

## Research Article

# Map Matching for Fixed Sensor Data Based on Utility Theory

Kangkang He,<sup>1</sup> Qi Cao,<sup>1</sup> Gang Ren ,<sup>1</sup> Dawei Li,<sup>1</sup> and Shuichao Zhang<sup>2</sup>

<sup>1</sup>School of Transportation, Southeast University, Nanjing 211189, China

<sup>2</sup>School of Civil and Transportation Engineering, Ningbo University of Technology, Ningbo 315211, China

Correspondence should be addressed to Gang Ren; [rengang@seu.edu.cn](mailto:rengang@seu.edu.cn)

Received 14 January 2021; Revised 20 February 2021; Accepted 5 March 2021; Published 27 March 2021

Academic Editor: Hai Ying Li

Copyright © 2021 Kangkang He 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.

Map matching can provide useful traffic information by aligning the observed trajectories of vehicles with the road network on a digital map. It has an essential role in many advanced intelligent traffic systems (ITSs). Unfortunately, almost all current map-matching approaches were developed for GPS trajectories generated by probe sensors mounted in a few vehicles and cannot deal with the trajectories of massive vehicle samples recorded by fixed sensors, such as camera detectors. In this paper, we propose a novel map-matching model termed Fixed-MM, which is designed specifically for fixed sensor data. Based on two key observations from real-world data, Fixed-MM considers (1) the utility of each path and (2) the travel time constraint to match the trajectories of fixed sensor data to a specific path. Meanwhile, with the laws derived from the distribution of GPS trajectories, a path generation algorithm was developed to search for candidates. The proposed Fixed-MM was examined with field-test data. The experimental results show that Fixed-MM outperforms two types of classical map-matching algorithms regarding accuracy and efficiency when fixed sensor data are used. The proposed Fixed-MM can identify 68.38% of the links correctly, even when the spatial gap between the sensor pair is increased to five kilometers. The average computation time spent by Fixed-MM on one point is only 0.067 s, and we argue that the proposed method can be used online for many real-time ITS applications.

## 1. Introduction

Map matching is the process of correctly identifying the path on which a vehicle is travelling [1]. It provides a promising opportunity to upgrade the service level of various intelligent traffic system (ITS) applications [2–4]. However, the current map-matching algorithms are generally designed for satellite-based GPS points that are provided by probe sensors mounted on probe vehicles. These probe vehicles provide spatial traffic information and direct measurements of travel time to monitor the traffic conditions in a citywide road network.

However, probe sensor data have limitations. The cost of purchasing GPS units and transferring data can severely limit the scale of probe samples. Only a biased estimation of the traffic information can be obtained because the probe data are usually collected from one type of vehicle, such as taxis. Additionally, a probe sensor system imposes an enormous computational burden on the system

administration owing to high polling frequency and positional noise [5].

Fixed sensor data show the potential to overcome the issues existing in the probe sensor data. Fixed sensors, such as cameras, loops, and microwaves, are widely used in urban traffic monitoring and management (with the development of ITS technology, camera sensors have been improved in terms of accuracy, cost, and ease of use. Therefore, the fixed sensor data considered in this paper refer specifically to the observations collected through camera-based sensors). The transit information of every vehicle approaching the fixed sensor station is captured. Consequently, the movement patterns of almost all vehicles running on a road network with fixed sensors can be recorded. This provides opportunities to reduce the estimation bias in traffic information. The fixed sensor system may also improve the efficiency of the map-matching process with a reduced polling frequency and more accurate location record, even for a large-scale urban traffic system.

Many map-matching methods have been developed, and their reviews can be found in [1, 6]. Quddus et al. classified the methods into four categories: geometric, topological, probabilistic, and advanced. However, such approaches can only perform well with high-frequency GPS data and may become less effective for low-frequency trajectory data [6]. In recent years, two groups of methods, namely, HMM-based algorithms and ST-Matching algorithms, have been developed to deal with the sparsity issue of low-polling frequency trajectory data.

- (i) HMM-based algorithms: Newson and Krumm [7] introduced a two-step map-matching algorithm based on a hidden Markov chain for a sparse GPS trajectory, called the HMM algorithm. First, this method finds a set of candidate links for each GPS point and defines a measurement probability to describe how the GPS point is aligned with each candidate link. Then, it connects each pair of consecutive candidate links with the shortest path to generate the candidate graph. Next, a transition probability defines the likelihood of the tracking vehicle moving along each candidate path. Finally, the best matching path sequence is identified using the Viterbi algorithm. The experimental results show that even with sampling intervals of 30 s, the accuracy of this algorithm is barely degraded. However, it has high computational complexity and becomes slow when working with long trajectories and extended search radii. Mohamed et al. [8] employed three filters (i.e., speed, direction, and  $\alpha$ -trimmed mean filters) to reduce the candidate sets for improving the efficiency of the map-matching process. Koller et al. [9] proposed a fast-HMM algorithm that replaces the Viterbi algorithm with the bidirectional Dijkstra to determine the optimal map-matching solution. This algorithm can avoid up to 45% of the costly routing operations without negatively affecting the map-matching result. Han et al. [10] partitioned road networks into approximate segments and then indexed the approximate segments into an optimised packed  $R$  tree to improve the road-network search duration. It has also been argued that mobility in a road network is non-Markovian. Jagadeesh and Srikanthan [11] complemented the HMM algorithms with the concept of drivers' route choice. The results show that this improves matching accuracy further, especially at high levels of noise.
- (ii) ST-Matching algorithms: Lou et al. [12] introduced a map-matching algorithm for low-polling frequency GPS trajectories based on both spatial and temporal analysis, called ST-Matching. It modelled temporal analysis using speed and travel time data to improve its accuracy. The experimental results show that ST-Matching is more robust to the decrease in sampling rate than the map-matching algorithm using only spatial information, indicating that temporal constraints are indeed useful in map matching with sparse trajectory data. Considering that this method

cannot handle the matching error well at junctions, Hsueh and Chen [13] introduced directional analysis to ST-Matching, called STD-Matching. It employs real-time directional motion with the directional analysis function to reflect the influence of a user's true movement over the GPS trajectories. The experimental results demonstrate that the STD-Matching algorithm significantly improves the matching accuracy. Liu et al. [14] proposed a spatial and temporal conditional random field map-matching method called the ST-CRF algorithm. The ST-CRF model considers both spatial and temporal accessibility between two GPS points, in addition to consistency in the direction of travel. A series of experiments showed that the ST-CRF method has better performance and robustness and solves the "label-bias" problem in the HMM algorithm.

The above-mentioned map-matching algorithms are mainly designed for low-frequency probe sensor data, such as GPS trajectories. They may become less effective for fixed sensor data because the fixed sensor data differ from probe sensor data in at least two aspects:

- (a) The fixed sensor data are much sparser than the probe sensor data. As shown in Figure 1(a), the distance between consecutive points recorded by fixed sensors is usually dozens of times that recorded by the probe sensor. Hence, there are too many possible paths to be matched between neighbouring fixed sensors. If only the shortest path length is considered (as in the current map-matching algorithms developed for probe sensors), the realistic paths may not be adequately evaluated.
- (b) The positions provided by the fixed sensors are fixed and accurate, while the probe sensors move along with the probe vehicle and generate GPS points with random errors [15]. Figure 1(b) presents a microscopic view of the trajectories between the fixed sensors 20200906 and 10203801. One easily finds that the fixed sensor data (green points) are located accurately on the road links, and the probe sensor data (red points) are always positioned several meters away from the true path.

