

## Research Article

# A New Model for Locating Plate Recognition Devices to Minimize the Impact of the Uncertain Knowledge of the Routes on Traffic Estimation Results

**Santos Sánchez-Cambronero** , **Fernando Álvarez-Bazo** , **Ana Rivas** ,  
and **Inmaculada Gallego** 

*Department of Civil and Building Engineering, University of Castilla-La Mancha, 13071 Ciudad Real, Spain*

Correspondence should be addressed to Santos Sánchez-Cambronero; [santos.sanchez@uclm.es](mailto:santos.sanchez@uclm.es)

Received 3 May 2020; Revised 22 June 2020; Accepted 10 July 2020; Published 28 August 2020

Academic Editor: Jinjun Tang

Copyright © 2020 Santos Sánchez-Cambronero et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A number of research papers have recently shown that the use of techniques based on the installation of vehicle identification devices allows us to address the observability problem of a traffic network in a much more efficient way than if it were done with traditional techniques. The use of such devices can lead to a better data set in terms of flows and therefore to a better definition of traffic flows, which is essential for traffic management in cities and regions. However, the current methodologies aimed at network modeling and data processing which are not fully adapted to the use of these devices in obtaining the necessary data for analyzing traffic and making network forecasts. This is because the essential variable in models which used data from plate scanning (as a particular case of AVI sensors) is composed of the route flows, while traditional methods are based on the observation of link and/or origin-destination flows. In this context, this paper proposes several practical contributions, in particular: (1) a traffic network design method aimed to use the plate scanning data to estimate traffic flows and (2) an algorithm for locating plate reader devices to reduce the effect of the uncertain knowledge of route enumeration. Next, using the well-known Nguyen-Dupuis network, a sensitivity analysis has been carried out to evaluate the influence of different parameters of the model on the final solution. These parameters are the considered routes, the degree of network simplification, and the available budget to install devices. Finally, the method has been applied to a real network.

## 1. Introduction

As is well known, estimating the origin-destination trip matrix, route flows, and link flows is essential to achieving efficient traffic management. Many authors have dealt with this problem, trying to estimate these traffic flows using either information from traditional sources such as traffic counts (see, among others, Castillo et al. [1, 2] and Perrakis et al. [3]) or information from more innovative sources such as mobile phones and GPS data (Huang et al. [4], Ibarra-Espinosa et al. [5], and Moreira-Matias et al. [6]), Big Data (Toole et al. [7] and Zin et al. [8]), or automatic vehicle identification (AVI) data (Castillo et al. [9], Fu et al. [10], and Fedorov et al. [11], among others).

Yang et al. pointed out [12], to deal with traffic flow estimation (not only for short-term predictions but for more

generalized studies too), classical statistical methods are widely applied, but also, machine learning methods are shown very useful due to their many advantages as, for example, problem adaptability, generalization, and also the learning ability, which is very important to estimate traffic flows using field data. For example, Sánchez-Cambronero et al. [13] used Bayesian networks; Bai and Chen [14] used neural networks; and Lui et al. [15] used deep learning. In any case, without going into detail concerning the models used to predict traffic flows, to conduct such traffic analysis both for static and dynamic studies, technicians need two things: a good representation of the traffic network and a good data set with which to simulate routes of the network and to predict the flows (Nigro et al. [16]). This means that the optimal number and locations of the sensors that can collect such data must be determined. The next sections of

this introduction deal with these problems in order to clarify concepts and describe problems that this paper faces.

*1.1. The Network Representation.* A traffic network is a pair  $(\mathcal{N}, \mathcal{A})$ , where  $\mathcal{N}$  is a set of nodes and  $\mathcal{A}$  is a set of directed links connecting these nodes. The links represent the streets of a city, and the nodes typically represent the intersections of these streets. To build a transportation network model, two aspects of the network must be considered (see Sheffi [17]):

- (i) The quantitative information associated with each link: this is the uncongested travel time, the travel cost, the parking places, the number of street lanes, the number of residents, etc. Each network link is associated with some impedance function derived from this information that, for a given flow, provides the “generalized cost” of using this link (see, for example, Spies [18], Huntsinger and Roupail [19], or Mtoi and Moses [20]).
- (ii) The graph representation: the role of the graph representation is to translate the physical structure of a city into a model of nodes and links. Some simplifications are needed:
  - (a) The division into traffic zones: the beginning of a commute is, for example, a person’s house, and the end is his or her workplace. To model this situation for all city inhabitants, thousands of origins and destinations would be needed. Thus, the transportation planning process is typically based on a partition of the city into traffic zones that are represented by nodes known as “centroids” from which all traffic routes are assumed to start and/or finish. They represent an aggregation of all the actual origins and all the actual destinations in each zone. Once the centroids are defined (and thus the set  $\mathcal{O}$  of origins and  $\mathcal{D}$  of destinations), the movement over an urban network can be expressed in terms of an origin-destination (OD) matrix  $\mathbf{T}$ , where  $t_{ks}$  is the number of trips originating from zone  $k$  and ending at zone  $s$ .
  - (b) The connection between zones and links: each zone (modeled by its centroid) is joined to the road network by special links called “connectors.” These links are fictitious links that do not represent any street of the city. The number of connectors depends on the level of detail with which an urban area is represented, but sometimes are chosen arbitrarily.

Indeed, the decision of how many centroids and then how many connectors must be used is closely related to the flow estimation error. For example, the links directly connected to the centroids may lead to incorrect flow and artificial congestion, or most importantly, for the purpose of this work, the resulting routes may be unrealistic.

Although the distribution of centroids and connectors seems crucial to obtain a coherent traffic flow estimation, it has received limited attention. Among the few studies found, Mann [21] presented a model in which every zone was divided into subareas with the aim of reducing assignment error; Friedrich and Galster [22] suggested methods for generating connectors based on geometric features in a microscopic reference scenario; Quian and Zhang [23] proposed a connector optimization algorithm to decide the number and location of connectors in order to minimize the maximum volume/capacity ratio in a given subset of network links by changing the connector travel time; and Jafari et al. [24] used a bilevel method to distribute each centroid demand both to its nearby nodes and to its peripheral nodes. Other methods for traffic network modeling are presented as in Hao and Yang [25] where they introduced the theory of granular computing to model the elements of the multilayer traffic network.

In summary and following some of the conclusions drawn by Quian and Zhang [23], building a transportation model with a good distribution of centroids and connectors is both a difficult and an important task because some problems may arise when any traffic assignment model is used:

- (i) The estimated link flows change significantly depending on the connector configurations
- (ii) If the network model is not well designed and the connectors’ travel time is not well defined, the final demand assignment can lead to a solution in which connectors are used to bypass congested links that would otherwise have to be used
- (iii) Too few connectors often lead to artificial congestion in those links adjacent to the connector
- (iv) Since the routes begin with the connectors, if few connectors are designed, the set of routes may be unrealistic or uncompleted

Up to this point, we have not discussed using anything but traditional methods to predict traffic flows, which are usually updated using observed representative link flows, in the network modeling problem. However, the model proposed by Castillo et al. [26] (and then extended by Mínguez et al. [27] or Sánchez-Cambroneró et al. [13]) suggests vehicle plate recognition as an alternative way to collect traffic data since this method is much more informative than traditional ones and can therefore be used more efficiently for traffic flow estimation. Other authors, such as Zhou and Mahmassani [28] or Liu et al. [29] or Li et al. [30], also used the information based on the automatic number of plate recognition in their models.

If little attention has been paid to build an appropriate traffic network using traditional methods, even less attention (to our knowledge) has had the development of models to build an appropriate network for traffic flow estimation using plate-recognition-based data, where the route flow is the key variable to be estimated (note that this procedure is just one of the possible techniques of automatic vehicle identification which can also be applied with the method

proposed in this paper). Sánchez-Cambronero et al. [31] addressed this problem by assuming that every node of the network can be the origin and destination of trips and built a node-based OD matrix used as a reference. With this, the authors proposed to use a route enumeration algorithm to build an “exhaustive set of routes” ( $\mathfrak{R}$ ) between the nodes of the network. Then, a route simplification algorithm is proposed based on transferring to adjacent nodes the generated or attracted (reference) demand of those nodes that generate or attract fewer trips than a given threshold ( $F_{\text{thres}}$ ). After the simplification process, a new set of routes  $\mathcal{Q}$  is obtained to be used in the plate scanning device location model. However, the authors did not mention the criteria to determine those sets of origin-destination nodes or the method to build the exhaustive set of routes, which, in practice, are two key points of network model design. As it will be shown, this paper deals with these problems.

*1.2. The Device Location for Traffic Flow Prediction.* Given a traffic network model  $(\mathcal{N}, \mathcal{A})$ , the device location for the traffic flow prediction problem consists of determining which subset of  $\mathcal{A}$  should be observed in order to estimate the traffic flow in the most reliable way. Due to the importance of device locations to achieving trustworthy flow predictions, many authors have addressed the issue of determining the optimal number and locations of traffic counts (see, for example, Yang and Zhou [32], Ehlert et al. [33], Gentili and Mirchandani [34], or Salari et al. [35]). Most of their models are formulated with the assumption that a set  $\mathfrak{R}$  of routes is given and fixed. However, we have found a lack of analysis of how the network model design (traffic zones, centroids, connectors, links, etc.) affects this route enumeration.

