Travel Time Estimation from Sparse Floating Car Data with Consistent Path Inference: A Fixed Point Approach

Mahmood Rahmani\textsuperscript{a}, Haris N. Koutsopoulos\textsuperscript{a,b}, Erik Jenelius\textsuperscript{a,*}

\textsuperscript{a}Department of Transport Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
\textsuperscript{b}Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115

Abstract

An important application of (sparse) floating car data (FCD) is the estimation of network link travel times, which requires map-matching and path inference. Path inference, in general, requires some a priori assumptions about link travel times to infer paths that are reasonable and temporally consistent with observations. Path inference and travel time estimation is thus a joint problem. However, this aspect has not been addressed adequately in the literature, despite the fact that it can bias the travel time estimation and subsequent identification of shortest paths. This paper proposes a fixed point formulation of the simultaneous path inference and travel time estimation problem. The methodology is applied in a case study to estimate travel times from taxi FCD in Stockholm, Sweden. In this case study, existing methods for path inference and travel time estimation, without any particular assumptions about path choice models or travel time distributions, are used. The results show that standard fixed point iterations converge quickly to a solution where input and output travel times are consistent. The solution is robust under different initial travel times assumptions and data sizes. The results highlight the importance of the simultaneous solution of the path inference and travel time estimation problem, in particular for accurate path finding and route optimization.

Keywords: floating car data, fixed point problem, travel time estimation, path inference.

*Corresponding author

Email addresses: mahmoodr@kth.se (Mahmood Rahmani), hnk@kth.se (Haris N. Koutsopoulos), jenelius@kth.se (Erik Jenelius)
1. Introduction

Floating car data (FCD) are becoming increasingly available as private automobile owners and fleet operators agree to share vehicle status data with services providing real-time traffic information, route planning, etc. FCD have the potential to provide high quality travel time information at a relatively low cost for urban traffic monitoring and management, as well as planning, policy, and provision of mobility services.

Taxi dispatching and delivery companies, who collect and monitor vehicle positions for internal purposes, are common sources of FCD in cities. Typically, such systems are designed to minimize the data transfer demand by reporting the positions of vehicles once every one or two minutes, and dropping data that are non-essential for the dispatching function such as instantaneous speed and direction (although this information is useful for traffic applications). In many cases, the only data being transferred are latitude, longitude, timestamp and ID of the vehicle.

There are a number of challenges when using sparse FCD for link travel time estimation. First, the paths taken by vehicles between probes are often unknown due to large gaps (in time and space) between probes and high density of urban road network links. Thus, there may exist more than one feasible paths connecting a sequence of probes. Second, even for a given path between two probes the traversal times on the links in the path are not observed. Hence, the allocation of the time between probes to the network links is an important task.

The majority of methods for link travel time estimation from sparse FCD address the problem through two steps: (a) map-matching and path inference, where the most likely network paths traversed by probe vehicles are inferred based on reported GPS positions, and (b) travel time allocation and estimation, where travel time observations between probes are allocated to links and aggregated across trips and vehicles for the estimation of link travel time profiles.

Various map-matching methods have been developed for navigation applications, but a new class of path inference methods has emerged with the increasing availability of sparse FCD (see for example Lou et al., 2009; Zheng and Quddus, 2011; Miwa et al., 2012; He et al., 2013; Rahmani and Koutsopoulos, 2013). A typical method first considers a number of consecutive probes, finds a set of candidate links for each probe, projects the probe to each candidate link, and calculates the likelihood of the probe belonging to each candidate link. Then, pairs of consecutive projections are connected through shortest paths. Finally, the path that maximizes an overall measure
of spatial and temporal consistency is selected as the most likely traversed path.

Following path inference, observed travel times from the probe timestamps are used to estimate link travel time distributions on traversed paths. Some studies consider allocation and estimation as sequential problems (Hellinga et al., 2008; Zheng and van Zuylen, 2013; Sanaullah et al., 2013; Rahmani et al., 2015). In these methods, link allocation is commonly performed proportionally the traversed distance on each link or some a priori link travel times. Other studies perform allocation and estimation simultaneously or iteratively (Hoffeitner et al., 2012; Westgate et al., 2013; Ramezani and Geroliminis, 2012; Jenelius and Koutsopoulos, 2013).

Many of the existing methods, during the path inference phase and for the calculation of the most likely path, make assumptions about a priori link travel times which, often, are not justified (for example, they assume free flow conditions). However, while travel time is an important factor in drivers' path choices (Ben-Akiva et al., 1984) good estimates of link travel times may not be available during the path inference step - indeed, they are the sought outcome of the estimation process. Therefore, it is common that an approximation of travel times, based on link lengths, free-flow conditions and/or speed limits, is used to infer paths (see for example Hunter et al., 2013; Rahmani and Koutsopoulos, 2013).

Path inference based on approximate link travel times implies that paths are generally not consistent with the subsequently estimated travel times. Given the assumption that travelers choose paths to minimize some function of travel time, it can be concluded that the path inference, and hence travel time estimation, are biased.

