



USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No. 158PUY2.2

Stochastic Network Vehicular Origin-Destination Demand Using Multi-Sensor Information Fusion Approaches

By

Han-Tsung Liou
Postdoc Researcher
Institute of Statistical Science
Academia Sinica, Taipei, Taiwan
iroya.liou@gmail.com

and

Shou-Ren Hu
Professor
Department of Transportation and Communication Management Science
National Cheng Kung University, Tainan, Taiwan
shouren@mail.ncku.edu.tw

and

Srinivas Peeta
Professor
School of Civil Engineering, Purdue University
peeta@purdue.edu

and

Yong Hoon Kim
Graduate student
School of Civil Engineering, Purdue University
kim523@purdue.edu

and

Choungryeol Lee
Graduate student
School of Civil Engineering, Purdue University
Lee1210@purdue.edu

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT12-G-UTC05 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.



USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

NEXTRANS Project No. 158PUY2.2

Final Report, April 21, 2017

Title

158PUY2.2/ Stochastic Network Vehicular Origin-Destination Demand Using Multi-Sensor Information Fusion Approaches

Introduction

An origin-destination (O-D) trip matrix characterizes the demand pattern in a vehicular traffic network. It is a crucial component for long-term transportation planning and short-term traffic management. Hence, the accurate estimation of the O-D matrix is a well-studied problem. Traditional approaches for estimating O-D matrices are based on manual surveys, such as household interview, license plate recording, and postcard mail-back. However, they suffer from high costs and potential data sampling or recording errors. To resolve the O-D data collection issues, methods have been proposed to estimate O-D trip matrices from link traffic counts or flows [1]. They include generalized least squares [2], maximum likelihood [3], entropy maximization [4], Kalman filter [5], and Bayesian inference [6] studies. However, using link flows to infer the O-D matrix can entail multiple solutions, as these problems are typically underdetermined. Two approaches are proposed in the literature to address this underdetermined problem. The first approach is to solve the problem using a bilevel model [7], [8] which estimates the O-D matrix at the upper level based on some pre-specified user route choice decision rules (for example, under user equilibrium (UE) or stochastic user equilibrium (SUE)) in the lower level model. The second approach is to determine the O-D matrix from link flows using a path flow estimator (PFE) [9], [10]. Both approaches assume that specific travelers' path choice decision rules are known based on some traffic assignment principles. However, the path choice decisions are not easy to obtain in practice.

With the rapid development of intelligent transportation systems (ITS) and sensor technologies, various sensors have been developed for traffic monitoring and data collection. In general, traffic sensors can be classified into two types: passive and active sensors. Passive sensors, such as vehicle detectors (VDs), are used to observe point measurements (e.g., link flow, occupancy, and speed) [11]. Active sensors, such as video cameras, are used for applications such as automatic vehicle identification (AVI) and license plate recognition (LPR). Through two-way communication between roadside equipment and an onboard unit, an active sensor based system provides point-to-point measurements (e.g., travel time, vehicle trajectory, and vehicle identification) [12], [13].

Past studies have incorporated link flow, path flow and/or vehicular flow pattern information provided by active sensors for the O-D matrix estimation problem [14], [15], [16], [17], [18], [19], [20]. Hence, they leverage vehicle trajectory information [14], [15] and path flow information [16], [17], [18], [19], [20], in addition to link flow information, for the network O-D matrix estimation problem.

The network sensor location problem (NSLP) [21], [22] seeks to determine the minimum number of traffic sensors and their installation locations to completely infer the network traffic conditions, and is similar to the observability problem [23]. From a network flow observability perspective, partial link flow information collected using strategic sensor deployment can infer full link flow information [22], [24], [25], and the upper bound on the number of required sensors without path enumeration can be derived [24], [25]. It has been applied in the context of route guidance [26], travel time data collection [27], travel time estimation [28], [29], and link flow inference [22], [23], [24], [30]. The quantity and content of the data collected from traffic sensors can substantially affect the performance of the O-D matrix estimation models [31], [32], [33]. Hence, the methods used to solve the NSLP can be adapted for the O-D matrix estimation problem [21], [22], [34]. However, past studies only sequentially solved the sensor location and O-D matrix estimation problems, using traffic information from a single sensor type [21], [35], [36], [37], [38] or from heterogeneous sensors [19], [39], [40], [41]. Castillo et al. used scanning-link information collected by plate scanning technique for link and/or route flow estimation [14], [35], [38], [39], [40]. In addition, Castillo et al. further defined the flow amount of information (FAO) to analyze the number of linearly independent scanning links for route flow reconstruction [39], [40]. Some of their studies solved the network O-D matrix estimation problem and the heterogeneous sensors deployment problem (HSDP) in an integrated manner either by algebraic based methods [39], [40] or a Bayesian approach [37], [38]. A strategic sensor deployment plan for different sensor types and installation locations has important implications for the performance of the network O-D matrix estimation model, and the discrepancy between the estimated and true O-D flows is crucial information to verify the appropriateness of the sensor deployment strategy. That is, solving for the sensor deployment strategy and then assuming it as given in the O-D matrix estimation problem in the sequential approach does not leverage information needs related to accurate O-D matrix estimation to guide the selection and location of deployed sensors. Hence, there is a key need to solve the heterogeneous sensor deployment problem and network O-D matrix estimation problem in an integrated manner.

The problem of solving the heterogeneous sensors deployment problem and network O-D matrix estimation problem is referred to as the HSDP-OD problem in this study. The corresponding problem using traffic information from a single sensor type is labeled the NSLP-OD problem. Past studies to solve the NSLP-OD or HSDP-OD problem usually use two steps in a sequential approach. The first step determines a strategy for locating sensors in the network, and the second

step solves the O-D matrix estimation problem based on the sensor deployment from the first step [14], [19], [21], [35], [36], [38], [39], [40], and [41]. Hence, the sequential approach is a one-shot procedure as the O-D matrix estimation results are not fed back to the NSLP/HSDP stage. Castillo et al. [37] propose a Wardrop-minimum variance (WMV) method assignment problem and Bayesian network (BN) approach for the link and O-D flow estimation problem. The NSLP-OD problem is solved by sequentially selecting one link as the equipped link to update the O-D matrix estimate until the budget constraint is violated. However, some specific input data or model assumptions are typically required in a sequential approach for the NSLP-OD or HSDP-OD problem. For example, a link-path incidence matrix is necessary for the developed sensor location and O-D matrix estimation models [14], [21], [35], [36], [37], [38], [40], [41]; this may entail solving a user equilibrium traffic assignment problem to identify the paths [37]. However, assumptions on the availability of some additional input data may not be realistic from a data availability perspective. For instance, the sensor location models require prior O-D/path flow information [35], [36], [37], [40], [41], or link flow proportions [37], [41] to determine the sensor locations. The O-D matrix estimation models assume the prior O-D/path information [14], [21], [35], [36], [37], [38], [41], or path assignment probability to be known [41], or that link flow proportions can be estimated [21], [37].

