

Owais, Mahmoud; Moussa, Ghada S.; Hussain, Khaled F.

Article

Sensor location model for O/D estimation: Multi-criteria meta-heuristics approach

Operations Research Perspectives

Provided in Cooperation with:

Elsevier

Suggested Citation: Owais, Mahmoud; Moussa, Ghada S.; Hussain, Khaled F. (2019) : Sensor location model for O/D estimation: Multi-criteria meta-heuristics approach, Operations Research Perspectives, ISSN 2214-7160, Elsevier, Amsterdam, Vol. 6, pp. 1-12, <https://doi.org/10.1016/j.orp.2019.100100>

This Version is available at:

<https://hdl.handle.net/10419/246393>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Sensor location model for O/D estimation: Multi-criteria meta-heuristics approach



Mahmoud Owais^{a,b,*}, Ghada S. Moussa^c, Khaled F. Hussain^c

^a Civil Engineering Department, Assiut University, Assiut, Egypt

^b Department of Civil and Environmental Engineering, College of Engineering, Majmaah University, Al-Majmaah, Saudi Arabia

^c Faculty of Engineering, Assiut University, Assiut, Egypt

ARTICLE INFO

Keywords:

Meta-heuristics
Multi-objective
Traffic sensors' location
Pareto optimal

ABSTRACT

With the progress in intelligent transportation systems, a great interest has been directed towards traffic sensors information for flow estimation problems. Nevertheless, there is a great challenge to locate such traffic sensors on a network to attain the maximum benefits from them. Considering the O/D matrix estimation problem, all traffic sensors location models depend crucially on the reliability of the estimated matrix compared with a priori flow information. Thus, the required sensors number (cost) and locations for a network vary according to the estimation technique (e.g. least square, minimizing entropy, maximum likelihood, etc.) as well as the reliability of the priori information. Alternatively, this study presents a robust traffic sensor location model, which produces different trade-offs between the potential accuracy of the estimated O/D matrix and the cost of sensors' installation in a polynomial time complexity. The proposed approach searches for the number and locations of sensors that minimize the boundary of the maximum possible relative error for the estimated O/D matrix. The traffic sensor location problem is formulated as a set covering problem, then a multi-criteria meta-heuristics algorithm is adopted. The pioneer of this work is that it targets the maximum possible relative error directly in the multi-objective design process, which is considered a robust criterion for evaluating a solution set. Moreover, the proposed approach is extended to incorporate the screen line problem in a straightforward manner. For the purpose of validating the feasibility and the effectiveness of the proposed approach, two real networks are used. The results show the capability of producing the Pareto optimal (near optimal) solutions for any network.

Abbreviations

<i>TSLP</i>	Traffic Sensor Location Problem
<i>TRE</i>	True Relative Error
<i>TOF</i>	Total Observed Flow
<i>SCP</i>	Set Cover Problem
<i>RaPS</i>	Randomized Priority Search
<i>RF</i>	Replication Factor
<i>PRV</i>	Preference Value
<i>PI</i>	Priority Index
<i>NOF</i>	Net Observed Flow
<i>MPRE</i>	Maximum Possible Relative Error

1. Introduction

Quantifying traffic flow information has received great attention over the last decade. This substantial attention arises from its practical

applications in transportation planning, traffic management, and traveler information systems [9]. Traffic sensors with their various types (e.g., inductive loops, video cameras, and radio frequency transponders) play a crucial role in estimating such information. However, the full installation of traffic sensors on all network's links is not a cheap process. In this context, the problem of seeking the optimal number and locations of traffic sensors on a network according to some objectives is known as Traffic Sensor Location Problem (TSLP) [53].

Along with the literature, the TSLP can be classified according to the aim of deploying the sensors into two-fold problems: sensor location flow-estimation problem and sensor location flow-observability problem [32]. The first fold problem identifies the sensors optimum locations on a network associated with high-quality target estimates (e.g., O/D matrix estimation, Path Flow Estimation (PFE), travel times estimation, etc.). The second fold problem identifies the sensors optimum locations on a network associated with the unique determination of the unobserved flow. While the sensor location for flow-estimation

* Corresponding author at: Assiut University, Assiut City Street, Assiut 71511, Egypt.

E-mail address: Maowais@aun.edu.eg (M. Owais).

<https://doi.org/10.1016/j.orp.2019.100100>

Received 29 August 2018; Received in revised form 22 January 2019; Accepted 22 January 2019

Available online 06 February 2019

2214-7160/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Symbols	
V	set of network vertices (nodes)
L	the network links set $L = \{l_1, l_2, \dots, l_n\}$
T_i	the historical demand per each O/D pair i
\bar{T}_i	the estimated demand per each O/D pair i
T_i^*	the true (unknown) demand per each O/D pair i
λ_i	the relative error per each O/D pair i
W	the set of all O/D pairs $W = \{1, 2, \dots, i, j, \dots, w\}$
w	the number of network O/D pairs $ W = w$
n	the number of network links
i, j	subscripts for O/D pairs. $i, j \in W$
a	subscript for links, $a \in L$
p_{ai}	portion of node pair i demand passes link a
z_a	dummy variable = 1 if sensor is located on link (a) , 0 otherwise
c_a	cost of installing traffic counter per link a , $\$/l$
v_a	link (a) flow
\bar{L}	the set of links equipped with traffic counters
η	no. of links equipped with traffic counters, $ \bar{L} = \eta$
H	the set of all generated paths $H = \{h_{11}, h_{21}, \dots, h_{m\bar{b}}, \dots, h_{kw}\}$
H_i	the set of paths with the same O/D pair i , $H_i = \{h_{i1}, h_{i2}, \dots, h_{i\bar{b}}, \dots, h_{ki}\}$
m	subscript for paths connecting an O/D pair i
k	no. of generated paths connecting node pair i
u	subscript for a candidate link to be in \bar{L}
s	subscript for a chosen link to be in \bar{L}
δ_{mi}^a	dummy variable = 1 if path h_{mi} contains link l_a , 0 otherwise
θ	depression factor
q_a	the link travel time as a function of link flow
q_a^0	the initial time associated with free flow condition
λ_{mi}	the travel time of path (m) connecting node pair (i)
Q_a	the link (a) capacity
α, χ	calibration parameters
β	multi-objective weight factor
ξ	search control factor
Λ	no. of iterations
Vectors	
C	the set covering matrix made by $(k \times w)$ rows \times n columns
T	the historical demand node pairs vector of size w
\bar{T}	the estimated demand node pairs vector of size w
V	the reference link flows vector of size η
V^*	the observed/real link flows vector of size η
Σ_T, Σ_V	are weight matrices of sizes $w \times w$ & $\eta \times \eta$

problems requires a priori information about the network's flows, the sensor location for flow observability problems relies crucially on the network's topological properties.

The majority of the attention is given for locating sensors in the attempt of estimating the O/D matrix. The dilemma here is to locate the minimum number of traffic sensors on a network which can result in the most accurate estimated O/D matrix. After observing some links flow, the process becomes reversible to the traffic assignment problem [57]. If there is an O/D matrix, the links flow can be obtained after an assignment process by any known assignment technique. However, reversing this process is not a trivial problem. A certain vector of links flow for a network could be obtained from an unlimited number of O/D matrices. Besides observing all links as mentioned before is not a practical action [2,42].

The PFE is considered as an alternative approach for the O/D estimation problem. It aims to infer all paths flow from traffic counting on links and consequently the O/D matrix. This alternative approach turns the O/D matrix estimation into a bi-level problem. The upper level is the PFE model whereas the lower level is a traffic assignment model concerned with maintaining the network at traffic equilibrium [4,17].

The flow observability problem shares the same importance as the estimation problems. It simply searches for the minimum number and locations of traffic sensors to observe all links flow through nodes conservation equation (inflow = outflow). For one origin centroid network, this trend stipulates that the number of unobserved links is equal to the number of (nodes - 1) [26,41]. Alternatively, in [6,27], the minimum number of sensors is searched to be installed in nodes (instead of arcs). It helps in observing the traffic turning ratios at nodes which are used for the full network observability problem.

This classification is very general since there are many related aspects. Optimizing the sensor placement on a link is an art itself. Questions such as what is the type of sensor to locate [28], where to locate the device on the link [21], what is the best time interval to choose and how to filter the data from the errors [22,23,43]. All these subjects are separate investigation matters.

For counting purpose, two main types of sensors could be identified, namely: passive and active sensors. In the passive mode of counting, only arcs flow are observed. Alternatively, in the active mode, other information would be obtained like; vehicle type (image sensors), path

information (path-ID sensors) and vehicle observation time (vehicle identification sensors) [27].