As Figure 1 illustrates, there are cases where a path can be inferred with certainty, e.g., the probe sequence \{a_1, a_2, a_3\}. The path inference in this case is not affected by the choice of a priori link travel times because there is only one possible (acyclic) path going through all three points. On the other hand, the sequence \{b_1, b_2\} can be connected by either path A or path B. Depending on assumptions about initial link travel times, one of the two paths may be more likely than the other. Therefore, using initial link travel times that are distant from actual ones may induce error in the path inference. Furthermore, selection of either of the two paths directly impacts the estimation of travel time for all links (in both paths).

Thus, path inference and travel time estimation is inherently a joint problem. However, only a few previous studies have approached it as such. Westgate et al. (2013) simultaneously estimate path probabilities and link travel time distributions using a Bayesian approach. The model assumes
that link travel times are independent and log-normally distributed, and that path probabilities follow a logit model based on mean path travel times. Parameters of the model are estimated using Markov chain Monte Carlo methods. Using both simulated and real data, the authors compare the iterative method with two non-iterative methods and conclude that it outperforms both of them. Similarly, Zhan et al. (2013) estimate mean link travel times and path probabilities from origin-destination travel time observations using a non-linear least squares approach. Path choices follow a logit model based on mean path travel time.

The aforementioned studies make strong parametric assumptions about travel time distributions and path choice models in order to formulate and solve the estimation problem jointly. The Monte Carlo simulation-based estimation method of Westgate et al. (2013) is computationally demanding and the assumption of independent, log-normal link travel time distributions is restrictive. The approach of Zhan et al. (2013) only estimates mean link travel times and implicitly assumes that route travel times have equal variance regardless of the links traversed and route length.

The aim of this paper is to address the joint path inference and travel time estimation problem with a more general approach. The path inference and estimation process is viewed as a mapping from some initial link travel times to estimated link travel times, i.e., a mapping from the space of link travel times onto itself. In this framework, the paper analyzes the sensitivity of estimated travel times to the initial travel times. In particular, the problem of achieving consistent paths and travel times is viewed as a fixed point problem, \( x = f(x) \). Thus, the paper investigates the feasibility of finding a set of link travel times that, if used as input to the process, will produce the same link travel times as output. In such a fixed point, estimated link travel times are consistent with inferred paths. As such, they are arguably less biased estimates of true link travel times compared to estimates based on approximated initial travel times.

The analysis focuses on the performance of the fixed point iteration.
using existing methods for path inference and travel time estimation. Few assumptions are made about the internal structure of the process. Hence the proposed fixed point formulation can be used with any other method in the literature. The fixed point methodology is applied in a case study using FCD from 1500 taxis in Stockholm, Sweden. Existing, efficient methods for path inference (Rahmani and Koutsopoulos, 2013) and travel time estimation (Rahmani et al., 2015) are used.

The convergence of a successive approximation scheme as well as the method of successive averages (MSA) is investigated. The impact on estimated link and route travel times is analyzed. Further, the study assesses the impact of the initial travel times and the size of the FCD set on the outcome, and the value of using historical travel times as initial input.

The remainder of the paper is organized as follows. Section 2 introduces the fixed point iteration methodology. Section 3 describes the case study for Stockholm with results presented in Section 4. Section 5 concludes the paper.

2. Methodology

2.1. Background

A fixed point of a function is an element of the function’s domain that is mapped to itself by the function. In other words, \( x^* \) is a fixed point of the function \( f(x) \) if and only if \( f(x^*) = x^* \). Brouwer’s theorem states that any continuous function from a compact convex set to itself has a fixed point (Brouwer, 1912); generalizations of the theorem to discontinuous functions also exist (e.g., Herings et al., 2008).

Stronger assumptions about the function and its domain are required in order to ensure that the fixed point is unique. For example, Banach’s fixed point theorem (Banach, 1922) states that there is a unique fixed point if \( f \) is a contraction mapping from a complete metric space (e.g., Euclidean space with the Euclidean distance metric) to itself; loosely speaking, \( f \) is a contraction mapping if the distance between \( f(x) \) and \( f(y) \) is strictly smaller than the distance between \( x \) and \( y \) for any two points \( x \) and \( y \) in the domain of \( f \).

In general, a fixed point \( x = f(x) \) can be found by using a successive approximation method: starting with an arbitrary vector \( x^1 \) and calculating \( \tilde{x}^1 = f(x^1) \). At each iteration, the new input \( x^{i+1} \) is calculated as a function of the previous output \( \tilde{x}^i \) and previous input \( x^i \); \( x^{i+1} = g(x^i, \tilde{x}^i) = g(x^i, f(x^i)) \), where \( g() \) is a smoothing function. Methods such as the method of successive averages (MSA) can be used to construct a sequence \( \tilde{x}^1, \tilde{x}^2, \cdots \)
in such a way that $\hat{x}^n$ will tend to $x^*$ in probability (Robbins and Monro, 1951).