Unlike previous studies for the NSLP-OD problem [14], [21], [35], [36], [37], [38], [39], [40], [41], this study specifically investigates the HSDP-OD problem. And, unlike the sequential approach used previously for the HSDP-OD problem [19], [41], this study addresses the HSDP-OD problem in an integrated manner using a two-stage optimization model where the error on the O-D matrix estimate in the second-stage model is fed back to modify the sensor deployment strategy in the first-stage model until some pre-specified error thresholds are met. In the first stage, the HSDP model determines the optimal numbers of active (camera-based license plate recognition), passive (vehicle detector) sensors, and their installation locations to maximize the traffic information available for the O-D trip matrix estimation. This traffic information consists of the observed link flows, path trajectories and path coverage information. In the second stage, the O-D matrix estimation model leverages this traffic information to determine the network O-D matrix that minimizes the error between the observed and estimated traffic flows (link, O-D and/or path). Two network O-D matrix estimation models are proposed. The link-based model determines the O-D matrix based on the link-node incidence matrix under flow conservation rules without requiring strong assumptions related to prior knowledge/data. The path-based model assumes a given link-path incidence matrix and leverages active sensor information. A specially designed feedback mechanism is proposed to update parameters (weights of the objective function terms) in the heterogeneous sensors deployment model of the first stage based on the performance of the O-D matrix estimates in the second stage. Through this feedback mechanism, the proposed two-stage model captures the interactions between the

heterogeneous sensor deployment strategy and the network O-D matrix estimation problem. This establishes a bridge between the HSDP and the O-D matrix estimation problem, and represents an integrated approach to solve the HSDP-OD problem. The results from an empirical study using the Sanmin network in Taiwan indicate that the proposed integrated optimization model can provide network O-D matrix estimates as well as the numbers and locations of the two sensor types consistent with the corresponding objectives at the two stages.

The remainder of this report is organized as follows. *Methodology* characterizes the heterogeneous sensor based traffic information and presents the integrated model for the HSDP-OD problem, including the formulations of the heterogeneous sensors deployment and O-D matrix estimation models. It describes the solution procedure with the feedback mechanism to solve the integrated model. *Findings* discusses the results of numerical experiments based on real road network. Finally, concluding remarks are presented in *Summary and Recommendations*.

Methodology

This study proposes an integrated two-stage optimization model for the HSDP-OD problem. The first stage is the heterogeneous traffic sensors deployment model, which seeks to determine the sensor deployment strategy that optimizes the traffic information available to the second-stage O-D matrix estimation problem. The weights of the objective function terms in this model are functions of the errors between the observed and estimated link and path/O-D flow data determined in the second-stage model. The second stage is the O-D matrix estimation model that leverages the traffic information obtained in the first-stage model to determine the O-D matrix that minimizes the errors between the observed and estimated traffic flow data. The traffic information from the first stage and the traffic flow data errors from the second stage integrate this two-stage optimization model. The two-stage model for the HSDP-OD problem is described hereafter.

A. Heterogeneous Sensor Based Traffic Information

The passive sensors used in this study are VDs and the active sensors are LPRs. The heterogeneous sensors deployment model determines the optimal heterogeneous sensor deployment strategy, in terms of the selection of the numbers of LPRs, VDs, and their installation locations, to maximize the available traffic information for the O-D matrix estimation problem. The observed traffic information using the VDs and LPRs can be categorized into three types: (i) link flow information, (ii) path trajectory information, and (iii) path coverage information.

Link flow information

Link flow information can be collected on links equipped with either VD or LPR sensors. Hence, the numbers of VDs and LPRs deployed can be regarded as the number of pieces of observed link flow information, and are formulated using 0-1 integer decision variables as follows.

$$\text{Link flow information: } \sum_{ij \in A} (x_{ij} + y_{ij}), \forall i, j \in N \quad (1)$$

where,

$$x_{ij} = \begin{cases} 1, & \text{if link } ij \text{ is equipped with a VD;} \\ 0, & \text{otherwise} \end{cases};$$

$$y_{ij} = \begin{cases} 1, & \text{if link } ij \text{ is equipped with an LPR;} \\ 0, & \text{otherwise} \end{cases};$$

A : link set;

N : node set.

Path trajectory information

For a given link-path incidence matrix, each element in this matrix indicates if a specific path passes through a link. Hence, such an element in a link-path incidence matrix can be expressed in Eq. (2).

$$\delta_{ij}^p = \begin{cases} 1, & \text{if path } p \text{ contains link } ij \\ 0, & \text{otherwise} \end{cases}. \quad (2)$$

The path trajectory information of a given O-D pair can be collected by LPR sensors installed at various links along this specific path, and is formulated by the LPR decision variable and the delta function as follows.

$$\text{Path trajectory information: } \sum_{p \in P} \sum_{ij \in A} y_{ij} \cdot \delta_{ij}^p \quad (3)$$

where P is the path set.

Besides its capability to collect link flow information, an LPR sensor is able to actively track a vehicle's identification. Thereby, installing LPR sensors at some strategic links would provide full or partial path flow information of a given O-D pair. As a result, the flow on a given O-D pair can be (partially) observed by summing up the corresponding (partially) collected path flows.

Path coverage information

When different paths share the same path trajectories, it is difficult to identify the observed path flows of these paths. Based on the path trajectory information of a given link-path incidence matrix, a distinction function [35], [42] is introduced to distinguish between paths, and expressed in Eq. (4).

$$d(p^0, p^1, \delta_{ij}^p) = \begin{cases} 1, & \text{if } \delta_{ij}^{p^0} \neq \delta_{ij}^{p^1} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The distinction function identifies the difference between two paths by comparing their trajectories. Specifically, if $d(p^0, p^1, \delta_{ij}^p)$ is 1, it means that link ij is equipped with an LPR sensor, and this LPR-equipped link is able to distinguish between paths p^0 and p^1 based on a given link-path incidence matrix.