In this study, we developed a map-matching algorithm designed specifically for fixed sensor data, called Fixed-MM. For this purpose, the conventional map-matching models for probe sensor data are abbreviated as Probe-MM. The contributions of Fixed-MM can be summarised as follows:

- (a) It combines both route choice preferences and temporal constraints to identify the true path of the fixed sensor data. The experimental results show that the proposed method significantly improves the matching accuracy.
- (b) Fixed-MM developed a candidate-path generation algorithm to search for a realistic path by relaxing the assumption that the location of each point is noisy. In this manner, the time-consuming candidate-path

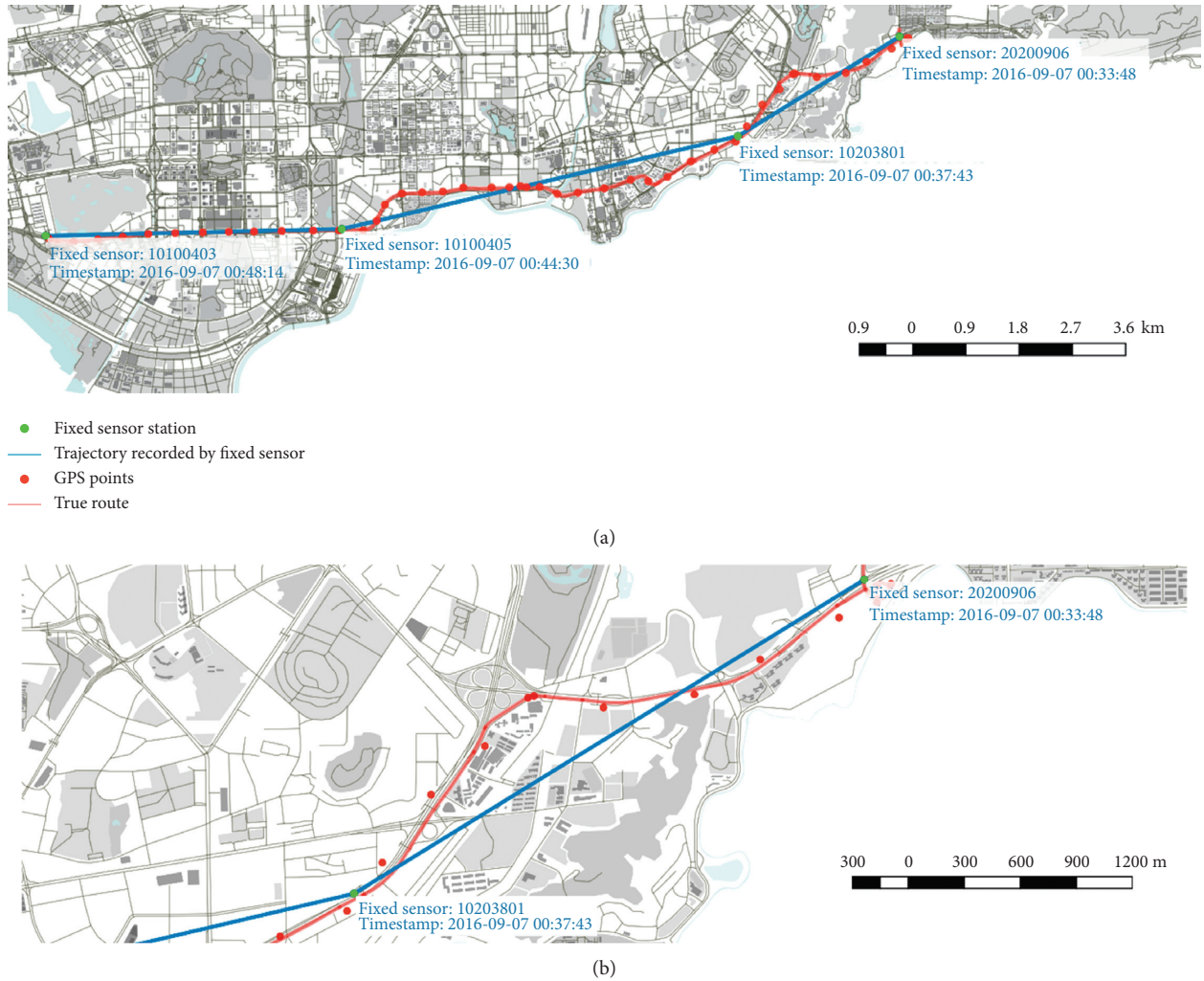


FIGURE 1: Comparison of fixed sensor and probe sensor data. (a) Macroscopic and (b) microscopic views of one vehicle's trajectories recorded by the probe and fixed sensors.

generation process can be conducted separately and in parallel, and average computation time of the matching process for a point is reduced to 0.067 s.

The remainder of this paper is organised as follows. The problem definition and overview of the framework are presented in the Preliminaries section. Then, the Fixed-MM algorithm and candidate-path set generation algorithm are proposed in the Methodology section. The Experiment section details the process and presents the experimental results. Finally, we conclude the paper in the last section.

## 2. Preliminaries

**2.1. Formulation of the Map-Matching Problem.** To better illustrate Fixed-MM, the definitions of variables and the problem are introduced in this section.

**Definition 1.** Road network: a road network (RN) consists of a set of road links  $\{l\}$  connected in a graph format. Each road link,  $l$ , is a directed edge with two terminal points, a length (l.len), a level (l.lev) (e.g., an expressway, a primary road, or a

secondary road), a direction (l.di) (e.g., one-way or bidirectional), and the number of lanes (l.lan).

**Definition 2.** Path: path  $P$  is represented by a sequence of connected road links,  $P: l_1, l_2, \dots, l_x, \dots, l_X$ , in an RN.

**Definition 3.** Fixed sensor trajectory: a fixed sensor trajectory,  $Tr$ , is a sequence of time-ordered points,  $Tr: F_{id^{(1)}}, F_{id^{(2)}}, \dots, F_{id^{(j)}}, \dots, F_{id^{(l)}}$ , where each point  $F_{id^{(j)}}$  has a unique identification number,  $id$ , geospatial coordinate,  $(F_{id^{(j)}} \cdot lon, F_{id^{(j)}} \cdot lat)$ , and timestamp,  $F_{id^{(j)}} \cdot t$ .

**Definition 4.** Sensor pair: a sensor pair is two neighbouring points in a  $Tr$ , namely,  $(F_{id^{(j)}}, F_{id^{(j+1)}})$ ,  $j = 1, 2, \dots, J-1$ , where  $F_{id^{(j)}}$  is the original fixed sensor point and  $F_{id^{(j+1)}}$  is the destination fixed sensor point.

**Definition 5.** Candidate path set: the candidate path set,  $\Phi_j$ , consists of all paths with a nonzero probability of matching between a given sensor pair  $(F_{id^{(j)}}, F_{id^{(j+1)}})$ , while all unrealistic paths have a probability of zero.

Now the problem of Fixed-MM is defined as follows.

**Problem 1.** Given a fixed sensor trajectory  $Tr$  and a road network  $RN$ , for each sensor pair  $(F_{id^{(j)}}, F_{id^{(j+1)}})$  in  $Tr$ , find a path  $P_i$  from  $\Phi_j$  with the highest probability of being a matched path.

**2.2. Framework.** The framework of Fixed-MM is illustrated in Figure 2. Three types of datasets, including fixed sensor data, probe sensor data, and road network data, are used as inputs. The trajectory of the fixed sensor data is first decomposed into separate sensor pairs. The probe sensor data are also matched with a specific path based on the Probe-MM algorithm. Meanwhile, a candidate path generation algorithm is used to search for possible paths for each sensor pair. Then, the matching probability for each candidate path is calculated, and the matching results can be attained by finding the candidate path with the highest matching probability.

### 3. Methodology

**3.1. Characteristics of the Data.** The key to Fixed-MM is finding the most likely path to connect the sensor pair. In this section, we provide two key observations of the true trajectories that lead to the proposed approach. Figure 3(a) illustrates the GPS trajectory of 1365 sample vehicles travelling between the sensor pair  $(F_{20507303}, F_{20501803})$ , and they are taken as examples to illustrate the observed laws.

**Observation 1.** The drivers prefer to travel along the path with high utility.

**Example 1.** Consider path A, path B, and path C visualised in Figure 3(a) with their attributes summarised in Table 1. Sixty-eight percent of the samples travel path A, while only 32% of the samples travel along the other two. Thus, it is reasonable to infer that drivers prefer to choose paths with less travel time, fewer intersections, and more high-level road links, which indicates that the higher the utility of the path, the more attractive the path is to the driver.

**Observation 2.** The observed travel time tends to be close to the expected travel time of the true path.

**Example 2.** Based on the Prob-MM algorithm, the GPS trajectories can be matched to three paths. The histograms of the observed travel times for the three paths are calculated in Figure 3(b). It is easily found that the histograms fit well to the normal distribution, which means that a path's observed travel time tends to be close to its expected travel time (average travel time). If the observed travel time of a sample is 18 min, we may infer that this trip is very likely to be matched with path C.

