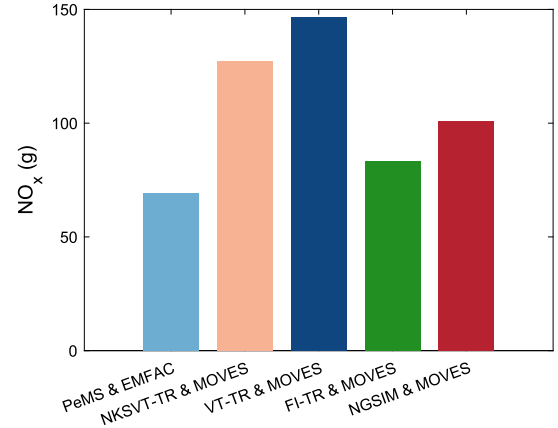
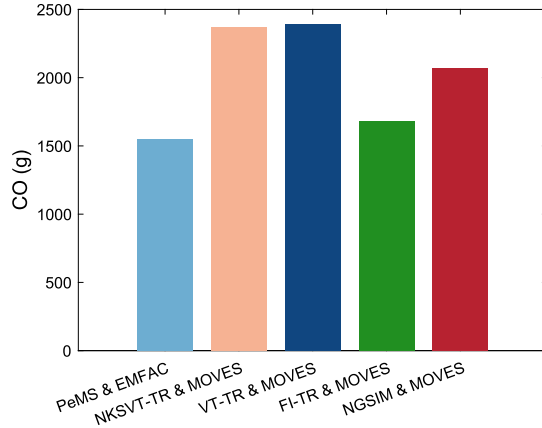
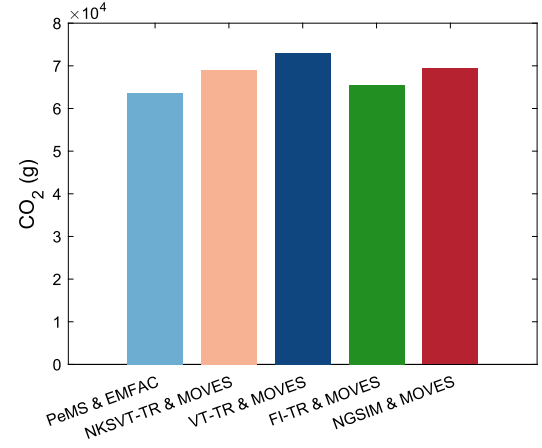
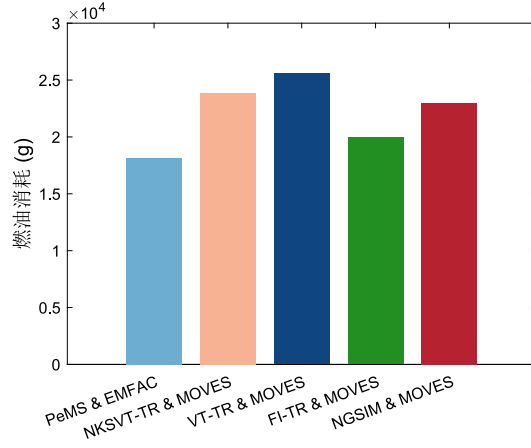


Fig. 13. Comparison of fuel consumption and emissions under different model combinations.

Table 2
Comparison of fuel consumption (FC) and emissions with real value.

type	Diff. with PeMS & EMFAC (%)			Diff. with NGSIM & MOVES (%)		
	NKSVT-TR	VT-TR	FI-TR	NKSVT-TR	VT-TR	FI-TR
FC	31.55	41.09	10.02	3.88	10.46	11.95
CO ₂	8.75	14.83	3.03	0.39	5.18	5.62
CO	53.03	54.46	8.41	14.48	15.55	18.90
NO _x	83.75	111.20	19.97	26.50	45.40	17.41

As shown in Fig. 13, it can be seen that the NKSVT-TR method has better estimation performance than the VT-TR and FI-TR methods, and the numerical results are closer to the calculated results of the microscopic combination NGSIM & MOVES. The VT-TR method overestimates the numerical results because it is gridded using the FD of the road, and this prevents it from accurately estimating the speed evolution of the entire spatio-temporal network. The results of the FI-TR method are closer to those of the macroscopic combination PeMS & EMFAC because the vehicle velocities calculated using the four-corner interpolation

method are based on the average velocities of the upstream and downstream detectors weighted by their distance, which essentially averages the speeds at both ends of the road, and is therefore closer to the fuel consumption emission results calculated using the macroscopic average speed. In summary, the results of this study are closer to the estimated fuel consumption and emission results from microscopic vehicle trajectories and provide more accurate estimates of the actual results.