This variety of traffic sensor types has added new dimensions to the conventional TSLP. Path-ID sensors are located to uniquely identify the paths flow on a network [13,14]. In [30], it has been proved that fewer vehicle identification sensors are required for the same path flow problem. Locating a mix of passive and active sensors (heterogeneous location problem) is also investigated in [24,33,35].

In this research, we focus on the TSLP for the O/D matrix estimation using the passive error-free sensors. In all reviewed design methods, although the selection procedure is clearly subjective, there is no quality guarantee of the estimated O/D matrix after the real deployment of sensors on the network. In addition, there is a blatant conflict between the required accuracy and the sensors installation cost. However, few studies dealt with the multi-objective term explicitly [18,56,58]. Therefore, the multi-objective approach is promoted since it has received little attention in the literature. We add to these studies by using a novel multi-objective meta-heuristic algorithm, besides a different perspective in the objective function.

The proposed design is based on a robust measure of the potential accuracy for O/D matrix estimation problem with respect to the number of used sensors. The problem of selecting the sensors is formulated as a Set Covering Problem (SCP). An innovative multi-criteria meta-heuristics algorithm is adapted for the problem. The structure of the paper is as follows. Section 2 presents the state of art for the TSLP. Section 3 provides the problem formulation and basic input data. Section 4 illustrates the proposed methodology. In Section 5, real case studies are used to evaluate the proposed methodology. Section 6 presents the conclusion.

2. Literature review

Considering the O/D matrix estimation, the accuracy term plays a pivotal rule in judging the effectiveness of the number of traffic sensors distributed on a network. In the literature, statistical measures are often used to quantify the quality of O/D estimates. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are popular examples of these statistical measures. They indicate the closeness between the observed (true) and estimated values, which could be links or O/D

flows [17]. In [37], to determine the priority of a link to be chosen for the O/D matrix estimation problem, the link flow and covering number of the O/D pairs are used. It is assumed that the true matrix is available. However, it may not be sufficient to quantify the quality of chosen links since the true information is often unknown in real life and consequently the true error of predicted flows is impossible to be estimated.

Therefore, extra information as priori information about either, the O/D matrix, links flow or paths flow is needed [16,35]. As an example, the historical O/D matrix was employed by Yim and Lam [61]. Consequently, the design process (locating the traffic sensors) based on such measurement tools is questionable (depends on the quality of the used matrix).

The pioneer work, which opened the gate to most of the TSLP art, is found in an analysis of the estimated O/D reliability from traffic counting in [55]. It presents a theoretical investigation of the reliability of an estimated O/D trip matrix employing the concept of Maximum Possible Relative Error (MPRE).

The MPRE represents the maximum possible relative deviation of the estimated O/D matrix from the true (unknown) one. It indicates the amount of space allowed by traffic counts for a particular fitted O/D matrix when this basis is measured based on the maximum value of a quadratic index [55].

To simplify the concept, let us imagine a solution search space made by all problem constraints constituting a sphere, see Fig. 1. If the estimated solution, for example, lies at the center of the sphere, one may realize that the MPRE of that solution is the sphere diameter. In other words, the maximum error comes when the worst case of the unknown true solution lies on the circumference (the farthest point).

For a network with O/D matrix made of (w) entries and (η) observed links, the MPRE mathematically is represented as follows;

$$\text{Maximize}_{\lambda_i} \sqrt{\sum_{i=1}^w \lambda_i^2 / w} \quad (1)$$

s.t

$$\lambda_i = \frac{T_i^* - T_i}{T_i^*} \quad \forall i \in W \quad (2)$$

$$\sum_{i=1}^w P_{ai} \bar{T}_i \lambda_i = 0 \quad \forall l_a \in L \quad (3)$$

$$\lambda_j \geq \frac{-\sum_{i=1, i \neq j}^w P_{ai} \bar{T}_i}{P_{aj} \bar{T}_j} \quad (4)$$

The objective function in Eq. (1) aims to maximize the possible error from the base (estimated) O/D matrix by maximizing the relative error λ all over the matrix entries. Eq. (2) denotes λ . Eq. (3) draws the feasible solution space made by all O/D matrices that produce the same link flow over the number of observed links (η). It reveals that the place and the number of chosen links (L) are the factors that draw the solution search space (determine MPRE value). Inequality (4) reflects the fact that λ is always ≥ -1 [55].

The MPRE is a robust way to check the accuracy of the observations coming from a link set with installed sensors. However and to the best of our knowledge, there is no study found in the literature that selects the sensors' location directly based on the MPRE value. Instead, in [57], the MPRE is decomposed into four heuristics rules to help in determining the best locations of sensors aiming to achieve the least MPRE. The rules are as follows:

- Rule 1: intercept at least one path for each O/D pair.
- Rule 2: maximize the portion of observed traffic from each node pair.
- Rule 3: maximize the total flow observed by each link.
- Rule 4: maximize the net flow observed by links.

Each rule is named to reflect the purpose of it. Rule 1 (coverage

rule) addresses a partial interception to each O/D pair flow which is a stipulation to acquire a finite value for the MPRE. Rule 2 (maximal flow fraction rule) aims to maximize the information obtained about each O/D pair, Rule 3 (maximal flow intercepting rule) targets the maximum flow to be observed through the network. Rule 4 (link independence rule) attempts to minimize the replicated traffic counting.

Interestingly, in [56], each rule is formulated as a separate objective function with its constraints. Hence, there is one solution for each rule, but there is no global solution that satisfies all the rules simultaneously. In other words, they are conflicting rules. Unfortunately, the resulted formulations of the rules are difficult to be solved with exact methods, especially for large size networks, therefore heuristic methods become a need. This conflict comes from the existence of two main objectives; first is minimizing the MPRE value (maximizing the potential accuracy). Second is minimizing the number of observed links (cost).

In [54], a reformulation is made to cover each demand pair flow. Trips between a particular O/D pair are considered to be fully observed if and only if there is no path that is able to bypass the selected traffic counting locations. The problem is a mathematical formulation of drawing the screen line problem (i.e. to observe every vehicle in the network once at least). Fortunately, we could also extend the proposed algorithm to solve that intrinsically.

Recently, the Compressed Sensing (CS) framework is used as a novel method to solve the stated problem [59]. Whereas, in [25], a new two-stage stochastic programming strategy is developed for the location problem to solve the path reconstruction problem with demand uncertainty. In [36], the TSLP is addressed by a graph-theoretic framework. It aims to monitor traffic emissions by a vehicle emission remote sensing system so that each vehicle should be monitored once through the network. The problem resembles the screen line problem. In [44], the number and locations of the sensors were optimized for the travel time estimation problem on a single road section. The state of the art is reviewed extensively in [5,12,26,27].

Regarding the multi-objective nature of the problem, there are two main approaches to handle this problem: the preference-based approach and the generating approach. The former approach simply converts the multiple objectives into a single objective, according to the preference structure (weight factors) supplied by the decision makers. This requires a good knowledge of the problem. Using improper weight factors will lead to the domination of one objective over the others [46]. The latter approach adopts the concept of Pareto optimality. The

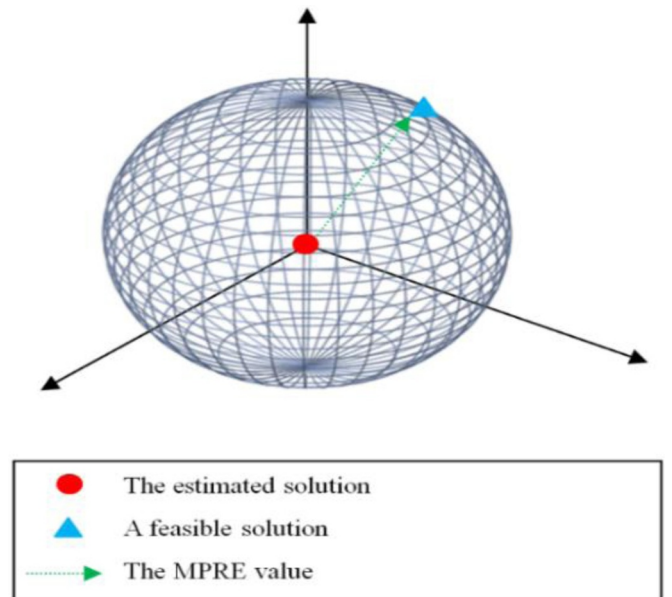


Fig. 1. Maximum possible relative error interpretation.

solution is "Pareto optimal" or "non-dominated" if there are no other feasible solutions that could improve some objectives without worsening at least one other objective. A set of non-dominated solutions generally forms the so-called "Pareto front" or "efficient frontier", which represents the relationship (trade-offs) among multiple objectives. This approach relies on changing the structure (values) of weight factors to produce the different solution sets.

