Optimal Traffic Plate Scanning Location for O-D Trip Matrix and Route estimation in Road Networks

R. Mínguez[1], S. Sánchez-Cambronero[2], E. Castillo[3], and P. Jiménez[2]

[1] *Department of Mathematics, University of Castilla-La Mancha, Spain*

[2] *Department of Civil Engineering, University of Castilla-La Mancha, Spain*

[3] *Department of Applied Mathematics and Computational Sciences, University of Cantabria, 39005 Santander, Spain*

Abstract

During the last decade, there has been a substantial interest in how to determine the optimal number and locations of traffic counters for origin-destination ($\mathcal{OD}$) trip matrices estimation. On the contrary, the optimal allocation of plate scanning devices has received very limited attention, even though several authors have demonstrated that plate scanning (route identification) techniques are much more informative than those based on traditional link count information. This paper provides techniques for obtaining the optimal number and location of plate scanning devices for a given prior $\mathcal{OD}$ distribution pattern under different situations, i.e. maximum route identifiability or budget constraints. Two rules analogous to the counting location problem are developed, and several integer linear programming models fulfilling these rules are proposed. The proposed methods are finally illustrated by their application into Nguyen-Dupuis and Cuenca networks.

Key Words: Plate scanning, OD trip matrix estimation, route identification, traffic count location problem.

1 Introduction

Origin-destination ($\mathcal{OD}$) trip matrices estimation is an essential process for efficient traffic control and management. Existing approaches combine, in an efficient way, the information obtained from link counts with other information (prior or target matrix, socioeconomic data, etc.). The most widely used methods for $\mathcal{OD}$ matrix estimation are based, among others, on mathematical programming techniques, such as, least squares methods (Cascetta and Nguyen (1988)), entropy or information based methods (Willumsen (1984)), classical statistical techniques (Hazelton (2000)), statistical methods based on Bayes theorem (Maher (1983)) or recent works which use Bayesian Networks to predict traffic flows (Sun et al. (2006); Castillo et al. (2008c,d)).

Data needed for these models are:

- A prior $\mathcal{OD}$ matrix given by the $\mathcal{OD}$ flows $\mathbf{T}^0 = [\ldots, t_i^0, \ldots]^T$, which is usually obtained from an out-of date matrix resulting from other studies or methods.
- The observed link flows $\hat{\mathbf{V}} = [\ldots, \hat{v}_a, \ldots]^T$, where a refers to links.
- The assignment map which describes the relationship between the observed link flows and the $\mathcal{OD}$ matrix:
$$v_a = \sum_r \sum_i p_r^i \delta_{ar}^i t_i, \tag{1}$$
or in matrix form:
$$\mathbf{V} = \mathbf{P}\boldsymbol{\Delta}\mathbf{T}, \tag{2}$$
where $\mathbf{V} = [\ldots, v_a, \ldots]^T$ are the link flows, $\mathbf{T} = [\ldots, t_i, \ldots]^T$ are the $\mathcal{OD}$ flows, both represented as column vectors, $\boldsymbol{\Delta} \equiv \{\delta_{ar}^i\}$ is the path-link incidence matrix, and $\mathbf{P} \equiv \{p_r^i\}$ is a matrix defining the probabilities of the users to select the different paths (routes) associated with all $\mathcal{OD}$ pairs.

Due to the importance of the traffic count locations to obtain good traffic flow predictions, several authors have focused on solving the problem of determining the optimal number and allocation of traffic counts. The target is to identify the number and links to be observed over a transportation network to reproduce a prior $\mathcal{OD}$ matrix as exactly as possible together with route flows. Castillo et al. (2008a) discuss the observability problem, and Yang et al. (1991), based on the maximal possible relative error (MPRE), examined the reliability of the estimated $\mathcal{OD}$ matrix with respect to the number and locations of counting points in the network. Yang and Zhou (1998) addressed the problem of how to determine the optimal number and locations of traffic counting points for a given prior $\mathcal{OD}$ distribution pattern, and proposed four location rules:

1. **$\mathcal{OD}$ covering rule.** Traffic counting points should be located so that a certain portion of trips between any $\mathcal{OD}$-pair will be observed.
2. **Maximal flow fraction rule.** For a particular $\mathcal{OD}$ pair, traffic counting points on a road network should be located at the links with the highest fraction of $\mathcal{OD}$ flow.
3. **Maximal flow-intercepting rule.** Given a certain number of links to be observed, the chosen links should intercept as many flows as possible.
4. **Link independence rule.** The traffic counting points should be located on the network so that the resulting traffic counts on all chosen links are linearly independent.

Yim and Lam (1998) recommend an optimal data collection method based on the results of the $\mathcal{OD}$ estimation problem sensitivity analysis. Chung (2001) added the purchasing and installing detector costs into the count location problem, proposing two different but equivalent models: (i) budget minimization subject to complete $\mathcal{OD}$ coverage, and (ii) maximization of $\mathcal{OD}$ coverage subject to budget limitations. Bianco et al. (2001) developed a two-stage procedure, which in its first stage minimizes the cost of traffic sensors installation on the network nodes, obtaining coefficients to get the observed flows over the network from the sensor information, and in the second stage uses common $\mathcal{OD}$ matrix estimation models. Bierlaire (2002) proposed a new model based on the total demand scale as a measure of the quality of the estimated $\mathcal{OD}$ flows from traffic counts. Ehlert et al. (2006) proposed several extensions to previous existing methods of practical relevance: (i) consideration of previously existing detectors, and (ii) adding the information content of the prior $\mathcal{OD}$ flows into the count location problem. Yang et al. (2006) considered that the traffic flow measurement is carried out at screen lines, through which all traffic movements with the origin, on one side of the screen line, and the destination, on the other, are intercepted. The basis of the model consists of the interesting idea of allocating the traffic counting stations so as to separate as many $\mathcal{OD}$ pairs as possible.

Note that the models above deal only with traffic counting points in links, and the only information available corresponds to a subset $\{\hat{v}_a | a \in \mathcal{OL} \subseteq \mathcal{A}\}$ of total link flows, where $\mathcal{A}$ is the set of links, and $\mathcal{OL}$ is the subset of observed links. However, Castillo et al. (2008b) combined the trip matrix and path flow reconstruction problem based on plate scanning and link flow observations, showing that the plate scanning method leads to better estimates in terms of $\mathcal{OD}$ and route flows because it provides much more information than link flow observations. Castillo et al. (2008b) also provided a method for selecting minimal sets of links to be scanned for predicting exact traffic flows for a given enumeration of possible routes.

Hellinga and Van Aerde (1994) use probe vehicles to determine network $\mathcal{OD}$ trip demands. In particular, the work is focused on determining the number of probe vehicles required to obtain some minimum quality in the estimation of the network flow, i.e. define the best population level for market penetration for each $\mathcal{OD}$ pair. Some authors, as for example Ben-Akiva et al. (1994), pointed out the importance of the market penetration in the quality of the results, in the sense that if the market penetration of probe vehicles is not exactly the same for all $\mathcal{OD}$ pairs, the estimation will be biased, and note that due to the spatial heterogeneity of urban land used patterns, the market penetration rates of probe vehicles may differ considerably among various $\mathcal{OD}$ pairs, leading to under or over estimation of the departure rates for some $\mathcal{OD}$ pairs.

Bianco et al. (2006) addressed the so called Sensor Location Problem, i.e. determine the minimum number of counting sensors on nodes of the network in order to obtain all arc flow volumes making a combinatorial analysis. Note that dealing with node sensors is more complicated than link sensors but, on the other hand, it has similar practical possibilities for real networks

Another research trend on locating active sensors is followed by Gentili and Mirchandani (2005), who developed two decision models under three different scenarios for locating active sensors so that: i) the minimum number of links and their locations in order to obtain flow volumes on specified paths is determined, and ii) if there exist total link flow counters (passive sensors), the question about how many and where should active sensors be located in order to get the maximum information on flow volumes on specified paths is answered. However, Gentili and Mirchandani (2005) use path-ID sensors and their work is based on the assumption that path-ID sensors on a given arc of the network is able to measure the flow of the paths (routes) to which the arc belongs. This is a strong assumption from the practical point of view, as indicated by the authors, because that information is not available from regular vehicles unless you stop them and ask drivers about their route. Nevertheless, as Gentili and Mirchandani (2005) pointed out, this assumption is reasonable for certain clases of vehicles, as for example, buses and trucks which normally has specified routes. Furthermore, toll tags which are monitored upstream (when entering the subnetwork) and downstream (when leaving the network), and license plate and blue-tooth readers can provide also route information.

The models proposed in this paper enrich the model proposed by Castillo et al. (2008b) considering alternative mathematical programming formulations to take into account some practical considerations. For example, the model proposed by Castillo et al. (2008b) worked for full observability, but in many situations the number of possible scanners is limited by budget constraints, or scanning devices can fail. Our models allow considering these limitations, constituting a novel and important contribution specially from the practical point of view.

The aim of this paper is to gain insight into the traffic plate scanning location problem, and to propose several methods for optimizing the location of traffic scanning devices considering different situations: (i) budget minimization subject to complete route identifiability, (ii) maximum route identifiability subject to budget constraints, and (iii) consideration of existing plate scanners.

The rest of the paper is organized as follows. Section 2 introduces the traffic plate scanning location problem and presents the location models and the corresponding implementations. Section 3 illustrates the models with an example of application. Section 4 presents the Cuenca network, that is, a realistic example. Finally, Section 5 provides some conclusions.

2 The traffic plate scanning location problem

In this section the plate scanning location problem is introduced. Firstly, the plate scanning technique is summarized. Next, proposed rules are described and some location models together with some implementation details are provided.

2.1 Plate scanning to predict traffic flow

Consider a traffic network $(\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of links. From $\mathcal{N}$ one can distinguish two subsets of nodes, $\mathcal{O}$ and $\mathcal{D}$, corresponding to origins and destinations, respectively.

