Network-wide on-line travel time estimation with inconsistent data from multiple sensor systems under network uncertainty

Hu Shao [a, b], William H.K. Lam [a, *], Agachai Sumalee [a, c] and Anthony Chen [a]

[a] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.*

[b] *School of Mathematics, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China.*

[c] *Department of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.*

Corresponding author
Prof. W.H.K. Lam,
Chair Professor of Civil & Transportation Engineering
Head
Department of Civil and Environmental Engineering
The Hong Kong Polytechnic University
Yuk Choi Road
Hung Hom, Kowloon
HONG KONG
Tel: +852-27666045
Fax: +852-23346389
E-mail address: william.lam@polyu.edu.hk

[*] Corresponding author. Tel.: +852-27666045; Fax: +852- 23346389.
E-mail address: william.lam@polyu.edu.hk

Network-wide on-line travel time estimation with inconsistent data from multiple sensor systems under network uncertainty

Hu Shao [a, b], William H.K. Lam [a, *], Agachai Sumalee [a, c] and Anthony Chen [a]

[a] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.*

[b] *School of Mathematics, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China.*

[c] *Department of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.*

This paper proposes a new modeling approach for network-wide on-line travel time estimation with inconsistent data from multiple sensor systems. With the development of Intelligent Transportation Systems (ITS), travel time can be estimated on some links or path segments in transportation networks by a variety of sensor systems. One limitation of most existing methods is that they focus only on estimating the travel times of some site-specific links or path segments through the use of observed data from one sensor system. Few attempts have been made to estimate network-wide travel times. To overcome this limitation, the proposed method makes full use of both the available data from multiple sensor systems (on-line data) and historical data (off-line data). The first- and second-order statistical properties of the on-line data are investigated together with the data inconsistency issue to estimate network-wide travel times. The proposed model is formulated as a generalized least squares (GLS) problem with non-linear constraints. A solution algorithm based on the penalty function method is adopted to solve the proposed model, whose application is illustrated by numerical examples using a local road network in Hong Kong.

Keywords: travel time estimation; Intelligent Transportation Systems; on-line data; off-line data; generalized least squares.

1. Introduction

1.1. Background

The travel time of each link (or path) in a territory-wide network provides a fundamental measure of system effectiveness in transportation networks (Vanajakshi, Williams, and Rilett 2009). In the Intelligent Transportation Systems (ITS) of transportation networks, travel time is one of the most commonly understood measures for road users, helping them to make informed decisions about travel choices to avoid unnecessary delay (Liu and Ma 2009). Therefore, the estimation of travel time is of increasing importance and has been extensively studied in the literature of transportation (Mori et al. 2015).

Conventionally, network-wide travel times can be estimated by traffic assignment models, as shown in Figure 1. This type of modeling approach requires both the specification of the origin-destination (O-D) matrix and the link performance function (e.g. the most widely used Bureau of Public Roads (BPR) function). With the aforementioned requirements, traffic

[*] Corresponding author. Tel.: +852-27666045; Fax: +852- 23346389.
E-mail address: william.lam@polyu.edu.hk

assignment models (e.g., the well-known user equilibrium (UE) and stochastic user equilibrium (SUE)) can calculate the resultant traffic flow and the travel time of each link (or path) in the territory-wide network. However, in reality, the actual O-D demand and the link performance function may be inaccurate or unavailable in a territory-wide network. It should be noted that BPR types link performance functions could not accurately measure realistic travel times when the link is very congested. As a result, the estimation of travel times in congested links could be problematic. In these circumstances, the traffic assignment-based travel time estimation approach may not be appropriate or applicable for estimating network-wide travel times.

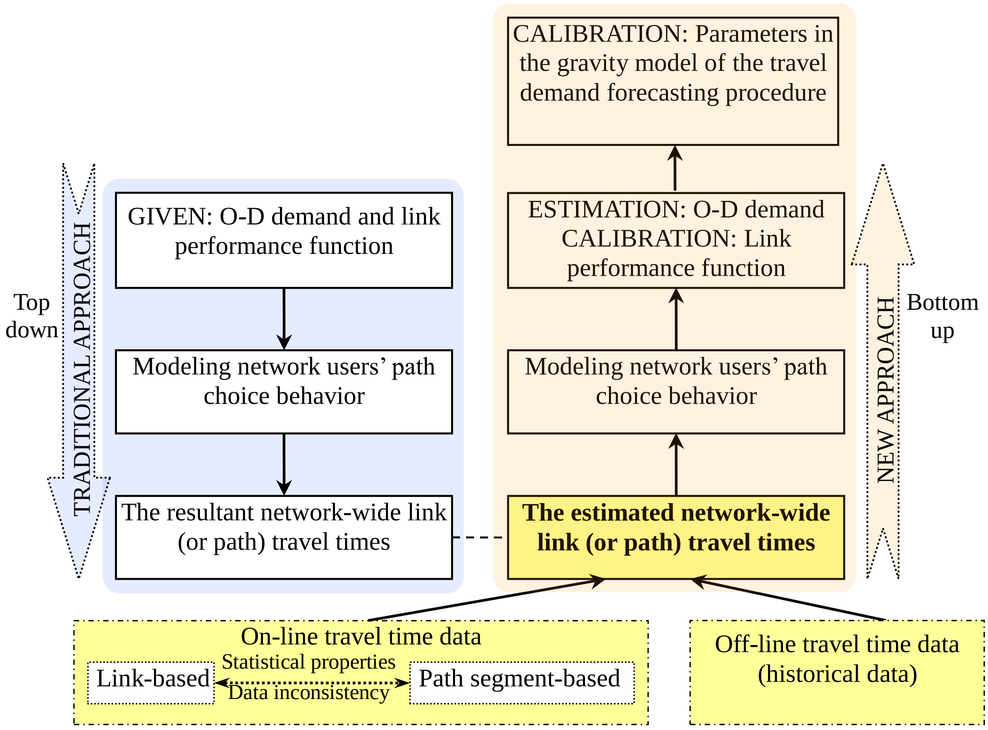


Figure 1. Basic motivation and framework for the proposed approach

Unlike the traditional approach, the proposed approach collects and observes travel times on links (or path segments) directly using a sensor system (e.g., loop detector, automatic vehicle identification (AVI) reader and Global Positioning System (GPS)-Geographical Information System (GIS)). For convenience, the term "sensor" is used hereafter to represent some specific devices that can collect the travel time data on links or path segments in the network. In reality, a network-wide travel time sensor system may not be realistic for practical application due to budgetary constraints and physical limitations (Hu, Peeta, and Chu 2009; Lam, Chan, and Shi 2002). Actually, sensor systems can only collect travel times on partial links or path segments in the network. Such travel time data are referred to as "on-line" travel time data. Meanwhile, historical travel time data (referred to as "off-line" travel time data) are usually available. Conditions such as these present an interesting and significant challenge, specifically how to make full use of on-line and off-line data to estimate network-wide travel times. This paper proposes a modeling approach that addresses the aforementioned problem by focusing on the inconsistency and statistical properties of the data, as shown in Figure 1. It should be pointed out that the top-down approach assumes some kind of user equilibrium situation (e.g. UE or SUE). The bottom-up approach (the proposed approach) does not need the equilibrium assumption. These two approaches assume different situations and they have different philosophy because bottom-up approach does not assume that the observed travel time data are generated from traffic equilibrium condition. Thus, the proposed approach is particularly suitable for the transportation network with disequilibrium flow patterns. For example, when transportation network is very congested, the equilibrium condition may not hold (e.g. UE or SUE). Thus, the top-down approach is inapplicable to estimate the network-wide travel time. Under such circumstance, the proposed method is investigated to address such kind of traffic situation.

Estimated network-wide travel times also have the potential for the estimation of the O-D demand and/or the calibration of the link performance function in addition to the network performance assessment. As Figure 1 illustrates, the traditional traffic assignment approach uses a "top-down" procedure that provides the O-D demand and link performance function, and the travel time can be estimated by carrying out the traffic assignment problem. This "top-down" mechanism could be adapted into a "bottom-up" procedure. Specifically, if the network-wide link travel time is known, then the O-D demand and/or link performance function can be adjusted to ensure that traffic assignment results can reproduce the observed network-wide travel times. Such an adjustment allows for the estimation of the O-D demand and/or the calibration of the link performance function. Similarly, the parameters in the gravity model of the travel demand forecasting procedure can be calibrated using the "bottom-up" procedure. This "bottom-up" method also necessitates the network-wide travel time estimation problem.

1.2. Literature review

Owing to its importance and wide applications, the travel time estimation problem has received a great deal of attention from transportation researchers. Previous works on the travel time estimation problem can be broadly divided into two types: link-based and path segment-based. In this paper, the path segment refers to a sequence of successive links in a path that may be a link or a complete path, as shown in Figure 2. For example, in the network, the travel time between two checkpoints can be observed by the sensor system (e.g., AVI or GPS-GIS), and this is called the path segment travel time. Following this definition, the arterial/road segment/route in some previous papers can be viewed as the path segment in this paper.

