Probe Vehicle Data Sampled by Time or Space: Consistent Travel Time Allocation and Estimation

Erik Jenelius
erik.jenelius@abe.kth.se
(Corresponding author)

Haris N. Koutsopoulos
haris.koutsopoulos@abe.kth.se

Tel: +46 (0)8 790 8032
Fax: +46 (0)8 212 899
Teknikringen 72, SE-100 44 Stockholm, Sweden

KTH Royal Institute of Technology
Dept. of Transport Science
Div. of Traffic and Logistics
Stockholm, Sweden

October 30, 2014

Abstract

A characteristic of low frequency probe vehicle data is that vehicles traverse multiple network components (e.g., links) between consecutive position samplings, creating challenges for (i) the allocation of the measured travel time to the traversed components, and (ii) the consistent estimation of component travel time distribution parameters. This paper shows that the solution to these problems depends on whether sampling is based on time (e.g., one report every minute) or space (e.g., one every 500 meters). For the special case of segments with uniform space-mean speeds, explicit formulae are derived under both sampling principles for the likelihood of the measurements and the allocation of travel time. It is shown that time-based sampling is biased towards measurements where a disproportionally long time is spent on the last segment. Numerical experiments show that an incorrect likelihood formulation can lead to significantly biased parameter estimates depending on the shapes of the travel time distributions. The analysis reveals that the sampling protocol needs to be considered in travel time estimation using probe vehicle data.

KEY WORDS: travel time estimation, travel time allocation, probe vehicle, floating car data, sampling, polling

1 Introduction

Information about historical, real-time and expected future traffic conditions is critical at all levels of travel planning, traffic management and transport policy. In recent years, GPS devices installed in vehicles and smartphones have emerged as a new type of traffic sensor. These probe vehicle (also called floating vehicle) sensors are often opportunistic in the sense that their original purpose is not to collect traffic data, but have a great potential for cost efficient traffic monitoring. Unlike stationary point sensors (e.g., loop detectors) and point-to-point sensors (e.g., automatic number plate recognition cameras), they can collect area-wide travel time data for any part of the network where equipped vehicles move (Antoniou et al., 2011).

However, a number of limitations mean that new sophisticated methods are needed to process the data and generate useful information, compared to traditional sensors (Leduc, 2008). One challenge is that sampling (also called reporting, or polling) frequencies are often low (less than one per minute), so that vehicles may have traversed significant distances between reports. A consequence of low sampling frequency is that the true path taken by the vehicle may be uncertain. Another consequence is that the spatial resolution of the individual travel time observations is typically lower than the resolution of the network travel time model to be estimated. For the path between two consecutive reports, only the average travel speed of the vehicle is known (derived from the traversed distance and the difference in the time stamps), while the finer details of the vehicle trajectory across different network components (e.g., network links) are unobserved. The travel times of individual network components are thus only measured in aggregated form through space-mean travel speeds.

The travel time estimation problem on probe vehicle data therefore consists of two parts: (i) allocating fractions of the observed travel times to different network components, and (ii) estimating the travel time distributions of the network components. Since the likelihood of a particular allocation depends on the travel time distributions of the components, allocation and estimation is fundamentally a joint problem. A natural requirement of an estimation method is that the true model parameters are recovered as the number of observations increase given that the model specification is correct, i.e., that the estimator is consistent. For consistent estimation, allocation and estimation should be performed simultaneously. Such an approach based on multivariate normal travel time distributions is proposed by Jenelius and Koutsopoulos (2013).

For computational efficiency, however, most proposed methods divide the process into two steps, performed either once or iteratively. First, each probe vehicle travel time observation is decomposed into a travel time for each traversed network component. Second, the travel time distributions for the network components are estimated. Heuristic methods for the travel time allocation are proposed by Hellinga et al. (2008), Zheng and van Zuylen (2013), and Sanaullah et al. (2013). Iterative procedures to estimate the parameters of the network travel time model (using Bayesian and maximum likelihood estimation, respectively), and to decompose the probe vehicle travel times according to the most likely allocation given the current parameter values, are proposed by Hunter et al. (2009), Westgate et al. (2013), and Hofleitner et al. (2012).

1.1 Probe vehicle sampling protocols

The times and locations where probe vehicles are sampled are determined by some form of sampling rule, or protocol. Two main sampling principles can be distinguished: *time-based sampling*, where the vehicle trajectory is sampled at certain time intervals, and *space-based sampling*, where the vehicle trajectory is sampled at certain distance intervals. Being opportunistic sensors, the protocol of the sampling typically cannot be controlled for the purpose of travel time estimation. In this situation, the method of analysis must instead be adapted to the sampling protocol.

Some studies comparing time-based and space-based probe vehicle data exist. As noted by Liu et al. (2007), Westgate et al. (2013) and others, a consequence of time-based sampling is that locations are sampled more densely in parts where the speed is low; with space-based sampling, on the other hand, the sampled locations are independent of the speed. Since congestion detection is an important application of probe vehicle measurements, this has been seen as an advantage of time-based sampling.

Liu et al. (2007) evaluate the benefits and drawbacks of time-based and space-based sampling of probe vehicle data in several dimensions: missing data ratio and GPS errors, number of useful records for travel time estimation, map-matching accuracy, and accuracy in link travel time or speed allocation. The analysis is performed using GPS data from 1570 taxis in Nagoya, Japan; the data set contained probe vehicle data sampled both by time (every 5 seconds) and by space (every 50 meters). Regarding the allocation of the observed space-mean speed to network links, the authors find that time-based sampling leads to somewhat smaller errors when the average sampling frequency is similar. However, definite conclusions are hard to draw from the study since the "true" link speeds were estimated from probe vehicle data with 5 second and 50 meter sampling intervals, respectively.

citewestgateetal2013 present a Bayesian model for estimating link travel time distributions based on GPS data from ambulances, utilizing position and instantaneous speed information and considering that the vehicle path is unknown. The model is applied to vehicle data from Toronto, Canada, and is evaluated against a reference method based only on the speed measurements from the GPS data. In an appendix, the authors show that the inverse of the harmonic mean of instantaneous speeds from probe vehicles is an unbiased and consistent estimator of the mean segment travel time when sampling by space, whereas it is biased upwards when sampling by time. This is a direct parallel to the case with stationary loop detectors, where the harmonic mean of the observed speeds is the appropriate quantity to use for travel time estimation. The paper does not derive an unbiased and consistent estimator of the mean segment travel time (nor other statistics of the travel time distribution) under time-based sampling.

Although the studies above have addressed certain aspects of the impact of the sampling protocol for travel time estimation, the two major problems of (i) travel time allocation and (ii) travel time estimation have not been analyzed systematically under different sampling protocols previously.

1.2 Objectives

The purpose of this paper is therefore to highlight the role of the sampling protocol for consistent travel time allocation and estimation. Using a general travel time model formulation for a network path, it is shown that whether sampling is space-based or time-based determines the proper specification of the likelihood function of the generated data. In particular, time-based sampling induces a selection bias in the observed travel times between two given positions which needs to be corrected for. This bias is not an effect of GPS measurement errors or uncertainty about the true traversed path but an integral feature of the time-based sampling process that exists also when true vehicle positions and paths are perfectly known.

The paper then considers the special case where the travel time model is partitioned into segments with uniform space-mean speed within each segment. Explicit expressions are derived under both sampling principles for the probability distribution of the allocation of the measured travel times among the traversed segments, and the likelihood function of the measurements. In particular, time-based sampling is biased towards measurements where a disproportionally large fraction of the travel time is spent on the last traversed segment, and requires adjustments of the likelihood function that have not been considered in previous studies.

Using numerical examples where the true segment travel time distributions are known, the optimal travel time allocation and the impact of model misspecification for the bias of parameter estimates are analyzed for several distribution families and parameter configurations. Given that most studies have implicitly treated the sampling as space-based although the actual sampling of the empirical data has been time-based, the estimation analysis focuses on this case. The performance of a reference travel time estimation approach based on explicit proportional allocation is also analyzed.

It is worth stressing that the purpose of the paper is not to propose or evaluate a specific model formulation or algorithm for travel time estimation, but to demonstrate that the sampling protocol needs to be properly taken into account for *any* model formulation and algorithm to avoid estimation bias, and to illustrate this for an important special case.

The remainder of the paper is organized as follows. The impact of space-based and time-based sampling protocols are analyzed in a general travel time model formulation in Section 2. Section 3 considers the special case with segments with uniform space-mean speeds. Sections 4 and 5 present numerical studies of the travel time allocation and the bias of MLE using correct and incorrect likelihood formulations, respectively. Section 6 discusses the implications of the results and Section 7 concludes the paper.

2 A general path travel time model

In this section, sampling protocols based on space and time are considered for a general probabilistic path travel time model. The results obtained here are thus valid for any more specific formulation. The special case where the path is discretized into segments with uniform space-mean speed is considered in Section 3.

Positions along a given path are uniquely identified by the offset from the start point of the path. The travel time $T(x, d) \geq 0$ between two positions x and $x + d$ is stochastic and described by a probability density function (pdf) $p_T(\tau \,|\, x, d, \theta)$, where θ is a set of parameters to be estimated, characterizing the pdf for any x and d. For simplicity in notation, travel time is assumed to be independent of the timing of the trip. This pdf for all x and d is referred to in the following as the *travel time model*.

A dual concept to the travel time $T(x, d)$ between two known positions x and $x + d$ is the travel distance $D(x, \tau) \geq 0$ from a known position x during a known time interval τ. The pdf of this distance for all possible start positions x and time intervals τ, denoted $p_D(d \,|\, x, \tau, \theta)$, is referred to as the *travel distance model*. As will be shown in Section 2.1, the travel time model is closely linked to space-based sampling of vehicle trajectories, while the travel distance model is closely linked to time-based sampling.