Let $\mathcal{OL} \in \mathcal{A}$ be the set of $n_{sc} \neq 0$ observed links, containing information about plate number (I_k), link (a_k), and time (t_k) of registration, i.e. the information provided consists of the set:

$$\mathcal{SI} \equiv (I_k, a_k, t_k); k = 1, \ldots, n_{\text{o}}, \tag{3}$$

where k is the k-th plate scanned, and n_{o} is the total number of plates scanned. Note that $a_k \in \mathcal{OL}$.

As shown in Castillo et al. (2008b), the plate scanning technique consists of registering plate numbers and the corresponding times of the vehicles at some subset of links to reconstruct vehicle routes by identifying identical plate numbers at different locations and times. Castillo et al. (2008b) also pointed out that the set of links to be scanned must be chosen adequately so that all different combinations of scanned links belong to a unique route, which means that the scanning process allows identifying uniquely the path of any scanned user. This allows us to summarize the scanned observations as

$$\hat{f}_r : \; r \in \mathcal{OR}, \tag{4}$$

where $\mathcal{OR}$ is the set of observed routes $\mathcal{OR} \equiv 1, \dots, n_r \in \mathcal{R}$, $\mathcal{R}$ is the set of all considered routes and n_r is the number of different C_r sets of scanned links which allows us identifying uniquely every observed route r.

This information is used for route flow estimation by means of the following model:

$$\underset{f_r;\, \forall r \in \mathcal{R}}{\text{Minimize}} \quad \sum_{\forall x \in \mathcal{R}} \sum_{\forall y \in \mathcal{R}} (f_x - f_x^0)\gamma_{xy}(f_y - f_y^0) \;\; , \tag{5}$$

subject to

$$\hat{f}_r \;=\; f_r; \; \forall r \in \mathcal{OR} \tag{6}$$

$$\hat{v}_a \;=\; \sum_{\forall r \in R} \delta_a^r f_r; \; \forall a \in \mathcal{OL}, \tag{7}$$

where f_r^0 and $\hat{f}_r$ are the prior and observed flows through route r, respectively, $\hat{v}_a$ is the observed flow in link a, and γ_{xy} are the weights (normally the elements of the inverse of the variance-covariance matrix). Note that constraint (7) allows us including in the estimation model the total link flows, which are also known from the scanning process. This constraint includes redundant information for links where all the passing routes are observable, but it improves the prediction of unobservable route flows.

Note that using this approach, the aim is to identify uniquely as many routes as possible through scanner devices in links. Castillo et al. (2008b) proposed a binary linear programming which selects the minimum number of links to distinguish the users of any pair of routes. The plate number observations over this set of links, supposing that the scanning process is error free[1], allow us to have a full identifiability of all path flows. The problem is formulated as follows:

$$\underset{\boldsymbol{z}}{\text{Minimize}} \quad n_{\text{sc}} = \sum_{a \in \mathcal{A}} z_a \tag{8}$$

subject to

$$\sum_{a \in \mathcal{A}} z_a d(r, r_1, a) \;\geq\; 1; \quad \forall (r, r_1) | r \neq r_1, \tag{9}$$

$$\sum_{a \in \mathcal{A}} z_a \delta_a^r \;\geq\; 1; \quad \forall r, \tag{10}$$

where z_a is a binary variable such that it takes value 1 if the link a is scanned, and 0, otherwise, r and r_1 are paths, $\boldsymbol{\Delta}$ is the incidence matrix with elements δ_a^r, and $d(r, r_1, a)$ are defined by

$$\delta_a^r \;=\; \begin{cases} 1 & \text{if} \quad \text{path r contains link } a \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$d(r, r_1, a) \;=\; \begin{cases} 1 & \text{if} \quad \delta_a^r \neq \delta_a^{r_1} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Note that constraint (9) guarantees that the selected subset of scanned links is able to distinguish the users of any given pair of paths r and r_1 based on their scanned plate numbers, i.e. there

[1] Due to the design character of the optimal plate scanning location problem, it is reasonable to assume that the scanning process is error free. This assumption may be no longer valid in the estimation process, however, in Castillo et al. (2008b) several procedures to avoid this shortcoming are presented.

exists at least one scanned link which is in path r and not in path r_1 or vice-versa. In addition, constraint (10) ensures that any route or path contains at least one scanned link, and therefore information, not only of all $\mathcal{OD}$ pairs but all the routes, becomes available.

The optimum n_{sc} (n^*_{sc}) is the minimum number of scanning device positions provided by model (8)-(10) that allows estimating the $\mathcal{OD}$ matrix exactly (error free) if all possible routes between any $\mathcal{OD}$ pair have been considered. The main shortcoming of model (8)-(10) is that it does not include budget considerations, so in case the number of possible links to be installed is limited or the scanning device costs are different between links this method is not suitable to get the best possible scanner locations reproducing as exactly as possible the $\mathcal{OD}$ matrix with minimum cost.

Hereinafter we will refer to the apparatus which allows identification of vehicles as *"scanning device"* instead of *"camera"*. The reason is simple, the route identification method was initially proposed to be used with plate scanning information but the method can be applied using any identification method capable of distinguishing uniquely between vehicles, for example, using transmitters installed inside vehicles, GPS, etc.

2.2 Location rules

In real life, the true error or reliability of an estimated $\mathcal{OD}$ matrix is unknown. Yang et al. (1991) proposed the concept of maximal possible relative error (MPRE), which represents the maximum possible relative deviation of the estimated $\mathcal{OD}$ matrix from the true one. Based on this concept Yang and Zhou (1998) proposed several location rules. In this paper, since the scanner location problem is of different nature to the counting location problem based on link flows, we derive analogous rules based on prior link and flow values and the following measure (RMSRE, root mean squared relative error):

$$\text{RMSRE} = \sqrt{\frac{1}{m}\sum_{i\in\mathcal{I}}\left(\frac{t_i^0 - t_i}{t_i^0}\right)^2}, \tag{13}$$

where t_i^0 and t_i are the prior and estimated flow of $\mathcal{OD}$-pair i, respectively, and m is the number of $\mathcal{OD}$-pairs belonging to the set $\mathcal{I}$. Note that we propose this alternative formulation because our model uses prior information and we also assume that the real network flows will be similar to those given by the prior approach, therefore our location rules try to reproduce through an estimation method the prior $\mathcal{OD}$ pair flows as exactly as possible, when other information is not available. Since the prior $\mathcal{OD}$ pair flows t_i^0 are known, they are used to calculate the relative error.

Given the set R of all possible routes, any of them corresponding to a unique $\mathcal{OD}$ pair, if R_i is the set of routes belonging to $\mathcal{OD}$-pair i, we have $t_i^0 = \sum_{r\in R_i} f_r^0$, and then the RMSRE can be expressed as:

$$\text{RMSRE} = \sqrt{\frac{1}{m}\sum_{i\in I}\left(\frac{t_i^0 - \sum_{r\in R_i} f_r^0 y_r}{t_i^0}\right)^2}, \tag{14}$$

where y_r is a binary variable equal to one if route r is identified uniquely (observed) through the scanned links, and zero otherwise. Note that the minimum possible RMSRE-value corresponds to $y_r = 1$; $\forall r \in R$, where $t_i = t_i^0$ and RMSRE=0.

However, if $n_{\text{sc}} = \sum_{\forall r\in R} y_r \le n_r$ then RMSRE> 0, and then, one interesting question is: how do we select the routes to be observed so that the RMSRE is minimized? From (14) we obtain

$$m \times \text{ RMSRE}^2 = \sum_{i\in I}\left(1 - \sum_{r\in R_i}\frac{f_r^0}{t_i^0} y_r\right)^2, \tag{15}$$

where it can be deduced that the bigger the value of $\sum_{r\in R_i}\frac{f_r^0}{t_i^0} y_r$ the lower the RMSRE. If the set of routes is partitioned into observed ($\mathcal{OR}$) and unobserved ($\mathcal{UR}$) routes associated with $y_r = 1$

or $y_r = 0$, respectively, (15) can be reformulated as follows

$$m \times \text{ RMSRE}^2 = \sum_{i \in I} \left(1 - \sum_{r \in (R_i \cap \mathcal{OR})} \frac{f_r^0}{t_i^0} \right)^2 = \sum_{i \in I} \left(\sum_{i \in (R_i \cap \mathcal{UR})} \frac{f_r^0}{t_i^0} \right)^2, \tag{16}$$

so that routes to be observed ($y_r = 1$) should be chosen minimizing (16).

The main shortcoming of equations (15) or (16) is their quadratic character which makes the RMSRE minimization problem to be nonlinear. Alternatively, the following RMARE (root mean absolute value relative error) based on the mean absolute relative error norm can be defined:

$$\text{RMARE} = \frac{1}{m} \sum_{i \in I} \left| \frac{t_i^0 - t_i}{t_i^0} \right| = \frac{1}{m} \sum_{i \in I} \left| \frac{t_i^0 - \sum_{r \in R_i} f_r^0 y_r}{t_i^0} \right|, \tag{17}$$

and since the numerator is always positive for error free scanners ($0 \leq \sum_{r \in R_i} f_r^0 y_r \leq t_i^0;\ \forall i \in I$), the absolute value can be dropped, so that the RMARE as a function of the observed and unobserved routes is equal to

$$\text{RMARE} = 1 - \frac{1}{m} \left(\sum_{i \in I} \sum_{r \in (R_i \cap \mathcal{OR})} \frac{f_r^0}{t_i^0} \right) = \frac{1}{m} \left(\sum_{i \in I} \sum_{r \in (R_i \cap \mathcal{UR})} \frac{f_r^0}{t_i^0} \right), \tag{18}$$

which implies that minimizing the RMARE is equivalent to minimizing the sum of relative route flows of unobserved routes, or equivalently, maximize the sum of relative route flows of observed routes. Note that this result derives in a rule that can be denominated the **Maximum Relative Route Flow** rule.

The above location rule has been derived by supposing that the prior trip distribution matrix is reasonably reliable and close to the actual true value, because the accuracy of the prior matrix has a great impact on the estimates of the true $\mathcal{OD}$ matrix. This is clear for a small number of scanned links, where the reliability of the $\mathcal{OD}$ estimations is dependent on the quality of the priors, but for increasing number of scanned links, this dependency decreases substantially, and vanishes for full identifiability of routes.