Based on the above observations, we propose a novel map-matching algorithm for fixed sensor data, namely, Fixed-MM that incorporates both (1) the utility of each route and (2) the travel time constraint to identify the path with the highest probabilities from the candidate path set as the

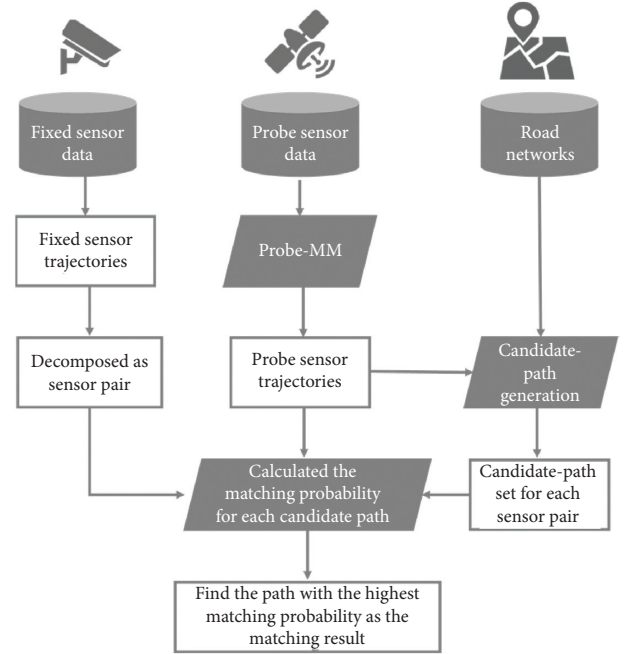


FIGURE 2: Framework of the proposed Fixed-MM Algorithm.

matched path. Details of the utility model, travel time constraint, and candidate path set generation algorithm are described in the following subsections.

**3.2. Utility Model.** Similar to the route choice model, the travel behaviour preference reflected in Observation 1 is modelled with utility theory. It assumes that the driver's preference for a path is captured by a value called utility, and the driver selects the path in the candidate set with the highest utility [16].

Let  $U_{i,j}$  be the utility of the  $i$ th path  $P_{i,j}$  belonging to the candidate set  $\Phi_j$  of the sensor pair:  $(F_{id^{(j)}}, F_{id^{(j+1)}})$ . It consists of a deterministic term  $V_{i,j}$  and a random term  $\varepsilon_{i,j}$  such that

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j}. \quad (1)$$

The random term  $\varepsilon_{i,j}$  is assumed to be independent and identically distributed (i.i.d.) as a Gumbel distribution. The deterministic term is assumed to have a linear relationship with path attributes, such that

$$V_{i,j} = \beta^{\text{FTT}} x_{i,j}^{\text{FTT}} + \beta^{\text{NSL}} x_{i,j}^{\text{NSL}} + \beta^{\text{PE}} x_{i,j}^{\text{PE}}, \quad (2)$$

where  $x_{i,j}^{\text{FTT}}$ ,  $x_{i,j}^{\text{NSL}}$ , and  $x_{i,j}^{\text{PE}}$  are vectors of the observed path attributes and  $\beta^{\text{FTT}}$ ,  $\beta^{\text{NSL}}$ , and  $\beta^{\text{PE}}$  are vectors of coefficients that represent drivers' preferences on path attributes. The descriptions of the path attributes are presented in Table 2.

Based on the above definitions of path utility, the matching probability of a candidate path  $P_{i,j}$  is given by [16]

$$\Pr(P_{i,j}) = \frac{e^{V_{i,j}}}{\sum_{P_{i',j} \in \Phi_j} e^{V_{i',j}}}. \quad (3)$$

Equation (3) can also be transformed as

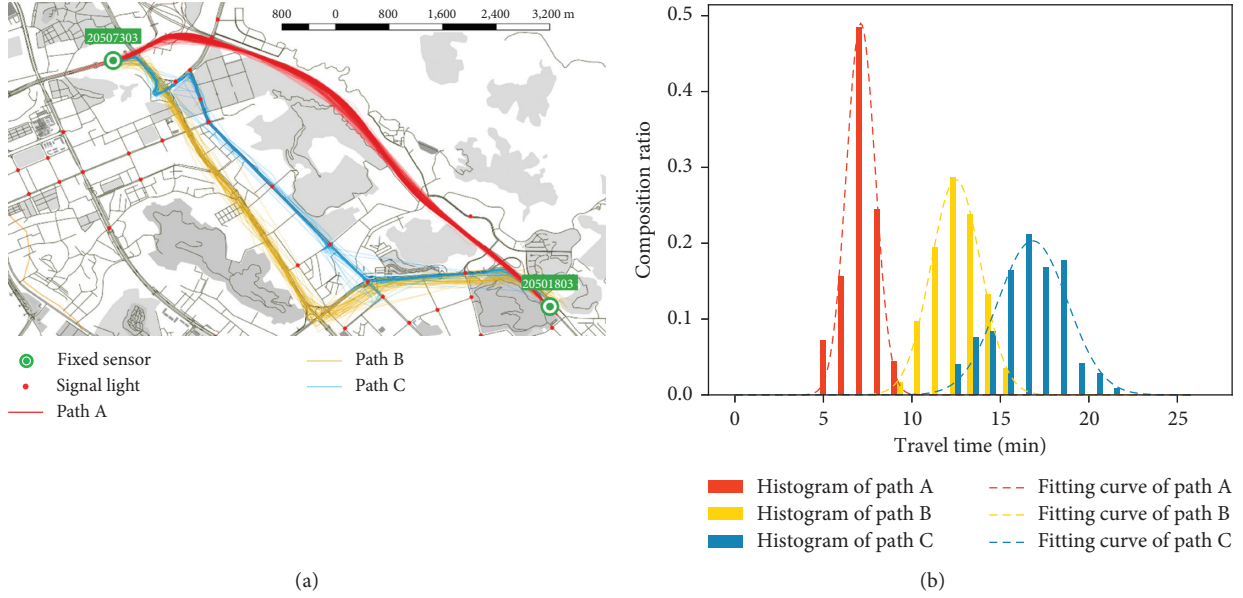


FIGURE 3: The sensor pair example. (a) The location of the sensor pair example and the GPS trajectories. (b) The distribution of travel times.

TABLE 1: Attributes of path A and path B.

Attributes	Path A	Path B	Path C
Length of route (m)	8239.25	9022.70	8971.15
Number of signal lights	0	0	7
Average travel time (min)	7.11	12.49	16.83
Proportion of expressway	1	1	0.37

$$\Pr(P_{i,j}) = \frac{e^{V_{i,j}}}{\sum_{P_{i',j} \in \Phi_j} e^{V_{i',j}}} = \frac{1}{\sum_{P_{i',j} \in \Phi_j} e^{V_{i',j} - V_{i,j}}}. \quad (4)$$

It is easy to find that the larger the difference between the utility  $V_{i,j}$  and the other  $V_{i',j}$ s, the higher the matching possibility,  $\Pr(P_{i,j})$ . This means that the candidate path with higher utility is more likely to be matched, which corresponds to the rule reflected in Observation 1.

**3.3. Temporal Constraint.** To consider Observation 2, the temporal constraint between the observed travel time and expected travel time of a candidate path must be modelled. Their definitions are as follows.

The observed travel time  $t_{j,n}$  is the time spent by the  $n$ th sample when travelling between sensor pairs ( $F_{id^{(j)}}$ ,  $F_{id^{(j+1)}}$ ) and can be obtained by calculating the difference between the transit timestamps recorded by  $F_{id^{(j)}}$  and  $F_{id^{(j+1)}}$ :

$$t_{j,n} = F_{id^{(j+1)}} \cdot t - F_{id^{(j)}} \cdot t. \quad (5)$$

The expected travel time  $\widehat{t}_{i,j}$  is the average travel time of the candidate path,  $P_{i,j}$ , where  $P_{i,j} \in \Phi_j$ . This can be calculated based on probe sensor data:

$$\widehat{t}_{i,j} = \sum_{l_x \in P_{i,j}} \frac{\sum_{n=1}^N t_{x,n}}{N_x}, \quad (6)$$

where  $t_{x,n}$  is the travel time spent by the  $n$ th sample on road link  $l_x$ , and  $N_x$  is the total number of probe vehicles traversing road link  $l_x$ .