When fully specified, the travel time model and the travel distance model contain the same information, and are characterized by the same set of parameters θ. Given a feasible specification of one of the models, the specification of the other model can, at least in principle, be derived.

2.1 Probe vehicle measurements

Each probe vehicle reports its position x_i and the corresponding time t_i at certain discrete points $i = 0, 1, \ldots, n$, producing a *trace* $(x_{0:n}, t_{0:n})$. The vehicle positions are assumed to be measured without errors. Uncertainty in the true vehicle positions due to GPS measurement errors is an important aspect of probe vehicle data but does not influence the present analysis. Further, it is assumed that the true path of the vehicle is known; in cases where the path is uncertain the analysis applies to each feasible path. Additional information such as instantaneous speed is not considered here.

In practice, the protocols that are used to determine the times and locations where vehicles are sampled differ depending on the design and configuration of the technical systems for measurement, transmission and storage of data. Two main and conceptually different principles can be distinguished:

1. Space-based sampling. The vehicle trajectory is sampled at certain locations or at a certain distance from the previous sample point. This may require that the vehicle has a separate device that measures distance, such as an odometer.

2. Time-based sampling. The vehicle trajectory is sampled at certain points in time or at a certain time interval since the previous sample point. This may be the simplest and most common sampling method.

The differences between space-based and time-based sampling are illustrated in Figure 1 for the case with constant sampling intervals. The curves represent the vehicle unit travel time, i.e., the inverse travel speed. The area under the curve between two positions represents the travel time between the positions. With space-based sampling, the sampled positions are exogenous to the travel speed. With time-based sampling, meanwhile, the positions are endogenous to the speed; low speed implies that sampled positions are concentrated in space, and vice versa.

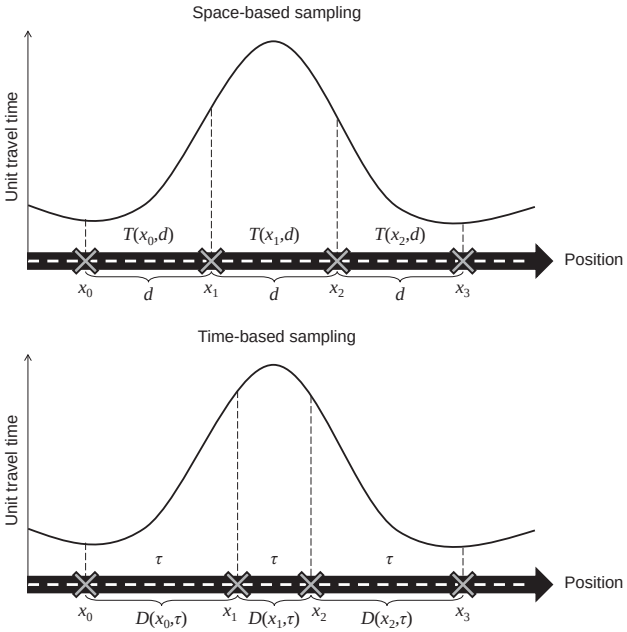

Figure 1: Illustration of space-based (top) and time-based (bottom) sampling, for the case with constant sampling intervals.

In the following a single pair of points (x_0, t_0) and (x_1, t_1) is considered. Rather than absolute positions and time stamps, it is convenient to express the pair of observations in terms of the distance and time interval between the reports, i.e.,

$$d = x_1 - x_0, \tag{1}$$

$$\tau = t_1 - t_0. \tag{2}$$

The first position and time stamp (x_0, t_0) are considered as given. Since the travel time distribution is assumed to be independent of the timing of the trip, t_0 does not influence the likelihood function. Given the first position $x = x_0$ and the parameters of the travel time model θ, the pdf of the subsequent measurement (d, τ) is denoted $p(d, \tau \,|\, x, \theta)$.

2.1.1 Space-based sampling

With a space-based sampling protocol, the probe vehicle trajectory is sampled at certain distance intervals according to some pdf $p_{\mathrm{sd}}(d \,|\, \beta_{\mathrm{sd}})$, where β_{sd} is the set of known parameters describing the sampling protocol. The sampling frequency may be fixed and deterministic, in which case $p_{\mathrm{sd}}(d \,|\, \beta_{\mathrm{sd}})$ is a point mass distribution. In practice, the sampling interval may be stochastic due to, e.g., transmission errors, so that each report has a certain probability of not being received by the data center, or odometer measurement errors.

Space-based sampling implies that the distance between reports are independent of the travel time model and its parameters θ. The likelihood function

$p(d, \tau \,|\, x, \theta, \beta_{\text{sd}})$ can thus be factored as

$$p(d, \tau \,|\, x, \theta, \beta_{\text{sd}}) = p_T(\tau \,|\, x, d, \theta) \cdot p_{\text{sd}}(d \,|\, \beta_{\text{sd}}). \tag{3}$$

$p_T(\tau \,|\, x, d, \theta)$ is the travel time model, i.e., the pdf of the travel time τ over the part of the path defined by (x, d). This implies that the MLE of the parameters of the travel time model corresponds directly to maximizing the pdf of the measurements.

2.1.2 Time-based sampling

With a time-based sampling protocol, the vehicle trajectory is sampled at a certain time interval according to some pdf $p_{\text{st}}(\tau \,|\, \beta_{\text{st}})$, where β_{st} is the set of parameters, independent of the travel time model. The pdf $p_{\text{st}}(\tau \,|\, \beta_{\text{st}})$ is a point mass distribution if the sampling frequency is fixed and deterministic. In practice, transmission errors and other technical issues may mean that the sampling interval is stochastic.

The likelihood function can thus be factored as

$$p(d, \tau \,|\, x, \theta, \beta_{\text{st}}) = p_D(d \,|\, x, \tau, \theta) \cdot p_{\text{st}}(\tau \,|\, \beta_{\text{st}}). \tag{4}$$

$p_D(d \,|\, x, \tau, \theta)$ is the travel distance model, i.e., the pdf of the travel distance d, starting from position x, over the time interval τ. This means that the MLE of the parameters θ corresponds to maximizing the pdf of the observed travel distances, as opposed to the travel times in the case with space-based sampling.

In other words, time-based sampling produces a biased sample of travel times between two given positions. Conditional on the traveled distance, the pdf of the travel time is obtained from Bayes' theorem,

$$p(\tau \,|\, x, d, \theta, \beta_{\text{st}}) = \frac{p_D(d \,|\, x, \tau, \theta)}{\int_0^\infty p_D(d \,|\, x, \tau', \theta) \cdot p_{\text{st}}(\tau' \,|\, \beta_{\text{st}}) \mathrm{d}\tau'} \cdot p_{\text{st}}(\tau \,|\, \beta_{\text{st}}), \tag{5}$$

which is in general not equal to the travel time model $p_T(\tau \,|\, x, d, \theta)$ of interest.

2.2 Remarks: Implications of sampling protocol

The analysis above shows that the principle for the sampling protocol determines the proper specification of the likelihood function. For a given observation (d, τ), the likelihood functions for space-based and time-based sampling protocols will not in general be maximized by the same parameter values θ. If the incorrect likelihood function is used, for example if the travel time model is estimated directly even though the sampling is time-based, the parameter estimates may be biased.

3 Special case: Uniform segment speeds

To apply and illustrate the general results from Section 2, a special case is considered where the path is partitioned into segments (for example, network links or some smaller unit), along each of which the vehicle space-mean speed is assumed to be uniform. Furthermore, the space-mean speeds are assumed to be independent between segments. Models of this form are used by, e.g., Hunter

et al. (2009) and Westgate et al. (2013). The assumption of independence between segments is mainly for notational convenience as it allows each segment pdf to be characterized by a separate set of parameters. The analysis can be extended to models with nontrivial joint distributions for the segments through the use of conditional pdfs in a straightforward way.

The path is divided into M segments. Without loss of generality the start position x is assumed to be on segment 1. Let L_j denote the length of segment j, except for $j = 1$, where L_1 denotes the distance from x to the end of segment 1. Further, let S_j and E_j be the distance from x to the beginning and the end locations of segment j, respectively, where $S_1 = 0$, $S_j = \sum_{i=1}^{j-1} L_i$ for $j = 2, \ldots, M$ and $E_j = S_j + L_j$. Measurements where the end position is beyond the end of segment M, i.e., $d > E_M$, are not considered. This has implications in the case of time-based sampling, as shown in Section 3.2.

The vehicle space-mean speed on each segment j is a stochastic variable V_j with pdf $p_{V_j}(v \,|\, \theta_j)$, where θ_j is a set of parameters. The reciprocal of the space-mean speed is referred to as the unit travel time. The unit travel time on each segment j is a stochastic variable $U_j = 1/V_j$ with pdf $p_{U_j}(u \,|\, \theta_j)$. The pdf of the space-mean speed is related to that of the unit travel time as

$$p_{V_j}(v \,|\, \theta_j) = \frac{1}{v^2} p_{U_j}\left(\frac{1}{v} \,\middle|\, \theta_j\right). \tag{6}$$

Given the start position x and the end position $x + d$, the distance traversed on each segment $j = 1, \ldots, M$, denoted ℓ_j, is known. With the end position $x + d$ on segment $m \in \{1, \ldots, M\}$,

$$\ell_j = \begin{cases} L_j & j < m, \\ d - S_m & j = m, \\ 0 & j > m. \end{cases} \tag{7}$$

In contrast, the time spent on each segment j, denoted τ_j, is not observed, expect in the case $m = 1$ where $\tau_1 = \tau$. In all cases, however, the sum of the times spent on each segment must be equal to the observed travel time, that is, $\sum_{j=1}^{m} \tau_j = \tau$. It is convenient to let $f_j = \tau_j / \tau$ denote the unobserved fraction of the total travel time τ spent on segment j. In the following, the vector $f_{1:m} = (f_1, \ldots, f_m)$ is referred to as the travel time *allocation*. The set of all feasible allocations is the unit simplex $\Delta_m = \{f_{1:m} \,|\, \sum_{j=1}^{m} f_j = 1\}$.