Note also that even though the knowledge of prior $\mathcal{OD}$ pair flows could be difficult in practical cases, the aim of the proposed formulation is determining which $\mathcal{OD}$ flows are more important than others in order to prioritize their real knowledge. In fact the prior $\mathcal{OD}$ matrix is only used as a weighting factor for O-D pairs flows. Alternatively, the MPRE criterion proposed by Yang et al. (1991) could be used for those cases where a prior O-D matrix is unavailable. Note that existing methods such us the one proposed by Yang and Zhou (1998), which is based on Hodgson (1990) and according to their maximal flow-interception rule, also use a flow pattern associated with a prior O-D matrix.

Analogously to Yang and Zhou (1998) who proposed the $\mathcal{OD}$ **covering rule** related to link observations, the equivalent rule could be stated in terms of route flows, so that at least one route for any $\mathcal{OD}$ pair should be observed. However, we consider more relevant from the practical point of view the problem of optimal route observation subject to budget constraints, which means that in many situations there are routes and/or $\mathcal{OD}$ pairs uncovered by the set of observed links.

Note also that since the proper identifiability of routes must be made through plate scanner devices in links, an additional rule related to links should be considered, which states that scanned links must allow us to identify uniquely the routes to be observed ($y_r = 1$) from all possible routes being considered. This rule can be denominated the **Full Identifiability of Observed Path Flows** rule.

2.3 Location models

In this section several models accounting for the derived rules and considering different assumptions are developed.

2.3.1 Budget considerations and partial covering

The first location model to be proposed in this paper considers full route observability, i.e. RMSRE= 0, but including budget considerations. Note that in the transport literature, each link, denoted by a, is considered independently of the number of lines of each link. Obviously, if we are trying to scan plate numbers it is likely to be more expensive in links with higher number of lanes (usually the number of scanning devices is bigger). Thus, problem (8)-(10) is reformulated as:

$$\mathrm{M}_1 = \underset{\boldsymbol{z}}{\text{Minimize}} \sum_{a\in\mathcal{A}} \mathcal{P}_a z_a \tag{19}$$

subject to (9)-(10), where $\mathcal{P}_a$ is the cost of plate scanning link a. Note that constraint (9) forces to select the scanned links so that every route is uniquely defined by a given set of scanned links (every row in the incidence matrix $\boldsymbol{\Delta}$ is different from the others) and (10) ensures that at least one link for every route is scanned (every row in the incidence matrix $\boldsymbol{\Delta}$ contains at least one element different from zero). Both constraints (9) and (10) force the *maximum relative route flow* and *full identifiability of observed path flows* rules to hold. Note also that all $\mathcal{OD}$ pairs are totally covered. In addition, this model allows the estimation of the required budget resources $\mathcal{B}^* = \sum_{a\in\mathcal{A}} \mathcal{P}_a z_a^*$ for covering all $\mathcal{OD}$ pairs in the network.

However, budget is limited in practice, meaning that some routes or even some $\mathcal{OD}$ pairs may remain uncovered, consequently based on the *maximum relative route flow* rule given by (18) the following model is proposed:

$$\mathrm{M}_2 = \underset{\mathbf{y},\boldsymbol{z}}{\text{Maximize}} \sum_{\forall i\in I}\sum_{r\in R_i} \frac{f_r^0}{t_i^0} y_r \tag{20}$$

subject to

$$\sum_{a\in\mathcal{A}} z_a d(r,r_1,a) \geq y_r; \quad \forall (r,r_1)|r\neq r_1, \tag{21}$$

$$\sum_{a\in\mathcal{A}} z_a \delta_a^r \geq y_r; \quad \forall r, \tag{22}$$

$$\sum_{a\in\mathcal{A}} \mathcal{P}_a z_a \leq \mathcal{B}, \tag{23}$$

where f_r^0 and t_i^0 are the route and $\mathcal{OD}$-pair flows, respectively, of a prior out-of-date $\mathcal{OD}$ matrix, y_r is a binary variable equal to 1 if route r can be distinguished from others and 0 otherwise, z_a is a binary variable which is 1 if link a is scanned and 0 otherwise, and $\mathcal{B}$ is the available budget.

Constraint (21) guarantees that the route r is able to be distinguished from the others if the binary variable y_r is equal to 1. Constraint (22) ensures that the route which is able to be distinguished contains at least one scanned link. Both constraints (21) and (22) force the model to fulfill the *full identifiability of observed path flows* rule, i.e. all routes such that $y_r = 1$ can be uniquely identified using the scanned links. It is worthwhile mentioning that using y_r instead of 1 in the right hand side of constraints (21) and (22) immediately converts into inactive the constraint (10) for those routes the flow of which are not fully identified.

The $\mathcal{OD}$ coverage rule has become the object of the optimization itself and it will be ensured or not depending on the available budget $\mathcal{B}$, i.e. depending on whether or not constraint (23) becomes active. For example, if the available budget equals the optimal value of the objective function given by model M_1 ($\mathcal{B} = \mathcal{B}^*$), model M_2 provides full $\mathcal{OD}$ coverage.

From the practical point of view constraint (21) can be replaced by the following alternative and equivalent constraint

$$\sum_{a\in\{\mathcal{A}\}} (\delta_a^r + \delta_a^{r_1})(1-\delta_a^r\delta_a^{r_1})z_a \geq 1; \quad \forall (r,r_1)|r<r_1 \text{ and } \sum_{a\in\mathcal{A}} \delta_a^r\delta_a^{r_1} > 0 \tag{24}$$

which has the following important advantages:

1. The parameter $d(r, r_1, a)$, using a huge amount of memory, needs not be stored.

2. It does not consider constraints relating routes without common routes $(\sum_{a \in \mathcal{A}} (\delta_a^{r_1} \delta_a^r) > 0)$.

3. It does not compare routes r and r_1 twice, because of the $<$ instead of the $\neq$ relation.

Alternatively, for those cases where no prior O-D matrix information is available and based on the MPRE concept derived by Yang et al. (1991), the following model is proposed:

$$\mathrm{M}_3 = \underset{\mathbf{y}, \boldsymbol{z}}{\text{Maximize}} \quad \sum_{\forall i \in I} \sum_{r \in R_i} y_r \tag{25}$$

subject to (21)-(23). Note that in this case the number of observed routes is maximized but no information is given about the prior route flows. This problem maximizes the number of routes that are exactly observed for a given budget without considering the amount of flow intercepted by those routes, however this solution is not necessarily optimal in terms of route and/or $\mathcal{OD}$ observed flows by the selected scanners. Nevertheless, this problem rarely happens in practice since traffic planners have information about which routes are more relevant to observe in terms of flow interception.

2.3.2 Practical considerations

Consideration of previously existing devices For real moderate to large problems, complete $\mathcal{OD}$ coverage is often imposible due to the number of origins and destinations in the network. In addition, the network configuration and its behavior may have changed due to several reasons: consideration of new $\mathcal{OD}$ pairs, construction of new links between nodes, or consideration of new or alternative routes between $\mathcal{OD}$ pairs. Under this circumstance, if some detectors are already installed and additional budget is available to increase observability in the network, new scanning devices should be located on unobserved links ($a \in \mathcal{UL}$) leading to the following models associated with models M_1 and M_2, respectively:

$$\mathrm{M}_4 = \underset{\boldsymbol{z}}{\text{Minimize}} \quad \sum_{a \in \mathcal{UL}} \mathcal{P}_a z_a \tag{26}$$

subject to (9)-(10) and

$$z_a = 1; \quad \forall a \in \mathcal{OL}, \tag{27}$$

where $\mathcal{OL}$ is the set of already observed links (links with scanning devices already installed), and

$$\mathrm{M}_5 = \underset{\mathbf{y}, \boldsymbol{z}}{\text{Maximize}} \quad \sum_{\forall i \in I} \sum_{r \in R_i} \frac{f_r^0}{t_i^0} y_r \tag{28}$$

subject to (21)-(22) and

$$\sum_{a \in \mathcal{UL}} \mathcal{P}_a z_a \leq \mathcal{B} \tag{29}$$

$$z_a = 1; \quad \forall a \in \mathcal{OL}. \tag{30}$$

Note that $\mathcal{A} = \mathcal{OL} \cup \mathcal{UL}$ and $\mathcal{OL} \cap \mathcal{UL} = \emptyset$, i.e. the sets of observed and unobserved links are disjoint and complementary.

Obtain redundant information to avoid mistakes Unfortunately, it is relatively easy for a scanning device to fail identifying the plates of some users. Therefore, these users, which belong to an specific route, can be assigned to another one (see section 3 for an illustrative example).

To deal with this problem in a simple way, redundant information can be included to cover possible failures by means of the following binary linear programming problem:

$$\mathrm{M}_6 = \underset{\boldsymbol{z}}{\text{Minimize}} \quad \sum_{a\in A}\sum_{i=1}^{n_a} \mathcal{P}_{a,i} z_{a,i} \tag{31}$$

subject to

$$\sum_{a\in\mathcal{A}|\delta_a^{r_1}\neq\delta_a^r} (\sum_{i=1}^{n_a} z_{a,i}) \quad \geq \quad q; \quad \sum_{a\in\mathcal{A}} (\delta_a^{r_1}\delta_a^r) > 0; \;\; \forall (r, r_1)|r < r_1 \tag{32}$$

$$\sum_{a\in A}\sum_{i=1}^{n_a} z_{a,i}\delta_a^r \quad \geq \quad 1; \quad \forall r, \tag{33}$$

where $z_{a,i}$ is a binary variable such that it takes value 1 if the link a is scanned using scanner i, and 0 otherwise, n_a is the maximum number of duplicate scanning devices to be installed in each link, and q is the redundancy level, i.e. the minimum number of failures in the scanning devices which are necessary for confusing two routes. Note that the larger the value of q, the larger the reliability of the system.