To highlight the relevant previous work to this study, Table 1 summarizes the studies that tackled the TSLP for the O/D matrix estimation. From Table 1, we could conclude the contributions of this study to the literature in considering the following points simultaneously;

- Targeting the MPRE directly in the design process, which is considered a robust criterion for evaluating a solution set.
- Tackling the multi-objective nature of the problem to produce the Pareto (near) optimal solution.
- Providing an effective Meta-heuristic for the problem with a polynomial execution time that simplifies solving the real case networks.
- Extending the methodology to incorporate the screen line problem.

3. Problem formulation

3.1. Input data

Consider a given directed road network, $G = (V, L)$, where V is the set of vertices which are connected by the set of links L . Each link is weighted with its travel time. A historical O/D matrix is also assumed given as a guide to the search process. The existence of such matrix is a common assumption in the literature [7,13–15,40,51,63,64].

The network path enumeration is a requirement for the developed sensor location model. The k-shortest path is used to generate all possible paths between each node pair. When it comes to the models that depend on path enumeration. For even a small network, a large number of paths could be generated for each node pair. Certainly, this would lead to a combinatorial problem for any methodology. Fortunately, it is common that 3 or 4 paths to carry the majority of the O/D flow, and rarely that more than 6 or 7 routes are utilized [8,31]. So, H_i has the maximum limit of paths $k = 7$. Also, paths with more than 1.5 the shortest path time are considered circuitous for travellers [49]. For each demand node pair (i) all h_{ki} paths are stored in the set H_i . At last, a link-path incidence matrix is generated.

3.2. TSLP objective function

Mathematically the considered objective function would be represented as follows;

Table 1
Traffic Sensors' location studies for O/D matrix estimation.

Author	Year	Objective function	Solution methodology	Evaluation method	Screen line problem
Yang and Zhou	1998	Single	Heuristic Greedy	The four rules ^a	— ^b
Yang et al.	2001	Single	Genetic Algorithm	The four rules	√
Chootinan et al.	2005	Multi ^c	Distance based Genetic Algorithm	Maximum flow captured	—
Yang et al.	2006	Single	Genetic Algorithm + Heuristic Greedy	The four rules	—
Chootinan et al.	2007	Single	Modified Path Flow Estimator + Genetic Algorithm	O/D coverage rule	√
Li and Ouyang	2011	Single	Lagrangian relaxation + Greedy algorithm	O/D coverage rule	—
Simonelli et al.	2012	Multi	Sequential heuristic	Variability of the posterior demand	—
Wang and Mirchandani	2013	Single	Greedy algorithm	Bayesian statistical procedure	—
Liu and Zhu	2014	Multi	Distance-based Genetic Algorithm	The four rules	—
Ye and Wen	2017	Single	Compressed Sensing	The four rules	—
This study	2018	Multi	Meta-RaPS ^d	MPRE ^e	√

^a The four location rules proposed in [57].

^b Not tackled.

^c Multi-objective.

^d Meta-heuristics based on Random Priority Selection.

^e Maximum Possible Relative Error.

$$\text{Minimize } \beta_1 \left(\text{Max}_{z_a} \sum_{i=1}^w \lambda_i^2 / w \right) + \beta_2 \sum_{a=1}^n c_a z_a \tag{5}$$

s.t

$$\sum_{m=1}^k \sum_{a=1}^n \delta_{mi}^a z_a \geq 1, \forall i \in W \tag{6}$$

$$\sum_{i=1}^w p_{ai} \bar{T}_i \lambda_i z_a = 0 \quad \forall l_a \in L \tag{7}$$

$$\lambda_i \geq -1, \forall i \in W \tag{8}$$

In the formulated objective function (5), it is aimed to incorporate the MPRE directly in the design of sensors' number and locations. It is to minimize the number of observed links (z_a) needed while the maximum possible error in estimating the O/D matrix by the chosen set is to be minimum. It reflects the multi-objective nature of the problem because the more accuracy is required (low value for the first term), the more links are to be observed (increasing sensors installation cost). The only decision variable in (5) is z_a , however, the formulation is very complex to be solved with standard integer solvers. To show the model complexity, it is sufficient to mention that the presence of inequality (6) turns the integer optimization into NP-hard problem [19]. Besides, the first part constitutes a min-max problem in which the decision variable z_a does not appear explicitly. The square root is omitted from the first term (MPRE) because it is not affecting the optimization results.

β denotes the relative importance of different terms in the objective function. Inequality (6) represents the partial coverage rule (Rule 1) in which it is necessary to select links to intercept one path in each set H_i . In other words, it is required to find the number of observed links to intercept a portion of all network demand pairs flow; this also guarantees a finite MPRE value. Eq. (7) determines the link selection effect on the MPRE value, whereas (8) is a relaxation to the inequality (4) [55].

It is obvious that the stated problem in Eqs. (5)–(8) is multi-objective which adds much more complexity. The direct approach is to calibrate the β_1 & β_2 based on the designer's preferences to obtain a single optimal solution. However, based on the multi-criteria analysis, it is aimed to produce a set of non-dominated solutions. A solution is recognized as non-dominated (Pareto optimal) if it is better when compared with each other solution in the set, at least, in one objective value. To achieve the Pareto optimal, the structure of β_1 & β_2 can be continuously changed to create a vast number of diverse solutions [45]. Generally, we would not resort to both approaches here because the proposed algorithm would diversify the search to obtain different solutions. These solutions would be filtered (ranked) to produce the Pareto-

set as will be shown next in the methodology section.

3.3. Screen line problem extension

The TSLP could be expanded to intercept the whole demand pairs flow. That problem is named as screen line problem. The links are chosen to cover all the generated paths. Mathematically, the problem is represented as follows:

$$\text{Minimize } \sum_{a=1}^n z_a \quad (9)$$

s.t

$$\sum_{a=1}^n \delta_{mi}^a z_a \geq 1, \forall m \in H_i, i \in W \quad (10)$$

3.4. Solution approach

It is obvious from the formulated problem that there is no exact solution could tackle large size networks. Therefore, a meta-heuristic becomes a need. The key to the solution methodology is that the two stated problems Eqs. (5)–(10) could be viewed as a Set Covering Problem (SCP), see Fig. 2. The link (column) selection procedure should cover the rows made of all network paths either partially or fully. Meta-heuristics with perturbation operators would be introduced for the SCP to generate a variety of solutions. One of the major advantages of viewing the problem as a SCP is the constant preserving of Rule 1. Consequently, all resulted solutions will have a finite MPRE value. These solutions would have a better performance rather than the solution sets which result in infinite MPRE. Any solution set (number of sensors distributed on a network) with infinite MPRE cannot be judged until it is deployed on the network, then used in estimating the O/D matrix and finally compared with the true (unknown) one.

4. Methodology

Now, the TSLP solution is viewed as solving the SCP which is defined as the problem of covering all the rows of a matrix made of $k \times w$ rows, n -columns, zero-one elements (δ_{mi}^a) by a subset of the columns at the minimum cost [1]. The SCP could be solved exactly by any existing integer solver, however, it is not profitable to our problem. This stems from two basic reasons; 1) it is classified as NP-hard, so it is difficult (impossible) to be solved for real scale networks (our target) [19,38]. 2) The SCP always addresses a single objective, unlike the stated objective function in Eq. (5). In other words, the optimal solution for the SCP does not imply the optimal solution for the TSLP. Therefore, we develop

$$\begin{matrix}
 & l_1 & l_2 & l_3 & & & & & & l_n \\
 h_{11} & \left(\begin{array}{cccccccc}
 \delta_{11}^1 & \delta_{11}^2 & \delta_{11}^3 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \delta_{11}^n \\
 \delta_{21}^1 & \delta_{21}^2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \delta_{21}^n \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 h_{k1} & \delta_{k1}^1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 h_{12} & \delta_{12}^1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \delta_{12}^n \\
 h_{mi} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \delta_{mi}^n \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 h_{kw} & \delta_{kw}^1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \delta_{kw}^n
 \end{array} \right)
 \end{matrix}$$

Fig. 2. The TSLP reformulation as a set covering problem.

a Meta-heuristic to control the solution process and direct it to the aimed multi-objective analysis.