Location models for plate recognition devices have been studied by Mínguez et al. [27], Castillo et al. [36], Yang and Sun [37], Sánchez-Cambronero et al. [38], Fu et al. [10], and Gentili and Mirchandani [39], among others. Again, all of these authors assume a given and fixed set of routes for the location model but do not analyze how to determine those routes and, even more important, how these location models (and the resulting flow estimates) are affected by uncertain knowledge of the routes. In addition, note that the device location problems do not have a unique solution, and the implications of such selection for later flow estimation are very relevant and important and hence deserve a deep analysis.

*1.3. Contributions of This Paper.* In the view of the above, this paper proposes a two-step methodology which leads to the following contributions compared with some of the previous studies in the same topic:

- (i) We propose (in the first step) a traffic network design method (based on the one proposed by Sánchez-Cambronero et al. [31]) to be used in traffic flow estimation models that use data collected by AVI sensors. This method is an alternative to the classic modeling of the network using centroids and

connectors as those proposed by Quian and Zhang [23] or Jafari et al. [24]. Note that since the estimation models that use this field information from AVI sensors try to reconstruct users’ routes, the precision to define network routes must be very high. This is one of the main advantages of the proposed method.

- (ii) We propose (in the second step) a new heuristic algorithm to obtain the plate scanning devices’ location aimed to obtain the best possible result in terms of traffic flow estimation. The main step forward of this contribution is that it deals with the uncertain knowledge of the network routes. Up to now, the existing models (proposed, for example, by Mínguez et al. [27] or Cerrone et al. [40]) assume the set  $\mathcal{Q}$  of modeled routes known and fixed. However, due to traffic conditions, the proposed set of scanned links ( $\mathcal{S}\mathcal{L}$ ) may give a set of observed combinations of scanned links ( $\mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ ) different from the expected set ( $\mathcal{E}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ ). Therefore, since the routes actually used by vehicles may not have been included in  $\mathcal{Q}$ , it may be possible that some of these  $\mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$  have no intersection with the routes in  $\mathcal{Q}$ , and hence, the observed flow cannot be assigned. For this, the algorithm expands this set of modeled routes to a bigger set  $\mathcal{C}$  of routes so that all of the combinations in  $\mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$  have intersections with at least one route.
- (iii) We propose a sensibility analysis to evaluate the influence of different parameters of the model on the final solution. This contribution is very important since it may give tools to the transportation planner to decide which is the best value for each parameter depending on the particular case to study.

The rest of the paper is organized as follows: in Section 2, the problem of the uncertain knowledge of network routes is explained; in Section 3, the model proposed by Sánchez-Cambronero et al. [31] is discussed and improved; in Section 4, the proposed algorithm is presented, described, and analyzed; Section 5 is devoted to performing the sensibility analysis of the model parameters in the solution using the well-known Nguyen-Dupuis network to next apply the method to a real network; and finally, some conclusions are provided in Section 6.

## 2. The Impact of the Uncertain Knowledge of Routes on the Traffic Flow Estimation Results

Traffic estimation models based on vehicle plate recognition (as a particular case of AVI sensors) are based on identifying the circulating vehicles on some subsets of links to reconstruct vehicle routes or partial routes, from which route, OD, and link flows can be derived. As mentioned before, plate recognition has become a useful technique because of the great amount of information it provides compared with that provided by other very common standard methods (see, among others, Castillo et al. [26]).

To illustrate the concepts involved, the simple network with 6 nodes and 18 links, shown in Figure 1, is going to be used. Let us consider the set of reference routes  $\mathfrak{R}$ , shown in Figure 2(a). Sánchez-Cambronero et al. [31] proposed to obtain this set from a  $k$ -shortest path enumeration algorithm using a node-based OD matrix (i.e., assuming that every single node is able to be an origin and destination). In fact, with this, they tried to cover all possible feasible routes in the network although its route flow may be negligible (this paper extends this procedure in order to give more tools to obtain them). From this set, the method proposed by the authors is applied, and a simplified set  $\mathcal{Q}$  (see Figure 2(b)) of routes is obtained. This set of routes  $\mathcal{Q}$  is assumed to be good enough to perform a traffic flow analysis, and we also assume that this set is the one that the existing methods would use both for locating the sensors and after for the flow estimation. Although the set of routes  $\mathcal{Q}$  may be considered very exhaustive after the simplification process, there is a great uncertainty on whether these routes represent reliably the routes actually used by the users.

Let us illustrate this with an example. Consider that the plate recognition devices are installed in the set  $SL = \{1, 2, 3, 4, 5, 7, 8, 9, 12, 13, 14, 15, 16, 17\}$ . With this information and using existing methods, Table 1 is developed which shows each set of expected combinations ( $s$ ) of scanned links ( $\mathcal{E}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L}$ ) as the intersection of each route in  $\mathcal{Q}$  with the set  $\mathcal{S}\mathcal{L}$ . Let us now consider that we develop a field test, i.e., we install plate reader devices (or AVI devices), so we can obtain the associated observed flow  $\hat{w}_s$  (see the last column in the table). This means, for example, that those vehicles identified in links 1 and 5 (60.31 in this case) belong to route 1. Also, vehicles identified in links 9 and 17 (5.62) belong to route 19. In this example, (which is not the general case), each set  $s$  is associated with just one route ( $q$ ) in set  $\mathcal{Q}$ , i.e., it is expected that the observed flow  $\hat{w}_s$  would be able to derive all the route flows by using the following relation:

$$\hat{w}_s = \sum_{q \in \mathcal{Q}} \lambda_s^q f_q, \quad \forall s \in \mathcal{E}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L}, \quad (1)$$

where  $\hat{w}_s$  is the observed flow in each set  $\mathcal{E}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L}$ ,  $f_q$  is the estimated flow in routes in set  $\mathcal{Q}$ , and  $\lambda_s^q$  is the element of the route-scanned combination incidence matrix for route  $q$ , which equals one if route  $q$  contains the subset  $s$  of scanned links and only those, and zero, otherwise. Therefore, the link flows can be calculated as follows:

$$v_a = \sum_{q \in \mathcal{Q}} \delta_r^a f_q; \quad \forall a \in \mathcal{A}, \quad (2)$$

where  $v_a$  is the flow for link  $a$  and  $\delta_r^a$  is the relation incidence matrix between link and route flows.

However, suppose that the related field test reveals that new sets of combinations of scanned links  $\mathcal{N}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L}$  appear in addition to those shown in Table 1, but in this case, we do not find any intersection with routes of set  $\mathcal{Q}$ . This new set together with its corresponding observed flow is shown in Table 2. For example, we have found that there were vehicles scanned only in link 16 (18.86) which do not match with any

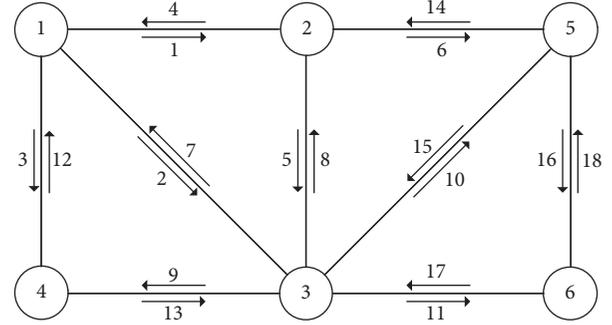


FIGURE 1: Elementary example network.

route in  $\mathcal{Q}$ . Note that  $\mathcal{E}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L} \cup \mathcal{N}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L} = \mathcal{O}\mathcal{S}\mathcal{C}\mathcal{S}\mathcal{L}$ , i.e., the observed set of combinations of scanned links.

To assign these new combinations, it is necessary to look for compatible routes with the new set of scanned links, for example, within set  $\mathfrak{R}$ . Taking the simplified sets in  $\mathcal{Q}$  plus the new ones gives us a new, larger set  $\mathcal{C}$  of routes (see Figure 2(c)). In other words, it is necessary to complete the set of routes once the field data were collected. Doing this, the uncertain knowledge of routes will be reduced. This paper proposes to include this procedure in the location model to improve the expected traffic flow estimation results.

With this last set of routes, it is already possible to solve a flow estimation problem using one of the methods proposed in the literature (see Mínguez et al. [27], which used a generalized least square method and Sánchez-Cambronero et al. [13], which used a Bayesian network model).

To compare estimation results using set  $\mathcal{Q}$  and using set  $\mathcal{C}$ , let us define the link relative absolute error (LRAE) as

$$\text{LRAE}(a) = \frac{|v_a - v_a^{\text{real}}|}{v_a^{\text{real}}}, \quad (3)$$

where  $v_a$  and  $v_a^{\text{real}}$  are the estimated flow and (assumed) real flow for link  $a$ . Such measure of the solution quality should be calculated over the link flows because the set of links  $\mathcal{A}$  will remain constant regardless of the studied simplification or the number of links in set  $\mathcal{S}\mathcal{L}$  (note that the number of routes in  $\mathcal{C}$  may vary depending on the field data collected). Therefore, once the estimation of flows has been made, it is possible to calculate the LRAE using equation (3). Table 3 compares the LARE after estimating the link flows using the set of routes in  $\mathfrak{R}$  versus using the set of routes in  $\mathcal{C}$ , proving the value of performing this calculation.