The temporal constraint can be calculated based on the deviation  $t_{j,n} - \widehat{t}_{i,j}$  between the observed  $t_{j,n}$  and the expected travel times,  $\widehat{t}_{i,j}$ . This is attributed to a combination of the natural variation in travel times and the error in the travel time estimate. The deviations of the three sample paths are shown in Figures 4–6 in Appendix A, respectively. The travel time varies significantly on different paths depending on the time of day, and all the histograms of  $t_{j,n} - \widehat{t}_{i,j}$  during the morning peak fit well to the normal distribution. Therefore, we can assume that the deviations have a Gaussian distribution  $t_{j,n} - \widehat{t}_{i,j} \sim N(\mu_s, \sigma_s)$ .  $\mu_s$  and  $\sigma_s$  are the mean and variance of  $t_{j,n} - \widehat{t}_{i,j}$  for the candidate path  $P_{i,j}$ , during period  $s$ . Then, the temporal constraint  $q(t_{j,n} - \widehat{t}_{i,j})$  can be defined as

$$q(t_{j,n} - \widehat{t}_{i,j}) = \frac{e^{-0.5(t_{j,n} - \widehat{t}_{i,j} - \mu_s/\sigma_s)^2}}{\sum_{P_{i',j} \in \Phi_j} e^{-0.5(t_{j,n} - \widehat{t}_{i',j} - \mu_s/\sigma_s)^2}}. \quad (7)$$

The denominator aims at normalizing the temporal constraint to one.

We added the temporal constraint as a correction term for the utility function. Then, the matching probability can be rewritten as

$$\Pr(P_{i,j}) = \frac{e^{V_{i,j} + \alpha \ln q(t_{j,n} - \widehat{t}_{i,j})}}{\sum_{P_{i',j} \in \Phi_j} e^{V_{i',j} + \alpha \ln q(t_{j,n} - \widehat{t}_{i',j})}}, \quad (8)$$

where  $\alpha$  is a scale parameter. The correct term  $\alpha \ln q(t_{j,n} - \widehat{t}_{i,j})$  in equation (8) describes the likelihood of compliance between the observed  $t_{j,n}$  and expected travel time  $\widehat{t}_{i,j}$ . When  $t_{j,n} - \widehat{t}_{i,j}$  is smaller (the observed travel time is closer to the expected travel time),  $q(t_{j,n} - \widehat{t}_{i,j})$  becomes larger.

TABLE 2: Attributes used for path utility.

Attributes	Parameters	Descriptions
$x_{i,j}^{\text{NSL}}$	$\beta^{\text{NSL}}$	Number of signal lights (NSL): the number of signal lights along the path
$x_{i,j}^{\text{FTT}}$	$\beta^{\text{FTT}}$	Free travel time (FTT): free flow travel time along the path (unit: s)
$x_{i,j}^{\text{PE}}$	$\beta^{\text{PE}}$	The proportion of expressway (PE): the proportion of expressway links

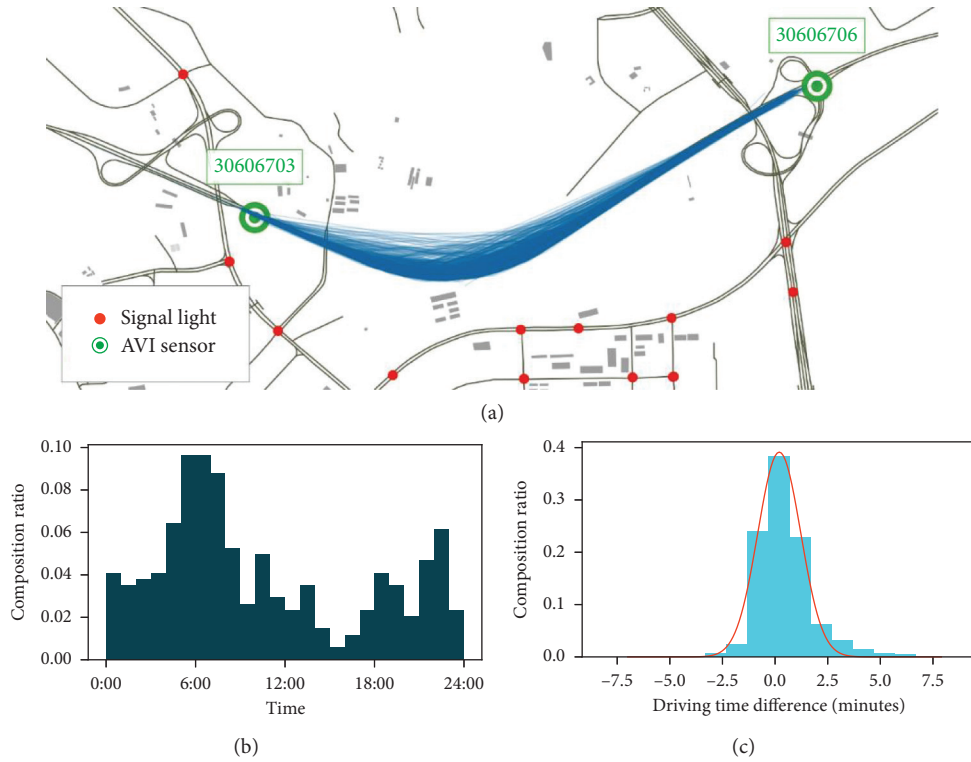


FIGURE 4: (a) GPS trajectories of samples. (b) Temporal distribution of the samples. (c) Histogram of the temporal constraint between 6:00 and 7:00 (fitting result:  $\mu = 0.20$  and  $\sigma = 1.02$ , goodness of fit: 0.99).

According to equation (4), the matching probability increases  $P_{i,j}$ . This is also in line with Observation 2 in the previous section.

**3.4. Generating Candidate Path Set.** Finding all possible paths that connect each sensor pair as candidates is another key step for Fixed-MM. The candidate path set is usually large, with a long distance between the paired sensors, and a dense urban road network. In addition, preferential and realistic paths should be included because comparing a path to a set of highly unattractive and unrealistic candidates would not provide much useful information [17]. In this study, we develop a protocol for generating a realistic candidate path set based on the following observations:

*Observation 3.* There may be certain detours on the candidate paths.

*Example 3.* Figure 7(a) illustrates the GPS trajectories of 620 samples that travel between sensors  $F_{20507301}$  and  $F_{20507302}$  near the Bao'an International Airport in Shenzhen, China.

Based on the map-matching algorithm designed for the probe data, each GPS point was projected onto a specific link. The observed number of samples on each link is represented by different colours in Figure 7(b). Most (92%) of the samples have a large offset against the shortest path, and the departure platform of the airport was chosen as a destination on the way. This indicates that there may be certain detours on these popular paths. These circuitous paths may be considered as unattractive alternatives for route choice models. However, they are popular candidates in the context of map-matching algorithms.

*Observation 4.* Trajectories captured by a sensor pair will not pass the links monitored by other fixed sensors.

*Example 4.* As shown in Figure 7(a), the road link monitored by the fixed sensor  $F_{20507403}$  has never been travelled by any vehicle captured by the sensor pair ( $F_{20507301}, F_{20507302}$ ). The reason for this phenomenon is that if a vehicle has travelled on the link where  $F_{20507403}$  located, the pass information will be recorded, and then the sensor pair