It was shown in Section 2 that the likelihood function is proportional to the travel time model with space-based sampling and proportional to the travel distance model with time-based sampling. The travel time model and the travel time allocation distribution under space-based sampling are derived in Section 3.1; the travel distance model and the travel time allocation distribution under time-based sampling are derived in Section 3.2.

3.1 Space-based sampling

The travel time $T(x, d)$ between the two given positions x and $x + d$ is the sum of the travel times on the traversed part of each segment,

$$T(x, d) = \sum_{j=1}^{m} \ell_j U_j. \tag{8}$$

For any x and $x+d$ on the same segment, i.e., for $m=1$, $T(x,d)=dU_1$ and the travel time model is

$$p_T(\tau \,|\, x,d,\theta) = \frac{1}{d} p_{U_1}\left(\frac{\tau}{d}\,\middle|\,\theta_1\right), \quad d \in (S_1, E_1]. \tag{9}$$

When x and $x+d$ are on different segments, i.e., $m \geq 2$, only the total travel time τ is observed, while the segment unit travel times U_j, $j=1,\ldots,m$, are latent variables. The joint pdf of the travel time τ between x and $x+d$ and allocation $f_{1:m}$ is the product of individual segment travel time models according to (9),

$$p(\tau, f_{1:m} \,|\, x,d,\theta) = \frac{\tau^{m-1}}{\prod_{j=1}^m \ell_j} \cdot \prod_{j=1}^m p_{U_j}\left(\frac{\tau}{\ell_j} f_j \,\middle|\, \theta_j\right), \tag{10}$$

where τ^{m-1} is a normalization factor. The travel time model is obtained by integration over all feasible allocations,

$$p_T(\tau \,|\, x,d,\theta) = \frac{\tau^{m-1}}{\prod_{j=1}^m \ell_j} \int_{\Delta_m} \prod_{j=1}^m p_{U_j}\left(\frac{\tau}{\ell_j} f_j \,\middle|\, \theta_j\right) \mathrm{d}f_{1:m}. \tag{11}$$

Conditional on the measured travel time, the pdf of the latent allocation of travel time to the traversed segments is obtained from (10) and (11) through Bayes' theorem,

$$\begin{aligned} p(f_{1:m} \,|\, x,d,\tau,\theta) &= \frac{p(\tau, f_{1:m} \,|\, x,d,\theta)}{p(\tau \,|\, x,d,\theta)} \\ &= \frac{\prod_{j=1}^m p_{U_j}\left(\frac{\tau}{\ell_j} f_j \,\middle|\, \theta_j\right)}{\int_{\Delta_m} \prod_{j=1}^m p_{U_j}\left(\frac{\tau}{\ell_j} f'_j \,\middle|\, \theta_j\right) \mathrm{d}f'_{1:m}}. \end{aligned} \tag{12}$$

The pdf of the travel time allocation thus depends only on the product of the segment unit travel time pdfs. It follows that the sequence in which the segments are traversed does not affect the travel time allocation. Another consequence of (12) is that the allocation pdf is symmetric in the segments if the segment unit travel time pdfs and parameters are identical.

3.2 Time-based sampling

The travel distance $D(x,\tau)$ from a given position x during a given time interval τ is

$$D(x,\tau) = \sum_{j=1}^{m-1} L_j + \left(\tau - \sum_{j=1}^{m-1} L_j U_j\right) V_m, \tag{13}$$

where m is the index of the segment such that

$$\sum_{j=1}^{m-1} L_j U_j < \tau, \tag{14}$$

$$\sum_{j=1}^{m} L_j U_j \geq \tau. \tag{15}$$

The number of traversed segments is an endogenous stochastic variable, which makes the travel distance model relatively more complicated than the travel time model. In order to derive the travel distance model and the pdf of the travel time allocation, an unnormalized pdf, or *potential*, of the travel distance d is first considered. For any x and $x+d$ on the same segment, i.e., for $m=1$, the traversed distance is $D(x,\tau)=\tau V_1=\tau/U_1$ and the travel distance potential is

$$\begin{aligned} g(d\,|\,x,\tau,\theta) &= \frac{1}{\tau} p_{V_1}\left(\frac{d}{\tau}\,\middle|\,\theta_1\right) \\ &= \frac{\tau}{d^2} p_{U_1}\left(\frac{\tau}{d}\,\middle|\,\theta_1\right), \quad d\in(S_1,E_1], \end{aligned} \tag{16}$$

where the second equality follows from (6). Since the travel distance may extend beyond the end of the segment, the potential (16) is not a proper pdf without normalization.

For a probe vehicle measurement ending on segment $m\geq 2$, the joint potential of the travel distance d and allocation $f_{1:m}$ is given by the product of the travel time models of the first $m-1$ segments and the travel distance model of segment m,

$$g(d,f_{1:m}\,|\,x,\tau,\theta) = \frac{\tau^{m-1}}{\prod_{j=1}^{m-1}\ell_j}\prod_{j=1}^{m-1} p_{U_j}\left(\frac{\tau f_j}{\ell_j}\,\middle|\,\theta_j\right)\cdot\frac{1}{\tau f_m}p_{V_m}\left(\frac{\ell_m}{\tau f_m}\,\middle|\,\theta_m\right), \quad d\in(S_m,E_m]. \tag{17}$$

Applying (6) to segment m, the potential can be formulated fully in terms of unit travel time pdfs,

$$g(d,f_{1:m}\,|\,x,\tau,\theta) = \frac{\tau^m f_m}{\ell_m\prod_{j=1}^{m}\ell_j}\prod_{j=1}^{m} p_{U_j}\left(\frac{\tau f_j}{\ell_j}\,\middle|\,\theta_j\right), \quad d\in(S_m,E_m]. \tag{18}$$

The marginal potential for travel distance d is obtained by integration over the latent travel time allocation,

$$g(d\,|\,x,\tau,\theta) = \frac{\tau^m}{\ell_m\prod_{j=1}^{m}\ell_j}\int_{\Delta_m} f_m\prod_{j=1}^{m} p_{U_j}\left(\frac{\tau f_j}{\ell_j}\,\middle|\,\theta_j\right)\mathrm{d}f_{1:m}, \quad d\in(S_m,E_m]. \tag{19}$$

The travel distance model is finally obtained by normalizing the potential $g(d\,|\,x,\tau,\theta)$ across all M segments of the path, i.e., from distance 0 to E_M. It follows after some algebra that the integral of $g(d\,|\,x,\tau,\theta)$ over segment m is

$$\begin{aligned} G_m(x,\tau,\theta) &= \int_{S_m}^{E_m} g(d\,|\,x,\tau,\theta)\mathrm{d}d \\ &= \frac{\tau^{m-1}}{\prod_{j=1}^{m-1}\ell_j}\int_{\Delta_m}\prod_{j=1}^{m-1} p_{U_j}\left(\frac{\tau f_j}{\ell_j}\,\middle|\,\theta_j\right)\left[1-P_{U_m}\left(\frac{\tau f_m}{L_m}\,\middle|\,\theta_m\right)\right]\mathrm{d}f_{1:m}, \end{aligned}$$

(20)

where $P_{U_m}(\cdot \,|\, \theta_m)$ is the cdf of the unit travel time on segment m. The travel distance model is thus

$$p_D(d \,|\, x, \tau, \theta) = \frac{g(d \,|\, x, \tau, \theta)}{\sum_{m=1}^{M} G_m(x, \tau, \theta)}. \tag{21}$$

Conditional on the measured travel distance d, the pdf of the allocation of travel time to the traversed segments is obtained from (18) and (19),

$$\begin{aligned} p(f_{1:m} \,|\, x, d, \tau, \theta) &= \frac{g(d, f_{1:m} \,|\, x, \tau, \theta)}{g(d \,|\, x, \tau, \theta)} \\ &= \frac{f_m \prod_{j=1}^{m} p_{U_j}\left(\frac{\tau f_j}{\ell_j} \,\middle|\, \theta_j\right)}{\int_{\Delta_m} f'_m \prod_{j=1}^{m} p_{U_j}\left(\frac{\tau f'_j}{\ell_j} \,\middle|\, \theta_j\right) \mathrm{d}f'_{1:m}}. \end{aligned} \tag{22}$$

Comparing (22) and (12), it is seen that the allocation pdf with time-based sampling puts the additional weight f_m on the last segment. Thus, while the allocation is independent of the order in which the first $m-1$ segments are traversed, the last segment is distinctly different.

3.3 Remarks: Implications of sampling protocol

The analysis above shows that the sampling protocol influences the proper allocation of measured travel times to the traversed segments and the estimation of segment travel time distributions. Regarding the travel time allocation, time-based sampling implies that all traversed segments are not equal, but that a larger fraction of the time tends to be spent on the last traversed segment. The reason for this is that the travel speed on one segment determines how much time is left for travel on subsequent segments before the next report is sent: the fact that the vehicle was not sampled on segments $(1, \ldots, m-1)$ is a signal that the vehicle speed was likely relatively high; conversely, the fact that the vehicle was sampled on segment m signals that the vehicle speed was likely relatively low. As a result, allocations with more travel time on the last segment are thus given proportionally higher weights in (21) compared to (11).