To check if the estimation of the link flows in the whole network is adequate, we can use the root mean absolute value relative error (RMARE), defined as

$$\text{RMARE} = \frac{1}{n} \sum_a \frac{|v_a - v_a^{\text{real}}|}{v_a^{\text{real}}}, \quad (4)$$

where  $n$  is the number of links in the network. The value obtained using equation (4) for the example given in this section is shown in the last row of Table 3. It indicates that although both sets of routes yield good flow estimations,

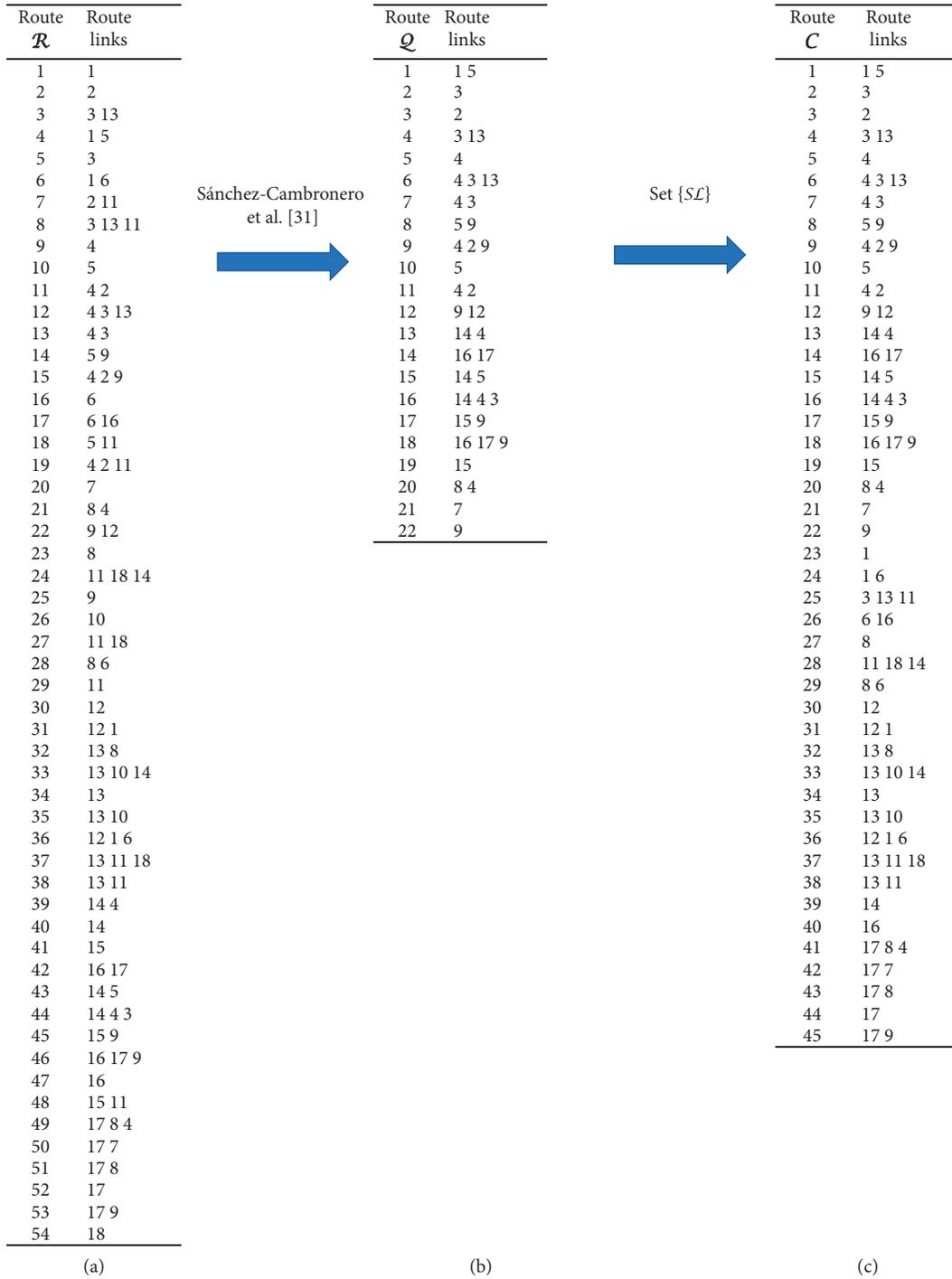


FIGURE 2: Operation scheme and sets of (a) reference routes  $\{\mathcal{R}\}$ , (b) simplified routes  $\{\mathcal{Q}\}$ , and (c) combination routes  $\{\mathcal{C}\}$ .

when the set of routes  $\mathcal{C}$  is used, the traffic estimation model performs better.

Note that the solution for this problem is not unique in terms of links included in  $\mathcal{SL}$  but constitutes a particular solution obtained through an optimization problem (see, for example, Castillo et al. [36]). Taking advantage of this fact, this paper proposes a heuristic algorithm to find set  $\mathcal{SL}$  that minimizes the RMARE obtained using equation (4) and that provides the best set of routes which are able to represent the traffic flow in the entire network.

### 3. Discussion and Improvement of the Sánchez-Cambronero et al. [31] Model

Let us now change the network used to the well-known Nguyen-Dupuis network shown in Figure 3 which will be used. It consists of 13 nodes and 38 links. Let us suppose that, to have a reference level for demand, we have used the data from a study where traditional methods were used. Figure 3 shows the network divided into 4 zones together with its associated origin-destination (OD) matrix and the traffic

TABLE 1: Set of expected combinations of scanned links ( $\mathcal{E}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ ) associated with simplified routes in set  $\mathcal{Q}$  and their values of observed flow.

| Routes, $\mathcal{Q}$ | Route links | Set $\mathcal{E}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ | Info obtained from set $\mathcal{S}\mathcal{L}$ |   |   |   |   |   |   |   |    |    |    |    |    |    | Observed flow |        |
|-----------------------|-------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|---------------|--------|
|                       |             |   | 1   | 2 | 3 | 4 | 5 | 7 | 8 | 9 | 12 | 13 | 14 | 15 | 16 | 17 |               |        |
| 1                     | 1 5         | 1   | X   |   |   |   |   | X |   |   |    |    |    |    |    |    |               | 60.31  |
| 2                     | 3           | 2   |   |   | X |   |   |   |   |   |    |    |    |    |    |    |               | 6.79   |
| 3                     | 2           | 3   |   | X |   |   |   |   |   |   |    |    |    |    |    |    |               | 83.80  |
| 4                     | 3 13        | 4   |   |   | X |   |   |   |   |   |    | X  |    |    |    |    |               | 31.85  |
| 5                     | 4           | 5   |   |   |   |   | X |   |   |   |    |    |    |    |    |    |               | 8.75   |
| 6                     | 4 3 13      | 6   |   |   | X | X |   |   |   |   |    | X  |    |    |    |    |               | 86.42  |
| 7                     | 4 3         | 7   |   |   | X | X |   |   |   |   |    |    |    |    |    |    |               | 1.04   |
| 8                     | 5 9         | 8   |   |   |   |   |   | X |   |   | X  |    |    |    |    |    |               | 1.00   |
| 9                     | 4 2 9       | 9   |   | X |   |   | X |   |   |   | X  |    |    |    |    |    |               | 1.09   |
| 10                    | 5           | 10  |   |   |   |   |   | X |   |   |    |    |    |    |    |    |               | 138.32 |
| 11                    | 4 2         | 11  |   | X |   |   | X |   |   |   |    |    |    |    |    |    |               | 49.39  |
| 12                    | 9 12        | 12  |   |   |   |   |   |   |   |   | X  | X  |    |    |    |    |               | 24.73  |
| 13                    | 14 4        | 13  |   |   |   |   | X |   |   |   |    |    |    | X  |    |    |               | 8.70   |
| 14                    | 16 17       | 14  |   |   |   |   |   |   |   |   |    |    |    |    |    | X  | X             | 45.68  |
| 15                    | 14 5        | 15  |   |   |   |   |   | X |   |   |    |    |    | X  |    |    |               | 44.35  |
| 16                    | 14 4 3      | 16  |   |   | X | X |   |   |   |   |    |    |    | X  |    |    |               | 3.15   |
| 17                    | 15 9        | 17  |   |   |   |   |   |   |   |   | X  |    |    |    | X  |    |               | 1.00   |
| 18                    | 16 17 9     | 18  |   |   |   |   |   |   |   |   | X  |    |    |    |    | X  | X             | 5.62   |
| 19                    | 15          | 19  |   |   |   |   |   |   |   |   |    |    |    |    | X  |    |               | 63.21  |
| 20                    | 8 4         | 20  |   |   |   |   | X |   |   | X |    |    |    |    |    |    |               | 49.65  |
| 21                    | 7           | 21  |   |   |   |   |   |   | X |   |    |    |    |    |    |    |               | 88.85  |
| 22                    | 9           | 22  |   |   |   |   |   |   |   |   | X  |    |    |    |    |    |               | 117.87 |

TABLE 2: New sets of combinations of scanned links ( $\mathcal{N}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ ) beside their values of observed flow.