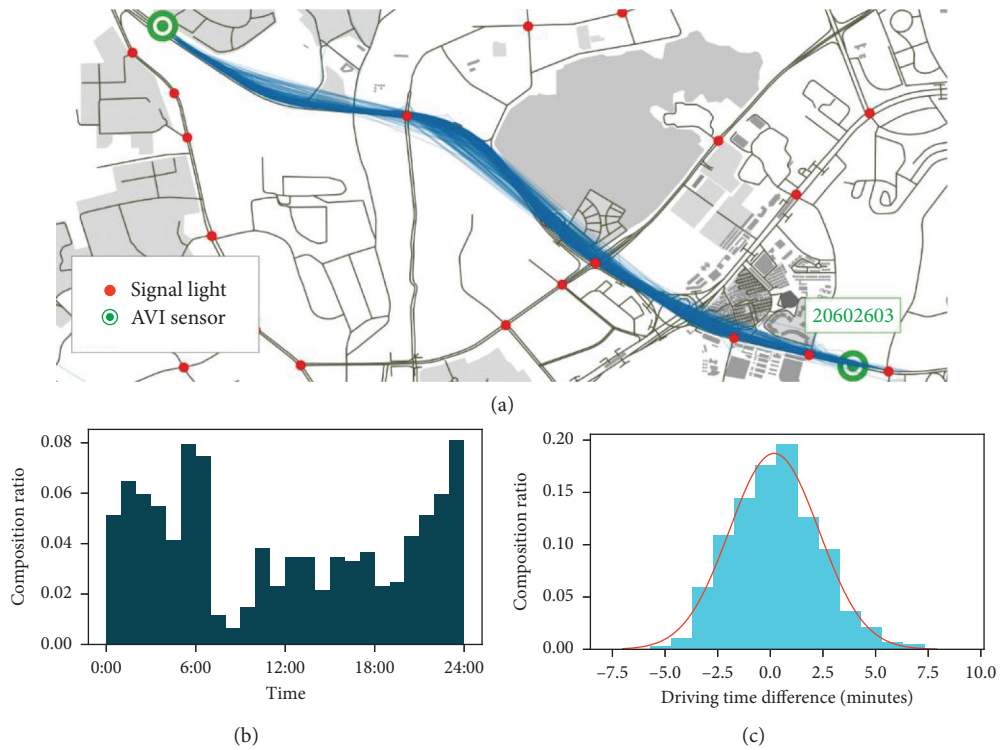


FIGURE 5: (a) GPS trajectories of samples. (b) Temporal distribution of the samples. (c) Histogram of the temporal constraint between 6:00 and 7:00 (fitting result:  $\mu = 0.18$  and  $\sigma = 2.13$ , goodness of fit: 0.98).

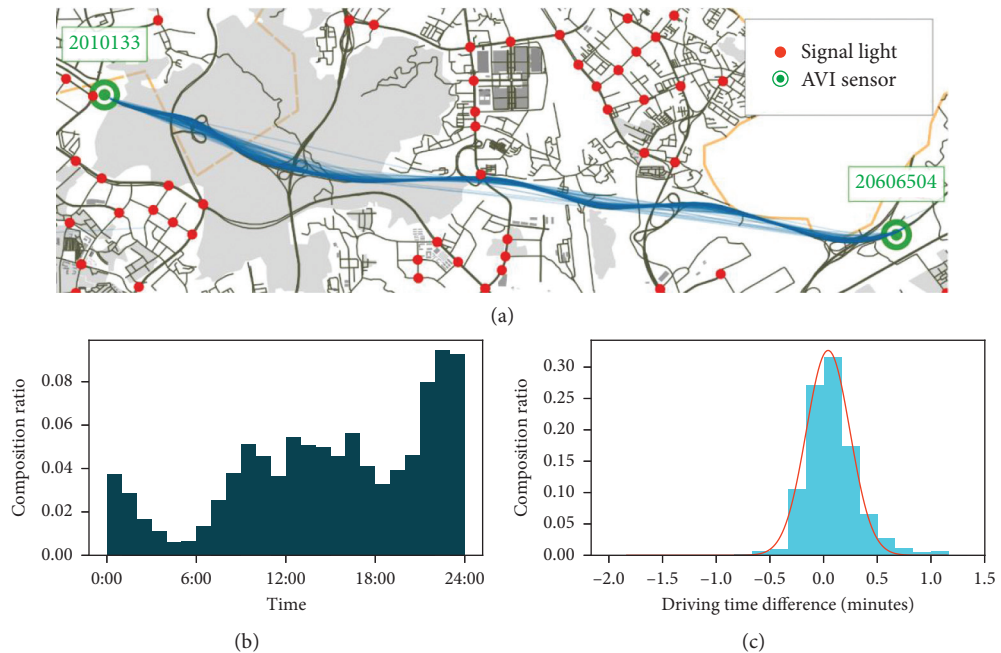


FIGURE 6: (a) Trajectories of samples. (b) Temporal distribution of the samples. (c) Histogram of the temporal constraint between 6:00 and 7:00 (fitting result:  $\mu = 0.25$  and  $\sigma = 1.22$ , goodness of fit: 0.99).

$(F_{20507301}, F_{20507302})$  will be decomposed into two sensor pairs, namely,  $(F_{20507301}, F_{20507403})$  and  $(F_{20507403}, F_{20507302})$ .

In this paper, we believe that historical GPS trajectories contain useful information about the composition of popular

candidates. Thus, the candidate path does not necessarily conform to behavioural assumptions but must be realistic; we use a biased random walk algorithm, which was first proposed by [17] to generate the candidate set. It draws a candidate path



FIGURE 7: Example of realistic path generation. (a) GPS trajectories. (b) Volumes of each link.

through a succession of random turns. The pseudocode of the candidate set generation algorithm is presented in Algorithm 1. The key to this algorithm is how the probability of turning is defined. In contrast to the original random walk algorithm, we set the turning probability of the links where other fixed sensors are located at 0 to satisfy the rule contained in Observation 4. In other situations, the turning probability is calculated based on field-test probe sensor data rather than the shortest path assumption. In this manner, the candidate path with the destination described in Observation 3 can be generated.

Based on the above analysis, the turning probability is defined as

$$\Pr(l_x, l_y) = \begin{cases} 0, & l_y \in \Phi_{\text{FS}} \text{ and } l_y \neq l_s, l_e, \\ \frac{N_{xy}}{\sum_{l_{y'} \in \Phi_x} N_{xy'}}, & \text{otherwise,} \end{cases} \quad (9)$$

where  $\Phi_{\text{FS}}$  is the set of links monitored by the fixed sensors,  $l_s$  is the start link where the origin fixed sensor is located,  $l_e$  is the end link where the destination fixed sensor is located,  $N_{xy}$  is the number of GPS trajectories traversing from link  $l_x$  to  $l_y$ , and  $\Phi_x$  is the set of outgoing links that connect the sink link  $l_x$ .

## 4. Experiment

**4.1. Experimental Dataset.** To examine the proposed Fixed-MM algorithm, both fixed and probe sensor data were used with the basic digital road network.

**Road Network:** the shapefile of the road network in Shenzhen, China, was used [18]. The network graph contained 237,440 vertices and 215,771 road links. As shown in Figure 8, the road network covers a  $40 \times 50$  km spatial area, with a total length of 21,985 km.

**Fixed sensor dataset:** A fixed sensor dataset generated by 715 cameras in Shenzhen from September 1 to October 31, 2016, was used. The transit information of vehicles was recorded, including license plate, timestamp, and detector ID.

**Probe sensor dataset:** we used a GPS trajectory dataset generated by 14,230 taxicabs during the same time range (from September 1 to October 31, 2016) as a probe sensor dataset. The GPS records include license plates, timestamps, and coordinates. The average sampling rate was set at 15 seconds per point.

With identical license plate information, we can extract the probe and fixed sensor data of the same taxicab as observed samples to train and test our model.

In the implementation, we removed noncontinuous driving trips. The main reason is that this noncontinuous driving part of the sample trips contains great uncertainty and will increase the estimation error of the Fixed-MM. Finally, 1,485,476 samples were extracted as a training dataset for estimation, while 156,192 samples were used as the testing dataset for evaluation. The estimation and evaluation of the Fixed-MM are introduced in the following sections.

**4.2. Model Estimation.** The coefficients of the Fixed-MM reflect the matching results' sensitivity to the variables. The



Input: The road network  $RN$  and the link pair  $(l_s, l_e)$ , where  $l_s$  and  $l_e$  are the links where the origin fixed sensor  $F_{id^{(j)}}$  and destination fixed sensor  $F_{id^{(j+1)}}$  are located.

Output: The candidate set  $\Phi_j$  for sensor pair  $(F_{id^{(j)}}, F_{id^{(j+1)}})$ .

Initialization

Set the candidate set:  $\Phi_j = \emptyset$

Set the size of the candidate set:  $DN$

Turning Probability

For  $l_x$  in road network  $RN$ :

Calculate the turning probability  $\Pr(l_x, l_y)$  based on equation (9).