Similar to the definition of the path trajectory information, the path coverage information is formulated by the LPR decision variable and the distinction function as follows.

$$\text{Path coverage information: } \sum_{ij \in A} y_{ij} \cdot d(p^0, p^1, \delta_{ij}^p), \forall p^0, p^1 \in \mathbf{P} \quad (5)$$

When the value of Eq. (5) is 0, it indicates that a current LPR sensor deployment configuration cannot distinguish the difference between paths p^0 and p^1 . When the value of Eq. (5) is 1 or greater than 1, it means that at least one LPR sensor can distinguish the difference between paths p^0 and p^1 . In addition, in order to maximize the capacity of the path differentiation capability of an LPR sensor, a path coverage variable, m_{p^0, p^1} is defined as follows.

$$\sum_{ij \in A} y_{ij} \cdot d(p^0, p^1, \delta_{ij}^p) \geq m_{p^0, p^1}, \forall p^0, p^1 \in \mathbf{P} \quad (6)$$

where,

$$m_{p^0, p^1} = \begin{cases} 1, & \text{if paths } p^0 \text{ and } p^1 \text{ can be differentiated by LPR sensors} \\ 0, & \text{otherwise} \end{cases}.$$

Eq. (6) depicts that the path coverage variable is the lower bound for the number of pieces of path coverage information, where the collected path coverage information for the entire network by deploying an LPR sensor at link ij is as large as possible.

B. Heterogeneous Sensors Deployment Model

The first-stage heterogeneous sensors deployment model is formulated as an integer program to determine the numbers of LPRs, VDs, and their installation locations to maximize the available

traffic information subject to constraints on the available budget, network topology, and set covering rules. To quantify the relative contributions of the three types of traffic information to the estimation of the O-D matrix in the second stage, weights are introduced for the corresponding terms in the first-stage objective function. Thereby, the heterogeneous sensor deployment strategy in the first stage, through the three types of traffic information observed using these sensors, is linked to the second-stage objective of determining the O-D matrix with the minimum amount of error. The first stage model formulation is as follows.

$$\begin{aligned} \text{Max } & \left\{ \sum_{ij \in A} w_{\alpha} [\alpha_{ij}, \varepsilon(\hat{v}_{ij})] \cdot (x_{ij} + y_{ij}) \right\} \\ & + \left\{ \sum_{p \in P} \sum_{ij \in A} w_{\beta} [\beta_{ij}^p, \varepsilon(\hat{t}^{rs})] \cdot y_{ij} \cdot \delta_{ij}^p \right\} \\ & + \left\{ \sum_{p^0, p^1 \in P} w_{\gamma} [\gamma_{t_{p^0}^{r_0 s_0}, t_{p^1}^{r_1 s_1}}, \varepsilon(\hat{t}^{r_0 s_0}, \hat{t}^{r_1 s_1})] \cdot m_{p^0, p^1} \right\} \end{aligned} \quad (7)$$

Subject to

$$\sum_{ij \in A} y_{ij} \cdot d(p^0, p^1, \delta_{ij}^p) \geq m_{p^0, p^1}, \forall p^0, p^1 \in P \mid p^0 \neq p^1 \quad (8)$$

$$\sum_{ij \in A} (x_{ij} + y_{ij}) \cdot \delta_{ij}^p \geq 1, \forall p \in P \quad (9)$$

$$x_{ij} + y_{ij} \leq 1, \forall ij \in A \quad (10)$$

$$C_D \cdot \sum_{ij \in A} x_{ij} + C_L \cdot \sum_{ij \in A} y_{ij} \leq C \quad (11)$$

where,

C_D : the cost of a VD (passive sensor);

C_L : the cost of an LPR (active sensor);

C : total budget;

α_{ij} : initial weight of link flow information for link ij ;

β_{ij}^p : initial weight of the p^{th} path trajectory information for link ij ;

$\gamma_{t_{p^0}^{r_0 s_0}, t_{p^1}^{r_1 s_1}}$: initial weight of the path coverage information for the p^0 path of $t^{r_0 s_0}$, and the p^1 path of $t^{r_1 s_1}$,

where $t^{r_0 s_0}$ and $t^{r_1 s_1}$ are the two different O-D pairs;

$w_{\alpha} [·]$: weight for link flow information;

$w_{\beta} [·]$: weight for path trajectory information;

$w_{\gamma} [·]$: weight for path coverage information;

$\varepsilon(\hat{v}_{ij})$: error function based on link flow estimate at link ij ,

where \hat{v}_{ij} is the estimated flow on link ij ;

$\varepsilon(\hat{t}^{rs})$: error function based on O-D flow estimate for O-D pair rs ,

where \hat{t}^{rs} is the estimated flow for O-D pair rs ;

$\varepsilon(\hat{t}^{r_0s_0}, \hat{t}^{r_1s_1})$: error function based on O-D flow estimates

for O-D pairs r_0s_0 and r_1s_1 .

Eq. (7) is the objective function whose goal is to maximize the number of pieces of observed traffic information on link flow, path trajectory, and path flow coverage. The decision variables are x_{ij} , y_{ij} and m_{p^0,p^1} . It is assumed here that maximizing the number of pieces of these three types of information collected using the VDs and LPRs will maximize the traffic information available to the second stage problem. The three terms in the objective function correspond to the observed link flow information, path trajectory information, and path coverage, respectively. The corresponding weights ($w_\alpha[\cdot]$, $w_\beta[\cdot]$, $w_\gamma[\cdot]$) reflect the relative importance of each type of information for the O-D matrix estimation problem. It is important to note here that each of the weights depends on a pre-determined initial weight and errors on link flow and O-D matrix estimates in the second-stage model. In the solution procedure described in Section III, these weights are iteratively updated when the O-D matrix estimation results are obtained in the second stage. Eq. (8) introduces the path coverage variable, m_{p^0,p^1} and states that if a path can be distinguished by at least one LPR sensor based on the distinction function from a given link-path incidence matrix, this path is covered. The number of the path coverage variables is dependent on the size of the path set. This number can be large, and some analytical approaches have been proposed to reduce this size [35]. Eq. (9) illustrates the set covering rule, which indicates that each path should be observed by at least one VD or LPR sensor [35], [42]. Eq. (10) reflects that a link is at most equipped with one sensor, which can be either active or passive. Eq. (11) is the budget constraint that incorporates the unit costs of both sensor types. Generally, the C_L/C_D ratio is approximately 100 since the system infrastructure of an active-type sensor system requires higher initial and maintenance costs.