The second difference between the two sampling protocols arises because the observed travel distances are constrained by the finite length of the path. With space-based sampling this has no impact on the estimation since the travel distance is independent of the parameters to be estimated. With time-based sampling, on the other hand, observations are only recorded if both positions are within the path, which creates a censoring of high-speed observations. In other words, segment unit travel times are not drawn from the full population but from the subpopulation of segment travel times such that the traveled distance is less than E_M. Hence, the travel distance model is normalized by the denominator in (21). The influence of the finite path length constraint depends on the length of the path in relation to the start position of the vehicle and the distance between reports.

The results suggest that applying the travel time model directly on probe vehicle measurements sampled by time will bias parameter estimates in two

ways: First, the constraint of the finite path means that low speeds are over-represented in the measurements compared to space-based sampling, so that all segment unit travel times will be overestimated. Second, the travel time model does not take into account that time-based sampling over-represents time spent on the last segment, and consequently allocates too little time to the last segment and too much time to the other segments. As a result, the unit travel time of all but the last traversed segment will be overestimated while the last traversed segment will be underestimated. The net effect of both mechanisms combined may depend on the length of the path, the characteristics of the segment unit travel time distributions and the probe vehicle sampling protocol. The influence of these factors is analyzed in Section 5.

4 Numerical analysis

In this section the significance of space-based versus time-based sampling protocols for the allocation of measured travel times to the traversed segments is studied numerically for the segment model in Section 3; the impact for travel time distribution parameter estimation is considered in the next section. To make the results as simple to interpret as possible, the path is split into $M = 2$ segments, both of length 0.5.

Three different distribution families for the segment unit travel times are considered: (i) lognormal, (ii) gamma, and (iii) inverse gamma. All distributions have two parameters, support on the positive real line and are unimodal. Furthermore, they have all been used in the literature to model travel time or speed distributions (Richardson and Taylor, 1978; Uno et al., 2009; Polus, 1979; Dandy and McBean, 1984; Haight and Mosher, 1962). Here, each distribution is parameterized in terms of its mean μ and variance σ^2. The three distributions are highly similar when the mean is high and the variance is low, while the differences are bigger with lower mean and higher variance. In general, the gamma distribution has the thinnest right tail and the inverse gamma distribution has the thickest right tail for the same mean and variance.

A number of parameter configurations are evaluated to analyze how the bias of the estimates depends on the combination of mean values and variances of the underlying segment unit travel time distributions. The mean parameters μ_i, $i = 1, 2$, for the two segments are given two possible levels; a baseline value 2 and a high value 10. In the same way, the variance parameters σ_i^2 can take a baseline value 1.5 and a high value 6. These are chosen to be fairly representative for the scale and shape of empirical unit travel time distributions while demonstrating some interesting features in the estimation results. For each segment, there are thus four different parameter combinations: $(\mu_i, \sigma_i^2) = (2, 1.5)$, $(10, 1.5)$, $(2, 6)$, and $(10, 6)$.

4.1 Travel time and travel distance distributions

Figure 2 illustrates the travel time model, more specifically the distribution of travel times from the fixed start position $x = 0$ at the beginning of segment 1 to the fixed end position $x + d = 1$ at the end of segment 2. The left diagram shows the distribution for the symmetric case $(\mu_i, \sigma_i^2) = (2, 1.5)$ for $i = 1, 2$; the right diagram shows the asymmetric case $(\mu_1, \sigma_1^2) = (10, 1.5)$, $(\mu_2, \sigma_2^2) = (2, 6)$.

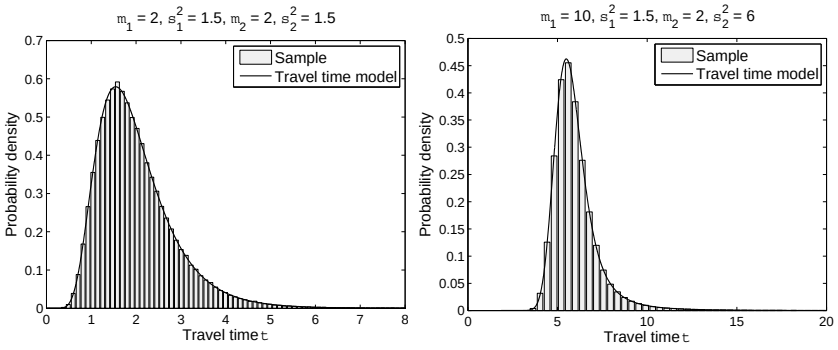

Figure 2: Illustration of the travel time model for two segments. Lognormal segment unit travel time distributions, start position $x = 0$, distance $d = 1$. Left: $(\mu_i, \sigma_i^2) = (2, 1.5)$, $i = 1, 2$. Right: $(\mu_1, \sigma_1^2) = (10, 1.5)$, $(\mu_2, \sigma_2^2) = (2, 6)$. Bars: empirical distribution, 20,000 samples. Solid line: theoretical travel time model.

The bars show the realized travel times of 20,000 sampled observations, while the line shows the corresponding theoretical travel time model according to (11). The travel time model does not have a closed form in either case but is calculated numerically. It can be seen that the distribution is smooth also for non-symmetric segment unit travel time distributions.

To illustrate the travel distance model, Figure 3 shows the distribution of travel distances from the fixed start position $x = 0$ during a fixed sampling time interval, $\tau = 1$ shown to the left and $\tau = 2$ to the right, respectively. The bars show the realized travel distances of 20,000 sampled observations while the line shows the theoretical travel distance model according to (21). The example illustrates the discontinuous behavior of the travel distance model at the transition point between the segments. In the right diagram in particular, there is a negative jump in the density at $x = 0.5$; this represents the fact that it is relatively unlikely that the vehicle only just makes it to segment 2 and then drives at low speed, compared to driving at lower speed on segment 1 and not reaching segment 2, or driving at higher speed on at least one of the segments. The example suggests that the shape of the distribution is highly dependent on the segment unit travel time distributions and the sampling time interval τ.

4.2 Travel time allocation

The probability density functions of the travel time allocation under space-based sampling (12) and time-based sampling (22) are compared for an identical probe vehicle measurement with fixed start position x, distance d and time interval τ. Specifically, the start position is set to $x = 0$ (beginning of segment 1), the distance to $d = 1$ (end of segment 2), and two different time intervals are considered, $\tau = 1$ and $\tau = 4$.

Figure 4 shows the pdf of the fraction of travel time allocated to segment 2 for three symmetric combinations of mean and variance parameters (μ_i, σ_i^2), $i = 1, 2$: (1) a moderately skewed configuration $(2, 1.5)$, (2) a more concentrated and non-skewed form $(10, 1.5)$, and (3) one with higher variance and skewness $(2, 6)$. The segment unit travel time distributions are lognormal. Space-based

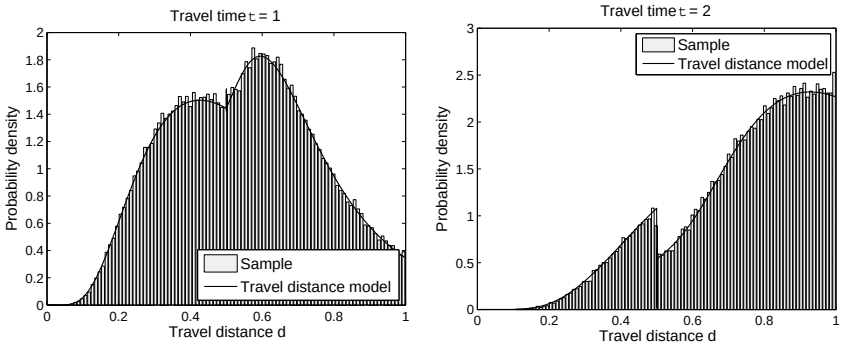

Figure 3: Illustration of the travel distance model for two segments. Lognormal segment unit travel time distributions with means 2 and variances 1.5, start position $x = 0$. Left: travel time $\tau = 1$. Right: travel time $\tau = 2$. Bars: empirical distribution, 20,000 samples. Solid line: theoretical travel distance model.

sampling is shown to the left, time-based sampling to the right. Results for observed time intervals $\tau = 1$ and $\tau = 4$ are shown at the top and the bottom, respectively.

The left column shows that the allocation pdf is symmetric around 50% with space-based sampling, which follows from (12). In other words, it is equally likely that fraction f or fraction $1 - f$ is spent on segment 2, for any $f \in (0, 1)$. The pdf is concentrated around 50% when the travel time distributions are concentrated around their means and increasingly dispersed when the variances around the means are greater. The bottom left figure shows that the allocation pdf may be bimodal if the observed travel time and the segment variances are high. Hence, it is more likely that most of the time is spent on one of the segments, than that the travel time is spent equally on both segments. This implies that the most likely allocation of travel time is not unique.

The right column shows that the allocation pdf is skewed to the right with time-based sampling, which means that it is more likely that $f > 50\%$ of the time is spent on segment 2 than $1 - f < 50\%$. This result follows directly from (22). The allocation pdf for the concentrated distribution $(10, 1.5)$ is almost non-skewed and very similar to the allocation with space-based sampling. The skewness increases with the relative variance and skewness of the travel time distributions. The skewness also increases with the travel time τ. Thus, the slower the vehicle is observed to travel, the higher the probability that most of the delay occurs on segment 2. As the bottom right figure shows, the most likely allocation with $\tau = 4$ and $(\mu_i, \sigma_i^2) = (2, 6)$ is that more than 90% of the time is spent on segment 2.