An effective meta-heuristic procedure is presented in [46] for the SCP. It depends on a randomized priority search (Meta-RaPS). It is accordingly modified to suit the proposed formulation of the problem as follows;

The pseudo-code of the proposed algorithm:

```

Input: All generated path sets ( $H_i$ ), All network links ( $L$ ), Neighbor_search_magnitude,
Tolerance%, No. of iterations ( $\Lambda$ ).
1. Build Set covering matrix ( $C$ )
2. Set the solutions set to be empty:  $S = \emptyset$ 
3. Set the  $\bar{L}$  set to be empty:  $\bar{L} = \emptyset$ 
4. Set  $I$  to be set of the currently uncovered rows:  $I = k \times w$ 
5. Let  $C^* = C$ 
6. Let  $L^*$  (new_set) =  $\emptyset$ 
7. For iter. = 1 to  $\Lambda$  do
8. Randomly remove links from  $L^*$ , the maximum number of links to be removed equals to:  $|L^*| \times (100 - \text{Neighbor\_search\_magnitude})$ 
9. While  $I \neq \emptyset$  do
10. Calculate  $PI_{max} = \max PI_a, a \in L / L^*$ 
11. Construct Candidate List ( $CL$ ) =  $\{u: u \in L / L^* \text{ and } PI_u \leq [PI_{min} \times (1 - \text{tolerance\%} / 100)]\}$ 
12. Randomly select  $l_s$  from  $CL$  and add it to  $L^*$ 
13. Mark each path group  $h_{mi}$  with  $\delta_{mi}^a = 1$  as ready for deletion
14. Mark each path in a group  $H_i$  containing a marked path as ready for deletion
15. Delete all marked paths (covered rows by  $l_s$ )
16. Delete empty columns
17. Update  $I$ 
18. end while
19.  $\bar{L} = L^*$ 
20. add  $\bar{L}$  to  $S$ 
21. end for
22. for each  $\bar{L} \in S$  Calculate MPRE and  $\eta$ 
23. sort and filter the solutions to produce Pareto solutions
24. End_of_Algorithm

```

The algorithm's operators are used to diverse the search process aiming to reach the near optimal Pareto solutions. The basic concept of the used algorithm is illustrated, in a very simple and generalized way, in Fig. 3. To give more insight into the proposed algorithm and its diversity operators, see the numerical example in Appendix A.

4.1. Link priority index

The core of the proposed algorithm is the selection of columns (links) in a stepwise manner according to a link Priority Index (PI_a), represented in Eq. (11). It reflects the potentiality of selecting each link (a) to the final solution set (\bar{L}).

$$PI_a = \xi_1 v_a + \xi_2 \sum_{m=1}^k \sum_{i=1}^w \delta_{mi}^a + \xi_3 c_a + PRV_a \quad (11)$$

where; δ_{mi}^a is dummy variable = 1 if path h_{mi} contains link l_a , 0 otherwise, c_a is the cost of installing traffic counter per link a , \$/l. The first term is the total link flow which comes from the historical O/D matrix data. It represents how much traffic information could be conveyed by this link selection. The second term denotes the number of unobserved paths passing through the link which depends on network structure. That term makes it more likely to cover all O/D pairs with a fewer number of links. The third term manages to incorporate different sensor installation cost for each link. PRV_a represents designer preference value in which he/she may direct the algorithm to choose particular links (e.g. links already equipped with sensors) in every solution by using a very large value. ξ controls the search direction. Different structures of ξ are used to create the required diversity of the final solutions set.

The link PI_a indicates that the links potentiality to be incorporated

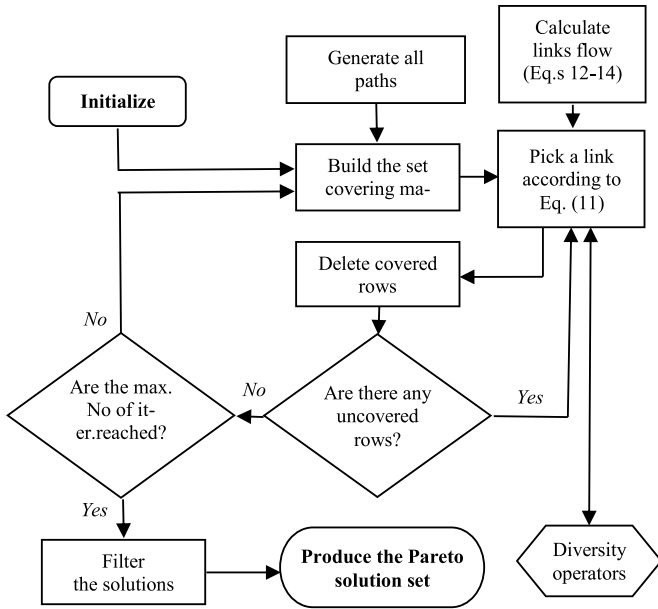


Fig. 3. The proposed algorithm general frame-work.

in the final solution is expressed in terms of the flow information obtained by this link. Therefore, after each link selection, the covered rows and observed demand are deleted. Then, the links' P_{I_a} are updated accordingly. Hence, there is a guarantee that every next selection would contribute to MPRE value.

4.2. Traffic assignment step

In (11), it is apparent that a link flow (ν_a) is an important element to determine the link priority to be selected. To find the flows over a given network an O/D matrix and an assignment method must be defined. The historical O/D matrix can substitute the real matrix since the link flow values are only reflection to the links importance in the network. However, the flows seem to be more sensitive to the used traffic assignment model.

In the strategic stage, the assignment problem has been traditionally treated mainly with two different methods: the user equilibrium (UE) model and the stochastic user equilibrium (SUE) model. In this study, the SUE approach is adopted, since the UE can be considered a special case of the SUE when users have perfect knowledge of their expected travel time. The SUE assignment is selected as an example however the assignment model should reflect the traffic condition for the network under study.

In [20], the concept of SUE is defined for the first time. Several versions are used as simple multinomial logit (MNL) model or multinomial probit (MNP) model. The SUE method considers paths travel time perceived by drivers as random variables. Therefore, there would be a variation in drivers' preferences for the real shortest path. Since all these items are considered random, the users make the appropriate choice of routes in a random manner [48]. The SUE based on the MNL choice solves the following;

$$\text{Minimize}_{\nu} Z_{SUE} = \frac{1}{\theta} \sum_{i \in W} T_i \log \left(\sum_{i \in W} \sum_{m \in H_i} \exp(-\theta \lambda_{mi}) \right) + \sum_{a \in A} \nu_a q_a(\nu_a) - \sum_{a \in A} \int_0^{\nu_a} q_a(\tau) d\tau \quad (12)$$

s.t

$$q_a(\nu_a) = q_a^0 \left(1 + \alpha_a \left[\frac{\nu_a}{Q_a} \right]^{\chi_a} \right) \quad (13)$$

$$\lambda_{mi} = \sum_{a \in A} q_a(\nu_a) \delta_{mi}^a \quad (14)$$

where θ is a parameter that reflects the knowledge of users. The link delay volume function is given by the well-known Bureau of Public Roads (BPR) formula in Eq. (13) [50]. α, χ are calibration parameters defining how the cost increases with traffic flow. The assignment model is solved based on the Method of Successive Averages (MSA) [47]. At each iteration, a search direction is found by carrying out a stochastic loading based on travel costs calculated from the current link flows.

4.3. Diversity operators

The key elements in the methodology are the perturbation operators. They are the tools that help in discovering new areas of the solution search space. Two main operators are tackled, namely: tolerance percentage and neighbor search magnitude. Tolerance percentage allows the algorithm to select a link that is neighbor to the highest priority one. It allows for continuously changing the solution structure at each iteration. Increasing its value makes the algorithm much more randomized. Neighbor Search Magnitude is a percentage of links to be moved from the current iteration to the next as part of the new solution. It helps the algorithm to touch new regions of the solution search space and to not get stuck in local optima. Like any Meta-heuristic procedure based on a stochastic mechanism, increasing the number of iterations may help in finding better solutions.

The methodology could be extended to incorporate the screen line problem stated in Eqs. (9), (10). The coverage steps 13–15 are responsible for this extension. It is sufficient to delete step. 14. Then, the algorithm searches how to locate the sensors to intercept all the paths and consequently the total demand flow.

In step. 22, the MPRE model presented in Eqs. (1)–(4) are calculated using the null-space active-set method [29]. It yields significant advantages in the computational effort and the storage requirements. The time complexity of the methodology for the overall procedure is $O[k \times w + \Lambda \sum_{m=1}^k \sum_{i=1}^w \sum_{a=1}^n \delta_{mi}^a]$.