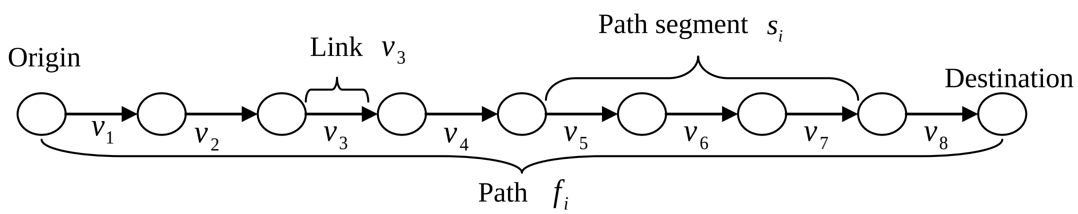


Figure 2. Illustration of path segment definition

- The link-based travel time estimation problem has been comprehensively studied in the literature. For example, Zhang (1999) studied the travel time estimation on urban arterial road links using loop detector data. California's Freeway Performance Measurement System (PeMS) is capable of collecting and analyzing real-time freeway traffic data and generating comprehensive performance measures, including freeway travel time (Chen 2002; Chen et al. 2001). Robinson and Polak (2005) utilized the k-nearest neighbor (k-NN) method for estimating the urban link travel time with inductive loop detector data. Guo and Jin (2006) proposed a method for estimating link travel time that used the cross-correlation analysis of traffic flow from single loop detector data. Yeon, Elefteriadou, and Lawphongpanich (2008) developed a model that can estimate travel time on a freeway using discrete time Markov chains, where the states correspond to whether or not the link is congested.

- With the development of electronic information and communication technologies, path segment-based travel time estimation methods have been investigated in transportation networks. Taylor, Woolley, and Zito (2000) developed a GPS-GIS system for collecting on-road traffic data, including path segment travel time from a probe vehicle. Using probe vehicle data, Lin, He, and Kornhauser (2008) developed the trip speed assignment technique to create speed distributions by time of day for many of the major roadway segments of the U.S. highway system. Wei and Lee (2007) presented an artificial neural network technique for path travel time estimation using observed data from GPS, vehicle detectors along the roadway, and the incident database. Furtlehner et al. (2007) proposed a belief propagation approach for real-time prediction of traffic conditions using probe vehicles. Tam and Lam (2008) estimated path segment travel time with the use of the AVI tag readers' information on time stamps. Liu and Ma (2009) proposed a virtual vehicle probe model for time-dependent travel time estimation on signalized arterial roads, and Kwong et al. (2009) proposed a practical system for the real-time estimation of travel time on arterial roads using data from wireless magnetic sensors. Xiong and Davis (2009) conducted a study on field evaluation of model-based estimation of arterial link travel times. Xia, Chen, and Huang (2011) proposed a multi-step method for corridor travel time prediction using loop detector data. Mazloumi et al. (2011) predicted bus travel time and its variability using an integrated framework that applied traffic flow data. Feng et al. (2014) used the information of GPS probes to augment less dynamic but available information estimating arterial travel times for real-time traffic monitoring.

However, most of the existing works were restricted to site-specific link or path segment travel time estimation problems by a variety of techniques. The existing methods were specifically developed for particular links (or path segments), which means they are inapplicable to the network-wide travel time estimation problem. The main difficulty inherent in network-wide travel time estimation is the limited data collection capabilities of real-world situations. It is unrealistic to hope that a territory-wide network will be completely covered by data collection sensors. Under such circumstances, there are two ways to overcome this difficulty without adding sensors.

The first is to fully utilize the observed travel time data from all kinds of available sensor systems for travel time estimation in the involved transportation network. Most of the previous models only take into account the observed data from one sensor system. Actually, different sensor systems for collecting travel time data (such as loop detector, GPS-GIS and AVI) may synchronously exist in a network. Different sensor systems are set up in the same network for different purposes. For example, AVI systems are usually used for electronic road pricing. GPS-GIS systems are widely used for route guidance. However, these sensor systems can also collect travel time data on the same or different links or path segments in the network during the same time period. The more observed data that can be obtained for use, the greater the accuracy and quality of the estimation. The observed data from multiple sensor systems may be inconsistent in value, however, which can make it difficult to use effectively in travel time estimation. For example, in Figure 2, the observed path segment travel time $\tilde{t}_{s_i}$ (e.g., from an AVI sensor system) should be theoretically equal to the summation of the observed link travel times $\tilde{t}_{v_5}$, $\tilde{t}_{v_6}$ and $\tilde{t}_{v_7}$ (e.g., from a loop detector system) as follows.

$$\tilde{t}_{s_i} = \tilde{t}_{v_5} + \tilde{t}_{v_6} + \tilde{t}_{v_7} . \quad (1)$$

Unfortunately, the foregoing equation does not usually hold in real-world situations due to the measurement errors of different sensor systems. Finding a way to handle such data inconsistency issues when using observed data from multiple sensor systems to solve network-wide travel time estimation problems constitutes a noteworthy and meaningful topic of investigation. Wang et al. (2014) investigated the citywide travel time estimation with GPS data with focus on addressing the issues of data sparsity and fragment of a path with trajectories. However, the travel time variations can not be estimated in their method.

The second way to address the limited data collection capabilities of real-world situations without adding additional sensors is to investigate the additional information from the observed data when managing a travel time estimation problem. In real-world situations, the observed travel times during a particular time period are not constant, varying day-to-day (Tam and Lam 2008; Uno et al. 2009). Thus, the observed travel time times are considered random variables (Guo and Jin 2006). Existing methods usually utilize the mean value of the observed data, which relates to its first-order statistical properties. Actually, the second-order statistical property of the data can offer some additional information, particularly for the spatial correlation of random variables. For example, the covariance between link travel times is an important indicator of how one link travel time varies with another. Such information is helpful, but less frequently considered in travel time estimation problems. In fact, covariance has already been used by Chandra and Al-Deek (2009), Dailey (1993), and Guo and Jin (2006), specifically for the cross-correlation analysis of traffic flow in a temporal manner for link travel time estimation. However, these authors' work focused on upstream and downstream traffic flow and did not consider network-wide travel time estimation. Lam et al. (2002), in contrast, investigated the covariance of link travel time in a spatial manner to estimate the network-wide travel times of links without sensors. However, their method adopted the traffic assignment-based method. The accuracy of such methods depends on the O-D matrix and the link travel performance function, which may not be available for a territory-wide network, as previously discussed.

1.3. Problem statement and contribution

This paper addresses the previously discussed travel time estimation problem by proposing a new model for estimating the on-line travel time and the travel time covariance for links in a territory-wide network with the use of on-line and off-line travel time data. The on-line data refer to the observed travel times from multiple sensor systems, including link-based and path segment-based observed travel times. The off-line data are the historical link travel time data, which are assumed to be known. Three critical issues are explicitly considered in the proposed model: (a) the data inconsistency of observed data from different sensor systems; (b) the first- and second-order statistical properties of the observed data; and (c) the relationship between observed and unobserved link travel times. The observed and unobserved link travel times refer to travel times on the links with sensors, and those without sensors, respectively. The proposed model is formulated as a generalized least squares (GLS) problem with non-linear constraints. A penalty function method is adopted to solve the proposed model. To the best of our knowledge, the network-wide travel time estimation problem has received little attention in the literature, with the notable exception of Lam et al. (2002), whose study employed the traffic assignment-based approach to estimate network-wide travel. In view of the aforementioned limitations, our study extends the previous work in this arena by making the following two contributions.

First, the proposed modeling approach is more applicable for practical use than conventional methods because it can estimate network-wide travel times without knowledge of O-D demand and the link performance function. This feature renders the proposed model applicable for real-world implementation in network performance assessment. More importantly, the estimated network-wide link travel time can be used to estimate the O-D traffic demand and/or calibrate the link performance function as input data for the traffic assignment problem and the parameter calibration problem for the gravity model. Thus, the proposed model has a useful implication in terms of those estimation and/or calibration problems found in transportation networks.

Second, the proposed modeling approach has the potential to obtain a highly accurate, quality estimation of network-wide travel times. One reason is that all the available data, including observed data from multiple sensor systems and historical data, are simultaneously considered. The data inconsistency presented by multiple sensor systems is also investigated. The proposed model further utilizes second-order statistical property, such as covariance, which can identify correlations between observed and unobserved link travel times (Lam et al. 2002). These correlations are particularly useful when inferring the unobserved link travel times in the territory-wide network.

The remainder of this paper is organized as follows. The next section formulates the network-wide travel time estimation problem as a GLS problem with non-linear constraints. Then, a solution algorithm based on the penalty function method is proposed. Numerical examples are provided to demonstrate the application characteristics of the proposed model. Finally, some conclusions and possible directions for future research are given.

2. Model formulation

2.1. Notations

The notations used throughout the paper are defined as follows unless otherwise specified. For notational consistency, capital letters in italics are used to denote random variables, and lower-case letters in italics are used to denote deterministic variables throughout the paper.