Random Walk

While  $n < DN$  do

$l_x = l_s$

$P = [l_s]$

While  $l_y \neq l_e$  do

Randomly select a next link  $l_y$  based on the turning probability  $\Pr(l_x, l_y)$

Update the generated path:  $P.append(l_y)$

Update the current link:  $l_x = l_y$

End while

$n++$

Update the candidate set:  $\Phi_j = \Phi_j \cup P$

End while

ALGORITHM 1: Candidate set generation algorithm.

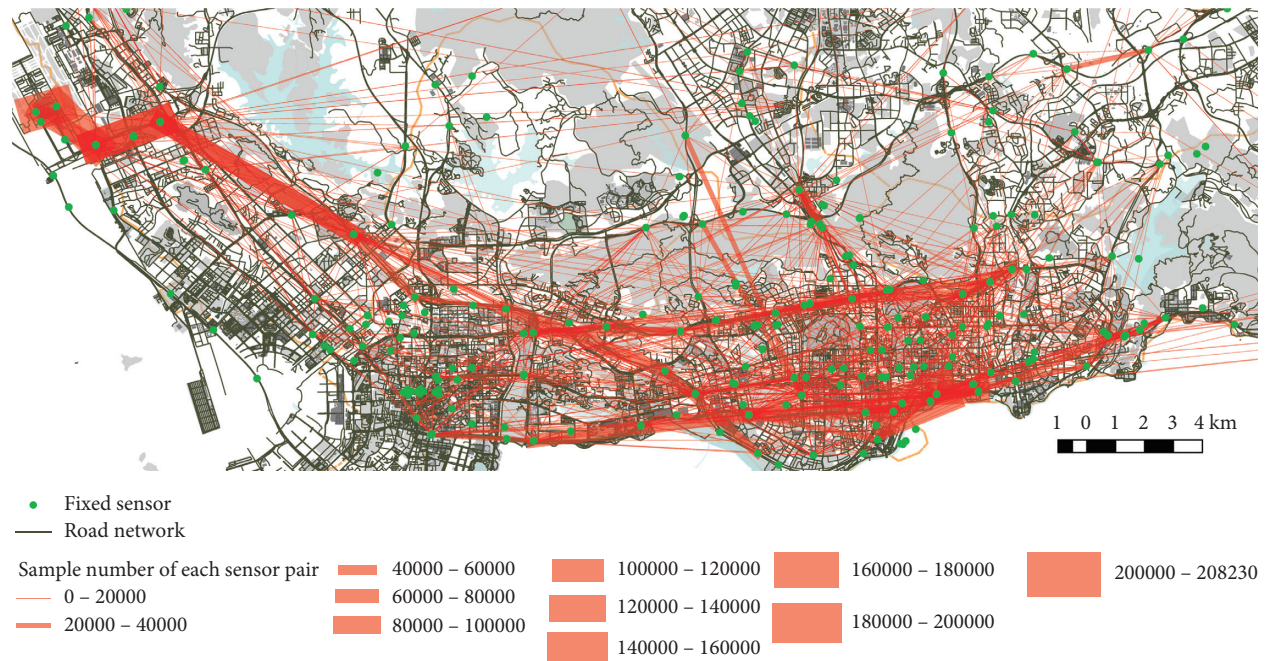


FIGURE 8: Distribution of observed trips.

values of the unknown parameters based on the training dataset must be identified. In this study, we consider the most widely used estimation procedure: the maximum likelihood technique [19].

Given the high number of sensor pairs, it is impossible to present detailed estimation results for each pair. Therefore, we only provide the detailed estimation results of the example sensor pair:  $(F_{2010002}, F_{1010403})$ . The GPS trajectories of the samples between this sensor pair are shown in Figure 9 in

Appendix B. The candidate path set generated by the algorithm proposed in this paper is illustrated in Figure 10.

Both the Fixed-MM model without temporal constraints (defined by equation (3)) and the Fixed-MM with temporal constraints (defined by equation (8)) are estimated. The estimation results of the two models are presented in Table 3, and several findings can be obtained.

Finding 1: as expected, the estimated parameter of “free travel time” and “number of signal lights” has a

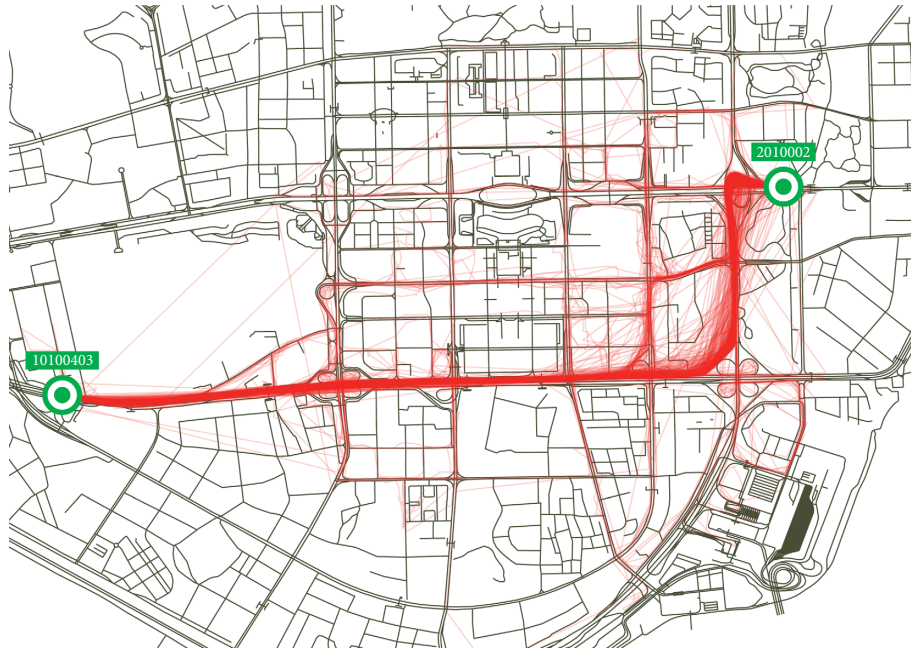


FIGURE 9: GPS trajectories between the example sensor pair:  $(F_{2010002}, F_{10100403})$ .