5. Conclusions

In this study, first, we use a nonparametric kernel smoothing method to fuse speed observations from multiple sources of data so as to overcome the drawbacks of the limited probe range of fixed sensors. This allows us to better reflect the traffic speed evolution between detectors, improve the accuracy of the road traffic state estimation, and effectively avoid the overlap of reconstructed trajectories. Second, we show that using mobile detector vehicle data as a reference can improve the accuracy of reconstructed trajectories under congested flow, better reflect the traffic kinematic wave characteristics caused by road congestion, and simulate the stop-and-go behavior of vehicles under congested flow conditions. The smaller the cell size and the higher the penetration rate of the probe vehicle, the higher the accuracy of the reconstructed trajectory will be. Third, the experimental results show that the proposed method is able to provide more accurate estimation, with a median reduction of 25.92%–75.73% in the RMSE of displacement, velocity, and acceleration compared to the VT-TR method, and 31.10%–71.16% in the three micro-variables compared to the FI-TR method, respectively. Next, the results of macro-micro combination models with different granularities show that the NKSVT-TR method reconstructs the trajectory with fuel consumption emission results closest to the micro-combination NGSIM & MOVES, and the relative errors of fuel consumption (CO_2 , CO, and NO_x) with the micro-combination NGSIM & MOVES are 0.39%–26.65%; in comparison with VT-TR, the errors were reduced by 6.88%–92.47%, and 23.39%–93.06% compared with FI-TR, thus indicating that the method proposed in this study is able to calculate fuel consumption emissions is closer to actual microscopic results.

It should be noted here, however, that the behaviors of overtaking, lane changing, and merging are not considered in data extraction. Accordingly, more fixed detectors will be set in bottleneck areas of the road in the next step. This could help to identify changes of vehicles in position and speed, and reconstruct different interaction behavior of vehicles, including lane-changing, overtaking and merging.

Funding

This work was supported by the National Key Research and Development Program of China (No.2021YFB2501200), Beijing Natural Science Foundation (No. L211027 and No. 9232003), the National Natural Science Foundation of China (No.61703053), the Shaanxi Province Key Research and Development Program (2022GY-300).