| Route, $\mathcal{Q}$ | Route links | Set $\mathcal{N}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ | Info obtained from set $\mathcal{S}\mathcal{L}$ |   |   |   |   |   |   |   |    |    |    |    |    |    | Observed flow |       |
|----------------------|-------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|---------------|-------|
|                      |             |   | 1   | 2 | 3 | 4 | 5 | 7 | 8 | 9 | 12 | 13 | 14 | 15 | 16 | 17 |               |       |
|                      |             | 23  | X   |   |   |   |   |   |   |   |    |    |    |    |    |    |               | 3.82  |
|                      |             | 24  |   |   |   |   |   |   |   |   |    |    |    |    |    | X  |               | 18.86 |
|                      |             | 25  |   |   |   |   |   |   | X |   |    |    |    |    |    |    |               | 2.43  |
|                      |             | 26  |   |   |   |   |   |   |   |   |    |    | X  |    |    |    |               | 4.56  |
|                      |             | 27  |   |   |   |   |   |   |   |   | X  |    |    |    |    |    |               | 1.34  |
|                      |             | 28  | X   |   |   |   |   |   |   |   | X  |    |    |    |    |    |               | 2.00  |
|                      |             | 29  |   |   |   |   |   |   | X |   |    | X  |    |    |    |    |               | 1.00  |
|                      |             | 30  |   |   |   |   |   |   |   |   | X  | X  |    |    |    |    |               | 1.33  |
|                      |             | 31  |   |   |   |   |   |   |   |   | X  |    |    |    |    |    |               | 6.91  |
|                      |             | 32  |   |   |   | X |   |   | X |   |    |    |    |    |    |    | X             | 1.22  |
|                      |             | 33  |   |   |   |   |   | X |   |   |    |    |    |    |    |    | X             | 1.24  |
|                      |             | 34  |   |   |   |   |   |   | X |   |    |    |    |    |    |    | X             | 1.55  |
|                      |             | 35  |   |   |   |   |   |   |   |   |    |    |    |    |    |    | X             | 2.36  |
|                      |             | 36  |   |   |   |   |   |   |   |   | X  |    |    |    |    |    | X             | 4.62  |

link parameters (BPR function parameters and attraction and generation capacities) that will be used and explained in the following.

As discussed in Section 1, the use of a centroid as an aggregation of all origins and destinations within a determined zone implies a flow estimation error. The same occurs with the use of connectors (see Quian and Zhang [23]). Therefore, the number and location of connectors and centroids should be determined carefully due to the errors they may cause during traffic flow estimation.

Trying to solve some of these problems, Sánchez-Cambronero et al. [31] proposed a model that allows to design a traffic network that minimizes the negative effects of the use of centroids and connectors by replacing them with

“origin nodes” and “destination nodes” in such a way that all trip origins and destinations are assigned to these nodes of the network in accordance with the vehicle paths and the network shape. An application of this can be observed in Figure 4. Suppose a vehicle actually performs the trip indicated in Figure 4(a), whose true origin is somewhere in link 34 and whose true destination is somewhere in link 5. This method assumes that every vehicle has its origin in the first node of its trip. In this example, node 1 is the first node, so it is the origin node (see Figure 4(b)). Similarly, the destination node is taken to be the last node, where the vehicle passes; in this example, the last node of the route is node 3, so it would be the destination node. This resulting route is finally included in the traffic model. Note that this

TABLE 3: Link flow in estimated and real situations and relative absolute error value.

| Link | Results using set $\mathcal{Q}$ |                     |      | Results using set $\mathcal{E}$ |                     |      |
|------|---------------------------------|---------------------|------|---------------------------------|---------------------|------|
|      | $v_a^{\text{est}}$              | $v_a^{\text{real}}$ | LRAE | $v_a^{\text{est}}$              | $v_a^{\text{real}}$ | LRAE |
| 1    | 60.31                           | 66.13               | 0.09 | 66.13                           | 66.13               | 0.00 |
| 2    | 134.29                          | 134.28              | 0.00 | 134.29                          | 134.28              | 0.00 |
| 3    | 129.26                          | 129.26              | 0.00 | 129.26                          | 129.26              | 0.00 |
| 4    | 208.19                          | 209.41              | 0.01 | 209.41                          | 209.41              | 0.00 |
| 5    | 243.97                          | 243.97              | 0.00 | 243.97                          | 243.97              | 0.00 |
| 6    | 0.00                            | 23.62               | 1.00 | 23.62                           | 23.62               | 0.00 |
| 7    | 88.85                           | 90.09               | 0.01 | 90.09                           | 90.09               | 0.00 |
| 8    | 49.65                           | 55.84               | 0.11 | 55.84                           | 55.84               | 0.00 |
| 9    | 151.32                          | 151.32              | 0.00 | 151.32                          | 151.32              | 0.00 |
| 10   | 0.00                            | 3.80                | 1.00 | 1.99                            | 3.80                | 0.48 |
| 11   | 0.00                            | 39.38               | 1.00 | 2.50                            | 39.38               | 0.94 |
| 12   | 24.73                           | 28.08               | 0.12 | 28.07                           | 28.08               | 0.00 |
| 13   | 118.28                          | 127.52              | 0.07 | 127.52                          | 127.52              | 0.00 |
| 14   | 56.20                           | 62.09               | 0.09 | 62.09                           | 62.09               | 0.00 |
| 15   | 64.21                           | 64.21               | 0.00 | 64.21                           | 64.21               | 0.00 |
| 16   | 51.30                           | 63.18               | 0.19 | 69.71                           | 63.18               | 0.10 |
| 17   | 51.30                           | 55.32               | 0.07 | 55.32                           | 55.32               | 0.00 |
| 18   | 0.00                            | 4.68                | 1.00 | 0.77                            | 4.68                | 0.83 |
|      | RMARE = 0.27                    |                     |      | RMARE = 0.13                    |                     |      |

trip is one of the 470 trips that go from zone 1 to zone 3 (see Figures 3 and 4(b)), and depending on the choice of connectors, the path of this trip may be wrongly defined if the traditional method based on centroids and connectors is used.

### 3.1. Characterization of Origin-Destination Nodes.

According to the assumption described above, every node of the network will generate or attract trips depending on the characteristics of the adjacent links, i.e., depending on the capability of the adjacent links to attract and generate trips (i.e., number of on-street parking spaces, number of private parking spaces, etc.). This can be quantified in terms of the capacity of link  $a$  to attract trips,  $CA_{la}$ , and to generate trips,  $CG_{la}$  (similar assumption was made in Levy and Benenson [41]). Then, the capacities of each node are calculated as follows:

$$CA_{ni} = \sum_{a \in \mathcal{A}_{out}^i} CA_{la}, \quad (5)$$

$$CG_{ni} = \sum_{a \in \mathcal{A}_{in}^i} CG_{la}, \quad (6)$$

where equation (5) expresses the capacity of node  $i$  to attract trips, depending on the capacities of the adjacent links leaving the node ( $\mathcal{A}_{out}^i$ ), and equation (6) expresses the capacity of node  $i$  to generate trips according to the capacities of the adjacent links arriving at the node ( $\mathcal{A}_{in}^i$ ). Such capacity values for each link in the network are shown in Figure 3.

Then, according to these capacity values, one can obtain the proportion of the total trips attracted and generated by zone  $Z$  that begins or ends at node  $i$ . To do this, we propose the following expressions:

$$PA_{ni} = \frac{CA_{ni}}{\sum_{j \in Z} CA_{nj}}, \quad (7)$$

$$PG_{ni} = \frac{CG_{ni}}{\sum_{j \in Z} CG_{nj}}, \quad (8)$$

where  $PA_{ni}$  and  $PG_{ni}$  are these proportions which are shown in Figure 5 for the case of the example with the Nguyen-Dupuis network.

With the values obtained using equations (7) and (8), the relationship between the number of trips made between the origin and destination nodes and the number of trips made between the zones to which these nodes belong can be established as follows:

$$T_{ij} = \hat{T}_{ZiZj} PA_{ni} PG_{nj}, \quad (9)$$

where  $T_{ij}$  is the number of trips from node  $i$  to node  $j$  and  $\hat{T}_{ZiZj}$  is the number of trips obtained through an out-of-date OD matrix (see Figure 5).

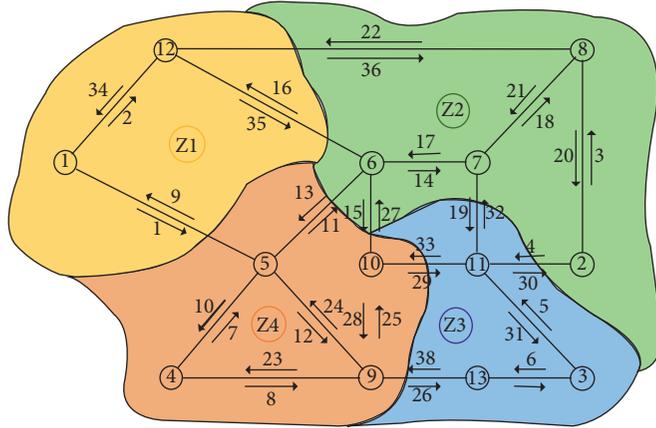
The completion of this step entails the definition of a new OD matrix defined by trips between nodes rather than a matrix defined by trips between zones. This means that, for the Nguyen-Dupuis network used in this example, the matrix with 8 OD flows is transformed in this step into a matrix with 84 OD flows (note that the simplification process proposed in this paper will reduce this number of OD flows. The process will be explained in the following section). This may be seen as a drawback, but note that, by doing this together with the simplification process proposed in Sánchez-Cambronero et al. [31], the problems associated with the use of connectors can be faced better. In addition, the key variables for the plate scanning technique models are the route flows (not the OD flows), whose number should be almost the same for both OD matrices.