Note that the model above does not consider scanner devices failure rates, which should be included in the model somehow in order to consider if it is more convenient to spend the money in increasing the redundancy ratio in scanner locations, or to allocate new scanners in unobserved links. The solution of this problem is out of the scope of this paper and is proposed for further research.

2.4 Model implementation

The proposed models M_1, M_2, M_3, M_4 and M_5 are mixed-integer linear programming problems (MIP), which are known to require dramatically more mathematical computation than those for similar sized pure linear programs. For problems with integer variables, Branch and Bound (BB) and Gomory Cuts (GC) methods are available. The BB is the most used and usually the most computationally efficient technique. However, more recently, a hybrid highly efficient technique of both methods, denominated branch-cut (BC), is being successfully applied (see Castillo et al. (2001)).

The solver used in this paper is CPLEX that uses a branch and bound algorithm with cuts (BC) which solves a series of LP subproblems. Note that the branching strategy used by this solver iteratively decreases the upper bound and increases the lower bound of the optimal solution, being the difference a measure of the proximity of the current solution to the optimal solution if it exists. The lower bound is obtained relaxing the integrality constraints, whereas the upper bound is obtained from any solution satisfying the integrality constrains. The solution holding the integrality constraints is looked for through an enumeration tree which increases exponentially with the number of integer variables. Due to this, reducing the integer variables is important from the practical point of view, and in models M_2, M_3 and M_5 it is possible, since the corresponding models with relaxed variables $y_r; \forall r \in \mathbb{R}$ and adding constraints

$$0 \leq y_r \leq 1; \forall r \in R, \tag{34}$$

provide a solution which satisfies the integrality constraints of variables $y_r \in \{0, 1\}; \forall r \in R$, reducing considerably the number of integer variables and the computational burden.

This is an important practical result that is precisely given in the following theorem.

Theorem 1 *If in the problem M_2 given by (20)-(23) and assuming, without loss of generality, $f_r \neq 0; \forall r$, the y_r variables are relaxed to continuous variables in the closed interval $[0,1]$, the optimal solution and the corresponding optimal value remains invariant (Schrijver (1987)).*

Proof. Assume that the optimal solution of problem M_2 with the variables y_r relaxed to be continuous in the interval $[0,1]$ is attained for $0 < y_s < 1$, where s is a given index value. We will prove that this is not possible. Since the left hand sides of constraints (21)-(22) can only get integer values since $\mathbf{z}, \mathbf{d}$ and $\mathbf{\Delta}$ take binary $\{0,1\}$-values, that is,

$$\sum_{a \in \mathcal{A}} z_a d(r, r_1, a) \in \{0,1,2,3,\ldots\}; \; \forall r_1 \neq r, \tag{35}$$

$$\sum_{a \in \mathcal{A}} z_a \delta_a^r \in \{0,1,2,3,\ldots\}, \tag{36}$$

and since $y_s > 0$, the associated constraints (20)-(22) are equivalent to constraints (9)-(10), i.e., with $y_s = 1$. Since y_s appears in no other constraint, and the objective function (20) increases with y_s because f_s is positive, a larger value of the objective function is obtained for $y_s = 1$ without changing the constraints from inactive to active or vice-versa. In other words, if in the assumed solution y_s is replaced by 1 while keeping unchanged the remaining variable values, we get another feasible solution with a better value of the objective function. Thus, the solution with $y_s < 1$ is not optimal. This concludes the proof. ∎

Note that the number of constraints in (9) and (21) increases exponentially with the number of routes making the use of CPLEX not possible for large size networks. In those cases alternative solvers and/or more computational resources, or even a more ad hoc branch and cut procedure could be suitable to solve the problem. Nevertheless, since the proposed models are devoted to scanning location design, computational times are not especially critical if they remain below a reasonable bound.

Note also that we use available mathematical programming solvers instead of a tailor-made ad hoc branch and cut procedure, which is almost always better than general purpose software, because the latter is more reliable, and besides, current state-of-the-art nonlinear programming solvers incorporate state-of-the-art treatment of sparse matrices and are both numerically robust and computationally efficient. Additionally, they may include parallelization features. Given the aforementioned characteristics, we think that these solvers constitute appropriate tools for model/algorithm development and model/algorithm testing.

3 Illustrative Example: the Nguyen-Dupuis network.

In order to understand the behavior of the proposed methods, consider the Nguyen-Dupuis network topology, with 13 nodes and 38 bidirectional links as shown in Figure 1. A total of 18 $\mathcal{OD}$ pairs are considered, whose origins and destination nodes are outlined in the figure and the corresponding origin-destination nodes are provided in Table 1 together with the corresponding out-of-date $\mathcal{OD}$ flows. Routes and associated route flows are shown in Table 2. They were obtained using a MNL logit assignment model (the link parameters used in this example are those exposed in Castillo et al. (2008b)) for the $\mathcal{OD}$ flows provided in Table 1.

To identify all path flows, a minimum number of links have to be scanned. To this end, problem (8)-(10) has been solved, resulting 18 links to be observed:

$$\{1,2,3,5,8,9,11,13,18,20,21,22,23,29,31,33,34,36\}, \tag{37}$$

which have been also outlined in Figure 1. Note that a set of 18 scanned links of a total of 38 possible are needed. Alternatively, model M_1 for $\mathcal{P}_a = 1; \; \forall a \in \mathcal{A}$ could have been solved resulting in the same solution.

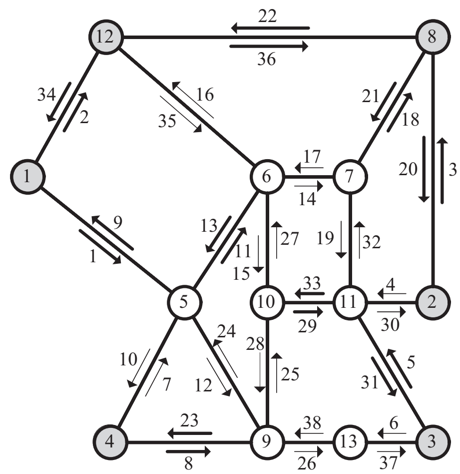

Figure 1: Nguyen-Dupuis traffic network used in the Example of application.

Table 1: $\mathcal{OD}$ pairs and corresponding flows used in the Nguyen-Dupuis example.

$\mathcal{OD}$	Origin $\mathcal{O}$	Destination $\mathcal{D}$	Prior or out-of-date $\mathcal{OD}$ flow
1	1	2	210
2	1	3	430
3	1	8	320
4	2	1	210
5	2	4	320
6	2	12	50
7	3	1	430
8	3	4	110
9	3	12	40
10	4	2	320
11	4	3	110
12	4	8	210
13	8	1	320
14	8	4	210
15	8	12	60
16	12	2	50
17	12	3	40
18	12	8	60

The model proposed by Gentili and Mirchandani (2005) with zero count information, i.e. none arc flow is known and path-ID sensors have to be located to determine all path flows, is as follows:

$$\underset{x_i}{\text{Minimize}} \quad \sum_{i=1}^{m} x_i, \tag{38}$$

subject to

$$\begin{aligned} \sum_{i=1}^{m} b_{ij} \geq & \quad 1;\ j = 1, \ldots, p \\ x_i \in & \quad \{0,1\};\ i = 1, \ldots, m, \end{aligned} \tag{39}$$

where x_i is 1 if a path-ID sensor is located in arc i and 0 otherwise, b_{ij} is 1 if the arc i belongs to route j, m is the number of arcs, and p is the number of routes. If this model is applied in the Nguyen-Dupuis network example, the path-ID sensors are located in the following arcs:

$$\{6, 14, 17, 22, 29, 33, 36, 37\}, \tag{40}$$

which is a completely different solution from (37). Note in Table 2 that solution in (40) ensures that using those 8 path-ID sensors, all routes have one of their arcs covered by one path-ID sensor, but nothing is said about how the sensor obtains route flows from link flows.

Table 2: Set of $\mathcal{OD}$-pairs and paths (routes) considered in the Nguyen-Dupuis network

Route (r)	Links (a)	Prior route flow (f_r)
$\mathcal{OD}$ pair 1		
1	1 11 14 18 20	27.87
2	2 35 14 18 20	13.87
3	2 36 20	168.26
$\mathcal{OD}$ pair 2		
4	1 11 14 19 31	67.84
5	1 11 15 29 31	76.72
6	1 12 25 29 31	31.22
7	1 12 26 37	182.29
8	2 35 14 19 31	33.75
9	2 35 15 29 31	38.18
$\mathcal{OD}$ pair 3		
10	1 11 14 18	42.47
11	2 35 14 18	21.13
12	2 36	256.40
$\mathcal{OD}$ pair 4		
13	3 21 17 13 9	29.17
14	3 21 17 16 34	15.44
15	3 22 34	165.40
$\mathcal{OD}$ pair 5		
16	3 21 17 13 10	47.54
17	3 21 19 33 28 23	19.95
18	4 33 28 23	252.51
$\mathcal{OD}$ pair 6		
19	3 21 17 16	4.27
20	3 22	45.73
$\mathcal{OD}$ pair 7		
21	5 32 17 13 9	64.95
22	5 32 17 16 34	34.37
23	5 33 27 13 9	72.04
24	5 33 27 16 34	38.13
25	5 33 28 24 9	37.50
26	6 38 24 9	183.01
$\mathcal{OD}$ pair 8		
27	5 33 28 23	18.71
28	6 38 23	91.29
$\mathcal{OD}$ pair 9		
29	5 32 17 16	18.96
30	5 33 27 16	21.04
$\mathcal{OD}$ pair 10		
31	7 11 14 18 20	49.93
32	8 25 29 30	250.84
33	8 25 29 32 18 20	19.23
$\mathcal{OD}$ pair 11		
34	8 25 29 31	16.09
35	8 26 37	93.91
$\mathcal{OD}$ pair 12		
36	7 11 14 18	151.61
37	8 25 29 32 18	58.39
$\mathcal{OD}$ pair 13		
38	21 17 13 9	44.45
39	21 17 16 34	23.52
40	22 34	252.03
$\mathcal{OD}$ pair 14		
41	21 17 13 10	147.92
42	21 19 33 28 23	62.08
$\mathcal{OD}$ pair 15		
43	21 17 16	5.12
44	22	54.88
$\mathcal{OD}$ pair 16		
45	35 14 18 20	3.81
46	36 20	46.19
$\mathcal{OD}$ pair 17		
47	35 14 19 31	18.77
48	35 15 29 31	21.23
$\mathcal{OD}$ pair 18		
49	35 14 18	4.57
50	36	55.43

Table 3 shows all routes defined by their order r, and the scanned links of each route (marked

Table 3: The set of scanned link and its combination in order to identify all the route flows.