Figure 5 shows the allocation pdf for two asymmetric combinations in which one segment has high mean and low variance and the other has low mean and high variance: (4) $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = (10, 1.5, 2, 6)$, and (5) $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = (2, 6, 10, 1.5)$. In all cases, as expected, the allocation pdf is skewed towards the segment with high mean unit travel time under both space-based and time-based sampling. Since the order in which the segments are traversed is irrelevant with space-based sampling, the two cases are identical under reflection around

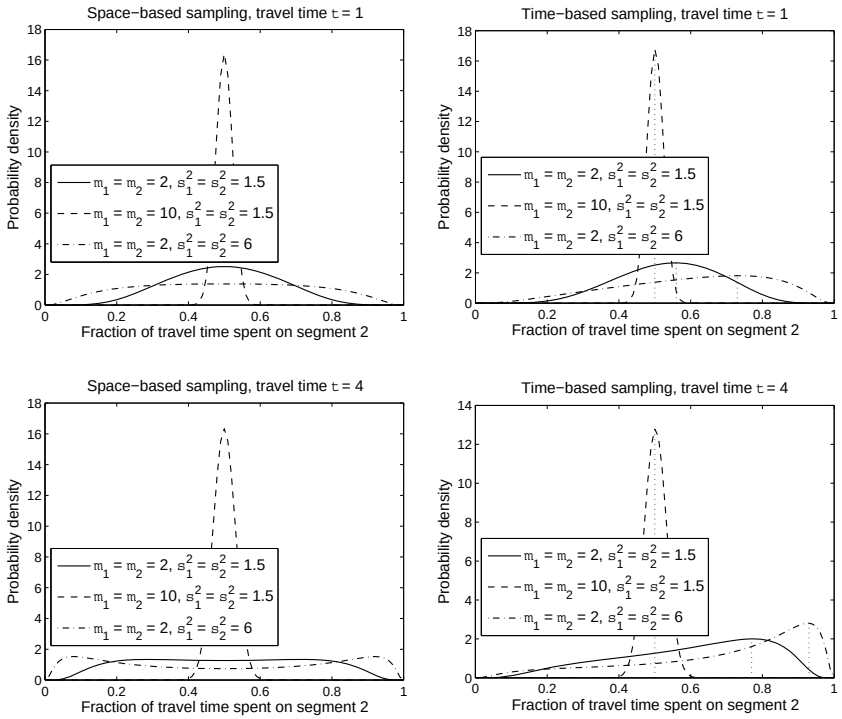

Figure 4: Pdf of the fraction of travel time allocated to segment 2. Lognormal segment unit travel times, traversed segment distances $\ell_1 = \ell_2 = 0.5$. Left column: space-based sampling. Right column: time-based sampling. Top row: travel time $\tau = 1$. Bottom row: travel time $\tau = 4$. Solid line: $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Dashed line: $\mu_1 = \mu_2 = 10$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Dash-dotted line: $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 6$.

the line $f = 0.5$. This is not the case with time-based sampling, and some probability mass is always shifted towards segment 2 as can be seen from the heights and widths of the peaks.

Figure 6 shows the allocation pdf for the lognormal, gamma and inverse gamma distributions for the same mean and variance parameters $(\mu_i, \sigma_i^2) = (2, 1.5)$, $i = 1, 2$. The solid line in each graph is identical to the solid line in the corresponding graph in Figure 4. As can be seen, the properties of the distribution beyond the mean and the variance affect the allocation as well. With space-based sampling, the left column shows that the allocation is unimodal for all distributions when $\tau = 1$, where the inverse gamma distribution has the most concentrated peak. For the long travel time $\tau = 4$ both the inverse gamma and the lognormal distributions exhibit non-unique most likely allocations. For the gamma distribution the allocation pdfs are identical for both travel times; in fact, the allocation pdf is the same for any travel time τ.

With time-based sampling, the right column shows that the level of right-skewness of the travel time allocation pdf depends on the distribution family. For $\tau = 1$ the gamma distribution leads to the most skewed allocation pdf followed by the lognormal distribution; the order is reversed for $\tau = 4$. For the

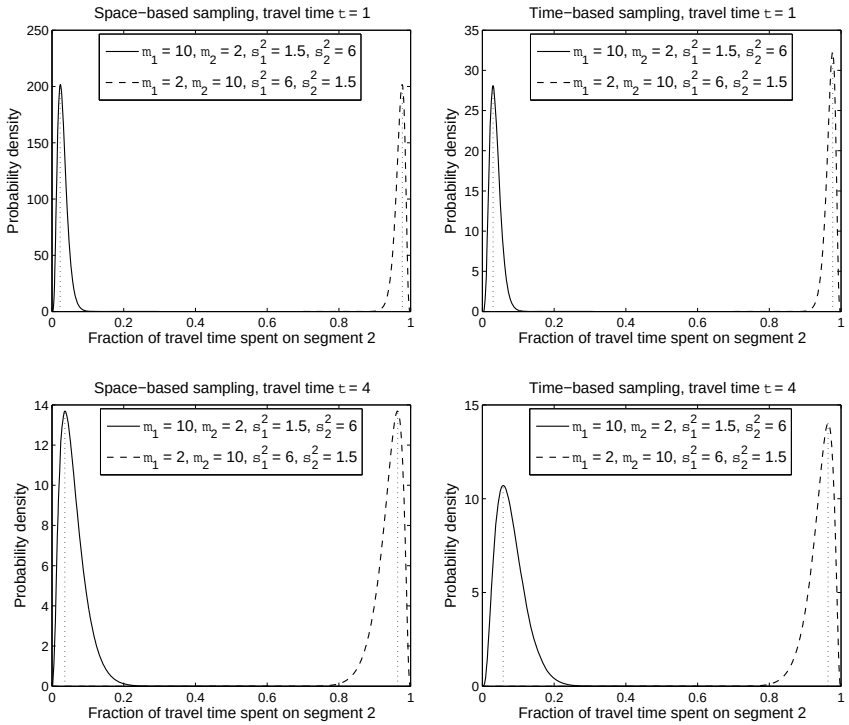

Figure 5: Pdf of the fraction of travel time allocated to segment 2. Lognormal segment unit travel times, traversed segment distances $\ell_1 = \ell_2 = 0.5$. Left column: space-based sampling. Right column: time-based sampling. Top row: travel time $\tau = 1$. Bottom row: travel time $\tau = 4$. Solid line: $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = (10, 1.5, 2, 6)$. Dashed line: $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = (2, 6, 10, 1.5)$.

gamma distribution the allocation pdfs are again the same for any travel time τ. In all considered cases the allocation pdf is unimodal, which implies that there is a unique most likely allocation. This property does not always hold, however.

5 Travel time estimation

This section considers the problem of estimating the parameters of the segment unit travel time pdfs using probe vehicle measurements sampled by time, and the consequences of using the correct travel distance model, the incorrect travel time model, or a simpler reference method. To distinguish between the impacts of travel time allocation and finite path censoring with time-based sampling, two different setups are considered: In the first setup, both segments have length 0.5 as in Section 4; in the second setup, segment 1 has length 0.5 and segment 2 has infinite length.

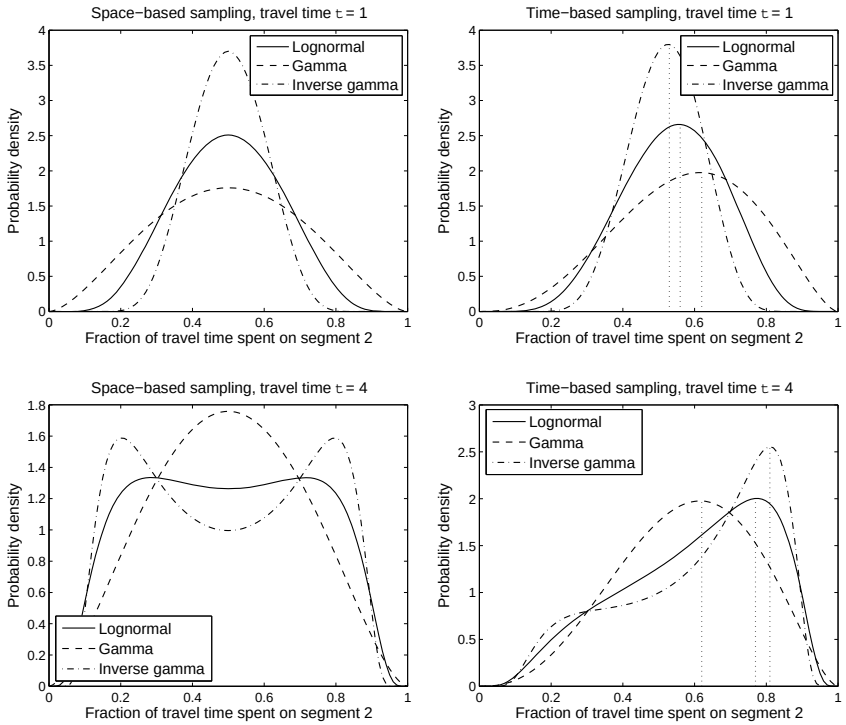

Figure 6: Pdf of the fraction of travel time allocated to segment 2. Mean and variance parameters $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1.5$, traversed segment distances $\ell_1 = \ell_2 = 0.5$. Left column: space-based sampling. Right column: time-based sampling. Top row: travel time $\tau = 1$. Bottom row: travel time $\tau = 4$. Solid line: lognormal. Dashed line: gamma. Dash-dotted line: inverse gamma.

5.1 Data generation and parameter estimation

The parameter estimation is based on K independently drawn observations (d_k, τ_k), $k = 1, \ldots, K$ sampled by time, each starting at a given position x_k sampled uniformly along segment 1, i.e., from a uniform distribution with lower bound 0 and upper bound 0.5. The sampling time interval is set to $\tau_k = 1$ for all observations. To avoid the trivial case where both reported positions are on the same segment, the end position $x_k + d_k$ is always on segment 2. The distances travelled on the two segments are thus $\ell_{1,k} = 0.5 - x_k$ and $\ell_{2,k} = x_k + d_k - 0.5$, respectively. The design is illustrated in Figure 7.