T_{v_i}	Travel time on link v_i.
t_{v_i}	Mean travel time on link v_i.
$\mathbf{T}_d^v$	Vector of observed link travel times, $\mathbf{T}_d^v = \{T_{v_1}, T_{v_2}, \cdots, T_{v_d}\}^T$.
$\tilde{\mathbf{t}}_d^v$	Sample mean of observed link travel times, $\tilde{\mathbf{t}}_d^v = E[\mathbf{T}_d^v] = \{\tilde{t}_{v_1}, \tilde{t}_{v_2}, \cdots, \tilde{t}_{v_d}\}^T$.
$\mathbf{T}_e^v$	Vector of unobserved link travel times, $\mathbf{T}_e^v = \{T_{v_d+1}, T_{v_d+2}, \cdots, T_{v_n}\}^T$.
$\mathbf{T}^v$	Vector of link travel times, $\mathbf{T}^v = \{T_{v_1}, T_{v_2}, \cdots, T_{v_n}\}^T = \{\mathbf{T}_d^v, \mathbf{T}_e^v\}^T$.
$\mathbf{t}^v$	Vector of mean link travel times, $\mathbf{t}^v = E[\mathbf{T}^v] = \{\mathbf{t}_d^v, \mathbf{t}_e^v\}^T$.
σ_{v_i,v_j}	Covariance of link travel times T_{v_i} and T_{v_j}, $\sigma_{v_i,v_j} = \mathrm{cov}(T_{v_i}, T_{v_j})$.
$\mathbf{\Sigma}^v$	Covariance matrix of link travel times, $\left(\mathbf{\Sigma}^v\right)_{n\times n} = \left\{\sigma_{v_i,v_j}\right\}_{n\times n}$.
$\tilde{\mathbf{\Sigma}}_d^v$	Sample covariance matrix of observed link travel times, $\left(\tilde{\mathbf{\Sigma}}_d^v\right)_{d\times d} = \left\{\tilde{\sigma}_{v_i,v_j}\right\}_{d\times d}$.
$\bar{\mathbf{t}}^v$	Vector of historical mean link travel times, $\bar{\mathbf{t}}^v = \{\bar{\mathbf{t}}_d^v, \bar{\mathbf{t}}_e^v\}^T$.

T_{f_i}	Travel time on path f_i.
t_{f_i}	Mean travel time on path f_i.
$\mathbf{T}^f$	Vector of path travel times, $\mathbf{T}^f = \{T_{f_1}, T_{f_2}, \cdots, T_{f_m}\}^T$.
$\mathbf{t}^f$	Vector of mean path travel times, $\mathbf{t}^f = E[\mathbf{T}^f] = \{t_{f_1}, t_{f_2}, \cdots, t_{f_m}\}^T$.
σ_{f_i, f_j}	Covariance of path travel times T_{f_i} and T_{f_j}, $\sigma_{f_i, f_j} = \mathrm{cov}(T_{f_i}, T_{f_j})$.
$\mathbf{\Sigma}^f$	Covariance matrix of path travel times, $\left(\mathbf{\Sigma}^f\right)_{m\times m} = \left\{\sigma_{f_i, f_j}\right\}_{m\times m}$.
T_{s_i}	Travel time on path segment s_i.
t_{s_i}	Mean travel time on path segment s_i.
$\mathbf{T}_s$	Vector of path segment travel times, $\mathbf{T}^s = \{T_{s_1}, T_{s_2}, \cdots, T_{s_q}\}^T$.
$\mathbf{t}_s$	Vector of mean path segment travel times, $\mathbf{t}_s = E[\mathbf{T}_s] = \{t_{s_1}, t_{s_2}, \cdots, t_{s_q}\}^T$.
σ_{s_i, s_j}	Covariance of path segment travel times T_{s_i} and T_{s_j}, $\sigma_{s_i, s_j} = \mathrm{cov}(T_{s_i}, T_{s_j})$.
$\mathbf{\Sigma}_s$	Covariance matrix of path segment travel times, $\left(\mathbf{\Sigma}_s\right)_{q\times q} = \left\{\sigma_{s_i, s_j}\right\}_{q\times q}$.
$\tilde{\mathbf{t}}^s$	Sample mean of observed path segment travel times.
$\tilde{\mathbf{\Sigma}}_s$	Sample covariance matrix of observed path segment travel time.

2.2. *Observed path segment- and link-based travel times*

The travel time on path segment s_i during a particular time period (e.g., morning peak) can be observed by the sensor system (e.g., AVI or GPS-GIS). Suppose that the travel time can be observed over l homogeneous and independent time periods. For instance, these observed periods could be 8:00 am-10:00 am on a sequence of weekdays (not necessarily consecutive). As discussed previously, the travel time on a path segment during the observed time period varies from day to day. $\tilde{T}^{s_i}$ is denoted as the observed travel time on path segment s_i. Thus, $\tilde{t}_{s_i(j)}$ is defined as the observed travel time on path segment s_i during the concerned time period on day j $(j = 1,2,\cdots,l)$. Then, $\tilde{t}_{s_i(1)}, \tilde{t}_{s_i(2)}, \cdots, \tilde{t}_{s_i(l)}$ is a sample of $\tilde{T}_{s_i}$ with sample size l. Denote $\tilde{\mathbf{t}}_{s(j)}$ as the vector of $\{\tilde{t}_{s_1(j)}, \tilde{t}_{s_2(j)}, \cdots, \tilde{t}_{s_q(j)}\}^T$. Then, the sample mean of the observed path segment travel time can be estimated as

$$\tilde{\mathbf{t}}_s = \frac{1}{l}\sum_{j=1}^{l} \tilde{\mathbf{t}}_{s(j)}. \tag{2}$$

The sample covariance matrix (unbiased estimator) of the observed path segment travel times can be estimated as

$$\tilde{\mathbf{\Sigma}}_s = \frac{1}{l}\sum_{j=1}^{l} \left\{ \left(\tilde{\mathbf{t}}_{s(j)} - \tilde{\mathbf{t}}_s\right)\left(\tilde{\mathbf{t}}_{s(j)} - \tilde{\mathbf{t}}_s\right)^T \right\}. \tag{3}$$

Similar to Equations (2) and (3), the sample mean $\tilde{\mathbf{t}}_d^v$ and sample covariance matrix $\tilde{\mathbf{\Sigma}}_d^v$ of the observed link travel times can also be calculated.

2.3. *Relationship between link and path segment travel times*

Denote $(\Delta)_{n\times m}$ as the link-path incident matrix. It follows that

$$\mathbf{T}^f = \Delta^T \mathbf{T}^v, \tag{4}$$

$$\mathbf{t}^f = \Delta^T \mathbf{t}^v \text{, and} \tag{5}$$

$$\mathbf{\Sigma}^f = \Delta^T \mathbf{\Sigma}^v \Delta . \tag{6}$$

Similarly, denote $(\Delta_s)_{n\times q}$ as the link-path segment incident matrix. It then follows that

$$\mathbf{T}_s = {\Delta_s}^T \mathbf{T}^v , \tag{7}$$

$$\mathbf{t}^s = {\Delta_s}^T \mathbf{t}^v \text{, and} \tag{8}$$

$$\mathbf{\Sigma}_s = {\Delta_s}^T \mathbf{\Sigma}^v \Delta_s . \tag{9}$$

2.4. Estimation of unobserved link travel times

Realistically, sensors are difficult to install on all of the links within a territory-wide network due to budget constraints and physical limitations. The order of the links, however, can be rearranged so that the observed links (links with sensors) appear first. Then, the link travel times and the covariance matrix can be partitioned according to whether or not the links are installed with sensors, as follows.

$$\mathbf{T}^v = \{T_{v_1}, T_{v_2}, \cdots, T_{v_n}\}^T = \{T_{v_1}, T_{v_2}, \cdots, T_{v_d} \;\vdots\; T_{v_d+1}, T_{v_d+2}, \cdots, T_{v_n}\}^T = \{\mathbf{T}_d^v, \mathbf{T}_e^v\}^T . \tag{10}$$

$$\mathbf{t}^v = E[\mathbf{T}^v] = \{t_{v_1}, t_{v_2}, \cdots, t_{v_d}, \;\vdots\; t_{v_d+1}, t_{v_d+2}, \cdots, t_{v_n}\}^T = \{\mathbf{t}_d^v, \mathbf{t}_e^v\}^T . \tag{11}$$

$$\mathbf{\Sigma}^v = \left[\begin{array}{ccc|ccc} \sigma_{v_1,v_1} & \cdots & \sigma_{v_1,v_d} & \sigma_{v_1,v_d+1} & \cdots & \sigma_{v_1,v_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{v_d,v_1} & \cdots & \sigma_{v_d,v_d} & \sigma_{v_d,v_d+1} & \cdots & \sigma_{v_d,v_n} \\ \hline \sigma_{v_d+1,v_1} & \cdots & \sigma_{v_d+1,v_d} & \sigma_{v_d+1,v_d+1} & \cdots & \sigma_{v_d+1,v_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{v_n,v_1} & \cdots & \sigma_{v_n,v_d} & \sigma_{v_n,v_d+1} & \cdots & \sigma_{v_n,v_n} \end{array}\right] = \begin{bmatrix} \mathbf{\Sigma}_d^v & \mathbf{\Sigma}_{d,e}^v \\ \mathbf{\Sigma}_{e,d}^v & \mathbf{\Sigma}_e^v \end{bmatrix} . \tag{12}$$

According to Lam et al. (2002), if the link travel time covariance matrix $\mathbf{\Sigma}^v$ is known, then the unobserved link travel times can be calculated as follows.

$$\mathbf{t}_e^v = \bar{\mathbf{t}}_e^v + \mathbf{\Sigma}_{e,d}^v \left(\mathbf{\Sigma}_d^v\right)^{-1} \left(\mathbf{t}_d^v - \bar{\mathbf{t}}_d^v\right), \tag{13}$$

where $\mathbf{t}_e^v$ is the vector of estimated travel times on unobserved links (links without sensors); $\mathbf{\Sigma}_{e,d}^v$ is a sub-matrix of $\mathbf{\Sigma}^v$; $\mathbf{\Sigma}_d^v$ is the covariance matrix of detected link travel times; and $\mathbf{\Sigma}_e^v$ is the covariance matrix of unobserved link travel times as shown in Equation (12).