	Scanned links																	
Route (r)	1	2	3	5	8	9	11	13	18	20	21	22	23	29	31	33	34	36
1	X						X		X	X								
2		X							X	X								
3		X								X								X
4	X						X								X			
5	X						X							X	X			
6	X													X	X			
7	X																	
8		X													X			
9		X												X	X			
10	X						X		X									
11		X							X									
12		X																X
13			X			X		X			X							
14			X								X						X	
15			X									X					X	
16			X					X			X							
17			X								X		X			X		
18													X			X		
19			X								X							
20			X									X						
21				X		X		X										
22				X													X	
23				X		X		X								X		
24				X												X	X	
25				X		X										X		
26						X												
27				X									X			X		
28													X					
29				X														
30				X												X		
31							X		X	X								
32					X									X				
33					X				X	X				X				
34					X									X	X			
35					X													
36							X		X									
37					X				X					X				
38						X		X			X							
39											X						X	
40												X					X	
41								X			X							
42											X		X			X		
43											X							
44												X						
45									X	X								
46										X								X
47															X			
48														X	X			
49									X									
50																		X

with an X). They include the combinations of scanned links that, considering all routes, must be simultaneously registered for a single user to identify his route. For example if a user is scanned in links $2, 18$ and 20, then he travels using route 2, because he is included in one and only one of the sets shown in the table. Thus, scanning the set of links resulting from model M_1 the identification of all route flows is possible. However, several routes can be confounded if the set of scanned links is not adequately chosen. For example, if link 34 is not scanned, the users of route 24 can be confounded with those of route 30. After the scan process is finished some standard procedure must be used for a correct estimation of the flows, as for example, solving problem (5)-(7).

As shown in previous sections, the application of this location problem to real situations often involves substantial investment in equipment and installation of scanning devices (cameras or video cameras, the plate reader, etc.). Therefore, a budget restriction should be included in the previous formulations to get the best and the maximum amount of information using a limited number of traffic scanner devices.

To deal with this problem, model M_2 has been solved. Table 4 shows the resulting sets of

Table 4: The set of scanned link using the proposed model for different available budget

	Scanned links			
Route (r)	2	7	20	36
1			X	
2	$\mathcal{X}$		$\mathcal{X}$	
3	$\mathcal{X}$		$\mathcal{X}$	$\mathcal{X}$
4				
5				
6				
7				
8	X			
9	X			
10				
11	X			
12	$\mathcal{X}$			$\mathcal{X}$
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31		$\mathcal{X}$	$\mathcal{X}$	
32				
33			X	
34				
35				
36		$\mathcal{X}$		
37				
38				
39				
40				
41				
42				
43				
44				
45			X	
46			$\mathcal{X}$	$\mathcal{X}$
47				
48				
49				
50				$\mathcal{X}$

	Scanned links							
Route	2	3	7	10	20	22	34	36
1					X			
2	$\mathcal{X}$				$\mathcal{X}$			
3	$\mathcal{X}$				$\mathcal{X}$			$\mathcal{X}$
4								
5								
6								
7								
8	X							
9	X							
10								
11	X							
12	$\mathcal{X}$							$\mathcal{X}$
13		X						
14		$\mathcal{X}$					$\mathcal{X}$	
15		$\mathcal{X}$				$\mathcal{X}$	$\mathcal{X}$	
16		$\mathcal{X}$		$\mathcal{X}$				
17		X						
18								
19		X						
20		$\mathcal{X}$				$\mathcal{X}$		
21								
22							X	
23								
24							X	
25								
26								
27								
28								
29								
30								
31			$\mathcal{X}$		$\mathcal{X}$			
32								
33					X			
34								
35								
36			$\mathcal{X}$					
37								
38								
39							X	
40						$\mathcal{X}$	$\mathcal{X}$	
41				$\mathcal{X}$				
42								
43								
44						$\mathcal{X}$		
45					X			
46					$\mathcal{X}$			$\mathcal{X}$
47								
48								
49								
50								$\mathcal{X}$

	Scanned links															
Route	1	2	3	5	8	9	11	13	19	20	22	23	29	33	34	36
1	$\mathcal{X}$						$\mathcal{X}$			$\mathcal{X}$						
2		$\mathcal{X}$								$\mathcal{X}$						
3		$\mathcal{X}$								$\mathcal{X}$						$\mathcal{X}$
4	$\mathcal{X}$						$\mathcal{X}$		$\mathcal{X}$							
5	$\mathcal{X}$						$\mathcal{X}$						$\mathcal{X}$			
6	$\mathcal{X}$												$\mathcal{X}$			
7	$\mathcal{X}$															
8		$\mathcal{X}$							$\mathcal{X}$							
9		$\mathcal{X}$											$\mathcal{X}$			
10	$\mathcal{X}$						$\mathcal{X}$									
11		$\mathcal{X}$														
12		$\mathcal{X}$														$\mathcal{X}$
13			$\mathcal{X}$			$\mathcal{X}$		$\mathcal{X}$								
14			$\mathcal{X}$												$\mathcal{X}$	
15			$\mathcal{X}$								$\mathcal{X}$				$\mathcal{X}$	
16			$\mathcal{X}$					$\mathcal{X}$								
17			$\mathcal{X}$						$\mathcal{X}$			$\mathcal{X}$		$\mathcal{X}$		
18												$\mathcal{X}$		$\mathcal{X}$		
19			$\mathcal{X}$													
20			$\mathcal{X}$								$\mathcal{X}$					
21				$\mathcal{X}$		$\mathcal{X}$		$\mathcal{X}$								
22				$\mathcal{X}$											$\mathcal{X}$	
23				$\mathcal{X}$		$\mathcal{X}$		$\mathcal{X}$						$\mathcal{X}$		
24				$\mathcal{X}$										$\mathcal{X}$	$\mathcal{X}$	
25				$\mathcal{X}$		$\mathcal{X}$								$\mathcal{X}$		
26						$\mathcal{X}$										
27				$\mathcal{X}$								$\mathcal{X}$		$\mathcal{X}$		
28												$\mathcal{X}$				
29				$\mathcal{X}$												
30				$\mathcal{X}$										$\mathcal{X}$		
31							$\mathcal{X}$			$\mathcal{X}$						
32					X								X			
33					$\mathcal{X}$					$\mathcal{X}$			$\mathcal{X}$			
34					X								X			
35					$\mathcal{X}$											
36							$\mathcal{X}$									
37					X								X			
38						$\mathcal{X}$		$\mathcal{X}$								
39															$\mathcal{X}$	
40											$\mathcal{X}$				$\mathcal{X}$	
41								$\mathcal{X}$								
42									$\mathcal{X}$			$\mathcal{X}$		$\mathcal{X}$		
43																
44											$\mathcal{X}$					
45										$\mathcal{X}$						
46										$\mathcal{X}$						$\mathcal{X}$
47									$\mathcal{X}$							
48													$\mathcal{X}$			
49																
50																$\mathcal{X}$

links to be scanned for three different budget levels, i.e using the necessary budget for $4, 8$ and 16 scanned links. Note that when the X in the same row appear as $\mathcal{X}$, it means that this specific route is able to be identified due to the plate scanning procedure. The tables on the left, center and right show that for $4, 8$ and 16 scanned links one can observe 7, 14 and 45 of 50 routes, respectively. Using this information with that in Table 2 (the routes of each $\mathcal{OD}$-pair), we find that the case of 16 scanned links permits having the perfect knowledge of 13 of a total of 18 $\mathcal{OD}$ pair flows.

Note also that it is easy to make scanning errors so that some users could be assigned to a wrong route. For example, from Table 2, if any user traveling through route 24 fails to be scanned at link 34, he will be assigned to route 30, which is wrong.

To solve this problem, model M_6 could be solved increasing the redundancy ratio, i.e. $q = 2$. In this case, the solution provides the following links to be scanned:

$$\{1, 1, 2, 2, 3, 3, 4, 5, 9, 10, 11, 13, 15, 17, 19, 20, 20, 22, 23, 28, 30, 31, 32, 34, 34, 35, 36, 37\}, \tag{41}$$

where 28 scanning devices are needed. Note that some links are scanned once, some links twice, and some links not included in the set of links shown in (37) are now included in order to have redundant information.

In order to illustrate the goodness of the proposed methods, a simulation with 1000 replications has been done.

In each simulation we have assumed as true $\mathcal{OD}$ flows different set of values, obtained multiplying the prior flows in Table 1 by a uniformly distributed random factor ($U(0.8, 1.2)$). To obtain the true route and link flows the corresponding SUE problem is solved. These values are used to get the subset of observed route flows $\hat{f}_r$, and the observed isolated link flows $\hat{v}_a$ if any. Finally, using these observed route and link flows and as prior the route flows[2] f_r^0, the problem (5)-(7) is solved to estimate the remaining route flows of the current simulation. Comparative results are provided decreasing the number of scanning devices from 18 (fully observability) down to 1, the positioning of scanning devices and therefore the knowledge of observed routes has been obtained using the proposed method (models M_2 and M_3 have been used in order to check the influence of prior information) and the following alternative methods, which consists of:

1. If the number of scanning devices n_{sc} is lower than the minimum number of counting points required to satisfy the $\mathcal{OD}$ covering rule (see Yang and Zhou (1998)), which in this case is equal to $l^* = 4$, the BIP-5 model proposed by Ehlert et al. (2006) is used.