4.4. Validation indicator

For a given link set installed with sensors, the Generalized Least Square (GLS) could be applied to estimate the O/D matrix from the counting on the equipped links [11]. GLS stipulates the estimated matrix to produce the same flows on the equipped links (flows obtained from real counting on the network) as follows [3]:

$$\text{Minimize}_T (\bar{T} - T)^t \sum_T^{-1} (\bar{T} - T) + (V - V^*)^t \sum_V^{-1} (V - V^*) \quad (15)$$

s.t

$$\sum_{i=1}^w p_{ai} T_i^* = \sum_{i=1}^w p_{ai} \bar{T}_i = \nu_a, \forall a \in \bar{L} \quad (16)$$

$$\bar{T}_i \geq 0 \forall i \in W \quad (17)$$

where; T is the historical demand node pairs vector, \bar{T} is the estimated demand node pairs vector, V is the vector of the reference link flows, V^* is the vector of the observed/real link flows (real link flow related to the true target demand), \sum_T, \sum_V are weight matrices represents the relative confidence between the reference demand and the observed flow.

The dilemma here is that we do not have any counting data at the design stage. If the real O/D matrix is assumed to be known (for validation purpose), we could consider the links flow come from the assignment of the true matrix as the real counted flow. The real O/D matrix existence also gives the opportunity to validate the obtained solutions against the True Relative Error (TRE). The TRE is calculated as follows;

$$TRE = \sqrt{\sum_{i=1}^w \lambda_i^2 / w} \tag{18}$$

$$\lambda_i = \frac{T_i^* - \bar{T}_i}{T_i^*}, \forall i \in W \tag{19}$$

5. Experimental study

To validate the proposed methodology, two real networks are adopted; Sioux Falls network and Ciudad network. The algorithm was written in Matlab, and was run on a workstation with two Intel® Xeon® Processor E5530, 2.40 GHz, 12 GB RAM. Preliminary runs were conducted to adjust the parameters. The number of maximum iterations is 100. The percentage of tolerance and neighbor search magnitude take the values (0, 0.25, and 0.5). The structure of ξ is continuously changed during the iterations to produce different trade-offs of the solutions. ξ_3 and PRV_a are set to zero. For both networks and for each O/D pair, paths are generated depending on k shortest path presented in [60].

The link flows are generated by assigning travel demands from the prior O/D matrix of each network according to the SUE presented in Section 4.2 with a dispersion parameter ($\theta=0.01$).

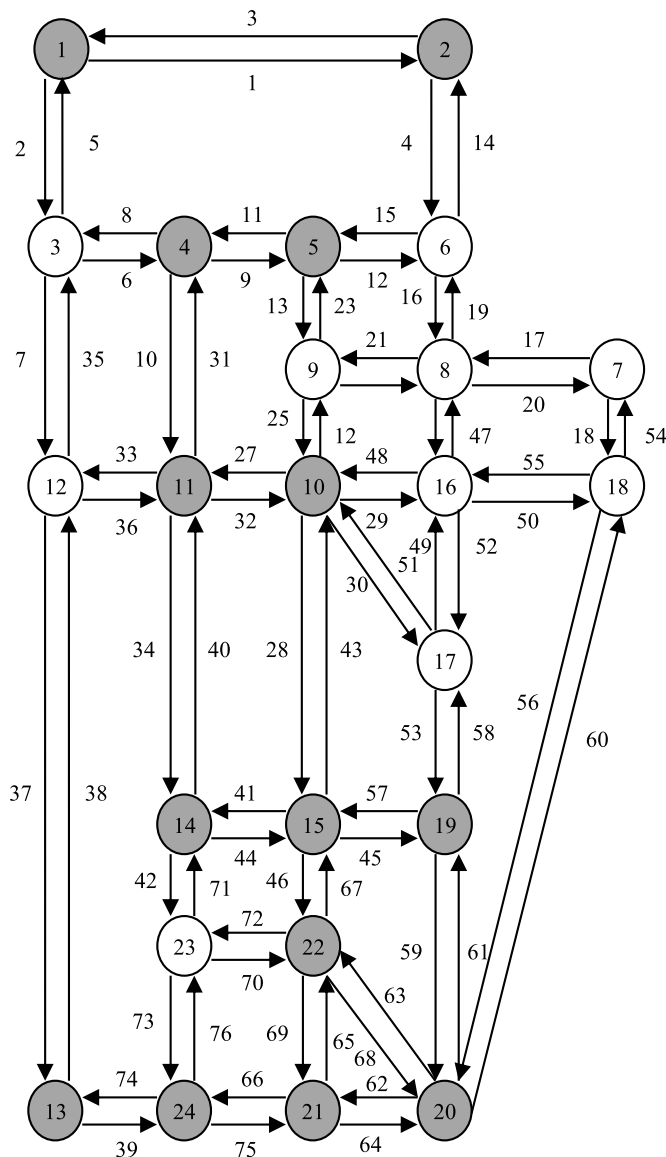


Fig. 4. Sioux Falls Network.

Table 2
Summary of the methodology results.

η	MPRE	TOF	NOF
10	4.85	48.25	33.57
11	4.52	42.47	32.53
17	3.72	79.97	54.23
18	3.62	77.17	53.67
19	3.11	84.51	54.85
20	2.79	89.51	56.80
27	2.69	108.94	63.48
30	2.63	134.30	71.61
38	2.32	145.09	69.24
49	2.27	209.38	80.77
50	2.26	214.25	80.77
52	2.24	211.41	80.25
53	2.14	216.88	80.17
58	2.09	226.95	80.77

5.1. Sioux Falls Network

It is first introduced for the TSLP problem by Yang and Zhou [57]. It is considered a benchmark problem for most of the TSLP literature [17,24,30,32,34,39,52-54,57,62]. The network consists of 182 O/D pairs, 76 links, and 24 vertices. Fig. 4 depicts the Sioux Falls network. The shaded nodes represent both trip origins and destinations.

5.1.1. Results and discussions

The results summary is presented in Table 2. Fourteen non-dominated solutions are obtained. Each solution is characterized by the number of selected links (included the distribution of these links through the network), the MPRE percentage, Total Observed Flow (TOF) and Net Observed Flow (NOF).

Fig. 5 portrays the inherent conflict between the MPRE value (potential accuracy) and the number of links to be equipped (total cost). It shows the multi-objective nature of the problem that precludes finding a unique optimal solution. Two extremely biased solutions (either to the accuracy or to the cost) could be obtained and other solutions are alternatives for various trade-offs. The structures are shown in Table 3. It is worth noting that we attained a finite MPRE (Rule1) at only 10 sensors, which is lower than other studies in the literature for the Sioux Falls network [18,54,56]. It is worth noting again that the number of obtained sensors is sensitive to the number of enumerated paths per each O/D pair and the applied traffic assignments model. Therefore, they should reflect the real traffic conditions for the network under study.

Also, this lower bound is attained when solving the formulated SCP considering only minimizing the number of sensors, Eqs. (9) and (10),

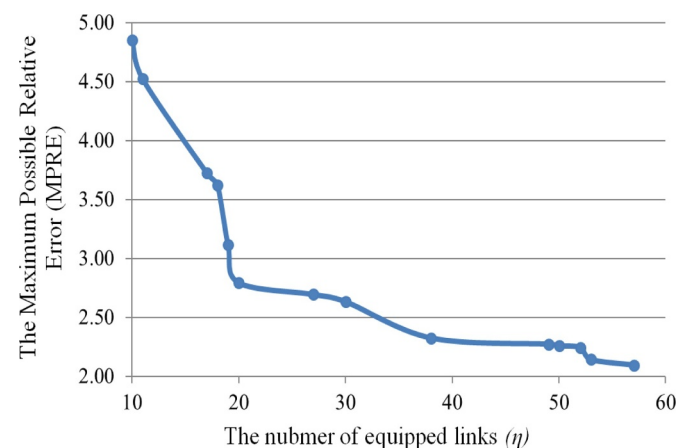


Fig. 5. The MPRE vs number of equipped links (η) for Sioux Falls Network (near Pareto optimal solutions).

Table 3
Link set structure for the two extreme solutions.

Solution I ($\eta = 10$)	Solution II ($\eta = 58$)
Links:	Links:
9, 10, 11, 12, 14, 15, 37, 39, 59, 73	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 22, 23, 25, 26, 27, 28, 31, 32, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 51, 53, 55, 56, 57, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 73, 74, 75, 76

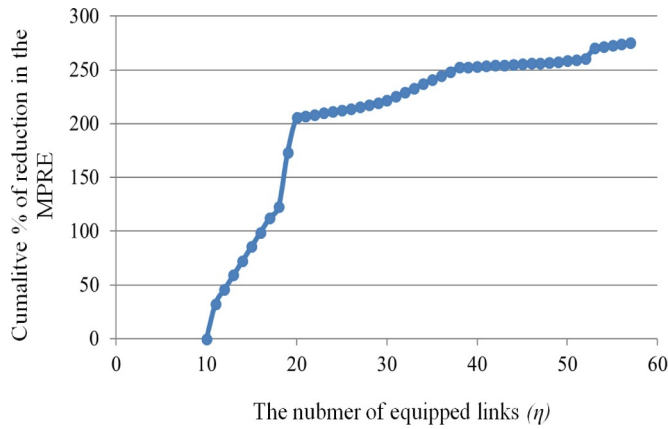


Fig. 6. The marginal effect of deploying traffic sensors for Sioux Falls Network.