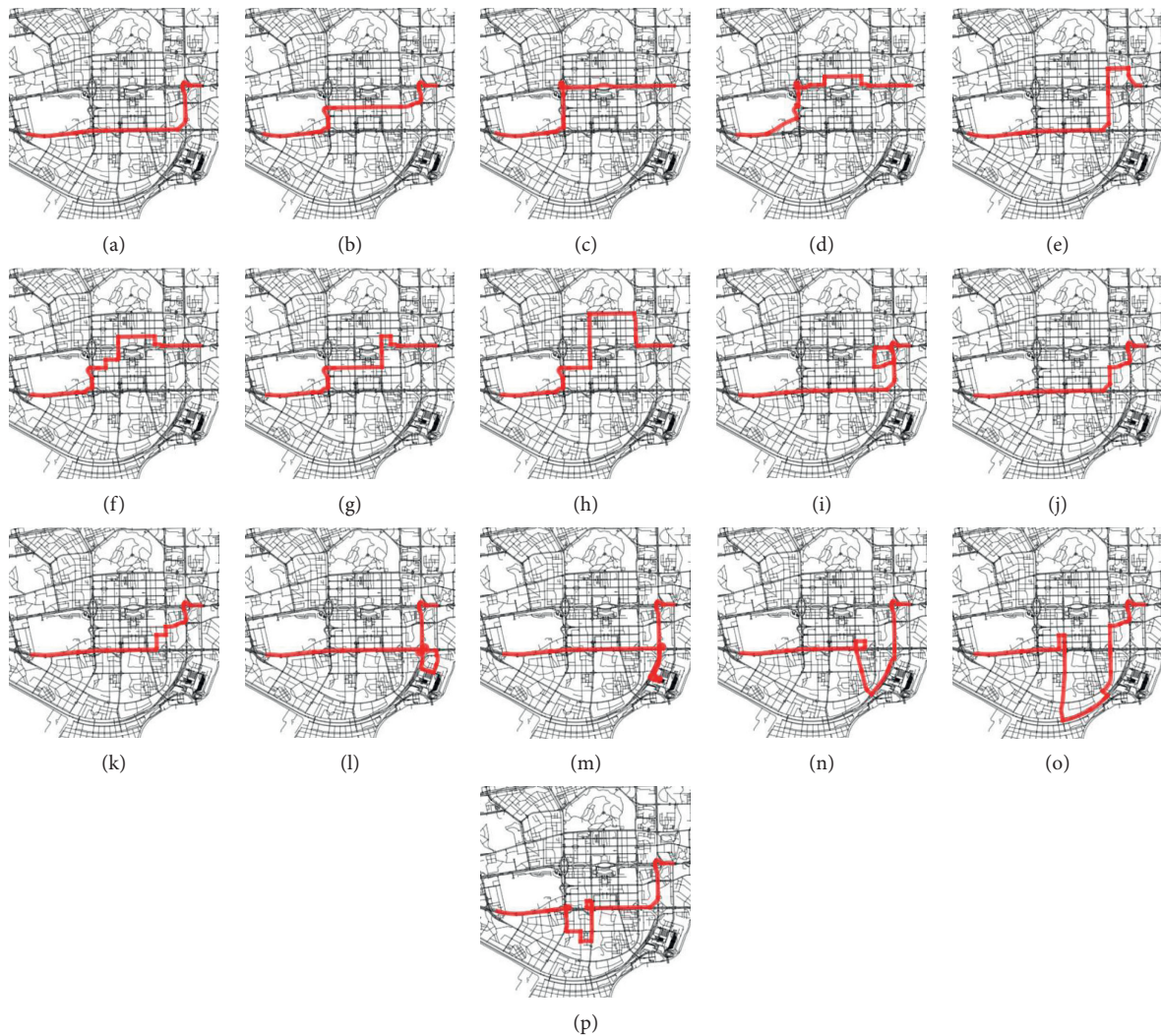


FIGURE 10: (a–p) Generated candidate paths between the example sensor pair:  $(F_{2010002}, F_{10100403})$ .

TABLE 3: Estimation results of Fixed-MM.

Parameters	Fixed-MM (without temporal constraint)	Fixed-MM (with temporal constraint)
Number of signal lights $\beta^{\text{NSL}}$	-0.185	-0.188
t-test	-21.442	-11.579
Free travel time $\beta^{\text{FTT}}$	-0.273	-0.101
t-test	-9.574	-1.839
Proportion of expressway $\beta^{\text{PE}}$	0.074	4.140
t-test	49.532	10.925
Temporal constraint $\alpha$	—	67.395
t-test	—	12.417
Log-likelihood	-17851.898	-1049.901
Adjusted likelihood ratio	0.511	0.971

negative sign and the “proportion of expressway” has a positive sign in each case. The negative sign and t-statistic of  $\beta^{\text{NSL}}$  and  $\beta^{\text{FTT}}$  suggest that the freer travel time and signal lights the path has, the less likely it is to be matched. The positive sign and t-statistic of  $\beta^{\text{PE}}$  imply that a path with a higher proportion of expressways will be more attractive to travellers.

Finding 2: the temporal constraint parameter,  $\alpha$ , is very large, which means that the correct term has a significant effect on the matching results.

Finding 3: when the temporal constraint term,  $\alpha \ln q(t_{j,n} - t_{i,j})$ , was considered, the Fixed-MM model with temporal constraints had a much lower log-likelihood. Thus, we can infer that it has a better model fit and is closer to the true model.

**4.3. Model Evaluation.** In this section, we describe our algorithm on the testing dataset. Two classical Probe-MM algorithms are used as benchmarks, details of which are introduced as follows:

HMM algorithm [7]: given that the positions of the fixed sensors are located without noise, the measurement probability is set to 1 and only the transmission probability is considered

ST-Matching algorithm [12]: similar to the HMM-based algorithm, the observation probability in the spatial analysis of this method was set to 1 because of the accurate positions of the fixed sensors

In this study, two indexes for expressing matching accuracy were used. One is the accuracy length ratio of paths (ALRP) index, defined as follows:

$$\text{ALRP} = \frac{\sum_{l_x \in P_{i,j}} \delta_x l_x \cdot \text{len}}{P_{\text{true}} \cdot \text{len}} \times 100, \quad (10)$$

where  $l_x \cdot \text{len}$  is the length of link  $l_x$  in the matched path,  $P_{i,j}$  is the set of paths between fixed sensors  $i$  and  $j$ ,  $P_{\text{true}} \cdot \text{len}$  is the total length of the true path, and  $\delta_x = 1$  if  $l_x$  is also in the true path, and otherwise is 0.

The other index is the accuracy number ratio of paths (ANRP) index, which is defined as

$$\text{ANRP} = \frac{\sum_{n=1}^{N_x} \delta_x}{N_x} \times 100, \quad (11)$$

where  $N_x$  is the total number of links in the true path  $P_{\text{true}}$ .

Figures 11(a) and 11(b) show the ALRR and ANRR of the proposed Fixed-MM algorithm and two classical Probe-MM algorithms with regard to the spatial gap between fixed sensors. It can be seen clearly that our Fixed-MM outperforms both HMM and ST-Matching significantly. Meanwhile, the performance of two Probe-MM algorithms degrades sharply when the spatial gap decreases while Fixed-MM is more robust to the change of spatial gap. The proposed Fixed-MM can correctly identify 68.38% of the links, even when the spatial gap between the sensor pair increases to 5 km.

Because the candidate generation process and model training process can be conducted separately and in parallel, a comparison of the latency of the matching process may be more meaningful for online applications. In this study, the computation time for one point (ACTOP) was used to measure the computational latency of the map-matching algorithm.

As shown in Figure 12, the ACTOP of the two Probe-MM approaches increases dramatically as the spatial gap between the fixed sensors increases. Conversely, the ACTOP of Fixed-MM increases slowly. The main reason, therefore, can be deduced from two factors. The HMM and ST-MM algorithms assume that the position of the sensor is stochastic and noisy, and the candidate set must be regenerated for every sensor pair. It involves several shortest path computations between states at the previous and current time steps, which consumes most of the computation time. Conversely, the candidate set generation of the proposed method can be run in parallel and does not increase the computation time because the projection of the fixed sensor data is known and fixed. In

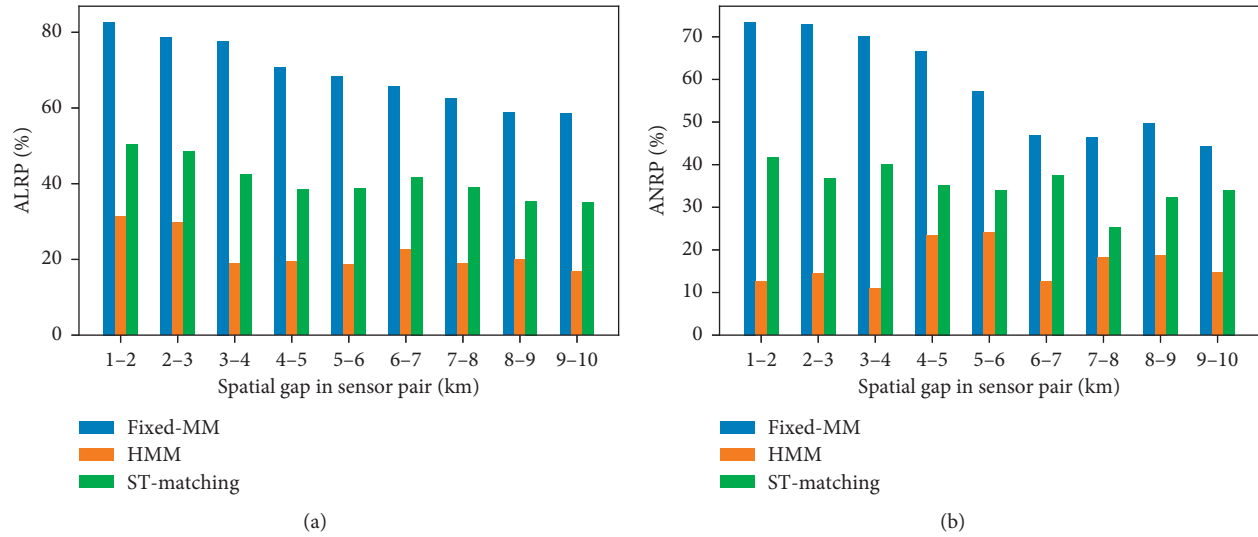


FIGURE 11: (a) ALRP and (b) ANRP with respect to the spatial gap.