CRediT authorship contribution statement

Wen-Long Shang: Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Mengxiao Zhang:** Writing – original draft, Methodology, Data curation. **Guoyuan Wu:** Writing – review & editing, Conceptualization. **Lan Yang:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Shan Fang:** Methodology, Writing – review & editing. **Washington Ochieng:** Writing – review & editing, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] Yang Z, Li Q, Shang W, Yan Y, Ochieng W. Examining influence factors of electric vehicle market demand based on online reviews under moderating effect of subsidy policy. *Appl Energy* 2022;326:120019.
- [2] Zhou P, Gao SZ, Lv Y, Zhao G. Energy transition management towards a low-carbon world. *Front Eng Manag* 2022;9(3):499–503.
- [3] Shang WL, Lv Z, Han. Low carbon technology for carbon neutrality in sustainable cities: a survey. *Sustain Cities Soc* 2023;92:104489.
- [4] Yang Y, Wang H, Löschel A, et al. Energy transition toward carbon-neutrality in China: pathways, implications and uncertainties. *Front Eng Manag* 2022;9:358–72.
- [5] Bi H, Shang WL, Chen Y, Wang K, Yu Q, Sui Y. GIS aided sustainable management for urban road transportation systems with a unifying queuing and neural network model. *Appl Energy* 2021;291:116818.
- [6] Samaras C, Tsokolis D, Toffolo S, Magra G, Samaras Z. Improving fuel consumption and co 2 emissions calculations in urban areas by coupling a dynamic micro traffic model with an instantaneous emissions model. *Transp Res Part D Transp Environ* 2017;65:772–83.
- [7] Yu Z, Li W, Liu Y, Zeng X, Zhao Y, Chen K, et al. Quantification and management of urban traffic emissions based on individual vehicle data. *J Clean Prod* 2021;328:129386.
- [8] Shao S, Tan Z, Liu Z, Shang W. Balancing the GHG emissions and operational costs for a mixed fleet of electric buses and diesel buses. *Appl Energy* 2022;328:120188.
- [9] Bramich D, Menendez M, Ambuehl L. Fitting empirical fundamental diagrams of road traffic: a comprehensive review and comparison of models using an extensive data set. *IEEE Trans Intell Transp Syst* 2021;23(9):14104–27.
- [10] Saffari E, Yildirimoglu M, Hickman M. Data fusion for estimating macroscopic fundamental diagram in large-scale urban networks. *Transp Res C* 2022;137: Article 103555.
- [11] Islam MR, Hadiuzzaman M, Barua S, Shimu TH. Alternative approach for vehicle trajectory reconstruction under spatiotemporal side friction using lopsided network. *Iet Intell Transp Syst* 2019;13(2):356–66.
- [12] Aljamal Mohammad A, Abdelghaffar Hossam M, Rakha Hesham A. Real-time estimation of vehicle counts on signalized intersection approaches using probe vehicle data. *IEEE Trans Intell Transp Syst* 2022;5:2719–29.
- [13] Misra A, Roorda MJ, MacLean HL. An integrated modelling approach to estimate urban traffic emissions[J]. *Atmos Environ* 2013;73:81–91.
- [14] Zhao B, Lin Y, Hao H, et al. Fuel consumption and traffic emissions evaluation of mixed traffic flow with connected automated vehicles at multiple traffic scenarios [J]. *J Adv Transp* 2022;2022:1–14.
- [15] Xiong C, Shahabi M, Zhao J, et al. An integrated and personalized traveler information and incentive scheme for energy efficient mobility systems. *Transp Res Pt C-Emerg Technol* 2020;113:57–73.
- [16] Xu J, Hilker N, Turchet M, Al-Rijleh M-K, Tu R, Wang A, et al. Contrasting the direct use of data from traffic radars and video-cameras with traffic simulation in the estimation of road emissions and PM hotspot analysis. *Transp Res Part D Transp Environ* 2018;62:90–101.
- [17] Delphine L, Arnaud C, Nicole S, Ludovic L. Accounting for traffic speed dynamics when calculating COPERT and PHEM pollutant emissions at the urban scale. *Transp Res Part D Transp Environ* 2018;63:588–603.
- [18] Grote M, Williams I, Preston J, Kemp S. A practical model for predicting road traffic carbon dioxide emissions using inductive loop detector data - sciencedirect. *Transp Res Part D Transp Environ* 2018;63:809–25.
- [19] Hao P, Wang C, Wu G, et al. Evaluating the environmental impact of traffic congestion based on sparse mobile crowd-sourced data[A]. In: 2017 IEEE conference on technologies for sustainability (SusTech)[C]. IEEE; 2017. p. 1–6.
- [20] Choi HW, Frey HC. Light duty gasoline vehicle emission factors at high transient and constant speeds for short road segments[J]. *Transp Res Part D Transp Environ* 2009;14(8):610–4.
- [21] Turkensteen M. The accuracy of carbon emission and fuel consumption computations in green vehicle routing[J]. *Eur J Oper Res* 2017;262(2):647–59.
- [22] Zegeye SK, De Schutter B, Hellendoorn J, et al. Integrated macroscopic traffic flow, emission, and fuel consumption model for control purposes[J]. *Transp Res Part C Emerg Technol* 2013;31:158–71.
- [23] Wang M, Daamen W, Hoogendoorn S, et al. Estimating acceleration, fuel consumption, and emissions from macroscopic traffic flow data[J]. *Transp Res Rec* 2011;2260:123–32.
- [24] Chen Z, Yang C, Chen A. Estimating fuel consumption and emissions based on reconstructed vehicle trajectories[J]. *J Adv Transp* 2014;48(6):627–41.
- [25] Hou Y, Edara P, Sun C. Situation assessment and decision making for lane change assistance using ensemble learning methods. *Expert Syst Appl* 2015;42(8):3875–82.
- [26] Herrera JC, Bayen AM. Traffic flow reconstruction using mobile sensors and loop detector data. University of California Transportation Center, Working Papers; 2007.
- [27] Marczak F, Buisson C. New filtering method for trajectory measurement errors and its comparison with existing methods. *Transp Res Rec J Transp Res Board* 2012; 2315(1):35–46.
- [28] Van Lint JWC, et al. Empirical evaluation of new robust travel time estimation algorithms. *Transp Res Rec* 2018;2160(1):50–9.