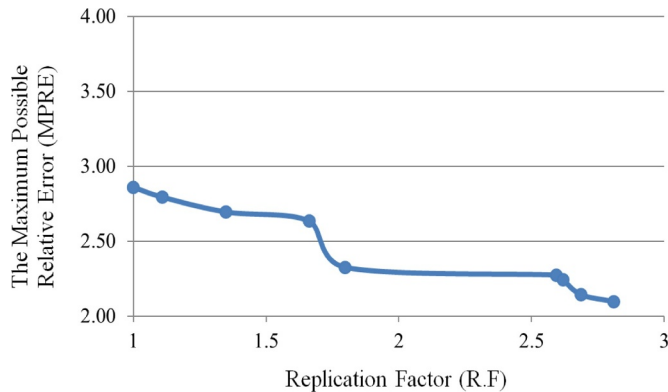


Fig. 7. Maximum possible relative error vs the replication factor.

with the branch and bound technique [10]. This match gives more validity to the algorithm when it is used in solving large networks for which the exact methods are not applicable.

Economically, it is better to know the marginal benefit of increasing the cost by deploying a new unit in your system. Fig. 6 depicts the reduction in MPRE corresponding to the increase in the number of used sensors. It could be noticed that the critical point is at 20 sensors. The rate of reduction before it is about 17.5% per installed sensor, and after it, is about 1.7% per sensor. Therefore, if the available budget manages to equip a number of links ranges from 10 to 30 sensors, 20 sensors would be the most economical choice. However, if there is an accuracy

Table 4
Link set structure comparison.

[7] $\eta = 55$ Links:	[56] $\eta = 48$ Links:	This study $\eta = 45$ Links:
1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 38, 39, 40, 41, 42, 43, 44, 45, 46, 49, 51, 52, 53, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 74, 75, 76	1, 2, 3, 4, 8, 9, 10, 11, 12, 13, 26, 27, 28, 31, 32, 34, 35, 36, 37, 39, 40, 41, 43, 44, 45, 46, 47, 50, 53, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 72, 73, 74, 75, 76	1, 3, 4, 5, 6, 9, 10, 11, 13, 14, 23, 27, 28, 31, 32, 34, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 53, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 74, 75, 76

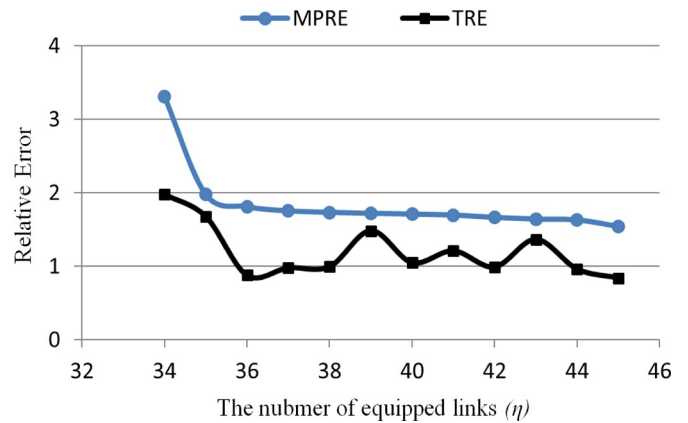


Fig. 8. The MPRE and TRE vs. number of equipped links (η) for Ciudad Real network.

stipulation, we would return to the multi-objective balance.

In the literature, there is a great concern given to place traffic sensors to intercept as maximum as possible of the net traffic flow. It aims to increase the information gained by the sensors. However, the results show that duplicating flow counting decreases the MPRE value. In Fig. 7, the relation between the observed flow Replication Factor (RF) and the MPRE value is drawn. RF is equal to the TOF divided by the NOF. For $RF = 2$ (i.e. observing each vehicle twice on average), the MPRE decreases by 65%. This supports the decision to prioritize each link according to its total flow in Eq. (11) rather than its net observed flow.

5.1.2. The screen line problem

Table 4 shows the location and the number of sensors in a comparison with the only two studies that reported their results considering the full flow interception problem. This study achieved the lowest number of sensors to fully separate all O/D pairs.

5.2. Ciudad Real Network

This subsection is provided to illustrate the application of the proposed methodology on a large real-size network. The Ciudad Real network consists of 380 O/D pairs, 218 links, and 105 vertices, see Fig. 9. The real O/D matrix and historical O/D matrix are reported in [13]. The paradox of sensors location design discussed earlier as for the O/D matrix estimation is that the estimation techniques depend on finding the matrix which reproduces the same flow on the observed

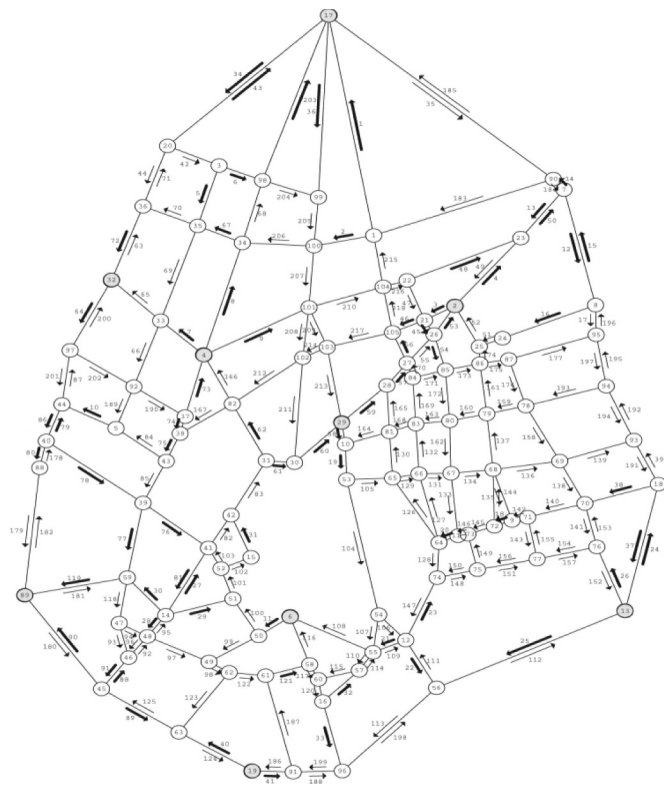


Fig. 9. Ciudad Real Network structure.

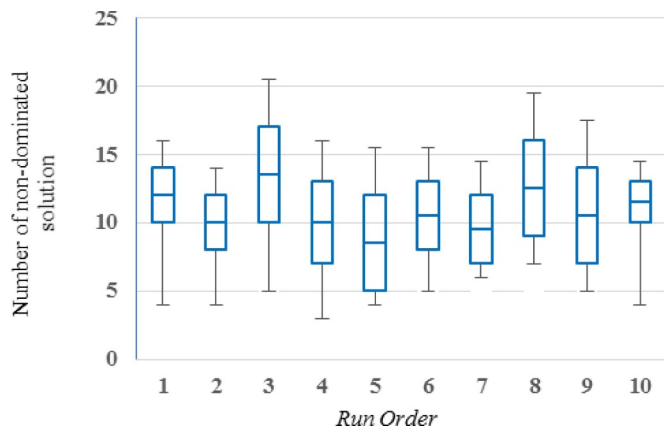


Fig. 10. Box plot draw for different ten runs.

links. However, in the design stage, real flow on links cannot be known until applying the sensors on the network. Reporting the real O/D matrix besides the historical one in [13] manages us to obtain the real flow of network arcs by the assignment step for the real O/D matrix.

Appendix A

In this section, it is considered the small network represented in Table A.1 which consists of 7 links and three demand node pairs. The historical O/D matrix resulted in the given link flows and the maximum number of paths for each node pair is limited to 3. The algorithm settings and the ξ structures is given in the Table 1.