Fixed point approaches have been applied to solve a wide range of transport problems. A common application is when the interaction between transport supply and demand is captured with equilibrium-like models such as deterministic or stochastic user equilibrium (e.g., Daganzo, 1983). Another application is to estimate origin-destination (OD) matrices (Cascetta and Postorino, 2001). The OD matrix solution, once assigned to the network, reproduces flows and costs consistent with the values used to compute the assignment matrix. The study investigates three fixed point algorithms: functional iteration, MSA, and MSA with decreasing reinitialization. It concludes that: (a) the three algorithms converge to the same solution, though with different speeds, and (b) the fixed point solutions outperform the solution without iterations.

Fixed point approaches have also been used for anticipatory route guidance, where travellers’ response to the route advice, which affects travel times and hence the attractiveness of different routes, is taken into account when guidance is provided (Kaufman et al., 1991; Bottom et al., 1999).

Bierlaire and Crittin (2006) propose a fixed point approach for systems of nonlinear equations and apply the algorithm on an OD matrix estimation problem with more than 120,000 variables. They show that the approach exhibits robust behavior in the presence of noise in the equations, which makes it particularly well suited for transport problems.

Ben-Akiva et al. (2012) use a fixed point approach to simultaneously calibrate the demand and supply parameters of DynamiT-P (a simulation-based DTA system for planning applications). The calibration problem is formulated as a constrained minimization problem. The optimization problem is solved using the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm. In each SPSA iteration, the output link travel times are used as input link travel times to the demand simulator for the next iteration. The difference between the input and output link travel times measures the convergence of the fixed point problem.

2.2. Problem formulation

The notation used in the paper is summarized below.
Let us consider a road network containing \( N \) links, indexed \( i \in \{1, \ldots, N\} \). Further, let \( D = \{d_j\}, j \in \{1, \ldots, M\} \) denote a set of FCD probes. Each data point \( d_j = (x_j, y_j, s_j, q_j) \) contains the \( x \)- and \( y \)-coordinates \( (x_j, y_j) \in \mathbb{R}^2 \), timestamp \( s_j \in \mathbb{R}_+ \), and vehicle ID \( q_j \in \mathbb{N} \) of an FCD probe. The set of GPS observations \( D \) typically is related to a specific time interval, for example 8:00-8:15. The problem considered in this paper is to estimate the link travel times \( \tau = (\tau_i) \in \mathbb{R}^N_+ \) for \( i \in \{1, \ldots, N\} \) of every network link based on the data in \( D \).

It is assumed that an estimation procedure with the following features is available:

1. The network paths (i.e., sequences of links) traversed by the vehicles are inferred, uniquely or probabilistically, based on some driver decision rule or choice model (e.g., shortest path, discrete choice model), which is a function of link travel times among other attributes.

2. The travel times between consecutive probes are allocated to the traversed links (of the inferred path), and used to estimate the mean travel time of each link (and possibly its distribution).

Given that the procedure requires initial link travel times as input for path inference and possibly the allocation of travel times to links, it can be regarded as a mapping from (initial) travel times to (estimated) travel times. Thus, considering travel times \( t \in \mathbb{R}_+^N \) as a variable, the estimation procedure can be described as a function \( f_D : \mathbb{R}_+^N \to \mathbb{R}_+^N \), i.e., \( \tau = f_D(t) \).

In general, two different FCD sets \( D, D' \) correspond to two different functions \( f_D, f_{D'} \), i.e., \( f_D(t) \neq f_{D'}(t) \) for some \( t \). An interesting question is how sensitive the output travel times are to the size of the dataset. Also, two different input travel time vectors \( t, t' \) generally produce different output travel times for the same FCD set, i.e., \( f_D(t) \neq f_D(t') \). This implies that estimated travel times are influenced by the initial link travel times \( t \).
which also raises the question how sensitive estimated travel times are to the prior travel times.

Furthermore, for the reason discussed in previous sections, the output travel time vector is in general different from the input, i.e., \( t \neq f_D(t) \) for some \( t \). Given the assumption that travelers choose paths to minimize some function of travel time, this implies that vehicle paths inferred from the initial travel times are generally not consistent with the subsequently estimated travel times. That is, path inference based on the estimated travel times may yield different results than based on the initial travel times. Hence, travel time estimation is also biased. To minimize bias, it is important to find paths and link travel times that are mutually consistent, i.e., to find a fixed point \( \tau^* \) such that \( \tau^* = f_D(\tau^*) \).

In general, the existence of a fixed point of the path inference and travel time estimation function is difficult to establish. The main complication stems from the fact that the most existing path inference methods identify the single most likely path. This discrete output means that small changes in input travel times can lead to discontinuous jumps in the output, as illustrated in Section 1.

In many other transport problems formulated as fixed point problems, the existence of fixed point solutions has not been proven, yet fixed point algorithms have been demonstrated to work well in practice and to converge to solutions within the required precision (e.g., Bottom et al., 1999). A further question to investigate is thus how fixed point iteration performs in practice, i.e., whether iterative schemes can be devised such that the process converges to a fixed point rather than oscillates or diverges.