C. Network O-D Matrix Estimation Models Using Heterogeneous Sensor Based Traffic Information

When a link is equipped with a VD or LPR sensor (x_{ij} or y_{ij}), the corresponding link flow information, $\bar{v}_{ij}(x_{ij}, y_{ij})$ can be collected and/or observed. If an LPR sensor deployment configuration is determined, the collection of y_{ij} 's and m_{p^0,p^1} 's, which are respectively denoted as \mathbf{Y} and \mathbf{M} , can identify the path flows based on the mapping of the observed path trajectory information and path coverage variable in a given link-path incidence matrix. The observed path flow information is defined as $\bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})$.

The heterogeneous sensor deployment model developed in the first stage of the integrated model can provide subsets of link and/or path flows. The network O-D matrix estimation model developed in the second stage leverages this traffic information on the observed link/path flows and path trajectory/coverage information. Accordingly, two versions of the network O-D matrix estimation model are developed. One is the link-based model, which abides by the link flow conservation rule based on a link-node incidence matrix. The other is the path-based model which estimates the path flows in light of a given link-path incidence matrix. Both models are formulated as nonlinear least squares (NLS) programs that minimize the errors on the estimated link/path flows and/or O-D matrix. The models are discussed hereafter.

Link-based O-D matrix estimation model

The link-based model is developed based on a non-proportional traffic assignment principle, and is formulated as a nonlinear program with linear constraints as follows.

$$\begin{aligned} \text{Min} \quad & \sum_{(i,j) \in A} \left[\frac{\hat{v}_{ij}}{\sum_{ij \in A} \hat{v}_{ij}} - \frac{\bar{v}_{ij}(x_{ij}, y_{ij})}{\sum_{ij \in A} \bar{v}_{ij}(x_{ij}, y_{ij})} \right]^2 \\ & + \sum_r \sum_s \left[\frac{\sum_p \hat{f}_p^{rs}}{\sum_r \sum_s \sum_p \hat{f}_p^{rs}} - \frac{\sum_p \bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})}{\sum_r \sum_s \sum_p \bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})} \right]^2 \end{aligned} \quad (12)$$

Subject to

$$\hat{\mathbf{V}} \cdot \mathbf{L} = 0 \quad (13)$$

$$\sum_i \hat{v}_{ir} + \sum_{s \in S} \hat{t}^{rs} = \sum_j \hat{v}_{rj}, \forall r \in \mathbf{R} \quad (14)$$

$$\sum_i \hat{v}_{is} = \sum_{r \in \mathbf{R}} \hat{t}^{rs} + \sum_j \hat{v}_{sj}, \forall s \in \mathbf{S} \quad (15)$$

$$\hat{t}^{rs} \geq \sum_{p \in \mathbf{P}} \bar{f}_p^{rs}, \forall r \in \mathbf{R}, s \in \mathbf{S}, p \in \mathbf{P} \quad (16)$$

$$\hat{t}^{rs} \geq 0, \forall r \in \mathbf{R}, s \in \mathbf{S} \quad (17)$$

$$\hat{v}_{ij} \geq 0, \forall ij \in A \quad (18)$$

where,

\hat{v}_{ij} : estimated flow on link ij ;

$\hat{\mathbf{V}}$: a vector of estimated link flows;

$\bar{v}_{ij}(x_{ij}, y_{ij})$: observed link ij flow based on the deployment

of VD or LPR sensors;

$\bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})$: observed p^{th} path flow between O-D pair rs
based on the deployment of LPR sensors;

\hat{t}^{rs} : estimated O-D flow between pair rs ;

\mathbf{L} : link-node incidence matrix;

\mathbf{R} : a set of origin nodes;

\mathbf{S} : a set of destination nodes;

\mathbf{Y} : the collection of y_{ij} ;

\mathbf{M} : the collection of m_{p^0, p^1} .

Eq. (12) is the objective function, which minimizes the errors between the (partially) observed and estimated link, and O-D flows. The flow estimation errors are normalized to circumvent the effect of difference in the orders (in terms of the absolute values) of the link and O-D flow estimates. In Eq. (12), two categories of the traffic information are collected. One is the observed link flow, $\bar{v}_{ij}(x_{ij}, y_{ij})$ which is provided by the VD and LPR sensors. The other is the observed path flow, $\bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})$ which is determined by the LPR sensor deployment configuration and path coverage variable conditions (i.e., Eq. (8)). The observed O-D flows are the summation of the partially or fully observed path flows from LPR sensors. Because sensors cannot be deployed on all links of a network due to the budget constraint, the observed traffic information may include some (partial) link and/or path flows for a given O-D pair. Eq. (13) is the link flow conservation constraint based on a link-node incidence matrix while origin and destination nodes are excluded from the flow conservation rule. When origin or destination nodes are intermediate nodes, the flow conservation rule for the origin and destination nodes is described by Eqs. (14) and (15), respectively. Eq. (14) states that the flow departing from an origin node is equal to sum of the flow originating at the origin and the pass-through flow(s). Similarly, for the destination nodes, Eq. (15) states that the incoming link flows to a destination node are composed of the destination flows and pass-through flow(s). Eq. (16) is the inequality constraint indicating that, for a particular O-D pair, the estimated O-D flow should be greater than or equal to the O-D flow partially or fully observed by LPR sensors. Eqs. (17) and (18) are the non-negativity constraints on the estimated O-D and link flows, respectively.