For each observation $k = 1, \ldots, K$, the unit travel time u_{ik} of the vehicle on each of the two segments $i = 1, 2$ is randomly drawn from the associated "true" distribution $p_{U_i}(u \,|\, \theta_i)$ with parameters according to a specific configuration as described in Section 4.2.

The time to reach the beginning of segment 2, $\tau_{1,k} = \ell_{1,k} u_{1,k}$, is calculated; if this exceeds τ_k, the observation is discarded. Otherwise, the distance traversed on segment 2 in the remaining time is $\ell_{2,k} = (\tau_k - \tau_{1,k})/u_{2,k}$. In the first setup where segment 2 has length 0.5, observations exceeding the length of the path are discarded.

The true unit travel time distributions are assumed to be known up to the

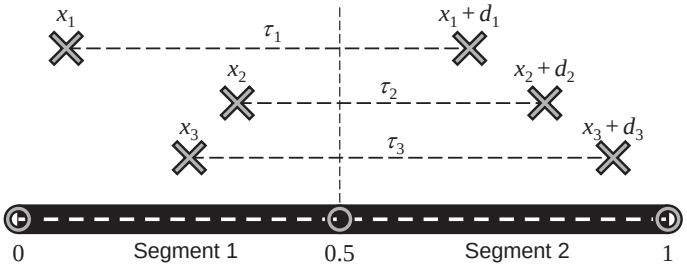

Figure 7: Illustration of the design for the numerical analysis.

parameters. In the experiments the parameter estimates $\hat{\theta}$ obtained with the different estimation methods are compared with the known, true parameter values θ, in order to investigate the errors introduced by misspecifying the likelihood function.

Three different maximum likelihood estimators are considered, based on (i) the travel time model (Section 3.1), (ii) the travel distance model (Section 3.2), and (iii) a reference approach using an explicit proportional travel time allocation. In all cases the maximization of the likelihood function is performed using a sequential quadratic programming procedure with an error tolerance in the parameters and the objective function of $1 \cdot 10^{-8}$. Integrals are evaluated numerically with recursive adaptive Simpson quadrature using an error tolerance of $1 \cdot 10^{-9}$. All calculations are performed in MATLAB (MATLAB, 2009).

MLE based on the travel time model The travel time model is the correct likelihood function for the measurements in the case of space-based sampling. It follows from (11) that the maximum likelihood parameter estimates are

$$\begin{aligned}\hat{\theta}_T &= \operatorname*{argmax}_{\theta} \prod_{k=1}^{K} p_T(\tau_k \,|\, x_k, d_k, \theta) \\ &= \operatorname*{argmax}_{\theta} \prod_{k=1}^{K} \frac{\tau_k}{\ell_{1,k}\ell_{2,k}} \int_0^1 p_{U_1}\left(\frac{\tau_k}{\ell_{1,k}}(1-f)\,\middle|\,\theta_1\right) \cdot p_{U_2}\left(\frac{\tau_k}{\ell_{2,k}} f \,\middle|\, \theta_2\right) \mathrm{d}f.\end{aligned} \tag{23}$$

MLE based on the travel distance model The travel distance model is the correct likelihood function for the measurements in the case of time-based sampling. In this application the likelihood function is conditional on that only measurements ending on segment 2 are used. The likelihood function is thus proportional to the conditional travel distance model $p_{D_2}(d \,|\, x, \tau, \theta) = g(d \,|\, x, \tau, \theta)/G_2(x, \tau, \theta)$. It follows from (21) that the likelihood function is

$$\begin{aligned}\mathcal{L}_D(\theta) &= \prod_{k=1}^{K} p_{D_2}(d_k \,|\, x_k, \tau_k, \theta) \\ &= \prod_{k=1}^{K} \frac{\tau_k}{\ell_{2,k}^2} \frac{\int_0^1 f \cdot p_{U_1}\left(\frac{\tau_k}{\ell_{1,k}}(1-f)\,\middle|\,\theta_1\right) \cdot p_{U_2}\left(\frac{\tau_k}{\ell_{2,k}} f \,\middle|\, \theta_2\right) \mathrm{d}f}{\int_0^1 p_{U_1}\left(\frac{\tau_k}{\ell_{1,k}}(1-f)\,\middle|\,\theta_1\right) \cdot \left[1 - P_{U_2}\left(\frac{\tau_k}{L_2} f \,\middle|\, \theta_2\right)\right] \mathrm{d}f},\end{aligned} \tag{24}$$

and the ML estimates are found as $\hat{\theta}_D = \text{argmax}_\theta \, \mathcal{L}_D(\theta)$.

MLE based on explicit proportional allocation For reference a sequential estimator is considered, in which the observed travel time τ_k is first split to the two segments proportionally to the traversed distances, i.e., $f_k = \ell_{2,k}/d_k$. This means that the vehicle unit travel times are assumed a priori to be equal to τ_k/d_k on both segments. The parameters of the travel time models are then estimated using MLE based on the allocated travel times,

$$\hat{\theta}_{\text{alloc},1} = \underset{\theta_1}{\text{argmax}} \prod_{k=1}^{K} p_{U_1}\left(\frac{\tau_k}{d_k} \,\middle|\, \theta_1\right), \tag{25}$$

$$\hat{\theta}_{\text{alloc},2} = \underset{\theta_2}{\text{argmax}} \prod_{k=1}^{K} p_{U_2}\left(\frac{\tau_k}{d_k} \,\middle|\, \theta_2\right). \tag{26}$$

The sequential approach with explicit allocation has the advantage that the parameters of the travel time models can be estimated separately for each segment, which increases computational efficiency. However, this estimator is in general not expected to be consistent under either space-based or time-based sampling.

5.2 Results

The parameters of the segment unit travel time distributions are estimated using maximum likelihood based on the travel time model, the travel distance model and the explicit proportional allocation. The estimation is based on $K = 20{,}000$ independently drawn samples; the large number is in general not realistic for real-world applications but is used to reduce standard errors and highlight the systematic bias of the estimators which is inherent also for smaller samples. The convergence of the parameter estimates as the number of observations increase is also investigated.

Table 1 shows the parameter estimates with time-based sampling, both segments lengths 0.5 and lognormal distributions for the five different parameter configurations studied in Section 4.2 (labeled 1–5).

As expected the parameter estimates from the travel distance model are not significantly different from the true values at the 0.05 level. Comparing the three symmetric configurations, it is seen that using the travel time model leads to greater bias the higher the variance is in relation to the mean. The parameter estimates for the concentrated distribution (2) are close to the true values. For the two other configurations (1 and 3), the mean travel time of segment 2 is overestimated while the mean travel time of segment 1 is closer to the true value; the travel time variance is underestimated for both segments, in particular segment 1. Meanwhile, the standard errors are in general smaller than for the travel distance model, which suggests that space-based sampling yields a better behaved likelihood function. This may be a reason to prefer space-based sampling of probe vehicles for travel time estimation.

Due to the explicit proportional allocation, the parameter estimates of the reference method are always the same for both segments. For configurations 1 and 3, the method overestimates the means and highly underestimates the variances. For configurations 4 and 5, the proportional allocation leads to significant bias in both mean and variance parameters, in particular for configuration 5.

Table 1: Estimated parameter values based on the travel distance model (correct), travel time model (incorrect), and proportional allocation (naive method). Time-based sampling, both segment lengths 0.5 (setup 1), lognormal unit travel time distributions, sample size $K = 20{,}000$. Standard errors are shown in parantheses.

	Unit travel time parameter configuration				
	(1)	(2)	(3)	(4)	(5)
True parameters					
μ_1	2	10	2	10	2
σ_1^2	1.5	1.5	6	1.5	6
μ_2	2	10	2	2	10
σ_2^2	1.5	1.5	6	6	1.5
Travel Distance Model (correct)					
$\hat{\mu}_{D,1}$	1.951	10.006	1.985	9.995	1.999
	(0.032)	(0.018)	(0.111)	(0.057)	(0.036)
$\hat{\sigma}^2_{D,1}$	1.488	1.501	6.596	1.548	5.891
	(0.111)	(0.029)	(1.462)	(0.084)	(0.378)
$\hat{\mu}_{D,2}$	2.042	10.025	1.940	2.070	10.012
	(0.037)	(0.016)	(0.078)	(0.041)	(0.024)
$\hat{\sigma}^2_{D,2}$	1.563	1.570	6.354	6.331	1.524
	(0.058)	(0.028)	(0.411)	(0.243)	(0.040)
$\mathcal{LL}_D$	16,120	65,803	14,564	19,095	48,753
Travel Time Model (incorrect)					
$\hat{\mu}_{T,1}$	1.977	9.836	2.114	9.910	1.557
	(0.010)	(0.015)	(0.012)	(0.024)	(0.012)
$\hat{\sigma}^2_{T,1}$	0.491	1.461	0.467	2.040	1.353
	(0.016)	(0.025)	(0.025)	(0.059)	(0.036)
$\hat{\mu}_{T,2}$	2.166	9.967	2.454	2.514	9.569
	(0.012)	(0.015)	(0.021)	(0.019)	(0.021)
$\hat{\sigma}^2_{T,2}$	0.982	1.492	3.949	4.199	1.608
	(0.027)	(0.026)	(0.180)	(0.115)	(0.038)
$\mathcal{LL}_D$	14,930	65,674	12,119	18,577	47,991
Prop. allocation (reference)					
$\hat{\mu}_{\text{alloc},1}$	2.107	9.894	2.300	4.930	4.063
	(0.005)	(0.007)	(0.007)	(0.019)	(0.014)
$\hat{\sigma}^2_{\text{alloc},1}$	0.435	0.980	0.935	6.989	4.210
	(0.005)	(0.010)	(0.013)	(0.109)	(0.063)
$\hat{\mu}_{\text{alloc},2}$	2.107	9.894	2.300	4.930	4.063
	(0.005)	(0.007)	(0.007)	(0.019)	(0.014)
$\hat{\sigma}^2_{\text{alloc},2}$	0.435	0.980	0.935	6.989	4.210
	(0.005)	(0.010)	(0.013)	(0.109)	(0.063)
$\mathcal{LL}_D$	11,718	64,752	9,122	13,732	29,683

The standard errors are small for the symmetric configurations but relatively large for the asymmetric configurations.