2. If $n_{sc} \geq 4$ and $n_{sc} \leq 12$ the TCL-P2 model in Yang and Zhou (1998) is used (see also Hodgson (1990)). Note that this model has a maximum limit of counting points ($n_{\max} = 12$) which does not allow the objective function to increase.

3. If $n_{sc} > 12$ we have chosen the set of links to be scanned selecting them randomly from the set given in (37), which allows fully observability.

Note that existing methods are specially suitable to work with link flows and comparison could seem to be unfair, however, since there is no alternative method for scanning device location, we have decided to use them for comparison purposes.

Figures 2(a), (c) and (e) show the simulation box plots (including the median, 25 and 75 quantiles and the outliers[3]) corresponding to the root mean squared relative error (RMSRE) of the $\mathcal{OD}$ predictions with different number of scanning devices selected through alternative methods, model M_2 and model M_3, respectively. Analogously, Figures 2(b), (d) and (f) show the same information but using the root mean absolute relative error (RMARE) (13) of the $\mathcal{OD}$ predictions. Note that errors are calculated with respect the assumed true flows.

Note that the performance of the proposed method using prior information (M_2) and without prior information (M_3) is statistically better specially when the number of scanning devices is in the range 3 to 17, where it is clearly shown that for any given number of scanning devices located, the mean is clearly smaller for the graphs c), d), e) and f) than those in a) and b), and also the interquartile range is smaller, which indicates that the variance of the errors is lower. Since results using RMSE and RMARE are analogous it can be concluded that the use of the root mean absolute value relative error for model derivations is justified.

These conclusions are confirmed from Tables 5 and 6, which provide the mean and standard deviation (in parenthesis) of the root mean squared relative errors for each $\mathcal{OD}$ flow and for different number of scanning devices installed, corresponding with the proposed method using prior information (model M_2) and alternative methods, respectively. Note that the mean and standard deviation tend to decrease when the number of scanning devices increases, being equal to zero (indicated in boldfaced) if the $\mathcal{OD}$ flow is completely observed. Note also that the percentage of times that the mean of the true error using the proposed method is lower than using existing methods in ≈55% of the times, whereas the standard deviation is lower in ≈61.5% of times.

Regarding results from models M_2 and M_3 shown in graphs (c)-(d) and (e)-(f), respectively. It can be observed that the use of prior information provides better results, as expected, which are

[2]These are the SUE route flows associated with the $\mathcal{OD}$ flows in Table 1

[3]Values above or below the mean 1.5 times the interquartile range are assumed to be outliers in this example.

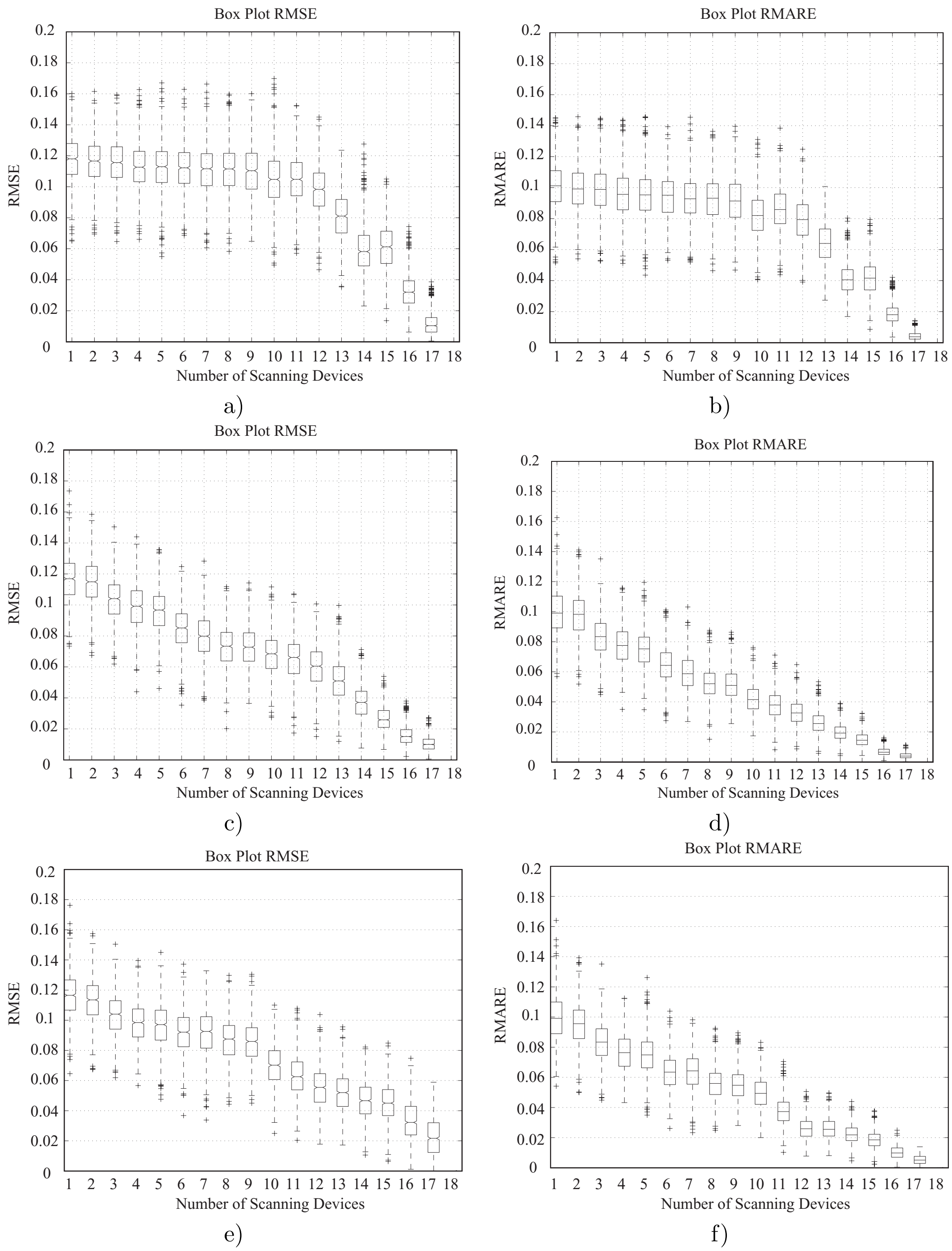

Figure 2: Box plots corresponding to 1000 different $\mathcal{OD}$ demand simulations, plotting the RMSRE and RMARE locating scanning devices through traditional methods (a and b) and by the proposed method: model M_2 (c and d) and model M_3 (e and f), respectively.

more clear with respect the RMSE. However, differences indicate that model M_3 is an appropriate alternative for those cases where no prior O-D matrix is available, providing also better results that traditional methods. Note that both models M_2 and M_3 for total observability (18 scanned links)

Table 5: Mean and standard deviation (between brackets) of the $\mathcal{OD}$ prediction errors obtained using 1000 simulations and locating the scanning devices using the proposed method.