Equation (13) reveals that the mean unobserved link travel times can be estimated on the basis of observed link travel times, historical unobserved link travel times, and the link travel time covariance matrix. This equation also implies that the covariance information between link travel times has significant influence on the accuracy of the travel time estimation (Lam et al. 2002).

2.5. GLS model for network-wide travel time estimation

To overcome the aforementioned data inconsistency issue, the GLS method is adopted to achieve a relatively accurate estimation of the mean link travel times and link travel time covariance matrix, as follows.

$$\min_{\mathbf{t}^v,\boldsymbol{\Sigma}^v}\left\{\omega_1\left\|\tilde{\mathbf{t}}_d^v-\mathbf{t}_d^v\right\|^2+\omega_2\left\|\tilde{\mathbf{t}}_s-\Delta_s{}^T\mathbf{t}^v\right\|^2+\omega_3\left\|\tilde{\boldsymbol{\Sigma}}_s-\Delta_s{}^T\boldsymbol{\Sigma}^v\Delta_s\right\|_F^2+\omega_4\left\|\tilde{\boldsymbol{\Sigma}}_d^v-\boldsymbol{\Sigma}_d^v\right\|_F^2\right\} \quad (14a)$$

$$\text{s.t. } \mathbf{t}_e^v=\bar{\mathbf{t}}_e^v+\boldsymbol{\Sigma}_{e,d}^v\left(\boldsymbol{\Sigma}_d^v\right)^{-1}\left(\mathbf{t}_d^v-\bar{\mathbf{t}}_d^v\right) \quad (14b)$$

$$\mathbf{t}^v=\{\mathbf{t}_d^v,\mathbf{t}_e^v\}^T\geq 0 \quad (14c)$$

$$\boldsymbol{\Sigma}^v\succeq 0, \quad (14d)$$

where $\|\cdot\|$ denotes the Euclidean norm for vectors; $\|\cdot\|_F$ is the Frobenius norm (e.g., $\left\|\boldsymbol{\Sigma}^v\right\|_F=\sqrt{\sum_{i=1}^n\sum_{j=1}^n\left(\sigma_{v_i,v_j}\right)^2}$) for the matrix; $\boldsymbol{\Sigma}^v\succeq 0$ represents the constraint that $\boldsymbol{\Sigma}^v$ should be a symmetric and positive semi-definite matrix ($\boldsymbol{\Sigma}^v$ is a symmetric and positive semi-definite matrix if and only if $\sigma_{v_i,v_j}=\sigma_{v_j,v_i}$ and all eigenvalues of $\boldsymbol{\Sigma}^v$ are non-negative); and ω_1, ω_2, ω_3, and ω_4 are the non-negative weighting parameters.

There are two kind of terms in the objective function (14a), which correspond to loop detector and AVI sensor systems. This objective function can be interpreted as a linear combination of two objectives corresponding to the two sensor systems. In other words, this problem can be regarded as a bi-objective optimization problem. The values of weighting parameters ω_1, ω_2, ω_3 and ω_4 quantify the trade-offs between the two kind of sensor data. In this paper, it is assumed that link travel time can be observed accurately by the detector, whereas travel time of ‘path segment’ can be observed accurately by the AVI. In reality, both kind of sensor systems have observed errors (or measurement errors). To present the essential idea of this paper, these observed errors are ignored in this paper. For real-world application of this model, the choice of value of ω_1, ω_2, ω_3 and ω_4 can be dependent on accuracy and sample size of the two sensor systems. How to investigate the Pareto efficiency (Tan et al., 2014) and choose the value of weighting parameters reveals significant extensions of this paper.

In this formulation, the mean link travel time $\mathbf{t}^v$ and the link travel time covariance matrix $\boldsymbol{\Sigma}^v$ are the decision variables. It should be noted that $\boldsymbol{\Sigma}^v$ is a $n\times n$ symmetric matrix, and only $\frac{n(n+1)}{2}$ entries in this matrix are the decision variables. The objective function is designed to obtain $\mathbf{t}^v$ and $\boldsymbol{\Sigma}^v$, which are expected to approximately reproduce the observed data from multiple sensor systems as much as possible. For this purpose, the first two items in the objective function represent the difference between the observed and estimated mean travel times in terms of the links and path segments, respectively. The last two items refer to the difference between the observed and estimated covariance matrix in terms of the links and path segments, respectively. It is known that observed data with a larger sample size should represent the real situation more reliably than those with a smaller size. Thus, the observed data with different sample sizes are treated differently by attaching a non-negative weighting parameter. For example, the value of weighting parameter ω_i corresponds to the sample size of the observed data. For instance, if the sample size of $\tilde{\mathbf{t}}_d^v$ is 100 and that of $\tilde{\mathbf{t}}_s$ is 50, then ω_1 and ω_2 can be said to be 100 and 50, respectively.

The decision variables should also satisfy certain constraints. The first constraint (14b) is a non-linear equality constraint, which represents the relationship between observed and unobserved link travel times. The second linear inequality constraint (14c) requires the estimated mean travel times to be non-negative. The third non-linear constraint (14d) restricts the decision variable $\mathbf{\Sigma}^v$ to be a symmetric and positive semi-definite matrix.

After obtaining the mean and covariance of link travel times from the GLS model (14), the mean path travel times and the path travel time covariance matrix can be calculated according to Equations (5) and (6), respectively.

3. Solution Algorithm

It should be noted that the GLS model (14) falls within the category of the non-linear constrained optimization problem. To the best of our knowledge, there are no universal and efficient solution algorithms that can directly solve this problem. The main difficulty in solving such problems arises from the non-linear constraints. In the optimization literature, the constrained optimization problem is usually transformed into an unconstrained one, which is much easier to solve.

The penalty method is one of the most widely used approaches for transforming a constrained optimization problem into an unconstrained one. The penalty method replaces a constrained optimization problem with a series of unconstrained problems whose solutions ideally converge to deliver a solution to the original constrained problem. The unconstrained problems are formed by adding a penalty term to the objective function, which consists of a penalty parameter and a measure of constraint violation. The measure of violation is nonzero when the constraints are violated, and zero in the regions in which they are not.

In this paper, the exterior penalty method is adopted due to its simplicity and efficiency. Regarding the three constraints in the GLS model (14), the addition of the following three penalty terms to the objective function is proposed.

$$\mathbf{t}_e^v = \bar{\mathbf{t}}_e^v + \mathbf{\Sigma}_{e,d}^v \left(\mathbf{\Sigma}_d^v\right)^{-1}\left(\mathbf{t}_d^v - \bar{\mathbf{t}}_d^v\right) \Rightarrow p_1^{(k)}\left\|\mathbf{t}_e^v - \left(\bar{\mathbf{t}}_e^v + \mathbf{\Sigma}_{e,d}^v\left(\mathbf{\Sigma}_d^v\right)^{-1}\left(\mathbf{t}_d^v - \bar{\mathbf{t}}_d^v\right)\right)\right\|^2 \tag{15}$$

$$\mathbf{t}^v \geq 0 \Rightarrow p_2^{(k)}\left\|\max\left(-\mathbf{t}^v, 0\right)\right\|^2 \tag{16}$$

$$\mathbf{\Sigma}^v \succeq 0 \Rightarrow p_3^{(k)}\left(\max\left(-\lambda_{\min}, 0\right)\right)^2, \tag{17}$$

where $p_1^{(k)}$, $p_2^{(k)}$, and $p_3^{(k)}$ are positive penalty coefficients at iteration k and $\lambda_{\min}$ is the minimal eigenvalue of $\mathbf{\Sigma}^v$. Then, the penalized objective function for the GLS problem (14) can be expressed as follows.