Path-based O-D matrix estimation model

The O-D matrix estimation model can alternatively be formulated as a path-based model under a given link-path incidence matrix by using path flows as the decision variables. It is developed by using a path flow estimator, and is formulated as a nonlinear program with linear constraints as follows.

$$\begin{aligned}
\text{Min } & \sum_{(i,j) \in A} \left[\frac{\hat{v}_{ij}}{\sum_{ij \in A} \hat{v}_{ij}} - \frac{\bar{v}_{ij}(x_{ij}, y_{ij})}{\sum_{ij \in A} \bar{v}_{ij}(x_{ij}, y_{ij})} \right]^2 \\
& + \sum_r \sum_s \left[\frac{\sum_p \hat{f}_p^{rs}}{\sum_r \sum_s \sum_p \hat{f}_p^{rs}} - \frac{\sum_p \bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})}{\sum_r \sum_s \sum_p \bar{f}_p^{rs}(\mathbf{Y}, \mathbf{M})} \right]^2
\end{aligned} \tag{19}$$

Subject to

$$\hat{\mathbf{F}} \cdot \boldsymbol{\delta} = 0 \tag{20}$$

$$\hat{f}_p^{rs} \geq \bar{f}_p^{rs}, \forall r \in \mathbf{R}, s \in \mathbf{S}, p \in \mathbf{P} \tag{21}$$

$$\hat{f}_p^{rs} \geq 0, \forall r \in \mathbf{R}, s \in \mathbf{S}, p \in \mathbf{P} \tag{22}$$

$$\hat{v}_{ij} \geq 0, \forall (i, j) \in A \tag{23}$$

where,

$\hat{\mathbf{F}}$: a vector of estimated path flows;

$\boldsymbol{\delta}$: link-path incidence matrix;

\hat{f}_p^{rs} : estimated path flow for the p^{th} path of O-D pair rs .

Eq. (19) is similar to Eq. (12) and includes two minimization terms: (i) the normalized errors between observed and estimated link flows, and (ii) the normalized errors between observed and estimated path flows. Eq. (20) is the flow conservation constraint under a given link-path incidence matrix. Eq. (21) is the inequality constraint indicating that, for a particular O-D pair, the estimated path flow should be greater than or equal to the path flow partially or fully observed by LPR sensors. Eqs. (22) and (23) are the non-negativity constraints for both path and link flows.

Note that the link-based O-D matrix estimation model estimates the network O-D flows using the simplified flow relationship between origin/destination nodes and adjacent links (see Eqs. (14) and (15)), and infers link flows based on a link-node incidence matrix (see Eq. (13)). For small networks, composed of a few nodes and links, a link-based O-D matrix estimation model could provide satisfactory solutions for both the link and O-D flow estimates. If a network is large or has a complex structure, the link-based model may not adequately capture the flow relationship between origin/destination nodes and intermediate links. Then, a path-based approach, such as the PFE method, which assumes a known link-path incidence matrix, could be used to solve the network O-D matrix estimation problem.

An iterative solution procedure is designed to determine the network O-D matrix and link flow estimates. Details can be found in Hu et al. (2016).

The real road network considered is located in the Sanmin district in Kaohsiung City, Taiwan, and is referred to as the Sanmin network in this study [44]. The Sanmin network consists of 72 nodes, 202 unidirectional links and 156 O-D pairs. The Sanmin network is located in the central district of Kaohsiung City in Southern Taiwan.

The true O-D flow data was obtained from a home survey of the Kaohsiung metropolitan area [45]. For model evaluation purposes, we obtained only the traffic data relevant to the Sanmin district. This dataset is used to evaluate the performance of the HSDP-OD model. Based on the cost databases from the RITA (U.S. Department of Transportation), the unit costs for an LPR and a VD were set at \$70,000 and \$600, respectively, and the total budget was assumed to be \$3,000,000. The termination criteria for the solution procedure are that the average mean absolute percent error (MAPE) of the link flow estimates is less than 15% and the average MAPE of the O-D flow estimates is less than 20%. As the Sanmin network is relatively large, the termination criterion is less tight for the O-D flow estimates compared to that of the fishbone network. Both the link-based model and the path-based O-D matrix estimation models are analyzed, and the termination criteria for both models are the same (15% - 20%).

B. Performance of the Link- and Path-based O-D Demand Estimation Models

Based on the above experimental setup and assumptions, Figure 2 shows the results of the link and O-D flow estimates across different iterations under the link-based model. Figure 3 shows the results for the path-based model.

Figure 2 illustrates that the accuracy of O-D flow estimates using the link-based O-D flow estimation model depends not just on the accuracy of the link flow estimates (see the seventh and eighth iterations). As the flow conservation rule between O-D and link flows (Eqs. (13)-(15)) is the only key characteristic of the link-based O-D matrix estimation model, it exhibits limited capability to identify the possible O-D flow estimates in a large bidirectional network. This is illustrated by the MAPEs in most iterations of Figure 2 being above 20%. Despite this deficiency, the link-based O-D estimation model terminates at the fifteenth iteration as the termination criterion is met. The average MAPE of the link flow estimates is 12.7%, and the average MAPE of the O-D matrix estimates is 18.5%.