	OD																	
n_{sc}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	0.2	0.1	0.6	0.4	-2.4	0.1	1.6	-0.6	-0.1	-0.8	0.4	-0.9	0.6	-1.0	-0.0	-0.1	-0.2	-0.2
	(24.6)	(49.1)	(36.9)	(24.1)	(25.4)	(5.7)	(48.6)	(12.4)	(4.6)	(37.3)	(12.6)	(24.2)	(35.2)	(24.6)	(6.9)	(5.8)	(4.7)	(7.0)
2	1.0	0.0	-1.1	-0.1	-1.7	-0.2	-1.8	-0.4	-0.1	-1.4	-0.9	1.3	-0.9	-0.5	0.3	-0.0	-0.0	-0.1
	(24.6)	(48.6)	(37.1)	(24.1)	(25.1)	(5.7)	(49.2)	(12.4)	(4.6)	(25.7)	(12.8)	(24.8)	(38.0)	(23.9)	(7.0)	(5.7)	(4.6)	(6.8)
3	-0.1	1.1	-0.9	1.3	-0.8	-0.0	0.1	0.1	-0.1	1.7	0.5	1.3	1.1	-0.6	-0.0	0.0	0.1	0.0
	(10.7)	(45.3)	(19.1)	(24.1)	(36.1)	(5.8)	(49.6)	(12.5)	(4.6)	(22.0)	(12.5)	(24.0)	(37.1)	(24.3)	(7.1)	(1.2)	(4.7)	(1.1)
4	-0.1	0.7	-0.1	-0.2	-1.3	-0.2	0.6	0.2	0.1	1.4	-0.7	-1.1	-1.5	-0.6	0.2	0.0	-0.2	0.0
	(5.2)	(44.5)	(19.5)	(24.2)	(37.0)	(5.8)	(50.1)	(12.9)	(4.6)	(16.0)	(12.8)	(9.6)	(37.0)	(24.3)	(7.0)	(0.6)	(4.6)	(1.1)
5	0.0	-0.9	0.2	0.8	-1.1	-0.3	-0.5	-0.4	0.0	0.5	-0.2	-0.8	-0.9	0.6	-0.1	0.1	0.1	0.1
	(5.3)	(44.6)	(18.9)	(24.1)	(25.8)	(5.7)	(49.9)	(12.6)	(4.6)	(15.9)	(12.8)	(9.4)	(37.4)	(24.1)	(6.8)	(0.6)	(4.6)	(1.0)
6	0.3	0.3	-0.1	0.6	-0.2	0.2	-1.9	-0.1	0.1	1.1	0.3	0.4	0.2	0.6	0.0	0.1	-0.0	0.0
	(10.6)	(45.0)	(18.7)	(10.6)	(21.8)	(1.3)	(45.0)	(12.6)	(4.5)	(21.7)	(12.7)	(24.1)	(19.4)	(24.5)	(1.1)	(1.2)	(4.6)	(1.0)
7	-0.3	-0.1	0.3	0.5	0.3	0.1	1.9	0.5	-0.1	2.0	-0.6	-1.1	0.2	-1.1	0.0	-0.0	0.3	0.1
	(5.2)	(43.9)	(19.7)	(11.2)	(22.6)	(1.3)	(44.5)	(12.7)	(4.7)	(16.0)	(12.5)	(9.5)	(19.6)	(24.3)	(1.1)	(0.6)	(4.7)	(1.1)
8	-0.0	1.3	-0.6	-0.1	1.8	0.0	-1.5	-0.3	-0.0	0.7	-0.6	-0.9	0.2	-1.3	0.1	0.0	-0.1	0.0
	(5.3)	(44.1)	(19.5)	(5.2)	(16.2)	(0.7)	(44.2)	(12.5)	(4.6)	(16.1)	(12.8)	(9.4)	(19.5)	(9.9)	(1.1)	(0.6)	(4.5)	(1.0)
9	-0.1	-2.2	-0.2	0.2	-0.1	-0.0	2.6	-0.8	0.2	1.3	-0.4	-1.2	0.6	-0.8	0.0	0.0	0.1	0.0
	(5.1)	(44.9)	(19.1)	(5.3)	(5.6)	(0.7)	(44.7)	(12.7)	(4.6)	(15.9)	(13.1)	(9.7)	(19.6)	(10.1)	(1.1)	(0.6)	(4.6)	(1.1)
10	0.7	0.1	-0.2	**0.0**	**0.0**	**0.0**	**0.0**	0.4	**0.0**	-0.1	0.3	-0.4	**0.0**	**0.0**	0.1	0.1	-0.1	0.1
	(11.0)	(45.0)	(19.3)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(8.9)	**(0.0)**	(22.2)	(13.0)	(23.7)	**(0.0)**	**(0.0)**	(6.5)	(1.2)	(4.6)	(1.1)
11	0.0	1.7	-0.5	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.3	-0.0	-0.8	**0.0**	**0.0**	0.1	0.1	0.1	0.0
	(11.2)	(45.0)	(20.2)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(22.4)	(12.7)	(24.4)	**(0.0)**	**(0.0)**	(6.3)	(1.2)	(4.6)	(1.1)
12	0.7	1.8	-0.3	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.1	-0.4	0.6	**0.0**	**0.0**	0.0	0.1	0.0	0.0
	(11.0)	(44.4)	(19.7)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(22.1)	(13.0)	(24.3)	**(0.0)**	**(0.0)**	0.0	(1.2)	(4.6)	(1.1)
13	-0.0	3.7	0.1	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.0	-0.1	-0.5	**0.0**	**0.0**	0.0	-0.0	-0.1	0.0
	(5.0)	(44.1)	(18.8)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(15.6)	(13.1)	(9.3)	**(0.0)**	**(0.0)**	0.0	(0.6)	(4.6)	(1.0)
14	**0.0**	0.0	-0.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.6	-0.3	0.7	**0.0**	0.0	0.0	0.1	0.1	0.0
	(0.0)	(15.5)	(15.5)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(13.1)	(3.8)	(24.3)	**(0.0)**	**(0.0)**	(1.1)	(1.7)	(2.5)	(1.0)
15	**0.0**	-0.5	0.5	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.0	-0.2	-0.8	**0.0**	**0.0**	0.1	0.1	**0.0**	0.1
	(0.0)	(14.1)	(14.1)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(12.5)	(3.9)	(9.5)	**(0.0)**	**(0.0)**	(1.1)	(1.2)	(2.5)	(1.0)
16	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.4	-0.2	-1.3	**0.0**	**0.0**	0.1	**0.0**	**0.0**	0.0
	(0.0)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(9.7)	(3.9)	(9.3)	**(0.0)**	**(0.0)**	(1.1)	**(0.0)**	**(0.0)**	(1.1)
17	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.1	-0.1	**0.0**	**0.0**	0.1	**0.0**	**0.0**	0.1
	(0.0)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(3.9)	(3.9)	**(0.0)**	**(0.0)**	(1.1)	**(0.0)**	**(0.0)**	(1.0)
18	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.0
	(0.0)	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**	(0.0)

provide the same results, intercepting 100% of the real route flow, and for example, considering 11 scanning devices, optimal locations from models M_2 and M_3 are, respectively:

$$\mathrm{M}_2 = \{2, 3, \mathbf{5}, 9, \mathbf{13}, 20, 21, \mathbf{23}, \mathbf{33}, 34, \mathbf{36}\}$$

and

$$\mathrm{M}_3 = \{2, 3, 9, \mathbf{16}, \mathbf{18}, \mathbf{19}, 20, 21, \mathbf{32}, 34, \mathbf{35}\},$$

which are different results (different links are boldfaced). Note that the first solution recognizes 29 routes, intercepting 63.87% of the total flow, and the second solution recognizes 32 routes, intercepting 51.76% of the total flow. Note that although the number of identified routes for the prior based solution is lower, the intercepted flow is greater. For this reason, in cases where the prior is known, results improve.

4 A Real example: the Cuenca network.

In order to apply the method to a realistic network, consider the Spanish city of Cuenca whose network topology representation is shown in Figure 3, consisting of 139 $\mathcal{OD}$ and 336 bidirectional

Table 6: Mean and standard deviation (between brackets) of the $\mathcal{OD}$ prediction errors obtained using 1000 simulations and locating the scanning devices using existing methods.

	OD																	
n_{sc}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	1.7 (23.4)	1.7 (48.8)	-0.9 (35.2)	-0.1 (25.0)	-0.3 (36.2)	0.2 (5.8)	0.6 (48.4)	0.2 (12.5)	-0.1 (4.6)	0.4 (34.8)	-0.6 (12.9)	1.2 (22.7)	0.2 (36.8)	-0.3 (24.4)	-0.1 (7.0)	0.1 (5.8)	-0.1 (4.9)	-0.1 (6.8)
2	0.5 (23.7)	2.1 (48.4)	-0.9 (35.0)	-0.2 (23.8)	-0.1 (35.1)	0.0 (5.7)	-0.3 (48.0)	-0.9 (12.8)	0.2 (4.7)	-1.8 (35.6)	-0.5 (12.9)	0.7 (22.8)	-1.4 (35.1)	0.5 (21.9)	0.3 (6.9)	-0.2 (5.8)	-0.1 (4.8)	-0.1 (6.9)
3	1.2 (24.1)	1.8 (45.9)	0.7 (36.4)	-0.1 (23.8)	-0.9 (35.3)	-0.2 (5.7)	0.2 (47.8)	-0.6 (12.7)	0.1 (4.9)	-0.9 (32.1)	-0.5 (12.4)	-1.2 (22.6)	-0.1 (35.1)	0.8 (22.8)	0.1 (6.9)	0.0 (5.8)	0.0 (4.6)	0.2 (7.0)
4	0.3 (22.1)	1.2 (43.8)	-0.7 (30.4)	1.1 (22.0)	-1.2 (35.9)	-0.3 (5.7)	0.1 (44.8)	0.1 (12.5)	0.2 (4.7)	2.1 (35.7)	0.3 (12.5)	-0.3 (23.2)	-0.9 (30.6)	-0.3 (24.1)	-0.1 (7.0)	0.1 (5.8)	0.2 (4.6)	-0.4 (7.0)
5	-0.1 (22.5)	0.3 (44.3)	0.7 (31.2)	1.0 (21.5)	1.4 (33.5)	0.1 (5.9)	-0.5 (42.6)	0.2 (12.6)	-0.2 (4.9)	-0.8 (35.5)	-0.1 (12.6)	0.2 (23.9)	-1.3 (29.9)	-0.4 (22.3)	-0.0 (6.9)	0.2 (5.7)	0.1 (4.7)	-0.1 (6.8)
6	0.3 (21.1)	-0.6 (43.3)	0.3 (30.0)	-0.9 (22.4)	-0.2 (33.8)	-0.2 (5.9)	3.2 (43.3)	0.3 (12.7)	-0.2 (4.9)	-0.4 (34.3)	0.3 (12.7)	0.3 (22.9)	0.4 (29.7)	-0.6 (22.1)	-0.1 (6.9)	-0.1 (5.9)	0.3 (4.8)	-0.1 (6.9)
7	-0.4 (22.0)	-1.6 (36.0)	0.2 (30.8)	0.3 (21.6)	-0.2 (34.2)	-0.1 (5.9)	0.4 (43.7)	-0.5 (12.6)	-0.1 (4.8)	1.4 (34.2)	0.5 (12.6)	-0.4 (22.4)	-0.7 (29.5)	1.3 (22.6)	-0.0 (6.8)	0.3 (5.8)	-0.0 (4.8)	-0.1 (7.1)
8	0.3 (22.4)	-1.0 (36.6)	1.5 (31.2)	0.2 (22.2)	0.7 (34.5)	-0.1 (5.9)	-0.7 (36.7)	0.0 (12.9)	0.1 (4.9)	0.6 (33.9)	-0.3 (13.0)	-0.4 (22.4)	-0.7 (29.9)	0.2 (22.6)	0.2 (6.9)	-0.2 (6.0)	-0.5 (4.9)	-0.1 (7.0)
9	-1.4 (22.0)	-0.5 (25.9)	1.4 (28.0)	1.1 (22.2)	0.6 (34.2)	-0.1 (5.9)	-1.4 (36.5)	-0.2 (12.9)	-0.1 (4.8)	-0.2 (34.7)	0.1 (12.9)	0.9 (22.1)	0.4 (29.3)	-0.3 (22.8)	0.0 (7.1)	0.4 (5.9)	0.2 (5.2)	-0.0 (7.3)
10	-0.5 (21.9)	0.2 (31.7)	0.5 (30.3)	0.3 (22.2)	-2.2 (24.3)	-0.1 (6.0)	-0.3 (6.3)	0.0 (0.0)	0.3 (6.3)	-1.0 (30.9)	-0.0 (13.0)	0.8 (22.3)	0.6 (28.6)	1.0 (21.2)	0.4 (7.0)	0.1 (5.8)	-0.1 (5.0)	0.0 (6.9)
11	0.6 (22.1)	1.1 (26.9)	-2.0 (28.8)	0.3 (22.5)	0.3 (32.0)	0.4 (5.8)	0.4 (35.5)	-0.2 (13.0)	0.3 (4.7)	0.1 (25.9)	-0.1 (4.0)	-0.0 (25.0)	-0.4 (30.4)	-1.2 (22.5)	-0.0 (6.9)	-0.1 (5.9)	-0.0 (4.6)	0.3 (6.9)
12	-0.3 (21.0)	0.7 (26.8)	-0.8 (28.8)	-0.2 (22.4)	0.1 (24.3)	-0.1 (5.8)	1.3 (26.5)	-0.3 (4.4)	0.3 (4.5)	1.0 (24.9)	-0.2 (3.9)	-0.8 (24.1)	-1.3 (28.7)	0.2 (23.4)	-0.0 (6.8)	0.4 (6.0)	0.2 (4.6)	-0.2 (7.1)
13	-1.4 (21.0)	0.0 (**0.0**)	1.4 (21.0)	-0.0 (21.0)	-0.1 (25.3)	-0.1 (4.8)	-0.2 (20.1)	-0.2 (4.1)	0.0 (2.3)	-1.9 (20.2)	-0.0 (11.2)	1.9 (20.2)	0.3 (25.7)	0.3 (24.6)	-0.0 (5.3)	-0.0 (4.4)	0.1 (2.5)	0.0 (4.4)
14	0.2 (10.6)	0.1 (22.1)	-0.5 (12.5)	0.4 (6.8)	1.7 (10.2)	-0.4 (6.8)	0.1 (16.9)	-0.2 (4.3)	0.1 (2.3)	-0.2 (10.6)	**0.0** (**0.0**)	0.5 (12.5)	-0.6 (19.1)	-1.4 (9.9)	0.3 (7.0)	0.0 (0.0)	0.1 (3.1)	0.0 (0.0)
15	-0.7 (20.3)	0.0 (0.0)	0.7 (20.3)	0.6 (10.9)	-0.5 (13.8)	**0.0** (**0.0**)	-0.4 (21.7)	-0.1 (7.2)	0.2 (4.0)	-0.9 (20.5)	**0.0** (**0.0**)	0.9 (20.5)	-0.4 (13.1)	0.4 (13.1)	**0.0** (**0.0**)	-0.1 (4.5)	**0.0** (**0.0**)	0.1 (4.5)
16	0.6 (10.8)	0.1 (21.2)	-0.8 (12.9)	**0.0** (**0.0**)	0.8 (9.6)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	-0.6 (10.8)	**0.0** (**0.0**)	0.8 (12.9)	**0.0** (**0.0**)	-0.8 (9.6)	**0.0** (1.1)	**0.0** (**0.0**)	**0.0** (3.0)	0.0 (0.0)
17	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	1.0 (9.3)	**0.0** (**0.0**)	-1.0 (9.3)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	0.0 (1.1)
18	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	**0.0** (**0.0**)	0.0 (0.0)