$$\begin{aligned}\min_{\mathbf{t}^v,\mathbf{\Sigma}^v}\Big\{&\omega_1\left\|\tilde{\mathbf{t}}_d^v - \mathbf{t}_d^v\right\|^2 + \omega_2\left\|\tilde{\mathbf{t}}_s - \Delta_s{}^T\mathbf{t}^v\right\|^2 + \omega_3\left\|\tilde{\mathbf{\Sigma}}_s - \Delta_s{}^T\mathbf{\Sigma}^v\Delta_s\right\|^2 + \omega_4\left\|\tilde{\mathbf{\Sigma}}_d^v - \mathbf{\Sigma}_d^v\right\|^2 \\ &+ p_1^{(k)}\left\|\mathbf{t}_e^v - \left(\bar{\mathbf{t}}_e^v + \mathbf{\Sigma}_{e,d}^v\left(\mathbf{\Sigma}_d^v\right)^{-1}\left(\mathbf{t}_d^v - \bar{\mathbf{t}}_d^v\right)\right)\right\|^2 + p_2^{(k)}\left\|\max\left(-\mathbf{t}^v,0\right)\right\|^2 + p_3^{(k)}\left(\max\left(-\lambda_{\min},0\right)\right)^2\Big\}\end{aligned} \tag{18}$$

For convenience, Equation (18) is re-written as follows.

$$\min_{\mathbf{t}^v,\mathbf{\Sigma}^v}\left\{g(\mathbf{t}^v,\mathbf{\Sigma}^v) + p\left(\mathbf{t}^v,\mathbf{\Sigma}^v,p_1^{(k)},p_2^{(k)},p_3^{(k)}\right)\right\}, \tag{19}$$

where

$$g(\mathbf{t}^v,\mathbf{\Sigma}^v) = \omega_1\left\|\tilde{\mathbf{t}}_d^v - \mathbf{t}_d^v\right\|^2 + \omega_2\left\|\tilde{\mathbf{t}}^s - \Delta_s{}^T\mathbf{t}^v\right\|^2 + \omega_3\left\|\tilde{\mathbf{\Sigma}}_s - \Delta_s{}^T\mathbf{\Sigma}^v\Delta_s\right\|^2 + \omega_4\left\|\tilde{\mathbf{\Sigma}}_d^v - \mathbf{\Sigma}_d^v\right\|^2 \tag{20}$$

$$p\left(\mathbf{t}^v, \mathbf{\Sigma}^v, p_1^{(k)}, p_2^{(k)}, p_3^{(k)}\right)$$
$$= p_1^{(k)} \left\| \mathbf{t}_e^v - \left(\bar{\mathbf{t}}_e^v + \mathbf{\Sigma}_{e,d}^v \left(\mathbf{\Sigma}_d^v\right)^{-1} \left(\mathbf{t}_d^v - \bar{\mathbf{t}}_d^v\right)\right) \right\|^2 + p_2^{(k)} \left\| \max\left(-\mathbf{t}^v, 0\right) \right\|^2 + p_3^{(k)} \left(\max\left(-\lambda_{\min}, 0\right)\right)^2. \quad (21)$$

According to objective function (19), the solution to the GLS problem (14) can be described by the following procedure.

Step 1: Set the initial values: $\mathbf{t}^{v(0)}$, $\mathbf{\Sigma}^{v(0)}$, $p_1^{(0)}$, $p_2^{(0)}$, $p_3^{(0)}$, $\xi > 1$ (enlarge parameter), $\tau > 0$ (stopping tolerance), and $k = 0$ (iteration number).

Step 2: Solve the unconstrained optimization problem (19) with the starting point $\mathbf{t}^{v(k)}$ and $\mathbf{\Sigma}^{v(k)}$:

$$(\mathbf{t}^{v(k+1)}, \mathbf{\Sigma}^{v(k+1)}) = \arg\min_{\mathbf{t}^v, \mathbf{\Sigma}^v} \left\{ g(\mathbf{t}^v, \mathbf{\Sigma}^v) + p\left(\mathbf{t}^v, \mathbf{\Sigma}^v, p_1^{(k)}, p_2^{(k)}, p_3^{(k)}\right) \right\}.$$

Step 3: If the penalty term is sufficiently small , i.e., $p(\mathbf{t}^v, \mathbf{\Sigma}^v, p_1^{(k)}, p_2^{(k)}, p_3^{(k)}) \le \tau$, then stop. Otherwise, set $p_1^{(k+1)} = \xi p_1^{(k)}$, $p_2^{(k+1)} = \xi p_2^{(k)}$, $p_3^{(k+1)} = \xi p_3^{(k)}$, and $k = k + 1$, and then go to Step 2.

It should be noted that the gradient of the objective function is difficult to obtain due to the penalty term's correspondence with the non-linear inequality constraint (Equation (17)). Therefore, some derivative-free optimization methods should be employed in Step 2, such as the simplex search method (Lagarias et al. 1998) and generalized pattern search methods (Audet and Dennis 2003; Torczon 1997). Details of the derivative-free optimization methods can be gathered from Conn, Scheinberg, and Vicente (2009). In this paper, the simplex search method (Lagarias et al. 1998) is used to solve the unconstrained optimization problem in Step 2, which is available in the Matlab optimization toolbox under the function "fminsearch."

The convergence of the proposed solution algorithm can be guaranteed by the following proposition.

Proposition: Suppose that each $\mathbf{t}^{v(k+1)}, \mathbf{\Sigma}^{v(k+1)}$ is the exact global minimizer of Equation (19); then, every limit point $\mathbf{t}^{v(*)}, \mathbf{\Sigma}^{v(*)}$ of the sequence $\{\mathbf{t}^{v(k)}, \mathbf{\Sigma}^{v(k)}\}$ is a global solution of the problem (14) (Theorem 17.1, Page 502, in Nocedal and Wright (2006)).

Remark: According to this proposition, it should be noted that the convergence to the global solution of the Equation (14) can be guaranteed if the global minimizer of Equation (19) can be found. However, global minimizer of Equation (19) may not be easily obtained. This is a non-linear optimization problem and the convexity property of this problem can not be guaranteed. Further investigation could be conduct to design efficient solution algorithm for solving this model.

4. Numerical Examples

The numerical examples presented in this section illustrate: (a) the effects of using observed data from multiple sensor systems; (b) the effects of using second-order statistical property; and (c) the effects of the number and location of sensors. A number of sensitivity tests and investigations are carried out for these purposes.

4.1. Preliminary

Hong Kong's Tuen Mun Road Corridor Network is used as the study network to demonstrate the performance of the proposed link travel time estimation model. This network consists of 3 zones, 4 nodes, 10 links, and 20 paths, as shown in Figure 3. The simulation method is used to generate the observed link travel time and path segment travel time. The number of iteration is set at 50, which represents an observed data sample size of 50. In other words, the travel time is observed over 50 homogeneous, independent workday time periods. A standard logit-based stochastic user equilibrium (SUE) traffic assignment model is adopted for the simulation, with a dispersion parameter of $\theta = 100$.

To capture the effects of demand and supply uncertainties on the simulation procedure, the stochastic link travel times are generated from the Bureau of Public Roads (BPR) type formula:

$$T_{v_a} = t_{v_a}^0 + \beta_a \left(\frac{V_a}{C_a} \right)^{\alpha_a}, \tag{22}$$

where $t_{v_a}^0$ is the free flow travel time on link a; V_a is the stochastic traffic flow on link a; C_a is the stochastic link capacity on link a, and $E[C_a] = c_a$; and α_a and β_a are parameters. The characteristics for the 10 links in the study network are given in Table 1, in which "pcu/hr" represents "passenger car unit/hour."

In each simulation, the link capacity is generated from a continuous uniform distribution, as shown in Table 1, and the O-D demand is generated from a normal distribution, as shown in Table 2. The historical mean link travel times, shown in Table 3, are generated from the standard SUE model, assuming that the O-D demand and the link capacity take the mean values.