Table 1 also shows the log likelihood value $\mathcal{LL}_D$ for each set of estimated parameter values evaluated using the correct travel distance model. The differences between the log likelihood values for each parameter configuration can be used to compare the performances of the three estimation methods. It can be seen that the log likelihood values are all similar for configuration 2, confirming that the errors from using the travel time model or the reference method are small. The differences are larger for the other configurations; the reference method performs particularly poorly on the asymmetric configurations 4 and 5.

In all cases, the travel time model performs better than the reference method.

The results follow the same general trends when segment unit travel time rates are drawn from the gamma and inverse gamma distribution families. The sign of the bias for a specific parameter configuration is almost always the same for all three distributions, although the magnitude varies.

Figure 8 shows the probability density function of the segment unit travel times corresponding to the estimated parameter values for configurations 1, 2 and 5. For configuration 1 (top) the modes of the distributions are markedly higher with the travel time model compared to the true distribution even though the bias in the mean parameter is moderate. For configuration 2 (middle) the distributions from the travel time models are very close to the true distributions, which is in line with the travel time allocation results in Section 4.2. For configuration 5 (bottom) the travel time model again works well for segment 1 but shows significant bias for segment 2. The reference method performs poorly in the case with asymmetric travel time distributions.

The convergence of the parameter estimates as the number of observations increase is shown in Figure 9 for configuration 1. For each considered sample size an independent sample of segment unit travel times is drawn; the same sample is used for both the travel distance model (left) and the travel time model (right). If an estimator is consistent the parameter estimates should approach the true values and the standard errors should tend to zero. With time-based sampling the travel distance model estimator is consistent by the general properties of MLE, and the figure indeed shows that the estimates are centered around the true values with standard errors decreasing in size. For the variance parameters in particular, however, the standard errors remain relatively large even for (unrealistically) large sample sizes. The estimates from the travel time model, meanwhile, converge faster and with smaller standard errors to stable values, but not the true values. It can thus be concluded that the travel time model estimator is not consistent in general. Analogous results hold for the other considered distributions and parameter configurations.

Figure 10 shows the probability density function of the segment unit travel times for the setup where segment 2 has infinite length and there is no censoring of observations due to a finite path. Based on the general remarks in Section 3.3, it is thus expected that the travel time model will allocate too little time to the last segment and therefore underestimate the travel time on segment 2 but overestimate it on segment 1. Indeed, compared to the first setup where segment 2 has length 0.5, the estimation bias is generally similar in magnitude but in some cases different in sign. For configurations 1 (top) and 5 (bottom) the mean is overestimated for segment 1 and underestimated for segment 2, in agreement with expectations. For configuration 2 (middle) the distributions from the travel time models are again close to the true distributions. The reference method performs better than the travel time model for configuration 1 in terms of likelihood, but again poorly for the asymmetric configuration 5.

6 Discussion

The analytical and numerical results suggest that the principles of the sampling protocol used by the probe vehicles must receive more attention from researchers than has been the case until now. In a sense, the difference between time-based

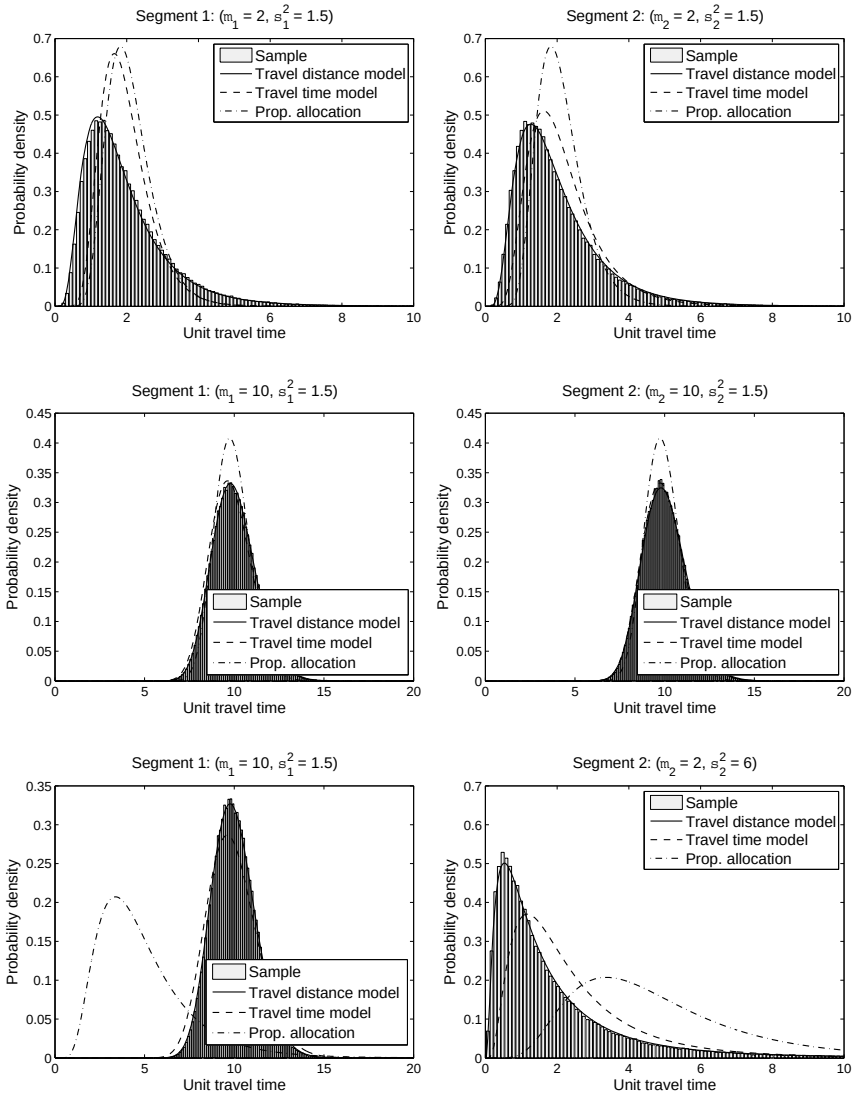

Figure 8: Estimated segment unit travel time pdfs with both segment lengths 0.5 (setup 1). Time-based sampling, lognormal unit travel time distributions, sample size $K = 20{,}000$. Left column: segment 1. Right column: segment 2. Top row: $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Middle row: $\mu_1 = \mu_2 = 10$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Bottom row: $\mu_1 = 10$, $\sigma_1^2 = 1.5$, $\mu_2 = 2$, $\sigma_2^2 = 6$. Bars: empirical sample distribution. Solid line: travel distance model (correct). Dashed line: travel time model (incorrect). Dash-dotted line: proportional allocation.

and space-based sampling is as important as the difference between the time-mean speed and the space-mean speed for loop detectors and other stationary sensors. The statistical model should incorporate the sampling protocol, so that the proper estimator will depend on the specific protocol used by the probe vehicle data source.

The numerical analysis here considers only a simple case where all obser-

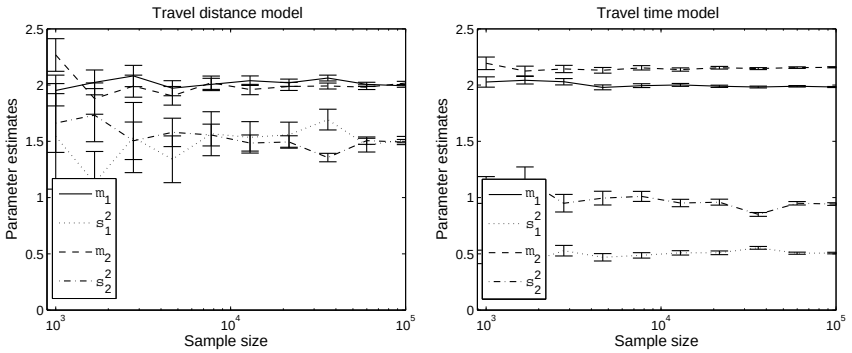

Figure 9: Estimated parameter values for different sample sizes K. Time-based sampling, lognormal segment unit travel time distributions, both segment lengths 0.5 (setup 1). Error bars show standard errors. True parameters: $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Left: travel distance model (correct). Right: travel time model (incorrect). Solid line: mean segment 1 ($\hat{\mu}_1$). Dotted line: variance segment 1 ($\hat{\sigma}_1^2$). Dashed line: mean segment 2 ($\hat{\mu}_2$). Dash-dotted line: variance segment 2 ($\hat{\sigma}_2^2$).

vations cover the same two links. The results show that while the estimation of the mean segment travel times is relatively manageable, very large sample sizes are required to precisely estimate the travel time variance parameters, and presumably other higher-order statistics, even when the correct travel distance model is used. Depending on the sampling frequency, network representation and the underlying variability in traffic conditions in practical applications, the situation may be more complicated in one sense and simpler in another sense. Specifically, vehicles may traverse more than two segments between reports, which makes the allocation and estimation problems more challenging. Section 3 showed that the distinct difference between space-based and time-based sampling persists for any number of traversed segments greater than one. On the other hand, probe vehicle observations often partially overlap each other along a given path, which could help reduce this problem. Further research should investigate how the precision and the misspecification bias depends on the sampling frequency and the level of overlap among observations.