links. A total of 528 routes, obtained using the Wardrop assignment model, are considered, which corresponds on average between 3 and four paths per Origin-Destination.

To identify all path flows, a minimum number of links have to be scanned. To this end, problem (8)-(10) has been solved, resulting 100 links to be observed. Note that the improved models proposed in this paper were, in fact, motivated by the interest in applying the scanning method proposed by Castillo et. al (2008) in this city, where budget limitations did not allow us to observe all route flows. For this reason, model M_2 is solved for different budgets, i.e., different number of scanning devices, $\{100, 95, 90, 85, 80, 75\}$. Note that we have included 100 which is the minimum number of scanning devices providing full observability.

In order to show computational time statistics, we have run 100 times problem M_2 (20)-(23) for the different number of scanning devices considered. All problems have been solved using CPLEX under GAMS (Brooke et al., 2008) on a Linux-based server using four processors clocking at 2.6 GHz and 32 GB of RAM.

Note that for this particular example, model M_2 has 38983 equations, 1201 continuous variables, 672 discrete variables, and 905351 non-zero elements. Simulation results are provided in Table 7 where the following comments are pertinent:

Figure 3: Traffic network representation of the city of Cuenca (Spain).

Table 7: Mean and standard deviation of the computational time (CPU time) in seconds considering different number of available scanning devices and using 100 simulations for the Cuenca network example.

	Number of scanning devices					
	100	**95**	**90**	**85**	**80**	**75**
Mean (sec)	89.4	183.4	113.5	330.3	422.0	984.0
Std (sec)	19.1	28.1	8.5	70.7	74.8	115.9

1. All problems considered in this example are solved within a reasonable amount of time (less than one 20 minutes of CPU time) for a design problem.

2. Computational time increases with the number of possible combinations of the available scanning devices n_{sc} (determined by the budget) in the existing links n_ℓ, given by the formula:

$$\left(\begin{array}{c} n_\ell \\ n_{sc} \end{array} \right) = \frac{n_\ell!}{n_{sc}!(n_\ell - n_{sc})!}. \tag{42}$$

5 Conclusions

The main conclusions that can be drawn from this paper are the following:

1. Flow covering is a very important rule to be considered in order to reproduce the $\mathcal{OD}$ flows as close as possible. In fact, location of plate scanning devices without consideration of this rule cannot lead to good predictions.

2. Including the prediction errors in the objective function leads to very important improvements in the quality of predictions, not only in the mean values (accuracy), but in the standard deviations (precision). This has been illustrated by simulations.

3. Prior O-D matrices provide better estimation results, nevertheless, for the cases where no prior information is available, model M_3 based on the MPRE concept derived by Yang et al. (1991) provide satisfactory results, giving an insight of the overall problem.

4. Several models have been presented for an adequate selection and location of plate scanning devices. They include budget constraints together with the consideration of already existing devices.

5. Some of the natural binary variables involving routes can be replaced by continuous variables in the range $[0, 1]$. This has been proved in Theorem 1 and leads to an important reduction in cpu and memory resources.

6. A technique to reduce the possibility of scanning errors by using redundant measures have been presented. However further research should be done on this topic using adequate statistical models to predict scanning devices failure behavior.

6 Acknowledgements

The authors acknowledge to Prof. Antonio Conejo (University of Castilla La Mancha) for providing computer resources.

References

Ben-Akiva, M., Ashok, K., Yang, Q., 1994. Commentaries: A statistical analysis of the reliability of using RGS vehicle probes as estimators of dynamic O-D departure rates. IVHS Journal 2(1), 92–97.

Bianco, L., Confessore, G., Gentili, M., 2006. Combinational aspects of the sensor location problem. Annals of Operations Research 144(1), 201-234.

Bianco, L., Confessore, L., Reverberi, P., 2001. A network based model for traffic sensor location with implications on O/D matrix estimates. Transportation Science 35(1), 50–60.

Bierlaire, M., 2002. The total demand scale: a new measure of quality for static and dynamic origin-destination trip tables. Transportation Research Part B 36(9), 837–850.

Cascetta, E., Nguyen, S., 1988. A unified framework for estimating or updating origin/destination matrices from traffic counts. Transportation Research Part B 22(6), 437–455.

Castillo, E., Conejo, A., Pedregal, P., García, R., Alguacil, N., 2001. Building and Solving Mathematical Programming Models in Engineering and Science. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs and Tracts. John Wiley & Sons Inc., New York.

Castillo, E., Conejo, J. A., Menéndez, J. M., and Jiménez, P. (2008a). The observability problem in traffic network models. Computer Aided Civil and Infraestructure Engineering, 23(3):208–222.

Castillo, E., Menéndez, J. M., Jiménez, P., 2008b. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. Transportation Research Part B 42(5), 455–481.

Castillo, E., Menéndez, J. M., Sánchez-Cambronero, S., 2008c. Predicting traffic flow using Bayesian networks. Transportation Research Part B 42(5), 482–509.

Castillo, E., Menéndez, J. M., Sánchez-Cambronero, S., 2008d. Traffic estimation and optimal counting location without path enumeration using Bayesian networks. Computer Aided Civil and Infraestructure Engineering 23(3), 189–207.

Chung, I.-H., 2001. An optimum sampling framework for estimating trip matrices from day-to-day traffic counts. PhD thesis, University of Leeds.

Ehlert, A., Bell, M. G. H., Grosso, S., 2006. The optimisation of traffic count locations in road networks. Transportation Research Part B 40(6), 460–479.

Brooke, A., Kendrick, D., Meeraus, A., Raman, R., Rosenthal, R. E.,2008 GAMS: a Users Guide. Washington, DC: GAMS Development Corporation, 2008. Available on: *http://www.gams.com/*.

Gentili, M., Mirchandani, P. B., 2005. Locating active sensors on traffic network. Annals of Operations Research 136(1), 229–257.

Hazelton, M. L., 2000. Estimation of origin-destination matrices from link flows on congested networks. Transportation Research Part B 34(7), 549–566.

Hellinga, D., Van Aerde, M., 1994. A statistical analysis of the reliability of using RGS vehicle probes as estimators of dynamic O-D departure rates. IVHS Journal 2(1), 21-44.

Hodgson, M., 1990. A flow-capturing location-allocation model. Geographical Analysis 22, 271-279.

Maher, M. J., 1983. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transportation Research Part B 17(6), 435–447.

Schrijver, A., 1987. Theory of Linear and Integer Programming. John Wiley & Songs, Chichester.

Sun, S. L., Zhang, C. S., Yu, G. Q., 2006. A Bayesian network approach to traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems 7(1), 124–132.

Willumsen, L. G., 1984. Estimating time-dependent trip matrices from traffic counts. Ninth International Symposium on Transportation and Traffic Theory, 397–411.

Yang, H., Iidia, Y., Sasaki, T., 1991. An analysis of the realiability of an origin-destination trip matrix estimated from traffic counts. Transportation Research Part B 25(5), 351–363.

Yang, H., Yang, C., Gan, L., 2006. Models and algorithms for the screen line-based traffic counting location problems. Computers and operations research 33(3), 836–858.

Yang, H., Zhou, J., 1998. Optimal traffic counting locations for origin-destination matrix extimation. Transportation Research Part B 32(2), 109–126.

Yim, P. K. N., Lam, W. H. K., 1998. Evaluation of count location selection methods for estimation of OD matrices. Journal of Transportation Engineering 124(4), 376–383.