2.3. Fixed point iteration

In the fixed point approach, initial link travel times, \( t^1 \), are used in the first iteration to process a set of FCD, \( D \), and estimate link travel times, \( \tau^1 = f_D(t^1) \) (note that what existing methods do is equivalent to this first iteration). At each subsequent iteration \( k + 1 \), the previous solution \( (\tau^k) \) and input travel times \( (t^k) \) are used to estimate the new input travel times: \( t^{k+1} = g(\tau^i, t^k) \). The iterative process continues until a termination criterion based on a measure of the difference between the input and output travel times, \( d(\tau^k, t^k) \), is met. Note that the same set \( D \) of FCD data is used throughout the process.

Algorithm 1 describes the iterative path inference-travel time estimation, and Figure 2 illustrate the steps of the algorithm.
Algorithm 1: Iterative path inference-travel time estimation.

\[
\text{input} : D ; \quad \text{// a floating car dataset} \\
\text{input} : t^1 ; \quad \text{// initial link travel time profiles} \\
\text{output}: \tau^* ; \quad \text{// estimated link travel time profiles} \\
k \leftarrow 1 ; \quad \text{// initialization} \\
\text{repeat} \\
\quad \tau^k = f_D(t^k) ; \quad \text{// path inference and travel time} \\
\quad \text{estimation} \\
\quad t^{k+1} = g(\tau^k, t^k) ; \quad \text{// smoothing} \\
\quad k \leftarrow k + 1 ; \\
\text{until } d(\tau^k, t^k) < \epsilon ; \quad \text{// termination criterion} \\
\quad \tau^* \leftarrow \tau^k ; \\
\text{halt} ;
\]

Figure 2: The main iteration of a fixed point algorithm of path inference and travel time estimation.
2.3.1. Initial conditions

Initial link travel times, $t^1$, can be obtained in different ways. One possibility is to use estimated travel times from a different dataset or a previous time interval. Another approach is to use travel times from planning or traffic simulation models. If estimated travel times are not available, free-flow travel times based on speed limits or some assumed speed may be used.

2.3.2. Updating schema

The updating schema determines how previous outputs contribute to the input of the next iteration. The choice is important, both for convergence as well as efficiency. The simplest scheme is the identity function, i.e., the output of one iteration is input to the next, $t^{k+1} = g_{id}(\tau^k, t^k) = \tau^k$. The result is a sequence of $\tau^{k+1} = f_D(\tau^k)$ which is called the sequence of successive approximations with the initial value $t^1$. It is also known as the Picard iteration starting at $t^1$ (Berinde, 2007).

A robust and commonly used scheme in transport applications is the Mann iteration, better known as MSA, where intermediate solutions contribute to the input,

$$t^{i+1} = g_{msa}(\tau^k, t^k) = t^k + \frac{1}{k} (\tau^k - t^k). \quad (1)$$

2.3.3. Termination criterion

Let $\{\tau^k\}_{k=1}^{\infty}$ be a given fixed point iteration sequence that converges to $\tau^*$. Since $\tau^k \to \tau^*$ as $k \to \infty$, for any $\epsilon > 0$, there exists a positive integer $\lambda$ such that

$$d(\tau^k, t^k) < \epsilon \text{ for } k \geq \lambda \quad (2)$$

where $d$ is a certain distance metric. If $\lambda$ can be practically determined (depending on $\epsilon$, the initial guess $t^1$ and the function $f_D$) then (2) serves as a stopping criterion for the iterative process (Berinde, 2007). In particular, the process is terminated after iteration $k^*$ such that the difference in input and output is less than $\epsilon$,

$$k^* = \arg \min_k \left\{ d(\tau^k, t^k) \right\} < \epsilon, \quad (3)$$

The distance measure $d()$ can be defined as, for example, the root mean
squared error (RMSE), or the mean absolute error (MAE):

$$\text{MAE}(\tau^k, t^k) = \frac{1}{N} \sum_{i=1}^{N} |\tau^k_i - t^k_i|.$$ (4)

In some cases it is of interest to determine convergence not only in the mean but across most or all links. For example, the $99^{th}$ percentile of the absolute difference (or even the maximum absolute difference) between input and output travel times across all links may be considered,

$$p_{99}(\tau^k, t^k) = 99^{th} \text{ percentile} \left\{ |\tau^k_i - t^k_i| \right\} \leq \eta,$$ (5)

where $\eta$ is a predefined threshold.

3. Application

The proposed fixed point approach to travel time estimation is applied to a real road network and FCD from about 1500 taxis from Stockholm, Sweden. The results are used to assess the benefits of fixed point iterations to achieve consistent travel times and investigate the impact of initial conditions. Convergence of the fixed point algorithm is investigated and the results after convergence are compared to a single sequence of path inference and travel time estimation.

3.1. Data and algorithms

The study area covers about 8 km$^2$ of the Stockholm inner city. The road network in the area consists of 3027 links (Figure 3a) with mean and median length of 61 m and 50 m respectively. FCD are collected by a GPS-equipped taxi fleet management system covering about 1500 taxis. Each taxi broadcasts its location (latitude, longitude), timestamp, id, and status (free/hired) once every two minutes on average.