Let us now go through 3 different iterations step by step:

Iter. no. 1:

- 1 use ξ str. 1 to calculate the PI_a for each link (input step)*
- 2 construct the CL based on the PI_4 (links in the vicinity of 15% for I_4 Priority Index) $CL = \{I_4, I_3\}$. (steps 1 to 6)
- 3 choose link randomly from the CL to be in the solution set $L^* = \{I_3\}$. Steps (7–12)
- 4 node pairs (2&3) are covered so delete rows 4 to 9. (steps 13–17)

Fig. 8 plots the MPRE and the TRE variation at different sensor numbers. The results indicate that the MPRE sets a boundary of the estimated TRE in which the obtained solutions cannot exceed. Therefore, the MPRE proves itself as a good criterion for the design of TSLP regarding the accuracy of the estimated O/D matrix.

To check the robustness of the methodology, other nine different runs were made. Each run contains 100 different iterations. The number of resulted non-dominated solutions per each iteration is depicted in the box plot in Fig. 10. The ANOVA test is performed to make sure if there is a significant difference in the results. The $F_{critical}$ ratio is equal to 1.98 with 9 degrees of freedom for the numerator and 90° for the denominator at significance level = 0.05. The obtained $F = 1.23$ ($F < F_{critical}$) which means that the null hypothesis that the results are statistically equivalent is accepted, i.e. there are no significant differences among the different runs.

6. Conclusion

This study presents a new multi-objective sensor location methodology considering the O/D matrix estimation, which depends on robust measurements of accuracy. It measures the maximum possible deviation that could be obtained by the worst estimated solution. Different trade-offs between potential accuracy of the estimated O/D matrix and the number of sensors are generated. The benefit of using the MPRE directly in the multi-objective design is shown for the first time. The polynomial time complexity of the proposed methodology manages to tackle real size networks. The results using two real networks show the generality and effectiveness of the proposed sensor location methodology. The obtained sensors number is plotted against the true relative error (TRE), which is bounded by the MPRE value. This ensures the reliability of any solution set when it is deployed on the network. The robustness of the methodology is measured by the ANOVA test which ensured the reliability of obtained solutions. Moreover, the methodology allows using the installation cost of sensors. This manages to choose the most suitable solution for a predetermined budget. The methodology is also extended to determine the minimum number of sensors required to solve the screen line problem. For future work, the proposed methodology can be developed to solve the location problem in terms of dynamic O/D matrix estimation using the time expanded network representation in which nodes and links are repeated through time slots. The heuristic nature of the methodology would help in dealing with this network expansion. The challenge would be how to determine the link priority to be selected over time and how to define the MPRE in terms of both dynamic O/D matrix entries and the expanded network.

Acknowledgment

The author would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. 1440 – 14.

Table A.1
The small network structure and basic data.

Row no.			l_1	l_2	l_3	l_4	l_5	l_6	l_7
1	Node pair 1	h_{11}	1	1	0	1	0	1	0
2		h_{21}	0	1	0	0	0	0	1
3		h_{31}	1	0	0	0	1	0	0
4	Node pair 2	h_{12}	1	1	1	0	1	0	0
5		h_{22}	1	0	0	0	0	1	0
6		h_{32}	0	0	1	0	1	0	1
7	Node pair 3	h_{13}	1	0	1	1	0	1	0
8		h_{23}	0	1	1	1	1	0	1
9		h_{33}	1	1	1	1	0	1	0
Covered paths sum.			6	5	5	4	4	4	3
Link volume v_a			20	40	120	150	60	70	30
PI_a			$\xi_1 = 1$						
$(\xi \text{ str. } 1)$			$\xi_2 = 20$						
PI_a			$\xi_1 = 0.8$						
$(\xi \text{ str. } 2)$			$\xi_2 = 22$						
PI_a			$\xi_1 = 0.6$						
$(\xi \text{ str. } 3)$			$\xi_2 = 25$						
.									
.									
.									
PI_a			$(\xi \text{ str. } A)$						

Algorithm settings: Tolerance% = 15 Neighbor_search_magnitude = 0.5.
 $L^* = \{l_4, l_5\}$, $[\eta = 2 \ \& \ MPRE = 1.12]$.

Table A.2
The updated matrix after first selection in iter. 1.

Row no.			l_1	l_2	l_3	l_4	l_5	l_6	l_7
1	Node pair 1	h_{11}	1	1	0	1	0	1	0
2		h_{21}	0	1	0	0	0	0	1
3		h_{31}	1	0	0	0	1	0	0
Covered paths sum.			2	2	0	1	1	1	1
Link volume v_a			15	10	0	10	5	8	6
PI_a			$\xi_1 = 1$						
$(\xi \text{ str. } 1)$			$\xi_2 = 20$						

- 5 construct the CL based on new best PI_a , see Table A.2, $CL = \{l_1, l_2\}$. (rerun in the loop to step 12)
- 6 choose link randomly from the CL to be in the solution set $L^* = \{l_1, l_3\}$.
- 7 node pair (1) is covered so delete rows 1 to 3.
- 8 the matrix is empty then produce the final solution set:

$L^* = \{l_1, l_3\}$ then estimate these link set performance $[\eta = 2 \ \& \ MPRE = 1.25]$ (steps 18 – 23)
 *the corresponding algorithm steps in page 6
 Iter. no. 2:

- 1 from the previous iteration keep randomly the half of the solution set (Neighbor_search_magnitude = 0.5), $L^* = \{l_1\}$.
- 2 node pairs (1&2&3) are covered so delete rows 1 to 9.
- 3 the matrix is empty then produce the final solution set: $L^* = \{l_1\}$, $[\eta = 1 \ \& \ MPRE = 2.22]$.

Iter. no. 3:

- 1 from the previous iteration keep randomly the half of the solution set (Neighbor_search_magnitude = 0.5) $L^* = \{\emptyset\}$.
- 2 use $\xi \text{ str. } 3$ to calculate the PI_a for each link
- 3 construct the CL based on the PI_3 (links in the vicinity of 15% for l_3 Priority Index) $CL = \{l_3, l_4\}$.
- 4 choose link randomly from the CL to be in the solution set $L^* = \{l_4\}$.
- 5 node pairs (1&3) are covered so delete rows 1 to 3 and 4 to 9.
- 6 construct the CL based on new best PI_a , see Table A.3, $CL = \{l_3, l_5\}$.
- 7 choose link randomly from the CL to be in the solution set $L^* = \{l_4, l_5\}$
- 8 node pair (2) is covered so delete rows 4 to 6.
- 9 the matrix is empty then produce the final solution set: $L^* = \{l_4, l_5\}$, $[\eta = 2 \ \& \ MPRE = 1.12]$

Table A.3
The updated matrix after first selection in iter. 2.

Row no.		l_1	l_2	l_3	l_4	l_5	l_6	l_7
4	Node pair 2	h_{12}	1	1	1	0	1	0
5		h_{22}	1	0	0	0	1	0
6		h_{32}	0	0	1	0	0	1
Covered paths sum.			2	2	2	0	2	1
Link volume v_a			0	5	30	0	10	10
Pl_a (ξ str. 3)		$\xi_1 = 0.6$	50	53	68	0	60	40
		$\xi_2 = 25$						31

References

[1] Beasley JE, Chu PC. A genetic algorithm for the set covering problem. *Eur J Oper Res* 1996;94(2):392–404.

[2] Bell MG. The estimation of an origin-destination matrix from traffic counts. *Transp Sci* 1983;17(2):198–217.

[3] Bell MG. The estimation of origin-destination matrices by constrained generalised least squares. *Transp Res Part B* 1991;25(1):13–22.

[4] Bell MG, Shield CM, Busch F, Kruse G. A stochastic user equilibrium path flow estimator. *Transp Res Part C* 1997;5(3):197–210.

[5] Bera S, Rao K. Estimation of origin-destination matrix from traffic counts: the state of the art. 2011.

[6] Bianco L, Cerrone C, Cerulli R, Gentili M. Locating sensors to observe network arc flows: exact and heuristic approaches. *Comput Oper Res* 2014;46:12–22.

[7] Bianco L, Confessore G, Reverberi P. A network based model for traffic sensor location with implications on O/D matrix estimates. *Transp Sci* 2001;35(1):50–60.

[8] Bonsall P, Firmin P, Anderson M, Palmer I, Balmforth P. Validating the results of a route choice simulator. *Transp Res Part C* 1997;5(6):371–87.

[9] Bosurgi G, Bongiorno N, Pellegrino O. A nonlinear model to predict drivers' track paths along a curve. *Int J Civil Eng* 2016;14(5):271–80.

[10] Caprara A, Toth P, Fischetti M. Algorithms for the set covering problem. *Ann Oper Res* 2000;98(1):353–71.

[11] Cascetta E. "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transp Res Part B* 1984;18(4–5):289–99.

[12] Castillo E, Grande Z, Calviño A, Szeto WY, Lo HK. A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks. *J Sens* 2015;2015:26. 903563.