### 3.2. The Definition of the Exhaustive Set of Routes $\mathfrak{R}$ .

Network path enumeration is a requirement for developing a model based on plate scanning data because it is needed for both traffic estimation and device location. This implies that, at this stage, we need an exhaustive set of routes  $\mathfrak{R}$ . To construct this set, we propose to find the  $k$ -shortest paths of the extended OD matrix. Although transportation planners usually recommend to use  $k=3$  and discard routes with more than 1.5 times the shortest path (see, for example, Sheffi [17]), in Section 4.1, a sensitivity analysis of various suitable values for  $k$  will be performed.

### 3.3. The Network Simplification and the Set of Routes $\mathcal{Q}$ .

The aim of the simplification process proposed by the authors is to transfer to adjacent nodes the generated or attracted demand of those nodes that generate or attract fewer trips than a given threshold ( $F_{\text{thres}}$ ). This is a good way to avoid problems derived from the use of connectors since the start point of the simplification process is a set of routes built based on the physical characteristics of the real network and not based on the definition of artificial links as are the



| Zone            | 1    | 2    | 3   | 4   | Generated trips |
|-----------------|------|------|-----|-----|-----------------|
| 1               | 0    | 640  | 470 | 0   | 1110            |
| 2               | 640  | 0    | 0   | 530 | 1170            |
| 3               | 470  | 0    | 0   | 110 | 580             |
| 4               | 0    | 530  | 110 | 0   | 640             |
| Attracted trips | 1110 | 1170 | 580 | 640 |                 |

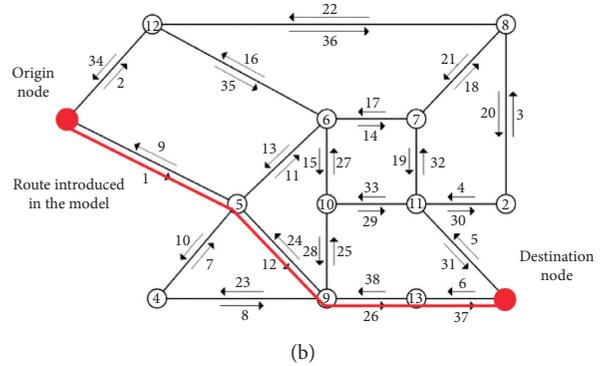
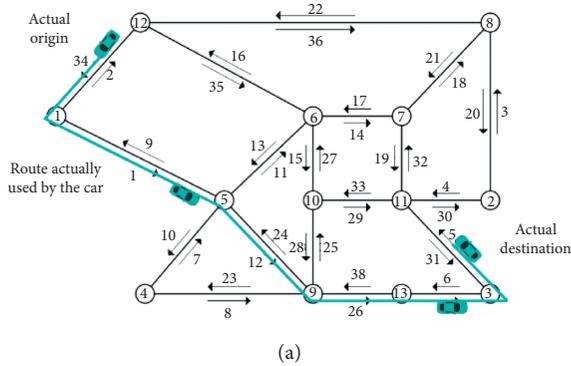
(a)

(b)

| Link | $c_a$ | $q_a$ | $\alpha_a$ | $\gamma_a$ | $CA_{la}$ | $CG_{la}$ | Link | $c_a$ | $q_a$ | $\alpha_a$ | $\gamma_a$ | $CA_{la}$ | $CG_{la}$ |
|------|-------|-------|------------|------------|-----------|-----------|------|-------|-------|------------|------------|-----------|-----------|
| 1    | 7     | 700   | 1          | 4          | 250       | 2         | 20   | 9     | 700   | 1          | 4          | 250       | 300       |
| 2    | 9     | 560   | 1          | 4          | 250       | 100       | 21   | 9     | 700   | 1          | 4          | 200       | 7         |
| 3    | 9     | 700   | 1          | 4          | 300       | 250       | 22   | 14    | 560   | 1          | 4          | 250       | 10        |
| 4    | 9     | 280   | 1          | 4          | 250       | 2         | 23   | 5     | 375   | 1          | 4          | 2         | 250       |
| 5    | 8     | 560   | 1          | 4          | 250       | 2         | 24   | 9     | 420   | 1          | 4          | 5         | 3         |
| 6    | 11    | 560   | 1          | 4          | 200       | 5         | 25   | 5     | 280   | 1          | 4          | 3         | 5         |
| 7    | 12    | 560   | 1          | 4          | 300       | 1         | 26   | 9     | 280   | 1          | 4          | 5         | 2         |
| 8    | 5     | 375   | 1          | 4          | 250       | 5         | 27   | 4     | 280   | 1          | 4          | 2         | 2         |
| 9    | 7     | 700   | 1          | 4          | 5         | 250       | 28   | 5     | 280   | 1          | 4          | 3         | 2         |
| 10   | 12    | 560   | 1          | 4          | 2         | 200       | 29   | 4     | 700   | 1          | 4          | 3         | 3         |
| 11   | 12    | 420   | 1          | 4          | 6         | 5         | 30   | 9     | 280   | 1          | 4          | 5         | 250       |
| 12   | 9     | 420   | 1          | 4          | 2         | 2         | 31   | 8     | 560   | 1          | 4          | 2         | 250       |
| 13   | 12    | 420   | 1          | 4          | 5         | 4         | 32   | 4     | 700   | 1          | 4          | 2         | 7         |
| 14   | 5     | 700   | 1          | 4          | 10        | 1         | 33   | 4     | 700   | 1          | 4          | 2         | 4         |
| 15   | 4     | 280   | 1          | 4          | 2         | 2         | 34   | 9     | 560   | 1          | 4          | 10        | 250       |
| 16   | 7     | 140   | 1          | 4          | 5         | 50        | 35   | 7     | 140   | 1          | 4          | 12        | 5         |
| 17   | 5     | 700   | 1          | 4          | 5         | 5         | 36   | 14    | 560   | 1          | 4          | 15        | 200       |
| 18   | 9     | 700   | 1          | 4          | 10        | 150       | 37   | 11    | 560   | 1          | 4          | 5         | 200       |
| 19   | 4     | 700   | 1          | 4          | 2         | 5         | 38   | 9     | 280   | 1          | 4          | 2         | 5         |

(c)

FIGURE 3: Nguyen-Dupuis network and its division into 4 traffic zones (Z) and the associated OD matrix.



(a)

(b)

FIGURE 4: Origin and destination nodes' definition: (a) route actually used by a certain vehicle; (b) route to be included in the traffic model.

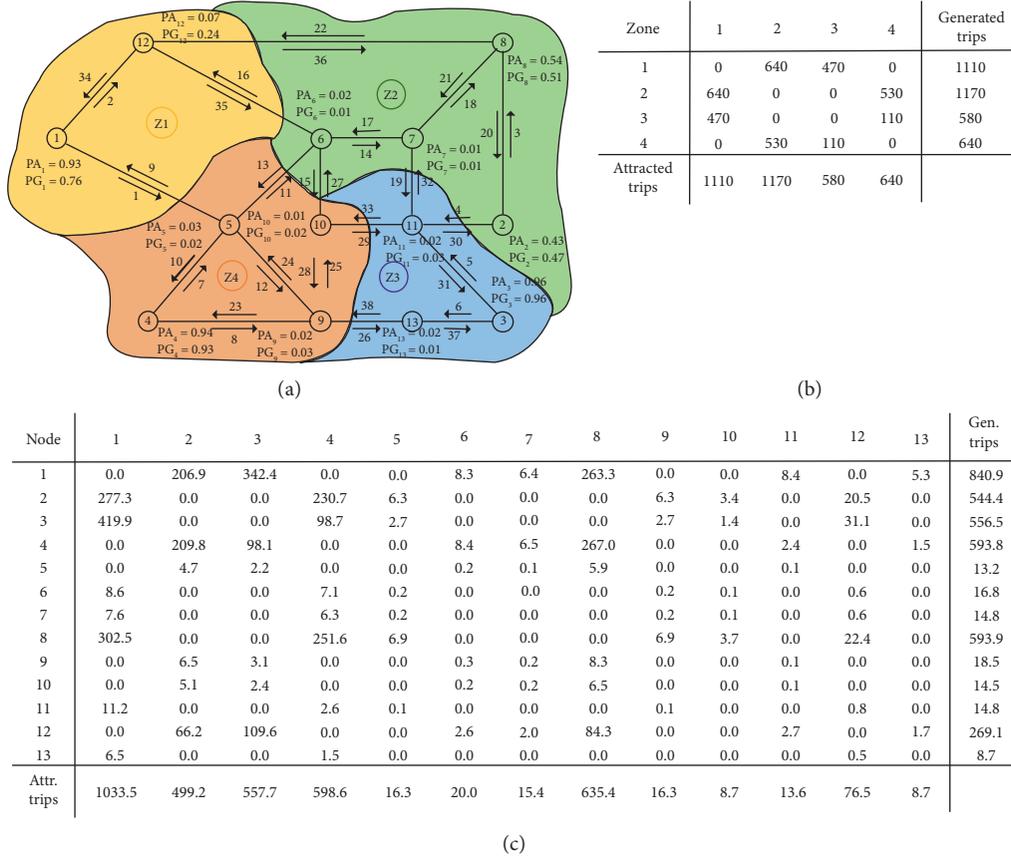


FIGURE 5:  $PA_{ni}$  and  $PG_{ni}$  proportions and the extended OD matrix for the example with the Nguyen-Dupuis network.

connectors. The next section deals with a detailed description of the process to obtain the simplified set of routes  $\mathcal{Q}$ .