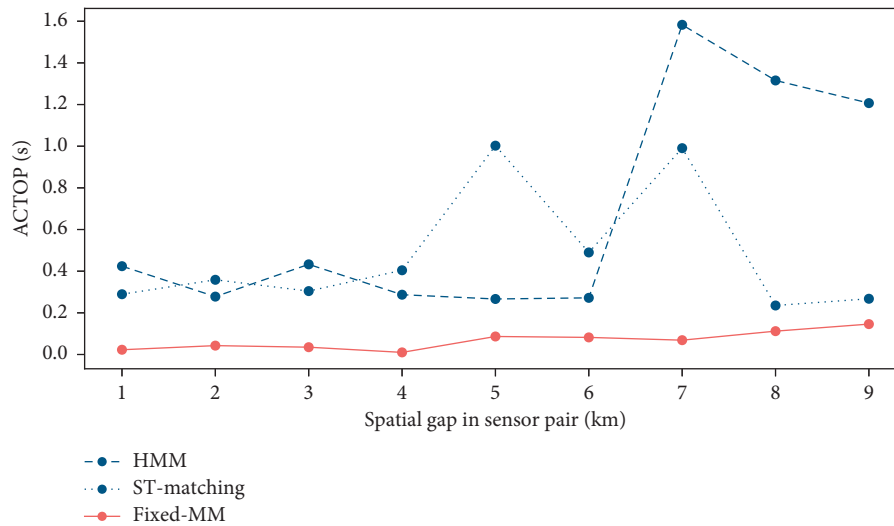


FIGURE 12: ACTOP with respect to the spatial gap (all algorithms are run on a Mac with 2.2 GHz Intel CPU and 16 GB RAM).

fact, the average ACTOP of Fixed-MM is only 0.067 s, and we argue that Fixed-MM can be performed online for many real-time ITS applications.

## 5. Conclusions

In this paper, we proposed a new map-matching algorithm called Fixed-MM to match vehicle trajectories recorded by fixed sensors onto a digital map. First, utility theory was employed to model the traveller's behaviour preference. Second, Fixed-MM was modified by adding a travel-time constraint term based on the observed and expected travel times. Moreover, a candidate path generation algorithm was designed for Fixed-MM.

Fixed sensor data and probe sensor data were collected as the experimental dataset. Both the Fixed-MM without a temporal constraint and Fixed-MM with a temporal

constraint were estimated. The statistical results of the estimated parameters prove that the path attributes correlate significantly to the true path, and the Fixed-MM with the temporal constraint having a better model fit. The Fixed-MM algorithm was also compared with two classical Probe-MM algorithms in terms of matching accuracy and computational efficiency. Fixed-MM outperforms the two Probe-MM algorithms in both number (ANRR) and length (ALRR) accuracy indexes. Meanwhile, the Fixed-MM is more robust to changes in the spatial gap between fixed sensors. Fixed-MM also has a huge improvement in computing efficiency and exhibits potential for online applications. The experimental results demonstrate that the proposed Fixed-MM algorithm is both effective and efficient.

More research is needed in the future to determine the potential application value of Fixed-MM. Although the travel time and speed can also be estimated by the Probe-

MM algorithm with probe sensor data, the Fixed-MM provides a more diverse and credible estimation of travel time and speed. This is because the fixed sensor data covers almost all types of vehicles using the road network, while the probe sensor data can only be collected from one type of vehicle, for example, taxicabs. Meanwhile, with the application of Fixed-MM, more traffic information can be mined from the fixed sensor data. If all the observed trips of every fixed sensor can be matched to the road network, the traffic volumes of each path or link can be estimated, which is the key input value for traffic planning and management. Thus, our next research focus is to utilize the Fixed-MM to mine more reliable and accurate traffic state information from fixed sensor data. Moreover, since the fare gate in the AFC system is fixed, applying the proposed map-matching algorithm to learn the route choice behavior of subway passengers [20, 21] also presents great practical application values and is worthy of further study.

## Appendix

### A. Estimated Results of Temporal Constraint

GPS trajectories of samples are presented in Figures 4, 5, and 6.

### B. Generated Candidate Path Set

GPS trajectories between the example sensor pair and generated candidate paths between the example sensor pair are presented in Figures 9 and 10, respectively.

### Data Availability

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

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The authors are grateful to the Shenzhen Urban Transport Center for the research data and the Big Data Center of Southeast University for facilitating the numerical calculations in this paper. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1600200, the National Natural Science Foundation of China under Grant 71971056, 16th Regular Meeting Communication Program of China-Bulgaria Science and Technology Cooperation Committee under Grant no. 16-4, the Science and Technology Project of Jiangsu Province, China, under Grant BZ2020016, and Jiangsu Province Graduate Research and Practice Innovation Plan under Grant 101010573.

### References

- [1] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: state-of-the art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 5, pp. 312–328, 2007.
- [2] K. Satyanarayana, A. D. Sarma, J. Sravan, M. Malini, and G. Venkateswarlu, "GPS and GPRS based telemonitoring system for emergency patient transportation," *Journal of Medical Engineering*, vol. 2013, Article ID 363508, 2013.
- [3] S. An, L. Wang, H. Yang, and J. Wang, "Discovering public transit riders' travel pattern from GPS data: a case study in harbin," *Journal of Sensors*, vol. 2017, Article ID 5290795, 2017.
- [4] M. N. Borhan, D. Syamsunur, N. Mohd Akhir, M. R. Mat Yazid, A. Ismail, and R. A. Rahmat, "Predicting the use of public transportation: a case study from putrajaya, Malaysia," *The Scientific World Journal*, vol. 2014, Article ID 784145, 2014.
- [5] T. Miwa, D. Kiuchi, T. Yamamoto, and T. Morikawa, "Development of map matching algorithm for low frequency probe data," *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 132–145, 2012.
- [6] M. Hashemi and H. A. Karimi, "A critical review of real-time map-matching algorithms: current issues and future directions," *Computers, Environment and Urban Systems*, vol. 48, pp. 153–165, 2014.
- [7] P. Newson and J. Krumm, "Hidden Markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, WA, USA, November 2009.
- [8] R. Mohamed, H. Aly, and M. Youssef, "Accurate and efficient map matching for challenging environments," in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, IEEE, Dallas, TX, USA, November 2014.
- [9] H. Koller, P. Widhalm, M. Dragaschnig, and A. Graser, "Fast hidden Markov model map-matching for sparse and noisy trajectories," in *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 2557–2561, IEEE, Gran Canaria, Spain, September 2015.
- [10] J. Han, X. Fu, L. Liu, and D. Jiang, "Online map matching by indexing approximate road segments," in *Proceedings: 2011 IEEE 2nd International Conference on Software Engineering and Service Science ICSESS 2011*, IEEE, Beijing, China, July 2011.
- [11] G. R. Jagadeesh and T. Srikanthan, "Online map-matching of noisy and sparse location data with hidden Markov and route choice models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2423–2434, 2017.
- [12] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, WA, USA, November 2009.
- [13] Y.-L. Hsueh and H.-C. Chen, "Map matching for low-sampling-rate GPS trajectories by exploring real-time moving directions," *Information Sciences*, vol. 434, pp. 55–69, 2018.
- [14] X. Liu, K. Liu, M. Li, and F. Lu, "A ST-CRF map-matching method for low-frequency floating car data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1241–1254, 2017.
- [15] S. M. Saab and Z. M. Kassas, "Power matching approach for GPS coverage extension," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 156–166, 2006.

- [16] M. Ben-Akiva and M. Bierlaire, "Discrete choice methods and their applications to short term travel decisions," *Handbook of Transportation Science*, vol. 26, 2000.
- [17] E. Frejinger, M. Bierlaire, and M. Ben-Akiva, "Sampling of alternatives for route choice modeling," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 984–994, 2009.
- [18] S. Coast, "Open street map data extracts, open street map community," 2019, <http://download.geofabrik.de/>.
- [19] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, 2009.
- [20] C. Yu, H. Li, X. Xu, and J. Liu, "Data-driven approach for solving the route choice problem with traveling backward behavior in congested metro systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, pp. 1–27, 2020.
- [21] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Systems with Applications*, vol. 95, pp. 324–332, 2018.