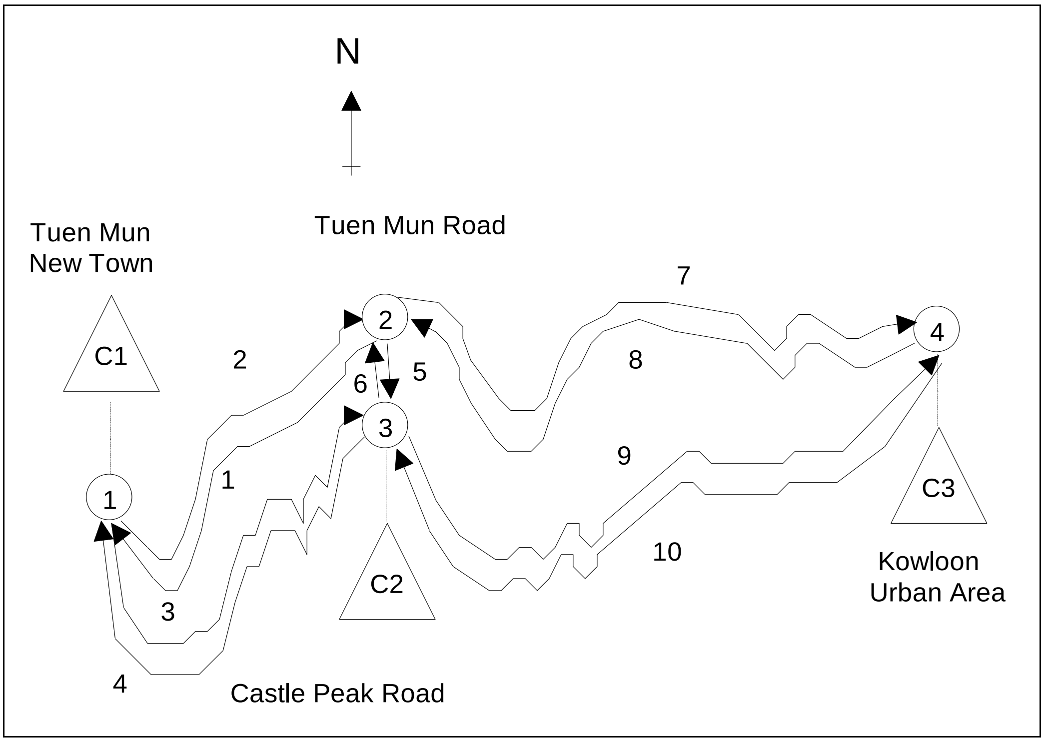


Figure 3. Tuen Mun Road Corridor Network

Table 1. Input Data of Link Travel Time Function

Link no.	$t^0_{v_a}$ (hour)	$C_a \sim U(c_{a-}, c_{a+})$		c_a (pcu/hr)	Parameters	
		c_{a-} (pcu/hr)	c_{a+} (pcu/hr)		α_a	β_a
1,2	0.0900	3105	5175	4140	3.5	0.1050
3,4	0.1106	510	850	680	3.6	0.1408
5,6	0.0056	690	1150	920	3.6	0.0071
7,8	0.0335	2880	4800	3840	3.6	0.0335
9,10	0.0767	600	1000	800	3.6	0.1073

Table 2. O-D Traffic Demand

O-D pair	Normal distributed traffic demand	
	Mean (pcu/hour)	Standard Deviation (pcu/hour)
C1 $\rightarrow$ C2	525	52.5
C1 $\rightarrow$ C3	5336	533.6
C2 $\rightarrow$ C1	50	5
C2 $\rightarrow$ C3	227	22.7
C3 $\rightarrow$ C1	3776	377.6
C3 $\rightarrow$ C2	97	9.7

This example assumes that the link travel times are observed by loop detectors and the path segment travel times are observed by AVI readers. The standard SUE model was solved 100 times. Each time, the link capacity and O-D demand were generated according to their probability distribution assumptions. Then, the corresponding SUE results were recorded to calculate the sample mean and covariance. The observed sample mean and covariance matrix of the link travel times were generated from the first 50 simulations, and are shown in Tables 3 and 4, respectively. It is assumed that three path segment travel times can be observed: s_1 (Link 3), s_2 (Links 2 and 5), and s_3 (Links 2, 7, and 10). The observed mean and covariance matrix of the path segment travel times were generated from the last 50 simulations, and are shown in Table 5. Because the link-based observed data and path segment-based observed data are generated from different simulation procedures, inconsistency occurs between these two types of data. For example, the mean link travel time on link 3 collected by loop detectors (0.3130 hour, link 3 in Table 3) is different from that collected by AVI readers (0.3450, path segment 1 in Table 5).

Table 3. Historical and Observed Mean Link Travel Times

Link no.	Historical mean link travel time (hour)	Observed mean link travel time (hour)
1	0.1399	0.1419
2	0.2742	0.2932
3	0.2897	**0.3103**
4	0.1506	0.1517
5	0.0118	0.0133
6	0.0086	0.0083
7	0.1146	0.1478
8	0.0590	0.0600
9	0.1285	0.1601
10	0.0804	0.0837

Table 4. Observed Link Travel Time Covariance Matrix (hour^2)

Link no.	1	2	3	4	5	6	7	8	9	10
1	0.001310	-0.000139	-0.000104	0.001244	0.000055	0.000002	0.000087	0.000194	0.000043	0.000176
2	-0.000139	0.014585	0.014444	-0.000239	0.000084	0.000063	0.003459	-0.000023	0.003427	-0.000044
3	-0.000104	0.014444	0.014324	-0.000219	0.000097	0.000053	0.003554	-0.000049	0.003502	-0.000053
4	0.001244	-0.000239	-0.000219	0.001205	0.000036	0.000010	-0.000052	0.000233	-0.000077	0.000192
5	0.000055	0.000084	0.000097	0.000036	0.000016	-0.000006	0.000142	-0.000018	0.000124	-0.000005
6	0.000002	0.000063	0.000053	0.000010	-0.000006	0.000007	-0.000026	0.000028	-0.000016	0.000014
7	0.000087	0.003459	0.003554	-0.000052	0.000142	-0.000026	0.004436	-0.000072	0.004206	-0.000018
8	0.000194	-0.000023	-0.000049	0.000233	-0.000018	0.000028	-0.000072	0.000244	-0.000050	0.000154
9	0.000043	0.003427	0.003502	-0.000077	0.000124	-0.000016	0.004206	-0.000050	0.004006	-0.000013
10	0.000176	-0.000044	-0.000053	0.000192	-0.000005	0.000014	-0.000018	0.000154	-0.000013	0.000107

Table 5. Mean and Covariance Matrix of Observed Path Segment Travel Times

Path segment no.	Mean (hour)	Path segment no.	Covariance (hour^2)		
			1	2	3
1	**0.3450**	1	0.03187	0.03213	0.03616
2	0.3421	2	0.03213	0.03240	0.03647
3	0.5563	3	0.03616	0.03647	0.04423

Due to the inconsistency of the observed data provided by different sensor systems, they may be inconsistent. Hence, the conventional measures (e.g., root mean square error) for quantifying estimate quality are not applicable in this example. The concept of confidence intervals is used as an alternative to examine reliability of the estimate. For a given confidence level $1-\eta$, the confidence interval for the mean and variance of the link travel times can be calculated as follows.

$$\left(\tilde{t}_{v_i} - \tilde{\sigma}_{v_i,v_i} t_{\eta/2}(l-1),\ \tilde{t}_{v_i} + \tilde{\sigma}_{v_i,v_i} t_{\eta/2}(l-1)\right) \quad i = 1,2,\cdots,n \tag{23}$$

$$\left(\frac{(l-1)\tilde{\sigma}^2_{v_i,v_i}}{\chi^2_{\frac{\eta}{2}}(l-1)}, \frac{(l-1)\tilde{\sigma}^2_{v_i,v_i}}{\chi^2_{1-\frac{\eta}{2}}(l-1)}\right) \quad i = 1,2,\cdots,n, \tag{24}$$

where l is the sample size; $t_{\eta/2}(l-1)$ is the inverse of the Student's T cumulative distribution function with $l-1$ degrees of freedom at the value $\frac{\eta}{2}$; and $\chi^2_{\frac{\eta}{2}}(l-1)$ is the inverse of the χ^2 cumulative distribution function with $l-1$ degrees of freedom at the value $\frac{\eta}{2}$. If the estimated mean link travel time falls within the confidence interval, then the estimated mean travel time is called a "reliable estimate." For simplicity, the results corresponding to the estimated variance are omitted from this paper. In this example, the confidence level is assumed to be 95% (i.e., $\eta = 0.025$).

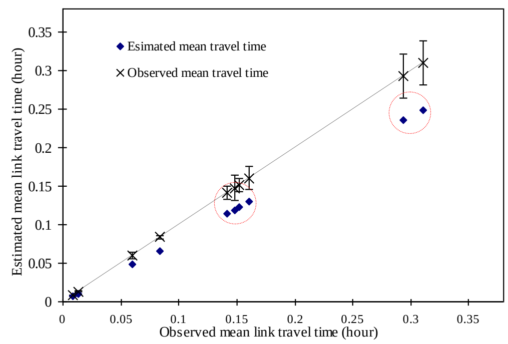
As the true value of the mean link travel time is unknown, the confidence interval explains how likely it is that the interval includes the true mean link travel time. It should be noted that if the estimated mean link travel time falls within the confidence interval, then this does not mean that the estimated mean link travel time is likely to be the true value with a probability of $1-\eta$. It simply means that the estimated mean travel time may be close to the true value because it falls within the interval, which offers the chance to include the true value with the probability of $1-\eta$. The same applies to the confidence interval of the link travel time covariance.