#### 4. The Proposed Algorithm

In this section, we present the proposed two-step algorithm that allows us to determine the traffic flow network to be used and the location of AVI devices to perform the traffic flow estimation.

Inputs:

- (i) Network model: sets  $\mathcal{A}$  of links and  $\mathcal{N}$  of nodes and links parameters  $C_a$  and  $Q_a$
- (ii) Cost and flow thresholds for the simplification process ( $C_{thres}$ ,  $F_{thres}$ )
- (iii) Capacities of links to attract and generate trips ( $CA_{la}$ ,  $CG_{la}$ )
- (iv) Number of trips  $\hat{T}_{ZiZj}$  based on network zones obtained through an out-of-date OD matrix
- (v)  $RMARE^{best}$  set to an initial value of 10 and  $iter^{max}$  set to a maximum of 1000 iterations to carry out

Step 0: obtaining reference routes:

- (i) Obtaining a new extended matrix  $T_{ij}$  of the number of trips from node  $i$  to node  $j$  using the procedure described in Section 2.1.1.

- (ii) Find the  $k$ -shortest paths using the extended matrix to obtain the set of reference routes ( $\mathfrak{R}$ ) with their respective reference route flows  $f_r^0$  that can be obtained by using, for example, MNL stochastic user equilibrium (see, for example, Sheffi [17] and Sánchez-Cambronero et al. [42]).

Step 1: network simplification:

The simplified network is obtained using the green part of the algorithm shown in Figure 6. It allows us to decide what nodes can be origin/destination or not, based on a demand threshold flow  $F_{thres}$  established by the transportation planner. The simplification process finishes when it does not exist any node which complies simplification criterion.

Step 1.0: initialization:

Initialize the set of simplified routes  $\mathcal{Q}$  as the exhaustive set  $\mathfrak{R}$ .

Step 1.1: search the least demand node:

The algorithm searches the node with the lower demand for both cases: origins and destination, i.e., least  $G_i = \sum_j T_{ij}$  or  $A_j = \sum_i T_{ij}$ .

Step 1.2: checking the simplification criterion:

Once the candidate node has been identified to lose its origin/destination condition, it is checked if it meets the simplification criteria established, i.e., if  $G_i \leq F_{\text{thres}}$  or  $A_j \leq F_{\text{thres}}$ , go to Step 1.3. Else, continue with Step 1.5.

Step 1.3: demand transmission:

If the candidate node is able to lose its OD condition, i.e., its OD demand is below the threshold flow  $F_{\text{thres}}$ , then the possibility that this demand has its origin or destination in some other node of the network is evaluated. Transmission will be made to the closer node of each route if and only if the node that could receive or emit demand is at a shorter distance than  $C_{\text{thres}}$ , and the involved routes will have to be modified accordingly. Otherwise, the candidate node and the demand will disappear, and the set of routes will be remade (see Sánchez-Cambronero et al. [31] for more details).

Step 1.4: set of routes' update:

Update the set of routes  $\mathcal{Q}$  and their corresponding flow values  $f_q^0$ . Go to Step 1.1.

Step 1.5: output of Step 1 and inputs to Step 2:

As a result of the application of this simplification algorithm, we will obtain a new set of routes  $\mathcal{Q}$  from the original set  $\mathcal{R}$  due to removal and re-enumeration of reference routes. Simultaneously, the route flows  $f_q^0$  for this new set will be updated. Proceed to Step 2.

Step 2: location and estimation problem:

The AVI device location problem is a complex problem that does not have a unique solution. In this step, we assume simulated "real flows" in order to obtain the values for the observed flow  $\hat{w}_s$ , depending on the device location. The main objective of this step of the algorithm is to obtain the subset of the set of scanned links  $\mathcal{S}\mathcal{L}$  that gives the best possible flow estimation.

To achieve this objective, an optimization problem is incorporated into the algorithm. This problem is based on previous problems studied by Mínguez et al. [27] or Cerrone et al. [40], but as an improvement, we have included a new restriction that examines different options for the device location in order to assess which of them lead to the best solution in terms of flow estimation.

Step 2.1: scan device location problem:

The scan device location problem is formulated as

$$\max_{z_a, \gamma_q} M = \sum_{q \in \mathcal{Q}} f_q^0 \gamma_q, \quad (10)$$

subject to

$$\sum_{a \in \mathcal{A}} P_a z_a \leq B, \quad \forall a \in \mathcal{A}, \quad (11)$$

$$\sum_{a \in \mathcal{A}} \delta_a^q z_a \geq \gamma_q, \quad \forall q \in \mathcal{Q}, \quad (12)$$

$$\begin{aligned} \sum_{a \in \mathcal{A}} \gamma_a^{qq_1} z_a + \sum_{a \neq b \in \mathcal{A}} \sigma_{ab}^{qq_1} x_{ab} &\geq \gamma_q, \quad \forall (q, q_1) \in \mathcal{Q}^2 | q > q_1; \\ \sum_{a \in \mathcal{A}} \delta_q^a \delta_{q_1}^a &> 0, \end{aligned} \quad (13)$$

$$2x_{ab} - z_a - z_b \leq 0, \quad \forall a \neq b \in \mathcal{A}, \quad (14)$$

$$\sum_{a \in \mathcal{A}} S_a^{\text{iter}} z_a \leq \sum_{a \in \mathcal{A}} S_a^{\text{iter}}, \quad \forall \text{iter} \in \mathbf{I}. \quad (15)$$

Objective function (10) maximizes the observed route flow in terms of  $f_q^0$ , which is the reference flow through route  $q$ ;  $\gamma_q$ , theoretically, is a binary variable which equals to 1 if a route can be distinguished from others and 0, otherwise; however, to speed up the model, it is set as a continuous variable  $\gamma_q \in [0, 1]$  (Mínguez et al. [27]).

Constraint (11) satisfies the budget requirement, where  $z_a$  is a binary variable that equals 1 if link  $a$  is scanned and 0, otherwise. This constraint guarantees that we will have a number of scanned links with cost  $P_a$  for link  $a$  that does not exceed the established limited budget  $B$ .

Constraint (12) ensures that any distinguished route contains at least one scanned link (for this reason, they are usually known as covering constraints). This constraint is indicated by the parameter  $\delta_a^q$ , which is the element of the link-route incidence matrix.

Constraints (13) are the diversification constraints. They indicate that route  $q$  must be distinguished from the other routes in at least one scanned link  $a$ . This happens if

- (i)  $\sum_a \gamma_a^{qq_1} z_a \geq 1$  because  $\gamma_a^{qq_1}$  is 1 if link  $a$  is contained either in  $q$  or in  $q_1$  (not in both) and 0, otherwise.
- (ii)  $\sum_a \sigma_{ab}^{qq_1} x_{ab} \geq 1$  because vehicles using  $q$  and  $q_1$  use the same links but in different order since  $\sigma_{ab}^{qq_1}$  is 1 if links  $a$  and  $b$  are both in routes  $q$  and  $q_1$ , but they appear in a different order.

Note that if  $\gamma_q = 0$ , constraints (13) always hold. The definition of these constraints taking into account that  $q > q_1$  and  $\sum_{a \in \mathcal{A}} \delta_q^a \delta_{q_1}^a > 0$  avoids the comparison of a great amount of routes without common links, resulting in an important reduction of the computational time. This is important for the analysis of real-size networks and is usually forgotten in many papers, for example, in Castillo et al. [9, 43] and Cerrone et al. [40], among others.