[13] Castillo E, Jiménez P, Menéndez JM, Conejo AJ. The observability problem in traffic models: algebraic and topological methods. *Intell Transp Syst IEEE Trans* 2008;9(2):275–87.

[14] Castillo E, Menéndez JM, Jiménez P. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transp Res Part B* 2008;42(5):455–81.

[15] Castillo E, Menéndez JM, Sánchez-Cambronero S. Traffic estimation and optimal counting location without path enumeration using Bayesian networks. *Comput-Aided Civil Infrastruct Eng* 2008;23(3):189–207.

[16] Castillo E, Nogal M, Rivas A, Sánchez-Cambronero S. Observability of traffic networks. Optimal location of counting and scanning devices. *Transportmetrica B: Transp Dyn* 2013;1(1):68–102.

[17] Chen A, Chootinan P, Recker WW. Examining the quality of synthetic origin-destination trip table estimated by path flow estimator. *J Transp Eng* 2005;131(7):506–13.

[18] Chootinan P, Chen A, Yang H. A bi-objective traffic counting location problem for origin-destination trip table estimation. *Transportmetrica* 2005;1(1):65–80.

[19] Chvatal V. A greedy heuristic for the set-covering problem. *Math Oper Res* 1979;4(3):233–5.

[20] Daganzo CF, Sheffi Y. On stochastic models of traffic assignment. *Transp Sci* 1977;11(3):253–74.

[21] Danczyk A, Liu HX. A mixed-integer linear program for optimizing sensor locations along freeway corridors. *Transp Res Part B* 2011;45(1):208–17.

[22] Eisenman S, Fei X, Zhou X, Mahmassani H. Number and location of sensors for real-time network traffic estimation and prediction: sensitivity analysis. *Transp Res Record J Transp Res Board* 2006(1964):253–9.

[23] Fei X, Mahmassani H, Eisenman S. Sensor coverage and location for real-time traffic prediction in large-scale networks. *Transp Res Record J Transp Res Board* 2007(2039):1–15.

[24] Fu C, Zhu N, Ling S, Ma S, Huang Y. Heterogeneous sensor location model for path reconstruction. *Transp Res Part B* 2016;91:77–97.

[25] Fu C, Zhu N, Ma S. A stochastic program approach for path reconstruction oriented sensor location model. *Transp Res Part B* 2017;102:210–37.

[26] Gentili M, Mirchandani P. Survey of models to locate sensors to estimate traffic flows. *Transp Res Record J Transp Res Board* 2011(2243):108–16.

[27] Gentili M, Mirchandani P. Locating sensors on traffic networks: models, challenges and research opportunities. *Transp Res Part C* 2012;24:227–55.

[28] Gentili M, Mirchandani PB. Locating active sensors on traffic networks. *Ann Oper Res* 2005;136(1):229–57.

[29] Gill PE, Murray W, Saunders MA, Wright MH. Procedures for optimization problems

with a mixture of bounds and general linear constraints. *ACM Trans Math Softw (TOMS)* 1984;10(3):282–98.

[30] Hadavi M, Shafahi Y. Vehicle identification sensor models for origin-destination estimation. *Transp Res Part B* 2016;89:82–106.

[31] Hazelton ML. Estimation of origin-destination matrices from link flows on uncongested networks. *Transp Res Part B* 2000;34(7):549–66.

[32] He S-x. A graphical approach to identify sensor locations for link flow inference. *Transp Res Part B* 2013;51:65–76.

[33] Hu S-R, Liou H-T. A generalized sensor location model for the estimation of network origin-destination matrices. *Transp Res Part C* 2014;40:93–110.

[34] Hu S-R, Peeta S, Chu C-H. Identification of vehicle sensor locations for link-based network traffic applications. *Transp Res Part B* 2009;43(8):873–94.

[35] Hu S-R, Peeta S, Liou H-T. Integrated determination of network origin-destination trip matrix and heterogeneous sensor selection and location strategy. *IEEE Trans Intell Transp Syst* 2016;17(1):195–205.

[36] Kang Y, Li Z, Zhao Y, Qin J, Song W. A novel location strategy for minimizing monitors in vehicle emission remote sensing system. *IEEE Trans Syst Man Cybern Syst* 2017;48(4):500–10.

[37] Lam W, Lo H. "Accuracy of OD estimates from traffic counts. *Traffic Eng Control* 1990;31(6):358–67.

[38] Lan G, DePuy GW, Whitehouse GE. An effective and simple heuristic for the set covering problem. *Eur J Oper Res* 2007;176:1387–403.

[39] Liu Y, Zhu N, Ma S, Jia N. Traffic sensor location approach for flow inference. *IET Intel Transp Syst* 2015;9(2):184–92.

[40] Mínguez R, Sánchez-Cambronero S, Castillo E, Jiménez P. Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks. *Transp Res Part B* 2010;44(2):282–98.

[41] Morrison DR, Martonosi SE. Characteristics of optimal solutions to the sensor location problem. *Ann Oper Res* 2015;226(1):463–78.

[42] Nie Y, Zhang H, Recker W. Inferring origin-destination trip matrices with a decoupled GLS path flow estimator. *Transp Res Part B* 2005;39(6):497–518.

[43] Okutani I, Stephanedes YJ. Dynamic prediction of traffic volume through Kalman filtering theory. *Transp Res Part B* 1984;18(1):1–11.

[44] Olia A, Abdelgawad H, Abdulhai B, Razavi S. Optimizing the number and locations of freeway roadside equipment units for travel time estimation in a connected vehicle environment. *J Intell Transp Syst* 2017;21(4):296–309.

[45] Owais M, Osman MK. Complete hierarchical multi-objective genetic algorithm for transit network design problem. *Expert Syst Appl* 2018;114:143–54.

[46] Owais M, Osman MK, Moussa G. Multi-objective transit route network design as set covering problem. *IEEE Trans Intell Transp Syst* 2016;17(3):670–9.

[47] Powell WB, Sheffi Y. The convergence of equilibrium algorithms with pre-determined step sizes. *Transp Sci* 1982;16(1):45–55.

[48] Prashker JN, Bekhor S. Route choice models used in the stochastic user equilibrium problem: a review. *Transp Rev* 2004;24(4):437–63.

[49] Sheffi Y. Urban transportation network. Prentice Hall; 1985.

[50] Spiess H. Technical note—Conical volume-delay functions. *Transp Sci* 1990;24(2):153–8.

[51] Wang N, Gentili M, Mirchandani P. Model to locate sensors for estimation of static origin-destination volumes given prior flow information. *Transp Res Record J Transp Res Board* 2012(2283):67–73.

[52] Xu X, Lo HK, Chen A, Castillo E. Robust network sensor location for complete link flow observability under uncertainty. *Transp Res Part B* 2016;88:1–20.

[53] Yang C, Chootinan P, Chen A. Traffic counting location planning using genetic algorithm. *J East Asia Soc Transp Stud* 2003;5:898–913.

[54] Yang H, Gan L, Tang W. Determining cordons and screen lines for origin-destination trip studies. *Proceedings of the Eastern Asia Society for Transportation Studies*. 3. 2001.

[55] Yang H, Iida Y, Sasaki T. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. *Transp Res Part B* 1991;25(5):351–63.

[56] Yang H, Yang C, Gan L. Models and algorithms for the screen line-based traffic-counting location problems. *Comput Oper Res* 2006;33(3):836–58.

[57] Yang H, Zhou J. Optimal traffic counting locations for origin-destination matrix estimation. *Transp Res Part B* 1998;32(2):109–26.

[58] Yang L, Ning Z. A multi-objective detector location optimization approach. 14th COTA international conference of transportation professionals. American Society of Civil Engineers; 2014.

[59] Ye P, Wen D. Optimal traffic sensor location for origin-destination estimation using

- a compressed sensing framework. *IEEE Trans Intell Transp Syst* 2017;18(7):1857–66.
- [60] Yen JY. Finding the k shortest loopless paths in a network. *Manag Sci* 1971;17(11):712–6.
- [61] Yim PK, Lam WH. Evaluation of count location selection methods for estimation of OD matrices. *J Transp Eng* 1998;124(4):376–83.
- [62] Zangui M, Yin Y, Lawphongpanich S. Sensor location problems in path-differentiated congestion pricing. *Transp Res Part C* 2015;55:217–30.
- [63] Zhou X, List GF. An information-theoretic sensor location model for traffic origin-destination demand estimation applications. *Transp Sci* 2010;44(2):254–73.
- [64] Zhou X, Mahmassani HS. Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Trans Intell Transp Syst* 2006;7(1):105–14.