Only probes from hired taxis are used. Excluding probes from free taxis reduces the influence from vehicles standing at taxi stations, searching for passengers, or changing passengers. Based on the results of map-matching and path inference, travel times are allocated to each link. The average travel times for links with at least seven observations are reported.

Three FCD sets (denoted $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$) are used in the study (Table 1). $\mathcal{D}_1$ and $\mathcal{D}_2$ are collected during different time periods, Oct-Nov 2012 and Dec 2012-Feb 2013, respectively. $\mathcal{D}_3$ is the union of $\mathcal{D}_1$ and $\mathcal{D}_2$. The data
Figure 3: The road network of the study area (a); bold lines are main arterial streets and thin lines side streets. The baseline set of FCD (dataset \( \mathcal{D}_1 \) in Table 1) (b).

cover Mondays through Thursdays from 08:00 to 08:15 am. The locations of the probes in dataset \( \mathcal{D}_1 \) are shown in Figure 3b, clearly highlighting the major arterial streets (shown with bold lines in Figure 3a) but also a set of commonly used minor streets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>From date</th>
<th>To date</th>
<th>Day of week</th>
<th>Time of day</th>
<th>Days</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{D}_1 )</td>
<td>Oct 2012</td>
<td>Nov 2012</td>
<td>Mon-Thu</td>
<td>08:00-08:15</td>
<td>44</td>
<td>12,000</td>
</tr>
<tr>
<td>( \mathcal{D}_2 )</td>
<td>Dec 2012</td>
<td>Feb 2013</td>
<td>Mon-Thu</td>
<td>08:00-08:15</td>
<td>65</td>
<td>12,000</td>
</tr>
<tr>
<td>( \mathcal{D}_3 )</td>
<td>( \mathcal{D}_1 \cup \mathcal{D}_2 )</td>
<td></td>
<td></td>
<td></td>
<td>109</td>
<td>24,000</td>
</tr>
</tbody>
</table>

The sequential path inference and travel time estimation methods used in the study serve as representatives of practically implemented methods that are applicable to large-scale problems. The method introduced in Rahmani and Koutsopoulos (2013) is used for map-matching and path inference. The approach is designed for sparse FCD where the only information available is latitude, longitude and timestamp. It is designed to be robust with respect to the frequency of probes, and appropriate for both off-line and real-time applications. Following path inference, link travel times are estimated using the computationally efficient, non-parametric method introduced in Rahmani et al. (2015). The method makes no a priori assumptions regarding the form of the link travel time distributions, and provides estimates not only of the mean but also any statistics of interest of the link travel time distribution.
3.2. Experimental design

Table 2 summarizes the experimental design of the case study. First, a base configuration is examined, in which the path inference and link travel time estimation process is initialized with free-flow travel times (calculated from link lengths and speed limits, Experiment 1). The estimation is based on $D_1$ using Picard iteration ($g_{id}$) for the solution of the fixed point problem.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Dataset</th>
<th>Initial travel times</th>
<th>Update rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$D_1$</td>
<td>Free-flow</td>
<td>$g_{id}$</td>
</tr>
<tr>
<td>2</td>
<td>$D_1$</td>
<td>Free-flow</td>
<td>$g_{msa}$</td>
</tr>
<tr>
<td>3</td>
<td>$D_1$</td>
<td>Speed of 40 km/h</td>
<td>$g_{id}$</td>
</tr>
<tr>
<td>4</td>
<td>$D_2$</td>
<td>Estimates from Exp. 1</td>
<td>$g_{id}$</td>
</tr>
<tr>
<td>5</td>
<td>$D_2$</td>
<td>Free-flow</td>
<td>$g_{id}$</td>
</tr>
<tr>
<td>6</td>
<td>$D_3$</td>
<td>Free-flow</td>
<td>$g_{id}$</td>
</tr>
</tbody>
</table>

Experiment 2 examines the effect of using a smoothing update rule, specifically MSA, in the fixed point iteration (Section 2.3.2). The hypothesis is that MSA helps the iterations to converge to the fixed point and avoid oscillation or divergence.

In Experiment 3, initial link travel times are based on a speed of 40 km/h common across all links. These are plausible initial conditions in cases where speed limit information is not readily available. All other aspects are the same as in Experiment 1. The impact of the initial link travel times is assessed by comparing the results from Experiments 1 and 3.

Experiment 4 studies the effect of using historical travel times (i.e. travel times presumably closer to the “true” times compared to free flow travel times) as initial travel times. In particular, travel times estimated using $D_1$, i.e., the output of Experiment 1, are used as initial travel times (Experiment 4 uses dataset $D_2$). The hypothesis is that better initial estimates of travel times may lead to less bias in the estimation process and faster convergence to a fixed point. Experiment 5 uses $D_2$ and free-flow initial link travel times, and is used for comparison with Experiment 4.