4.2. Effects of using observed data from multiple sensor systems

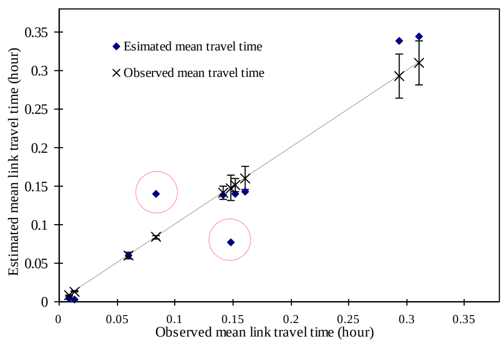
Three cases are used to illustrate the effects of using observed data from multiple sensor systems (Cases I, II, and IV), as shown in Table 6. In this example, it is assumed that links 2, 3, 5, 7, and 10 are observed links. It can be seen that if only loop detector data are used for travel time estimation (Case I), then the number of reliable estimates is zero. If AVI data alone are used (Case II), then the number of reliable estimates is 2. However, if both loop detector and AVI data are used, then the number of reliable estimates is 5, which is better than the other two cases. The corresponding comparisons between observed and estimated mean link travel times are also depicted in Figure 4. The quality of estimation in Case IV is much better than that in Cases I and II because the estimates in Cases I and II (Figures 4(A) and 4(B)) are far from the confidence intervals. However, the estimates in Case IV (Figure 4(D)) are nearer. This example illustrates that the greater the amount of sensor data used, the greater the accuracy and quality of the achieved network-wide travel time estimation.

Table 6. Effects of Multiple Sensor Systems and Second-order Statistical Property

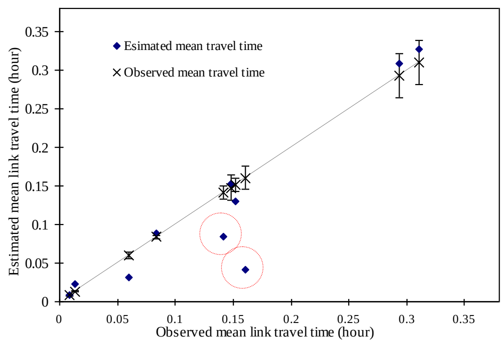
Case no.	Figure no.	Parameters				Description			Number of reliable estimates
		ω_1	ω_2	ω_3	ω_4	Loop detector data	AVI data	Second-order statistical property	
I	4(A)	1	0	0	1	√	×	√	0
II	4(B)	0	1	1	0	×	√	√	2
III	4(C)	1	1	0	0	√	√	×	4
IV	4(D)	1	1	1	1	√	√	√	5



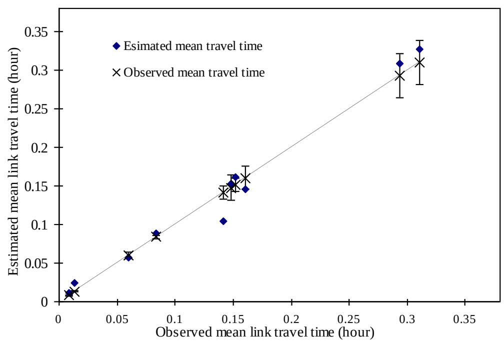
4(A)



4(B)



4(C)



4(D)

Figure 4. Comparison between observed and estimated mean link travel times with use of different sensor systems.

4.3. Effects of using second-order statistical property

To demonstrate the effects of using second-order statistical property, two cases are assumed (say Cases III and IV) in this example, as shown in Table 6. If the second-order statistical property (say covariance) is not used for travel time estimation (Case III), the number of reliable estimates is 4. This number increases to 5 if the covariance is considered in the objective function. Using second-order statistical property could improve the accuracy and quality of travel time estimation. This merit is more obvious in the comparison between Figures 4(C) and 4(D). As the two figures show, some estimates in Case III (Figure 4(C)) are far from the confidence interval, whereas those in Case IV are very close to or fall within it. This example demonstrates how using a second order statistical property of the observed data benefits network-wide travel time estimation.

4.4. Effects of number and location of sensors

For real-world applications of the proposed modeling approach, the number and location of the sensors must be determined in advance for data collection. To examine the effects of the number and location of the sensors, four scenarios are considered in Table 7. The other parameters are set in the same way as those in Case IV in the previous subsection. The number of reliable estimates increases with the number of observed links, a reasonable phenomenon considering that more observed links provide more information for travel time estimation. Meanwhile, the location of the sensors can also influence the quality of travel time estimation. For Scenarios C and D, although the numbers of observed links are significantly different, the numbers of reliable estimates are equal. That is to say, the location of the sensors needs to be optimized to save on the costs of data collection in real-world applications. It should be noted that the location of the sensors is also restricted to the network topology and physical conditions. In view of this, the sensor location problem considers the number and location of the sensors, the physical conditions of the networks, and the combination of the sensor systems. This type of sensor location problem deserves further investigation, which can be achieved by extending the proposed modeling approach (for examples of locating traffic counting stations and AVI readers, see Chen, Chootinan, and Pravinvongvuth (2004); Chen, Pravinvongvuth, and Chootinan (2007); Chen et al. (2010); Chootinan, Chen, and Yang (2005); and Eisenman et al. (2006)).

Table 7 Effects of Number and Location of Sensors

Scenario no.	Number of observed links	Combinations of observed links	Number of reliable estimates
A	1	1	1
B	5	2, 3, 5, 7, 10	5
C	**6**	3, 4, 5, 6, 7, 8	**8**
D	**10**	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	**8**

5. Conclusions and further study

This paper proposes a new model for the estimation of network-wide, on-line travel times using inconsistent data from multiple sensor systems. Traditionally, network-wide travel times can be estimated by traffic assignment models with given input data, such as the network-wide O-D matrix and the link performance function. However, such input data may be outdated or unavailable in real-world circumstances. To overcome this difficulty, the proposed model estimates the network-wide travel times by making full use of available traffic data, rather than using the traditional method. On-line traffic data, including link-based and path segment-based observed travel times from different sensor systems, are considered together with their first- and second-order statistical properties. Off-line data (historical data) are also considered to estimate the link travel times without sensors. In view of the data inconsistencies that stem from the use of different sensor systems, a GLS problem with non-linear constraints is formulated for network-wide travel time estimation. A solution algorithm based on the penalty function method is suggested to solve the proposed model. Evidence from numerical examples is provided to show that the use of data from multiple sensor systems and second-order statistical property has the potential to enhance the accuracy and quality of the network-wide travel time estimates. Also, the number and location of sensors are also shown to influence the accuracy and quality of these estimates.

On the basis of the model proposed in this paper, the following extensions can be envisaged.

- The network-wide travel times estimated by the proposed model can be used as input data for updating the O-D demand and calibrating the link performance function and the parameters in the gravity model for travel demand forecasting. This is the “bottom-up” procedure shown in Figure 1.
- The last numerical example indicates that the number and location of the sensors have a significant impact on the accuracy and quality of the estimates. Thus, the sensor location problem merits further study (Chen et al. 2004, 2007, 2010; Chootinan et al. 2005).
- The proposed model falls within the category of a static model for the travel time estimation problem at the strategic level. From a transportation management perspective, the proposed model can be extended into the dynamic or time-dependent dimension for travel time prediction (Ban, Li, Skabardonis, and Margulici 2010; Ni and Wang 2008; Liao and Davis 2007; Long et al. 2011; Ye et al. 2012). The data aggregation method (Tan et al. 2009; Soriguera and Robusté 2011) can also be incorporated into the network-wide travel time estimation model to consider the temporal covariance relationships of the multiple sources of observed data.
- The Maximum likelihood method developed for O-D estimation problem from link and path traffic data (Parry and Hazelton 2012) is worthy of being extended for network-wide travel time estimation problem to account for data inconsistency issue from multiple sensor systems.
- It should be pointed out that the proposed model may not be easily to solve particularly for large scale transportation networks. This is because that there are a large number of elements (or parameters) in the covariance matrix of link travel times to be estimated. For example, for a $n \times n$ covariance matrix, there are actually $(n+1) \times n/2$ elements needed to be estimated. However, in reality, some elements in the travel time covariance matrix may be zero. In particular, if the distance of two links are very long in spatial manner, the travel times of these two links should be independent on each other. And the corresponding elements in the covariance matrix should be zero. Due to the identifiability difficulty, the reasonable solutions of these covariances (i.e., zero) may not be identified as multiple solutions may exist. To overcome this difficulty, the shrinkage estimator, a technique that is useful for estimating large-dimensional parameters with comparatively fewer observations, has the potential to overcome this difficulty in practice. Lasso method, a widely used shrinkage estimator, does variable selection and shrinkage by solving the L1-penalized least squares (or linear regression) (Tibshirani, 1996 and 2011). The lasso method has been used by Shao et al. (2015) for OD covariance matrix estimation. Similar method is expected to be adopted in the proposed estimation model of this paper.

Acknowledgements

The work described in this paper was jointly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. PolyU 152074/14E and PolyU 152628/16E), National Natural Science Foundation of China (Project Nos. 71671184, 71271205 and 71371175), Research Institute for Sustainable Urban Development (RISUD) of The Hong Kong Polytechnic University (Project Nos. 1-ZVBY and 1-ZVBZ), and Qing Lan Project of Jiangsu Province.

References

Audet, C., and J.E. Dennis. 2003. "Analysis of Generalized Pattern Searches." *SIAM Journal on Optimization* 13 (3): 889-903.

Ban, X.J., Y.W. Li, A. Skabardonis, and J.D. Margulici. 2010. "Performance Evaluation of Travel-Time Estimation Methods for Real-Time Traffic Applications." *Journal of Intelligent Transportation Systems* 14 (2): 54-67.