- [29] Ni D, Wang H. Trajectory reconstruction for travel time estimation. *J Intell Transp Syst* 2008;12(3):113–25.
- [30] Wu G, Boriboonsomsin K, Barth M. Development and evaluation of intelligent energy management strategies for plug-in hybrid electric vehicles 2012;15(3): 1091–100.
- [31] Krol L. The reconstruction of vehicle trajectories with dynamic macroscopic data. Master Thesis, University of Twente; 2009.
- [32] Coifman B. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. *Transp Res Part A Policy Pract* 2002;36(4):351–64.
- [33] Treiber M, Helbing D. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooperative Transp Dyn* 2002;1. 3.1–3.24.
- [34] Van Lint JWC, Hoogendoorn S. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Comput Aided Civ Inf Eng* 2009;24:1–17.
- [35] Chen X, Yin J, Tang K, Tian Y, Sun J. Vehicle trajectory reconstruction at signalized intersections under connected and automated vehicle environment. *IEEE Trans Intell Transp Syst* 2022;23(10):17986–8000.
- [36] Chen P, Wei L, Meng F, Zheng N. Vehicle trajectory reconstruction for signalized intersections: a hybrid approach integrating Kalman filtering and variational theory. *Transp B Transp Dyn* 2021;9(1).
- [37] Wei L, Wang Y, Chen P. A particle filter based approach for vehicle trajectory reconstruction using sparse probe data. *IEEE Trans Intell Transp Syst* 2020;22(5): 2878–90.
- [38] Feng Y, Sun J, Chen P. Vehicle trajectory reconstruction using automatic vehicle identification and TrafficCount data. *J Adv Transp* 2015;49(2):174–94.
- [39] Xu X, Lint HV, Verbraeck A. A generic data assimilation framework for vehicle trajectory reconstruction on signalized urban arterials using particle filters. *Transp Res Part C Emerg Technol* 2018;92:364–91.
- [40] Mehran B, Kuwahara M. Fusion of probe and fixed sensor data for short-term traffic prediction in urban signalized arterials. *Int J Urban Sci* 2013;17(2):163–83.
- [41] Daganzo CF. A variational formulation of kinematic waves: solution methods. *Transp Res Part B Method* 2005;39(10):934–50.
- [42] Daganzo CF. A variational formulation of kinematic waves: basic theory and complex boundary conditions[J]. *Transp Res Part B Method* 2005;39(2):187–96.
- [43] Sun Z, Hao P, Ban XJ, Yang D. Trajectory-based vehicle energy/emissions estimation for signalized arterials using mobile sensing data. *Transp Res D* 2015; 34:27–40.
- [44] Tsanakas N, Ekström J, Olstam J. Generating virtual vehicle trajectories for the estimation of emissions and fuel consumption. *Transp Res Part C Emerg Technol* 2022;138:103615.
- [45] Chen X, Yin J, Qin G, et al. Integrated macro-micro modelling for individual vehicle trajectory reconstruction using fixed and mobile sensor data. *Transp Res Part C* 2022;145.
- [46] He Z, Lv Y, Lu L, Guan W. Constructing spatiotemporal speed contour diagrams: using rectangular or non-rectangular parallelogram cells? *Transp.B* 2019;7:44–60.
- [47] Frey HC. Trends in onroad transportation energy and emissions. *J Air Waste Manage Assoc* (1995) 2018;68(6):514–63.
- [48] Montanino M, Punzo V. Making ngSim data usable for studies on traffic flow theory: multistep method for vehicle trajectory reconstruction. *Transp Res Rec J Transp Res Board* 2013;2390:99–111.
- [49] He Z, Zhang W, Jia N. Estimating carbon dioxide emissions of freeway traffic: a spatiotemporal cell-based model. *IEEE Trans Intell Transp Syst* 2019;21(5): 1976–86.