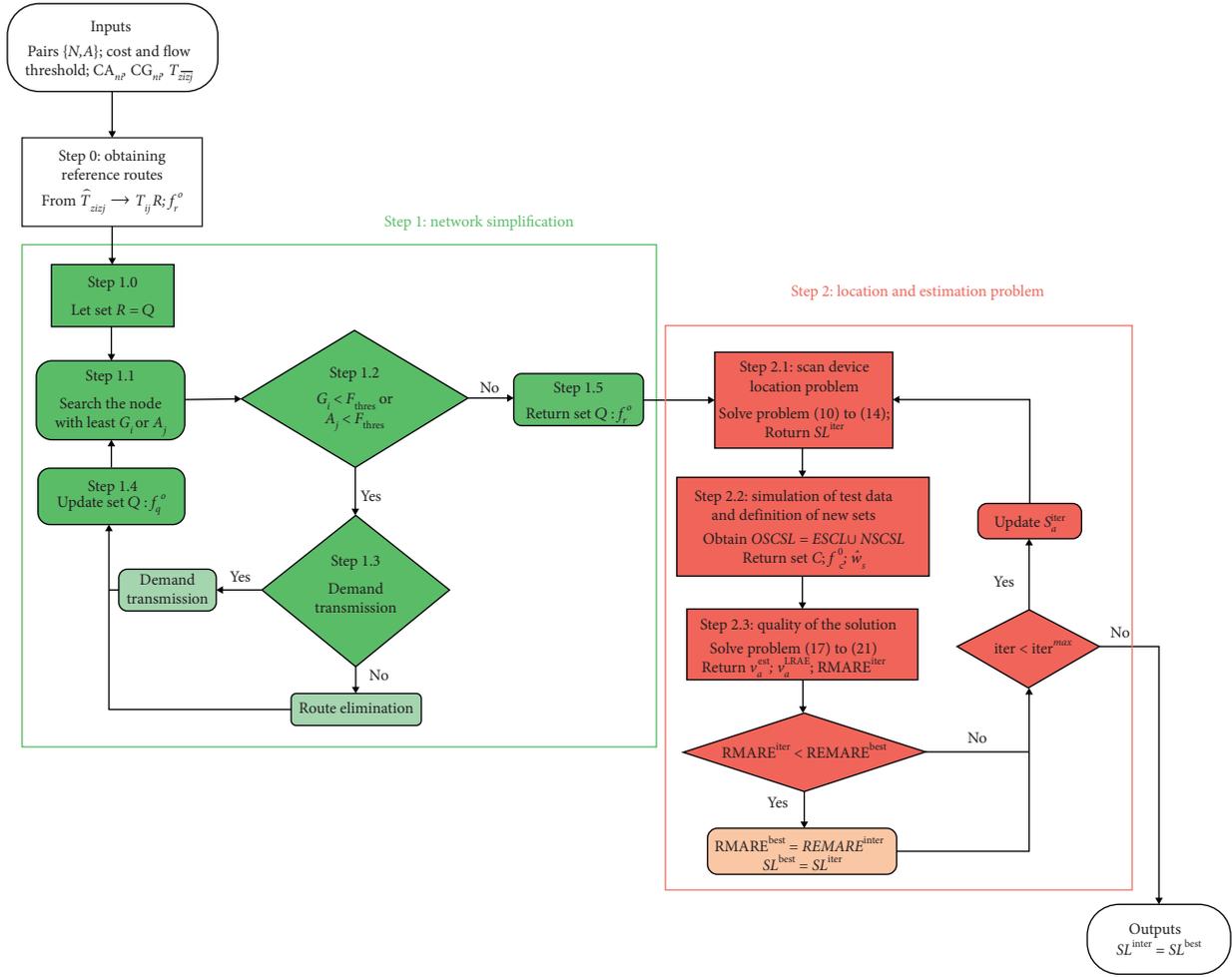


FIGURE 6: Flowchart of the algorithm.

Constraint (14) is the logical constraint linking the binary variable  $x_{ab}$  to  $z_a$  and  $z_b$  (see Cerrone et al. [40]).

Finally, since this model is a part of an iteration process, we propose additional constraints (15).  $S_a^{iter}$  is a matrix that grows with the number of iterations, in which each row reflects the set  $\mathcal{SL}$  resulting from each iteration carried out up to then by the model in such a way that  $S_a^{iter}$  is 1 if link  $a$  was proposed to be scanned in the solution provided on iteration  $iter$  and 0, otherwise. This ensures that each iteration keeps the previous solutions and does not permit the process to repeat a solution in future iterations. That is, each iteration carried out by the algorithm is forced to search for a different solution with the same value of objective function (10).

Step 2.2: simulation of test (“real”) data and definition of the new set of routes  $\mathcal{E}$ :

To test the quality of the solution, we have simulated “real” route data to carry out the process

described in Section 1.3, which through the new set of combinations of scanned links  $\mathcal{NSCSL}$ , observed by subset  $\mathcal{SL}$  (obtained in Step 2.1), searches compatible combinations for these new observed routes and forms the observed set of combinations of scanned links ( $\mathcal{ESCSL} \cup \mathcal{NSCSL} = \mathcal{OSCSL}$ ) and finally obtains set  $\mathcal{E}$  of all available routes.

The real data were obtained multiplying each  $T_{ij}$  of the extended matrix in Step 0 by a random number  $U(0.8-1.2)$  and executing a MNL SUE assignment to a  $k$ -shortest path enumeration set by using  $k=7$ . From this set of “real” routes, we obtain  $\hat{w}_s$ , where  $s \in \mathcal{OSCSL}$ .

Step 2.3: measuring the quality and updating the solution:

Once we have the  $\hat{w}_s$  and  $\hat{v}_a$  “observed flow” (note that plate scanning devices also allow us to observe link flows on links of  $\mathcal{SL}$ ), a traffic flow estimation can be carried out. As we have shown in Section 1.3,

this can be done by several mathematical methods. In this paper, we have used a generalized least square method as follows:

$$\min_{V, Fz} = \sum_{c \in \mathcal{C}} U_c^{-1} \left( \frac{f_c - f_c^0}{f_c^0} \right)^2 + \sum_{a \in \mathcal{S}\mathcal{L}} Y_a^{-1} \left( \frac{v_a - \hat{v}_a}{\hat{v}_a} \right)^2, \quad (16)$$

subject to

$$\hat{w}_s = \sum_{c \in \mathcal{C}} \beta_s^c f_c, \quad \forall s \in \mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}, \quad (17)$$

$$v_a = \sum_{c \in \mathcal{C}} \delta_a^c f_c, \quad \forall a \in \mathcal{A}, \quad (18)$$

$$f_c \geq 0, \quad \forall c \in \mathcal{C}, \quad (19)$$

$$v_a \geq 0, \quad \forall a \in \mathcal{A}, \quad (20)$$

where  $U_c^{-1}$  and  $Y_a^{-1}$  are the inverses of the variance-covariance matrices corresponding to the flow in routes in  $\mathcal{C}$  and the observed flow in  $a$ , respectively,  $\hat{w}_s$  is the observed flow in each set  $\mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$ ,  $f_c$  is the estimated flow of routes in set  $\mathcal{C}$ , and  $\beta_s^c$  is the element of the route-scanned combination incidence matrix corresponding to route  $c$ , which is 1 if route  $c$  contains subset  $s$  of scanned links and only those, and 0 otherwise.

With this traffic estimation, one can compare the quality of the results obtained for the link flows using equation (3). As shown in Section 1.3, the quantification of the error is made using the results of the link flows since the set of these remains constant in all the iterations of the proposed algorithm. Additionally, the use of equation (4) allows us to check the global quality of the modeled network. For each iteration carried out by the algorithm, the value of  $\text{RMARE}^{\text{iter}}$  is evaluated with respect to the best value  $\text{RMARE}^{\text{best}}$ . If  $\text{RMARE}^{\text{iter}} < \text{RMARE}^{\text{best}}$ , then the algorithm updates  $\text{RMARE}^{\text{iter}} = \text{RMARE}^{\text{best}}$  and considers the set of scanned links  $\mathcal{S}\mathcal{L}^{\text{iter}}$  to be the best solution up to that iteration. If  $\text{iter} < \text{iter}^{\text{max}}$ , then the algorithm updates  $\mathcal{S}_a^{\text{iter}}$  and goes to Step 2.1; otherwise, return the solution  $\mathcal{S}\mathcal{L}^{\text{iter}}$  as the best solution.

The complete process of this step can be observed on the right side of Figure 6.

## 5. A Sensitivity Analysis of the Model Results

**5.1. The Nguyen-Dupuis Network.** In this section, a sensitivity analysis of the model results depending on the value of some parameters is presented. In particular, we have used the Nguyen-Dupuis network shown in Figure 4, and we have analyzed the influence of (i) the partial knowledge of the routes in terms of the  $k$  value for obtaining the reference set of routes, (ii) the degree of network simplification in terms of  $F_{\text{thres}}$ , and (iii) the available budget  $B$  for locating AVI devices.

For comparison purposes, a set of initial values was considered, and then the algorithm was applied. As a base situation, it was assumed that  $P_a = 1, B = 16$  (hence, the number of installed cameras is 16), and the threshold flow  $F_{\text{thres}}$  required for the simplification network method (Step 1 in Figure 6) is 50. In all studied cases, the parameters on different links in the network, shown in Figure 3, remain constant throughout the iterations carried out by the model.

### 5.1.1. Influence of the Partial Knowledge of the Routes.

To perform this analysis, a route enumeration algorithm was used to check the effect of considering  $k$  routes on each OD pair for the reference set  $\mathfrak{R}$ . The algorithm used here is based on Yen's  $k$ -shortest path algorithm (Yen [44]). This algorithm introduces into the model an initial reference route matrix  $\mathfrak{R}$  that varies in size according to the value of  $k$ . Then, in Step 1, shown in Figure 6, the reference route matrix is simplified to  $\mathcal{Q}$ , taking into account  $F_{\text{thres}} = 40$ , i.e., those nodes which attract or generate flow which is below 40 lose their OD condition, its demand is transferred to other adjacent nodes, and the corresponding routes are grouped and hence simplified. This reduces the set of routes  $Q$  used in the model, (10)–(15), as shown in the third column of Table 4.