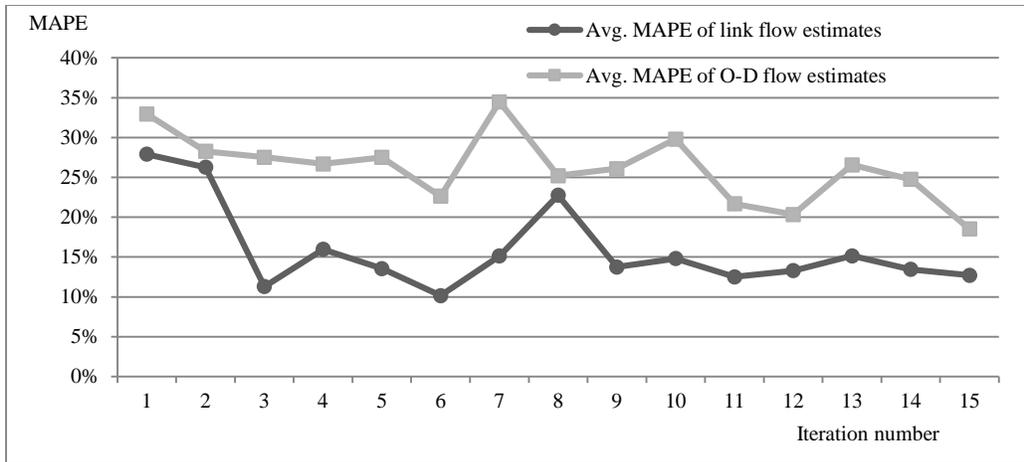


Fig. 2. MAPEs of the link-based model

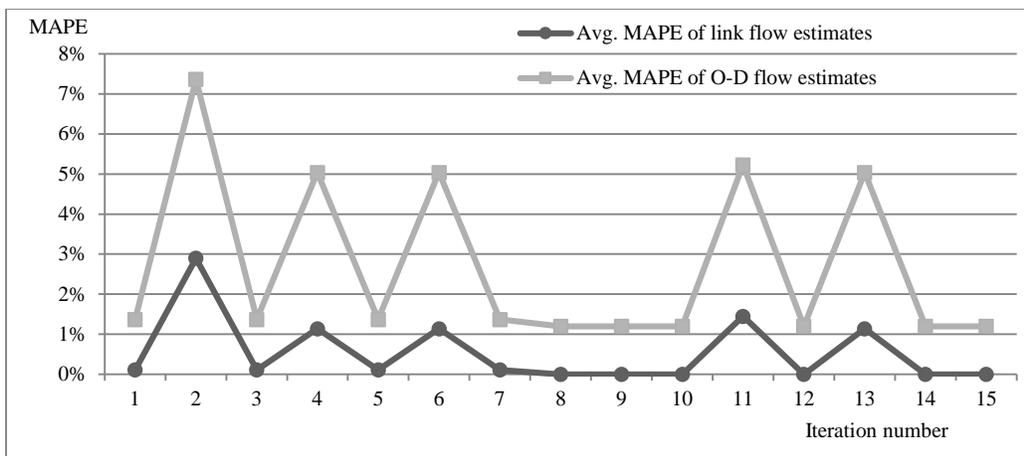


Fig. 3. MAPEs of the path-based model

The results for the path-based O-D matrix estimation model are shown in Figure 3. The performance in terms of the estimation accuracy of both the link and O-D flows is satisfactory. The similarity in the trends of the two MAPE plots in Figure 3 indicates that the accuracy of the O-D flow estimates depends on that of the link flow estimates as the spatial relationship between links and paths is pre-determined by the given link-path incidence matrix. The path-based model terminates at the eighth iteration; the average MAPE of the link flow estimates is 0.0% and that of the O-D flow estimates is 1.2%. The sensor deployment configuration corresponding to the link-based model is shown in Figure 4. It indicates that 142 sensors (42 LPR sensors and 100 VD) are required, and the sensor deployment rate is 70%. The sensor deployment configuration corresponding to the path-based model is shown in Figure 5, and again 142 sensors (42 LPR sensors and 100 VD) are required. However, the sensor location deployment pattern for the path-based model is different from that of the link-based model.

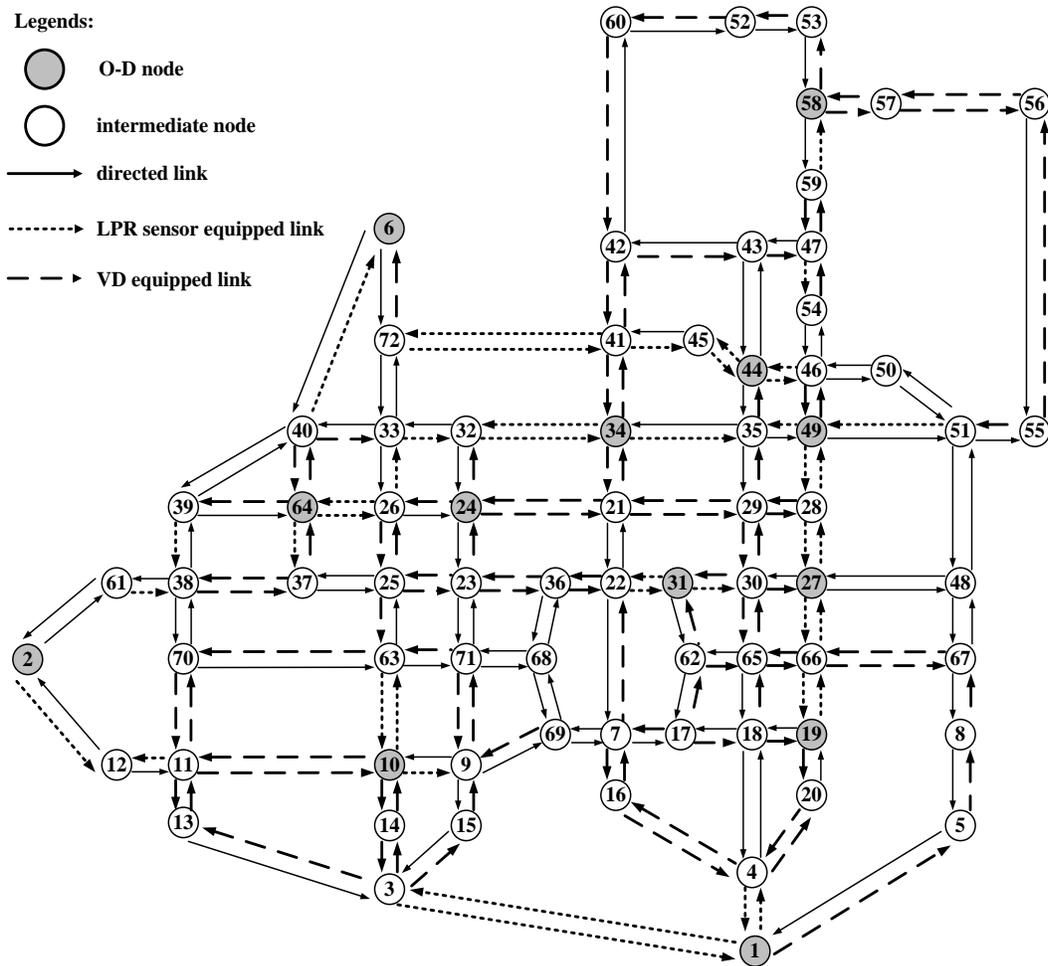


Fig. 5. Optimal sensor deployment configuration for the path-based model

Table I illustrates that under the link-based O-D matrix estimation model, using both VDs and LPRs reduces the average MAPE for O-D flow estimates from 63.8% (partial VD only: replacement), 98.0% (partial VD only: fixed) or 74.6% (full VD only) to 18.5%. Similarly, under the path-based O-D matrix estimation model, using both VDs and LPRs reduces the average MAPE of O-D flow estimates from 31.8% (partial VD only: replacement), 46.6% (partial VD only: fixed), or 31.8% (full VD only) to 1.2%. The underlying reasoning is similar to that discussed for the fishbone network case related to the valuable data that LPRs can additionally observe compared to VDs. Further, the path-based O-D matrix estimation model, which explicitly captures path-related information by using the additional path flow information provided by the LPR sensors, outperforms the link-based O-D matrix estimation model.