Chandra, S.R., and H. Al-Deek. 2009. "Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models." *Journal of Intelligent Transportation Systems* 13 (2): 53-72.

Chen, A., P. Chootinan, and S. Pravinvongvuth. 2004. "Multiobjective Model for Locating Automatic Vehicle Identification Readers." *Transportation Research Record* 1886: 49-58.

Chen, A., S. Pravinvongvuth, and P. Chootinan. 2010. "Scenario-Based Multiobjective AVI Reader Location Models Under Different Travel Demand Patterns." *Transportmetrica* 6 (1): 53-78.

Chen, A., S. Pravinvongvuth, P. Chootinan, M. Lee, and W. Recker. 2007. "Strategies for Selecting Additional Traffic Counts for Improving O-D Trip Table Estimation." *Transportmetrica* 3 (3): 191-211.

Chen, C. 2002. Freeway Performance Measurement System (PeMS). (Unpublished doctoral dissertation), University of California, Berkeley, CA.

Chen, C., K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia. 2001. "Freeway Performance Measurement System: Mining Loop Detector Data." *Transportation Research Record* 1748: 96-102.

Chootinan, P., A. Chen, and H. Yang. 2005. "A Bi-objective Traffic Counting Location Model for Origin-Destination Trip Table Estimation." *Transportmetrica* 1 (1): 65-80.

Conn, A.R., K. Scheinberg, and L.N. Vicente. 2009. *Introduction to derivative-free optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Dailey, D.J. 1993. "Travel-Time Estimation Using Cross-Correlation Techniques." *Transportation Research Part B* 27 (2): 97-107.

Eisenman, S., X. Fei, X. Zhou, and H.S. Mahmassani. 2006. "Number and Location of Sensors for Real-Time Network Traffic Estimation and Prediction: A Sensitivity Analysis." *Transportation Research Record* 1964: 260-269.

Feng, Y.H., J. Hourdos, and G.A. Davis. 2014. "Probe Vehicle Based Real-Time Traffic Monitoring on Urban Roadways." *Transportation Research Part C* 40: 160-178.

Furtlehner, C., J.M. Lasgouttes, and A. De La Fortelle. 2007. "A Belief Propagation Approach to Traffic Prediction using Probe Vehicles." *Proceedings of 10th International IEEE Conference on Intelligent Transportation Systems*: 1022-1027.

Guo, H.R., and J.H. Jin. 2006. "Travel Time Estimation with Correlation Analysis of Single Loop Detector Data." *Transportation Research Record* 1968: 10-19.

Hu, S.R., S. Peeta, and C.H. Chu. 2009. "Identification of Vehicle Sensor Locations for Link-Based Network Traffic Applications." *Transportation Research Part B* 43 (8-9): 873-894.

Kwong, K., R. Kavaler, R. Rajagopal, and P. Varaiya. 2009. “Practical Scheme for Arterial Travel Time Estimation Based on Vehicle Reidentification Using Wireless Sensors.” *Proceedings of Transportation Research Board 88th Annual Meeting*, Washington DC.

Lagarias, J.C., J.A. Reeds, M.H. Wright, and P.E. Wright. 1998. “Convergence Properties of The Nelder-Mead Simplex Method in Low Dimensions.” *SIAM Journal of Optimization* 9 (1): 112-147.

Lam, W.H.K., K.S. Chan, and J.W.Z. Shi. 2002. “A Traffic Flow Simulator for Short-Term Forecasting of Travel Times.” *Journal of Advanced Transportation* 36 (3): 265-291.

Liao, C.F., and G.A. Davis. 2007. “Simulation Study of Bus Signal Priority Strategy - Taking Advantage of Global Positioning System, Automated Vehicle Location System, and Wireless Communications.” *Transportation Research Record* 2034: 82-91.

Lin, I., R. He, and A.L. Kornhauser. 2008. “Estimating Nationwide Link Speed Distribution Using Probe Position Data.” *Journal of Intelligent Transportation Systems* 12 (1): 29-37.

Liu, H.X., and W.T. Ma. 2009. “A Virtual Vehicle Probe Model for Time-Dependent Travel Time Estimation on Signalized Arterials.” *Transportation Research Part C* 17(1): 11-26.

Long, J.C., Z.Y. Gao, and W.Y. Szeto. 2011. “Discretised Link Travel Time Models Based on Cumulative Flows: Formulations and Properties.” *Transportation Research Part B* 45 (1): 232-254.

Mazloumi, E., G. Rose, G. Currie, and M. Sarvi. 2011. “An Integrated Framework To Predict Bus Travel Time and Its Variability Using Traffic Flow Data.” *Journal of Intelligent Transportation Systems* 15 (2): 75-90.

Mori U., A. Mendiburu, M. Álvarez, and J.A. Lozano. 2015. “A Review of Travel Time Estimation and Forecasting for Advanced Traveller Information Systems. *Transportmetrica A* 11 (2): 119-157.

Ni, D.H., and H.Z. Wang. 2008. “Trajectory Reconstruction for Travel Time Estimation.” *Journal of Intelligent Transportation Systems* 12 (3): 113-125.

Nocedal, J., and S.J. Wright. 2006. *Numerical Optimization* (2nd ed.). New York: Springer.

Parry, K., and M.L. Hazelton. 2012. “Estimation of Origin–Destination Matrices from Link Counts and Sporadic Routing Data.” *Transportation Research Part B* 46 (1): 175-188.

Shao, H., W.H.K. Lam, A. Sumalee, and M.L. Hazelton. 2015. “Estimation of mean and covariance of stochastic multi-class OD demands from classified traffic counts.” *Transportation Research Part C* 59: 92-110.

Robinson, S. and J.W. Polak. 2005. “Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the K-NN Method.” *Transportation Research Record* 1935: 47-56.

Soriguera, F., and F. Robusté. 2011. “Highway Travel Time Accurate Measurement and Short-Term Prediction Using Multiple Data Sources.” Transportmetrica A 7 (1): 85-109.

Tam, M.L., and W.H.K. Lam. 2008. “Using Automatic Vehicle Identification Data for Travel Time Estimation in Hong Kong.” *Transportmetrica* 4 (3): 179-194.

Tan, M.C., S.C. Wong, J.M. Xu, Z.R. Guan, and P. Zhang. 2009. “An Aggregation Approach to Short-Term Traffic Flow Prediction.” *IEEE Transactions on Intelligent Transportation Systems* 10 (1): 60-69.

Tan, Z.J., H. Yang, and R.Y. Guo. 2014. “Pareto efficiency of reliability-based traffic equilibria and risk-taking behavior of travelers.” *Transportation Research Part B,* 66(8):16-31.

Taylor, M.A.P., J.E. Woolley, and R. Zito. 2000. “Integration of The Global Positioning System and Geographical Information Systems for Traffic Congestion Studies.” *Transportation Research Part C* 8 (1-6): 257-285.

Tibshirani, R. 1996. “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society Series* B 58(1): 267-288.

Tibshirani, R. 2011. “Regression shrinkage and selection via the lasso: A retrospective.” *Journal of the Royal Statistical Society Series* B 73(Part 3): 273-282.

Torczon, V. 1997. “On The Convergence of Pattern Search Algorithms.” *SIAM Journal on Optimization* 7 (1): 1-25.

Uno, N., F. Kurauchi, H. Tamura, and Y. Iida. 2009. “Using Bus Probe Data for Analysis of Travel Time Variability.” *Journal of Intelligent Transportation Systems* 13 (1): 2-15.

Vanajakshi, L.D., B.M. Williams, and L.R. Rilett. 2009. “Improved Flow-Based Travel Time Estimation Method from Point Detector Data for Freeways.” *Journal of Transportation Engineering* 135 (1): 26-36.

Wang, Y., Y. Zheng, and Y. Xue. 2014. “Travel Time Estimation of a Path using Sparse Trajectories.” *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 25-34.

Wei, C.H., and Y. Lee. 2007. “Development of Freeway Travel Time Forecasting Models by Integrating Different Sources of Traffic Data.” *IEEE Transactions on Vehicular Technology* 56 (6): 3682-3694.

Xia, J.X., M. Chen, and W. Huang. 2011. “A Multistep Corridor Travel-Time Prediction Method Using Presence-Type Vehicle Detector Data.” *Journal of Intelligent Transportation Systems* 15 (2): 104-113.

Xiong, H., and G.A. Davis. 2009. “Field Evaluation of Model-Based Estimation of Arterial Link Travel Times.” *Transportation Research Record* 2130: 149-157.

Ye, Q., W.Y. Szeto, and S.C. Wong. 2012. “Short-Term Traffic Speed Forecasting Based on Data Recorded at Irregular Intervals.” *IEEE Transactions on Intelligent Transportation Systems* 13 (4): 1727-1737.

Yeon, J., L. Elefteriadou, and S. Lawphongpanich. 2008. “Travel Time Estimation on A Freeway Using Discrete Time Markov Chains.” *Transportation Research Part B* 42 (4): 325-338.

Zhang, H.M. 1999. “Link-journey-speed model for arterial traffic.” *Transportation Research Record* 1676: 109-115.