Finally, Experiment 6 studies the impact of using a larger dataset than in the basic experiment. With more data, network coverage as well as the number of observations for each link are expected to increase. The variability of travel times may also increase due to the fact that the data come from days with different traffic and weather conditions. Thus, $D_3$, which amounts to about twice as many probe observations as in the base experiment, is
used. The main hypothesis to be investigated is that a larger dataset may produce more observations for each link and reduce the impact of fixed point iterations. The effect is assessed by comparing the results from Experiments 1 and 6.

4. Results

4.1. Convergence

The convergence of the fixed point iteration is assessed by the MAE and the 99th percentile of the absolute difference between input and output travel times criteria. The stopping criterion is set to \( p_{99} < 0.5 \) seconds.

Experiment 1 uses a Picard iteration \( (g_{id}) \). Link travel times converge quickly in terms of MAE (Figure 4, left) and \( p_{99} \) (Figure 4, right). In ten iterations, MAE decreases from the initial value of 3 sec to 0.004 sec; the stopping criterion of \( p_{99} < 0.5 \) sec is satisfied from iteration 5 onward. In iteration 10, 99% of all links have zero error and there are only two links out of 3027 (0.07%) with MAE larger than 1 sec. Thus, after 10 iterations travel times for the majority of the network links have converged.

Experiment 2 is a Mann iteration, i.e., MSA. Compared to iteration using \( g_{id} \), convergence with \( g_{msa} \) is slower (Figure 5). The stopping criterion of \( p_{99} < 0.5 \) sec is satisfied in iteration 14 (compared to iteration 5 without smoothing). In general, MSA and other updating schema may help the fixed point iteration to converge, at the price of a slower convergence. In this particular case, however, the results suggest that \( g_{id} \) is sufficient to yield fast and stable convergence of link travel times. In the following, the fixed point iterations are performed using \( g_{id} \).
Figure 5: Convergence of Experiment 2 (MSA) in MAE (left) and 99th percentile absolute error (right) of $\delta \tau^k = \tau^k - t^k$.

As Figure 6 illustrates, the link travel times estimated in Experiment 2 are similar to those in Experiment 1.

Figure 6: Comparison of link mean travel times estimated in Experiments 1 and 2 after convergence. MAE (in seconds) refers to the difference between estimated travel times of Experiments 1 and 2.

The main consequence of the fixed point iterations is that shortest paths between consecutive probes (path inference) are based on more consistent travel times as the iteration progresses. Hence, the paths inferred for a given FCD set can vary from one iteration to another due to changes in the estimated travel times of the underlying links.

Experiment 1 shows that the inferred path sets of two consecutive iterations become increasingly similar as the fixed point iteration converges. Figure 7 illustrates the number of changes of inferred paths as a function of
the number of iterations. For 83% of the observations (i.e., pairs of consecutive probes), the paths inferred in the first and the second iterations are identical. For the remaining 17% of the observations the paths inferred in the first iteration are not consistent with the subsequently estimated travel times. The impact on path inference is therefore potentially significant, compared to the more moderate impact on estimated link travel times. Between iterations 9 and 10, meanwhile, about 99.9% of the paths remain unchanged. Thus, the iterative process has essentially converged to a fixed point also in terms of inferred paths.

Figure 7: Number of changes in inferred paths between iterations (Experiment 1). The total number of paths is 10,611.

4.2. Impact on link travel times

Estimated link travel times after convergence, \( \tau^* \), differ from travel times estimated in the first iteration, \( \tau^1 \) (note that the first iteration is equivalent to how existing methods perform). The magnitude of the difference, and subsequently the importance of addressing the problem of initial travel time assumptions in practice, is expected to be a function of the complexity of the network, the degree of congestion, the quality of the initial assumptions on travel times, and the frequency of probes. Furthermore, it is expected that differences are bigger on links on local roads compared to links on major roads, arterials, and freeways (since there the path inference solution is quite robust and there are more probes).

Table 3 summarizes the differences, \( \Delta \tau^{10,1} \), between the link travel times without any iteration (current methods) and the travel times corresponding
to the fixed point solution $\tau^{10} - \tau^1$ (Experiment 1). The absolute change in mean travel time is less than 1 sec for 90% of the network links. Thus, the impact on travel time is small for the majority of links. The change in mean travel time is more than 5 sec for about 3% of the links, and more than 10 sec for about 1% of the links. Overall, the mean travel time is reduced for 15% of the links and increased for 33% of the links, while the remaining 55% of the links are not affected.

Table 3: Distribution of differences in estimated link travel times between iterations 10 and 1 for Experiment 1, where $\Delta \tau^{i,j} = \tau^i - \tau^j$.