TABLE I O-D and Link Flow Estimates for the Sanmin Network

Data source	O-D flow estimation model	Average MAPE for O-D flow estimates	Average MAPE for link flow estimates
142 VDs (<i>Replacement</i>)	Link-based model	63.8%	28.3%
142 VDs (<i>Fixed</i>)	Link-based model	98.0%	21.1%
202 VDs	Link-based model	74.6%	0.0%
42 VDs & 100 LPRs	Link-based model	18.5%	12.7%
142 VDs (<i>Replacement</i>)	Path-based model	31.8%	3.2%
142 VDs (<i>Fixed</i>)	Path-based model	46.6%	1.3%
202 VDs	Path-based model	31.8%	0.0%
42 VDs & 100 LPRs	Path-based model	1.2%	0.0%

C. Discussion

The link-based O-D matrix estimation model expresses the relationship between O-D and link flows by using the flow conservation rules (Eqs. (13)-(15)). The link-based model has two issues. First, it cannot adequately identify possible O-D flows in a large bidirectional network. Second, the accuracy of the O-D flow estimates does not depend only on the accuracy of link flow estimates. The flow conservation rules cannot completely describe the spatial relationships between O-D and link flows. Two approaches can be used to address these issues. The first approach is to incorporate multiple traffic data sources that can provide information not only on link flows, but also on other aspects such as path trajectories/flows (note the reduced MAPEs for the link-based model in TABLE I when both VDs and LPRs are used). The second approach leverages the knowledge of a known link-path incidence matrix as part of a path-based O-D matrix estimation model, which can capture the causal relationships among the O-D, path and link flows (note the performance of the path-based models in TABLE II using VDs only or VDs and LPRs).

Summary

This study proposes a two-stage optimization model for the HSDP-OD problem. The heterogeneous sensors deployment model in the first stage incorporates three sources of observed traffic information (link flow, path trajectory and path coverage) into an integer program. The O-D matrix estimation model in the second stage is constructed as link-based and path-based NLS programs. The following summarizes the key findings of this study.

- The usage of LPR sensors in addition to the VDs can significantly enhance the accuracy of the O-D matrix estimation. That is, there is value to considering heterogeneous sensors that

provide observations of traffic data beyond just link flows, such as path trajectories, and partial path or O-D flows, etc.

- Integrating the determination of the sensor deployment strategy with that of the accuracy of estimating the O-D matrix enables a more holistic perspective to addressing both problems by leveraging their interactions.

Recommendations

In this study, we solved the heterogeneous sensors deployment and network O-D matrix estimation problems in a two-stage iterative model. Two potential future research directions are as follows.

- First, the feedback criteria used in the two-stage model can be based on other representations of the errors.
- Second, because the partially measured link or path flows by different sensors are associated with various degrees of measurement errors, and the assignment mappings between a set of unknown O-D flows and path/link flows are random variables determined by travelers' route choice decisions, the HSDP-OD problem can be modeled as a state estimation problem under these traffic and/or users' route choice uncertainties.

References

- [1] P. Robillard, "Estimating the O-D matrix from observed link volumes," *Transportation Research*, vol. 9, no. 2-3, pp. 123-128, 1975.
- [2] E. Cascetta, "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator," *Transportation Research-B*, vol. 18, no. 4-5, pp. 289-299, 1984.
- [3] H. Spiess, "A maximum likelihood model for estimating origin-destination matrices," *Transportation Research-B*, vol. 21, no. 5, pp. 395-412, 1987.
- [4] H. J. Van Zuyen, and L. G. Williumsen, "The most likely trip matrix estimated from traffic counts," *Transportation Research-B*, vol. 14, no. 3, pp. 281-293, 1980.
- [5] N. L. Nihan, and G. A. Davis, "Recursive estimation of origin-destination matrices from input / output counts," *Transportation Research-B*, vol. 21, no. 2, pp. 149-163, 1987.
- [6] J. M. Maher, "Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach," *Transportation Research-B*, Vol. 17, No. 6, pp. 435-447, 1983.

- [7] H. Yang, T. Sasaki, Y. Iida, and Y. Asakuba, "Estimation of origin-destination matrices from link traffic counts on congested networks," *Transportation Research-B*, vol. 26, no. 6, pp. 417-434, 1992.
- [8] H. Yang, "Heuristic algorithm for the bilevel origin-destination matrix estimation problem," *Transportation Research-B*, vol. 29, no. 4, pp. 231-242, 1995.
- [9] H. D. Sherali, R. Sivanandan, and A. G. Hobeika, "A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes," *Transportation Research-B*, vol. 28, no. 3, pp. 213-233, 1994.
- [10] M. G. H. Bell, C. M. Shield, F. Bush, and G. Kruse, "A stochastic user equilibrium path flow estimator," *Transportation Research-C*, vol. 5, no. 3-4, pp. 197-210, 1997.
- [11] L. A. Klein, *Data requirements and sensor technologies for ITS*. Massachusetts: Artech House, 2001.
- [12] A. Toppen, and K. Wunderlich, *Travel Time Data Collection for Measurement of Advanced Traveler Information Systems Accuracy*, DTFH61-00-C-00001. FHWA, US Department of Transportation, 2003.
- [13] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdener, *Travel Time Data Collection Handbook*, FHWA-PL-98-035. FHWA, US Department of Transportation, 1998.
- [14] E. Castillo, J. M. Menendez, and P. Jimenez, "Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations," *Transportation Research-B*, vol. 42, no. 5, pp. 455-481, 2008.
- [15] M.P. Dixon, and L.R. Rilett, "Real-time OD estimation using automatic vehicle identification and traffic count data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 17, no. 1, pp. 7-21, 2002.
- [16] M.P. Dixon, and L.R. Rilett, "Population origin-destination estimation using automatic vehicle identification and volume data," *Journal of Transportation Engineering*, vol. 131, no. 2, pp. 75-82, 2005.
- [17] S.M. Eisenman, and G. F. List, "Using probe data to estimate OD matrices," *The 7th IEEE International Conference on Intelligent Transportation Systems*, pp. 291-296, 2004.
- [18] N. J. Van Der Zijpp, "Dynamic OD-matrix estimation from traffic counts and automated vehicle identification data," *Transportation Research Record 1607*, pp. 87-94, 1997.