It is worth noting that several previous studies have reported good empirical results regarding travel time allocation on time-sampled probe vehicle data even though they did not consider the implications of the sampling protocol. Hellinga et al. (2008) propose an analytical model of link travel time and delay to decompose observed travel times, while Zheng and van Zuylen (2013) use artificial neural networks (ANN) for the same problem. The performance of these heuristic methods was evaluated based on the square and absolute errors between the estimated and the true travel time allocations for each individual observation; the focus is thus somewhat different than in the present paper, and it is difficult to judge to what extent the resulting segment travel time distributions after aggregating the allocated observed travel times are biased. To some extent, a machine learning approach such as ANN may automatically account for the sampling protocol effect, which would imply that it should be trained on and applied to data from the same sampling protocol.

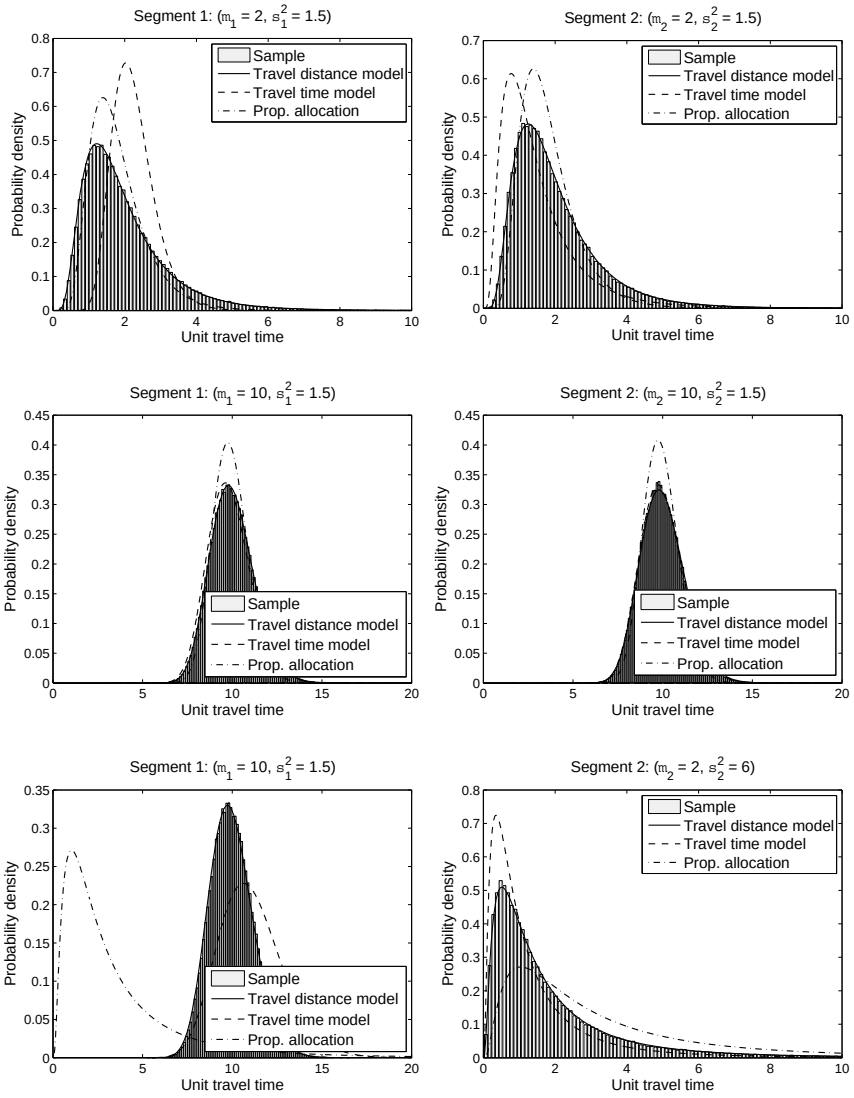

Figure 10: Estimated segment unit travel time pdfs with infinite length on segment 2 (setup 2). Time-based sampling, lognormal unit travel time distributions, sample size $K = 20{,}000$. Left column: segment 1. Right column: segment 2. Top row: $\mu_1 = \mu_2 = 2$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Middle row: $\mu_1 = \mu_2 = 10$, $\sigma_1^2 = \sigma_2^2 = 1.5$. Bottom row: $\mu_1 = 10$, $\sigma_1^2 = 1.5$, $\mu_2 = 2$, $\sigma_2^2 = 6$. Bars: empirical sample distribution. Solid line: travel distance model (correct). Dashed line: travel time model (incorrect). Dash-dotted line: proportional allocation.

Further work is needed to develop estimators for the specific mixed protocols based on distance, time and sometimes other events that are used in practice, and apply them to real probe vehicle data. Another topic for further research is unbiased and efficient travel time estimation considering both the sampling protocol and the uncertainty in positions and paths due to GPS measurement errors and low sampling frequency.

7 Conclusion

This paper considered the estimation of travel time distributions on probe vehicle data sampled by time and space. A general path travel time model was analyzed, highlighting the fundamental difference between time-based and space-based sampling protocols. It was shown that space-based sampling corresponds to a likelihood function for the travel time over a given distance, while time-based sampling corresponds to a likelihood function for the travel distance over a given time interval. Thus, estimates of the travel time distribution parameters may be biased if the sampling protocol is not properly considered in the estimation.

The special case in which the path is partitioned into segments with uniform space-mean speed was considered, and explicit formulae for likelihood function for each of the two sampling principles were derived. The probability distribution of the allocation of travel time to the traversed segments was also derived under both sampling principles. Unlike space-based sampling, time-based sampling has a selection bias where a disproportionally large fraction of the observed time is spent on the last traversed segment. Using probe vehicle measurements sampled by time to directly estimate the travel time model will bias parameter estimates in two ways: First, all segment unit travel times are overestimated since low speed observations are overrepresented due to the finite path constraint. Second, the unit travel time of all but the last traversed segment will be overestimated while the last traversed segment will be underestimated. The combined effect of both mechanisms may depend on the length of the path, the characteristics of the segment unit travel time distributions and the probe vehicle sampling protocol.

The bias of the travel time model estimator under time-based sampling was investigated numerically for several different travel time distribution forms and parameter configurations. The results indicate that the bias in the mean and variance parameters is small if the unit travel time distributions are similar among the segments and concentrated around the means, but increases as the distributions become more skewed and asymmetric between segments. Further, the reference approach where each travel time observation is first split proportionally to the distance traversed distance on each segment in general tends to underestimate variance and performs worse than both the correct travel distance model and the incorrect travel time model, in particular for asymmetric parameter configurations.

Acknowledgements

The authors would like to thank two anonymous reviewers for helpful comments and suggestions on an early version of the paper. We further want to thank participants at the 17th International Conference of the Hong Kong Society for Transportation Studies, December 15–17 2012, Hong Kong, and the 2nd Symposium of the European Association for Research in Transportation (hEART), September 4–6 2013, Stockholm, Sweden, for valuable feedback on earlier versions of the paper. The work was funded by the Swedish Transport Authority through the Mobile Millennium Stockholm project, the Swedish Governmental Agency for Innovation Systems (Vinnova) through the iQFleet project, and

TRENoP Strategic Research Area.

References

Antoniou, C., Balakrishnan, R., Koutsopoulos, H. N., 2011. A synthesis of emerging data collection technologies and their impact on traffic management applications. European Transportation Research Review 3, 139–148.

Dandy, G. C., McBean, E. A., 1984. Variability of individual travel time components. ASCE Journal of Transportation Engineering 110, 340–357.

Haight, F. A., Mosher, W. H., 1962. A practical method for improving the accuracy of vehicular speed distribution measurements. Highway Research Record 341, 92–116.

Hellinga, B., Izadpanah, P., Takada, H., Fu, L., 2008. Decomposing travel times measured by probe-based traffic monitoring systems to individual road links. Transportation Research Part C 16, 768–782.

Hofleitner, A., Herring, R., Abbeel, P., Bayen, A., 2012. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. IEEE Transactions on Intelligent Transportation Systems 13, 1679–1693.

Hunter, T., Herring, R., Abbeel, P., Bayen, A., 2009. Path and travel time inference from GPS probe vehicle data. Presented at the Neural Information Processing Systems Foundation (NIPS) Conference, Vancouver, Canada.

Jenelius, E., Koutsopoulos, H. N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B 53, 64–81.

Leduc, G., 2008. Road traffic data: collection methods and applications. JRC Technical Notes, Working Papers on Energy, Transport and Climate Change, N.1.

Liu, K., Yamamoto, T., Morikawa, T., 2007. Comparison of time/space polling schemes for a probe vehicle system. Proceedings of the 14th World Conference on Intelligent Transport Systems, Beijing, China.

MATLAB, 2009. MATLAB version 7.8.0. Natick, Massachussetts: The MathWorks Inc.

Polus, A., 1979. A study of travel time and reliability on arterial routes. Transportation 8, 141–151.

Richardson, A. J., Taylor, M. A. P., 1978. Travel time variability on commuter journeys. High-Speed Ground Transportation 6, 77–99.

Sanaullah, I., Quddus, M., Enoch, M., 2013. Estimating link travel time from low frequency GPS data. Transportation Research Board 92nd Annual Meeting Compendium of Papers, #13-3909.

Uno, N., Kurauchi, F., Tamura, H., Iida, Y., 2009. Using bus probe data for analysis of travel time variability. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 13 (1), 2–15.

Westgate, B. S., Woodard, D. B., Matteson, D. S., Henderson, S. G., 2013. Travel time estimation for ambulances using Bayesian data augmentation. Annals of Applied Statistics 7 (2), 1139–1161.

Zheng, F., van Zuylen, H., 2013. Urban link travel time estimation based on sparse probe vehicle data. Transportation Research Part C 31, 145–157.