<table>
<thead>
<tr>
<th>Range</th>
<th># of link</th>
<th>%</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\Delta \tau^{10,1}</td>
<td>\leq 1$</td>
<td>2739</td>
</tr>
<tr>
<td>$1 &lt;</td>
<td>\Delta \tau^{10,1}</td>
<td>\leq 5$</td>
<td>193</td>
</tr>
<tr>
<td>$5 &lt;</td>
<td>\Delta \tau^{10,1}</td>
<td>\leq 10$</td>
<td>60</td>
</tr>
<tr>
<td>$10 &lt;</td>
<td>\Delta \tau^{10,1}</td>
<td>$</td>
<td>35</td>
</tr>
<tr>
<td>$\tau^{10} = \tau^1$</td>
<td>1583</td>
<td>52.30</td>
<td>-</td>
</tr>
<tr>
<td>$\tau^{10} &lt; \tau^1$</td>
<td>443</td>
<td>14.63</td>
<td>-</td>
</tr>
<tr>
<td>$\tau^{10} &gt; \tau^1$</td>
<td>1001</td>
<td>33.07</td>
<td>-</td>
</tr>
</tbody>
</table>

In addition to the main impact on path inference, which in turn impacts travel time calculation, another important consequence is on the set of links that are “active”. Active links refer to links that have FCD observations assigned to them. The link travel time can thus be estimated from the assigned observations, and the estimated link travel time is used in subsequent iterations rather than the initial link travel time. The number of links with estimated travel times (“active” links) thus accumulates over the iterations.

In Experiment 1, the fraction of active links grows from 47% of the total links in the network in iteration 1 to 52% in iteration 10. Hence, the fixed point iteration increases the coverage of the travel time estimation compared to a single iteration of path inference and travel time estimation. Figure 8 shows links that are active after one iteration (in blue) and after convergence (blue and red). The majority of the new active links are side streets. In the first iteration path inference is based on free-flow travel times, and few observations are assigned to these links. Later in the process, however, the estimation captures the congestion on the main routes, and a sufficient number of observations is assigned to the side street links. In subsequent iterations the path inference may return to the main streets, but the estimated travel times on the side street links are maintained.
4.3. Impact on path travel times and shortest paths between OD pairs

An important application of link travel time estimation is trip planning (for example, through a travel planner service) and path finding. It is thus of interest to investigate the impact of the fixed point iteration when travel times or shortest paths (in travel time) between two arbitrary network locations are sought.

The estimated travel times on a number of significant arterial routes are analyzed (Table 4 and Figure 9). The number of observations on main routes is typically high. Furthermore, path inference is in general more stable on main routes compared to lesser routes with fewer observations. As a result, changes in mean route travel times are moderate.

Figure 10 shows an example in which the k-shortest path problem between two locations for $k = 2$ is solved based on (left) estimated link travel times after iteration 1, and (right) estimated link travel times after convergence. Based on only one iteration, the two shortest paths are A (mean travel time 8.30 min.) and B (mean travel time 8.42 min.). After convergence, travel times on paths A and B increase to 10.40 min. and 10.63 min., respectively, which represents an increase of over 25%. Even more importantly, they are no longer the shortest and second shortest paths. Instead, the two shortest paths are two different paths: C (8.87 min.) and D (8.92 min.)
Table 4: Changes in estimated route travel times for some main routes in the study area, Experiment 1, where $\Delta \tau^{ij} = \tau^i - \tau^j$. Routes are shown in Figure 9.

<table>
<thead>
<tr>
<th>ID</th>
<th>$\tau^1$ (sec)</th>
<th>$\Delta \tau^{10,11}$ (sec)</th>
<th>$\Delta \tau^{10,11}/\tau^1 \cdot 100%$</th>
<th>Length (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>223</td>
<td>17</td>
<td>8%</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>154</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>60</td>
<td>232</td>
<td>0</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>61</td>
<td>224</td>
<td>5</td>
<td>2%</td>
<td>1.2</td>
</tr>
<tr>
<td>64</td>
<td>173</td>
<td>-14</td>
<td>-8%</td>
<td>0.7</td>
</tr>
<tr>
<td>77</td>
<td>158</td>
<td>-1</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>95</td>
<td>297</td>
<td>18</td>
<td>6%</td>
<td>1.8</td>
</tr>
<tr>
<td>96</td>
<td>326</td>
<td>-1</td>
<td>0</td>
<td>1.9</td>
</tr>
<tr>
<td>116</td>
<td>338</td>
<td>-1</td>
<td>0</td>
<td>1.9</td>
</tr>
<tr>
<td>119</td>
<td>293</td>
<td>-3</td>
<td>-1%</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 9: Routes in the study area used for travel time estimation (Table 4).
min.). The fixed point iteration has significant impact on both, path travel times and shortest path finding.

![Figure 10: Example of the effect of fixed point iterations on shortest path finding. Left: iteration 1 (path A 8.30 min., path B 8.42 min.). Right: after convergence (path C 8.87 min., path D 8.92 min.)](image)

4.4. Sensitivity analysis

A number of questions regarding the sensitivity of the results presented in the previous section to various aspects of the fixed point methodology are of practical interest. Such aspects include the impact of the initial travel times assumed for path inference and the size of the available dataset.