- [19] S. R. Hu, and H. T. Liou, "A generalized sensor location model for the estimation of network origin-destination matrices," *Transportation Research-C*, vol. 40, pp. 93-110, 2014.
- [20] X. Zhou, and H. S. Mahmassani, "Dynamic OD demand estimation using automatic vehicle identification data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 105-114, 2006.
- [21] L. Bianco, G. Cofessore, and P. Reverberi, "A network based model for traffic sensor location with implications on O/D matrix estimates," *Transportation Science*, vol. 35, no. 1, pp. 50-60, 2001.
- [22] S. R., Hu, S. Peeta, and C. H. Chu, "Identification of vehicle sensor locations for link-based network traffic applications," *Transportation Research-B*, vol. 43, no. 8-9, pp. 873-894, 2009.
- [23] E. Castillo, A. J. Conejo, R. E. Pruneda, and C. Solares, "Observability in linear systems of equations and inequalities: applications," *Computer and Operations Research*, vol. 34, no. 6, pp. 1708-1720, 2007.
- [24] M. Ng, "Synergistic sensor location for link flow inference without path enumeration: a node-based approach," *Transportation Research-B*, vol. 46, no. 6, pp. 781-788, 2012.
- [25] E. Castillo, A. Calviño, J. M. Menéndez, P. Jiménez, and A. Rivas, "Deriving the upper bound of the number of sensors required to know all link flows in a traffic network," *IEEE Transactions on Intelligent Transportation Systems*, vol.14, No. 2, pp. 761-771, 2013.
- [26] K. S. Chan, and W. H. K. Lam, "Optimal speed detector density for the network with travel time information," *Transportation Research-A*, vol. 36, no. 3, pp. 203-223, 2002.
- [27] B. Yang, and E. Miller-Hooks, "Determining critical arcs for collecting real-time travel time formation," *Transportation Research Record 1783*, pp. 34-41, 2002.
- [28] A. Danczyk, and H. X. Liu, "A mined-integer linear program for optimizing sensor locations along freeway corridors," *Transportation Research-B*, vol. 45, no. 1, pp. 208-217, 2011.
- [29] T. Xing, X. Zhou, and J. Taylor, "Designing heterogeneous sensor networks for estimating and predicting path travel time dynamics: An information-theoretic modeling approach," *Transportation Research-B*, vol.57, pp. 66-90, 2013.
- [30] S. X. He, "A graphical approach to identify sensor locations for link flow inference," *Transportation Research-B*, Vol. 51, pp.65-76, 2013.
- [31] W. H. K. Lam, and H. P. Lo. "Accuracy of O-D estimates from traffic counting stations," *Traffic Engineering and Control*, vol. 31, no. 6, pp. 358-367, 1990.

- [32] P. K. N. Yim, and W. H. K. Lam, "Evaluation of count location selection methods for estimation of O-D matrices," *Journal of Transportation Engineering*, vol. 124, no. 4, pp. 376-383, 1998.
- [33] L. Gan, H. Yang, and S. Wong, "Traffic counting location and error bound in origin-destination matrix estimation problems," *Journal of Transportation Engineering*, vol. 131, no. 7, pp. 524-534, 2005.
- [34] H. Yang, and J. Zhou, "Optimal traffic counting locations for origin-destination matrix estimation," *Transportation Research-B*, vol. 32, no. 2, pp. 109-126, 1998.
- [35] R., Minguez, S. Sanchez-Cambroner, E. Castillo, and P. Jimenez, "Optimal traffic plate scanning location for O-D trip matrix and route estimation in road networks," *Transportation Research-B*, vol. 44, no. 2, pp. 282-298, 2010.
- [36] N., Wang, M. Gentili, and P. Mirchandani, "Model to locate sensors for estimation of static origin-destination volumes given prior flow information," *Transportation Research Record 2280*, pp. 67-73, 2012.
- [37] E. Castillo, J. M. Menendez, and S. Sanchez-Cambroner, "Traffic estimation and optimal counting location without path enumeration using Bayesian network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 3, pp. 189-207, 2008.
- [38] E. Castillo, P. Jiménez, J. M. Menéndez, and M. Nogal, "A Bayesian method for estimating traffic flows based on plate scanning," *Transportation*, vol. 40, no. 1, pp. 173-201, 2013.
- [39] E. Castillo, A. Rivas, M. P. Jiménez, and J. M. Menéndez, "Observability in traffic networks. plate scanning added by counting information," *Transportation*, vol. 39, no. 6, pp. 1301-1333, 2012.
- [40] E. Castillo, M. Nogal, A. Rivas, and S. Sánchez-Cambroner, "Observability of traffic networks. Optimal location of counting and scanning devices," *Transportmetrica B*, vol. 1, no. 1, pp. 68-102, 2013.
- [41] X. Zhou, and G. F. List, "An information-theoretic sensor location model for traffic origin-destination demand estimation applications," *Transportation Science*, vol. 44, no. 2, pp. 254-273, 2010.
- [42] E. Castillo, M. P. Jiménez, J. M. Menéndez, A. Rivas, and I. A. Gallego, "Ternary-arithmetic topological based algebraic method for networks traffic observability," *Applied Mathematical Modelling*, vol. 35, no. 11, pp. 5338-5354, 2011.

[43] C. D. Lewis, Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworth Scientific, London, 1982.

[44] T. Y. Hu, C. C. Tong, S. R. Hu, C. N. Chen, L. W. Chen, P. S. Shiu, W. M. Ho, C. Y. Chen, C. H. Hsu, Y. Y. Jan, T. Y. Wu, and H. Y. Guo, Demonstration and Outreach of the Real-Time and Dynamic Module for Traffic Analysis and Prediction (II). Institute of Transportation, Ministry of Transportation and Communications. Project No. 97-TDB001, Taipei, Taiwan, 2008.

[45] Sinotech Engineering Consultants. The First Stage Comprehensive Planning of Yanchao Light Rail Transit System in the Kaohsiung Metropolitan Area. Kaohsiung County Government, Kaohsiung County, 2008

Contacts

For more information:

Dr. Srinivas Peeta
Purdue University - Discovery Park
3000 Kent Ave.
West Lafayette, IN 47906
(765) 496-9724
Fax Number
peeta@purdue.edu
www.purdue.edu/dp/nextrans

NEXTRANS Center
Purdue University - Discovery Park
3000 Kent Ave.
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9724
www.purdue.edu/dp/nextrans

Dr. Shou-Ren Hu
National Cheng Kung University
1, University Road, Tainan City, 70101, Taiwan
+886-6-2757575 ext. 53203
+886-6-2753882
shouren@mail.ncku.edu.tw
www.tcm.ncku.edu.tw