4.4.1. Impact of initial travel times

In order to investigate the effect of initial link travel times on estimated travel times, Experiment 3 is carried out with the same dataset as the basic experiment (D1), but initial link travel times are set assuming a constant speed of 40 km/h as opposed to the speed limit (with speeds ranging from 30 to 50 km/h). The fixed point algorithm converges at a rate similar to Experiment 1 (cf. Figure 4). Comparison of Experiments 1 and 3 after convergence shows that 12% of the paths inferred during the path inference step differ. Figure 11 compares estimated link travel times in the two experiments. For the majority of links, estimated travel times are identical after convergence. The result suggests that the process as expected is not sensitive to the choice of initial link travel times (free-flow or 40 km/h).
4.4.2. Initialization with historical travel times

Once link travel times have been estimated, they may be used as initial travel times for processing future FCD sets with potentially less bias than if free-flow travel times are used. Experiment 4 utilizes estimated travel times from dataset $D_1$ as initial travel times to estimate travel times corresponding to dataset $D_2$. Figure 12 shows that the error between input and output is considerably lower even after one iteration compared to the case where free-flow travel times are used (Experiment 5). Convergence is also faster; the stopping criterion $p_{99} < 0.5$ sec is met after 3 iterations, and input and output are identical after iteration 9. The results suggest that using reasonable link travel times as the initial solution for path inference, has the potential to lead to quick convergence after a very small number of iterations. This is a useful result for both the fixed point approach and the conventional single iteration estimation approach.

4.4.3. Impact of dataset size

To evaluate the impact of dataset size, estimated travel times based on FCD from dataset $D_3$ are compared with the baseline experiment (dataset $D_1$). Dataset $D_1$ is a subset of $D_3$. Convergence using $D_3$ is similar to the convergence observed in Experiment 1; thus, there is no support for the hypothesis that a larger dataset leads to faster convergence. This also implies that using an extensive dataset does not negate the need for solving the path inference and travel time estimation problem simultaneously.

As expected, network coverage after convergence increases when the
larger dataset is used: the number of active network links is 59% compared to 52% for the smaller dataset. The results show that for 11% of the observations in $D_1$ the inferred paths differ depending on if $D_1$ is used alone or together with more data. Thus, the size of the FCD set is relatively influential for path inference. The reason for this is mainly that the additional data, with broader coverage of links, alter the shorter paths between the same pair of probes.

5. Conclusion

Path inference and travel time estimation from sparse floating car data is a joint problem: inferred vehicle paths are needed to estimate link travel times, and link travel times are needed to infer vehicle paths. This paper proposes a fixed point formulation of the problem of the simultaneous path inference and travel time estimation problem, which results in consistent path inference. Current methods correspond to iteration 1 of the fixed-point method. For evaluation purposes, a case study is used to estimate travel times from taxi FCD in Stockholm, Sweden. The method converges to a fixed point where input and output link travel times are consistent.
The solution for the fixed point formulation increases the network coverage of links with estimated travel times compared to the existing methods in the literature.

In the particular case study, the impact of the fixed point iterations is minor for many links, especially on major facilities. The main reason is that FCD observations on major links to a larger extent have unambiguous paths for which inference is not influenced by link travel times (for example probes on freeways do not usually face the problem of path inference). However, for links representing local and side streets the impact can be significant. Therefore, the need for fixed point iterations is expected to be even more important if the sparsity of the FCD probes increases. In general, as discussed in a previous section, it is expected that the importance of the fixed point formulation will increase with the complexity of the network, the degree of congestion, the quality of the initial assumptions on travel times, and the frequency of probes. Furthermore, it is expected that differences are bigger on local roads compared to links on major arterials and freeways (since for those the path inference solution is quite robust).

The analysis in the particular case study shows that for many links the estimated travel times are relatively insensitive to the initial conditions in terms of assumed link travel times. Meanwhile, the use of good historical travel times, as opposed to free-flow travel times as initial solution, can reduce bias and decrease the number of iterations.

The results reveal that the shortest paths between two network points can differ considerably in geometry and travel time. This has important implications for applications involving path finding, such as journey planning and route optimization.

The fixed point algorithm was applied using a specific method for path inference and travel time estimation from the literature. The method used has no features that give reason to expect that the fixed point estimation approach would work particularly well for these methods. Therefore, it is reasonable to expect that it would work well with other path inference and travel time estimation methods.

The case study further suggests that a Picard iteration converges fast (in less than 10 iterations) for varying dataset sizes. Thus, if the path inference and travel time estimation methods are efficient, the fixed point iteration does not increase computing times significantly. Using the Mann iteration (MSA) slows down the convergence. However, the situation may be different for other networks or sources of FCD.

Although the method is scalable, its computational performance can be improved. The computationally heaviest part of the process is path
inference. On the other hand, many paths never change during the iterative process. Therefore, path inference may be performed more efficiently by focusing on pairs of probes based on their frequency of path change.

Acknowledgments

The authors would like to thank the Swedish National Transport Authority for funding part of this work through the Mobile Millennium Stockholm project, and Trafik Stockholm for their support and provision of data. The first author would also like to acknowledge the many useful comments and suggestions from colleagues at the Swedish ITS Graduate School.

References