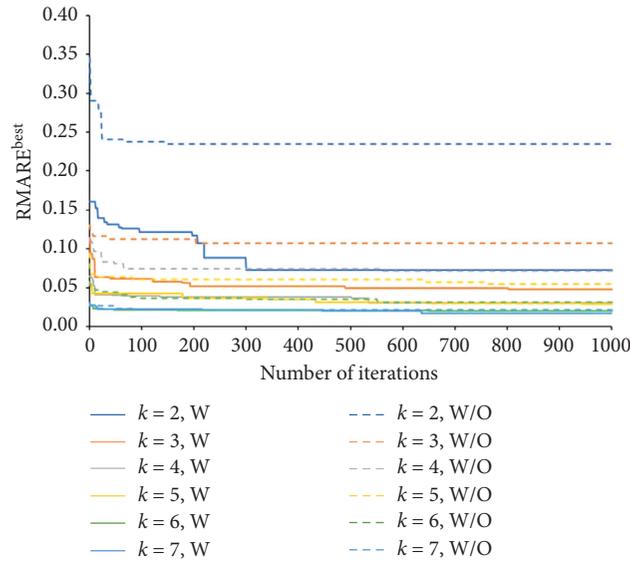
It is well known that the enumeration of routes is essential to solve this location problem. In this analysis, the  $k$ -shortest path is used to generate routes between each pair of nodes, but it may exist other routes (actually used by the vehicles) which they do not have been taken into account in the model (this is the reason why the “real situation” has been simulated using  $k=7$ ) (see Section 1). In this sensitivity analysis, a range of values of  $k$ , which is usually used in this kind of transportation analysis, has been selected, with values of  $k=3$  or 4 (Owais et al. [45]). Although values such as  $k=5, 6$  or 7 are rarely used to enumerate routes (Bonsall et al., [46]; Hazelton, [47]), because it is not normal that it exists such amount of routes for each O-D pair and because it increases the computation time and complexity of the problem, we have checked its influence in the simple example network. According to some authors, a value of  $k$  less than 2-3 would be unreasonable when working with this type of model (Sheffi [17]).

Figure 7 shows the evolution of  $\text{RMARE}^{\text{best}}$  for all iterations performed, considering different  $k$  routes by each OD pair to construct the set of routes  $\mathfrak{R}$ . The continuous line represents the results obtained with the proposed algorithm. Each step in the lines indicates that a new solution has been found that is better than the previous best solution, i.e., with lower RMARE. As expected, better global solutions are found for higher number of  $k$  (higher number of reference routes in  $\mathfrak{R}$ ).

Note also that the first solution in all cases provides the worst results, and it is improved with the iteration process, corresponding to a step forward compared with the previous models. Note that models of Mínguez et al. [27] or Cerrone et al. [40] used a fixed set of routes (let us say set  $\mathcal{Q}$ ), so their estimation would be worse than the one provided in the first iteration since they do not work with the improved set  $\mathcal{C}$ . In any case and in order to explicitly probe this, we have applied our proposed Step 2 but with the set of routes in  $\mathcal{Q}$  assumed

TABLE 4: Number of routes that appear with the application of the proposed algorithm according to the  $k$  value in the best solution in each case study.

| $k$ value | Number of routes in $\mathfrak{R}$ | Number of routes in $\mathcal{Q}$ | Added routes compatible with $\mathcal{SL}$ | Number of routes in $\mathcal{E}$ |
|-----------|------------------------------------|-----------------------------------|---|-----------------------------------|
| 2         | 117                                | 44                                | 46  | 90                                |
| 3         | 150                                | 66                                | 25  | 91                                |
| 4         | 162                                | 72                                | 20  | 92                                |
| 5         | 168                                | 76                                | 20  | 96                                |
| 6         | 172                                | 80                                | 26  | 106                               |
| 7         | 174                                | 82                                | 27  | 109                               |

FIGURE 7: Evolution of the error  $\text{RMARE}^{\text{best}}$  with the number of iterations made, considering different values of  $k$  with (W) or without (W/O) the proposed model.TABLE 5: Number of routes that appear with the application of the proposed algorithm according to the  $F_{\text{thres}}$  value.

| Threshold flow, $F_{\text{thres}}$ | Number of routes in set $\mathfrak{R}$ | Number of routes in set $\mathcal{Q}$ | Added routes with $\mathcal{SL}$ | Number of routes in $\mathcal{E}$ |
|------------------------------------|--|---------------------------------------|----------------------------------|-----------------------------------|
| 0                                  | 162                                    | 162                                   | 9                                | 171                               |
| 10                                 | 162                                    | 152                                   | 9                                | 161                               |
| 20                                 | 162                                    | 114                                   | 13                               | 127                               |
| 30                                 | 162                                    | 80                                    | 18                               | 98                                |
| 40 and 50                          | 162                                    | 72                                    | 20                               | 92                                |
| 60                                 | 162                                    | 63                                    | 20                               | 93                                |

fixed, i.e., assuming that  $\mathcal{E} = \mathcal{Q}$  in any case. The results, presented in dashed lines in Figure 7, show that the proposed model clearly outperforms the estimations given by other methods. This difference is more evident for lower values of  $k$  because difference in the number of routes in  $\mathcal{Q}$  and in  $\mathcal{E}$  is bigger.

In addition, higher improvements in terms of lowering the RMARE occur in the first 100–200 iterations and remain almost constant (with lower steps) till the end of the experiment.

The effect of assigning different values of  $k$  is observed in the second column of Table 4. When the value of  $k$  increases, the number of routes in  $\mathfrak{R}$  also increases. The third column collects the number of routes that appear in  $\mathcal{Q}$  after the reference set has been simplified by means of the simplification algorithm. For the best solution after applying the

algorithm, the fourth column shows the number of added routes compatible with the  $\mathcal{OSCSL}$  set. The fifth column shows the total number of routes finally included in  $\mathcal{E}$ .

Indeed, the last column of Table 4 indicates that we are able to reinclude many routes once  $\mathcal{OSCSL}$  is obtained from  $\mathcal{SL}$ .

In summary, a high value for  $k$  yields a low error due to the existence of more routes, and this implies that more information is available for the estimation problem; however, the difference in terms of the quality of the final solution is not very large. This means that, for large networks, lower values of  $k$  can be chosen to avoid problems related to computational costs since the algorithm would operate with a smaller number of routes in each iteration.

Finally, from the results shown in Figure 7, we confirm that, as other authors have noted, a value of  $k=3$  or 4 is a

TABLE 6: Sets of routes resulting after the application of Step 1 according to  $F_{thres} = 30$ .

| OD pair | Path ID | Links         | $f_q$  |
|---------|---------|---------------|--------|
| 1-2     | 1       | 1 11 14 19 30 | 37.82  |
|         | 2       | 2 35 14 19 30 | 22.68  |
|         | 3       | 1 11 15 29 30 | 39.78  |
|         | 4       | 1 11 14 18 20 | 106.60 |
| 1-3     | 5       | 1 11 14 19 31 | 57.75  |
|         | 6       | 2 35 14 19 31 | 34.63  |
|         | 7       | 1 12 26 37    | 189.25 |
|         | 8       | 1 11 15 29 31 | 60.74  |
| 1-6     | 9       | 1 11          | 14.40  |
|         | 10      | 2 35          | 8.64   |
| 1-8     | 11      | 1 11 14 18    | 12.40  |
|         | 12      | 2 35 14 18    | 7.44   |
|         | 13      | 2 36          | 243.47 |
| 2-1     | 14      | 3 22 34       | 190.42 |
|         | 15      | 3 21 17 13 9  | 53.66  |
|         | 16      | 4 33 27 16 34 | 17.49  |
|         | 17      | 3 21 17 16 34 | 15.71  |
| 2-4     | 18      | 4 33 28 23    | 230.68 |
| 2-6     | 19      | 4 33 27       | 1.95   |
|         | 20      | 3 21 17       | 1.75   |
|         | 21      | 4 32 17       | 1.60   |
| 2-12    | 22      | 3 22          | 16.42  |
|         | 23      | 4 33 27 16    | 1.51   |
|         | 24      | 3 21 17 16    | 1.35   |
|         | 25      | 4 32 17 16    | 1.24   |
| 3-1     | 26      | 5 32 17 13 9  | 131.62 |
|         | 27      | 6 38 24 9     | 202.75 |
|         | 28      | 5 33 27 16 34 | 46.99  |
|         | 29      | 5 32 17 16 34 | 38.53  |
| 3-4     | 30      | 6 38 23       | 70.11  |
|         | 31      | 5 33 28 23    | 28.61  |
| 3-6     | 32      | 5 32 17       | 0.61   |
|         | 33      | 5 33 27       | 0.75   |
| 3-12    | 34      | 5 32 17 16    | 14.22  |
|         | 35      | 5 33 27 16    | 17.33  |
| 4-2     | 36      | 8 25 29 30    | 209.82 |
| 4-3     | 37      | 8 26 37       | 75.31  |
|         | 38      | 8 25 29 31    | 22.83  |
| 4-6     | 39      | 7 11          | 8.32   |
|         | 40      | 8 25 27       | 4.56   |
| 4-8     | 41      | 8 25 29 32 18 | 82.13  |
|         | 42      | 7 11 14 18    | 119.51 |
|         | 43      | 8 25 27 14 18 | 65.40  |
| 6-2     | 44      | 14 18 20      | 2.24   |
|         | 45      | 15 29 30      | 0.84   |
|         | 46      | 14 19 30      | 0.79   |
| 6-3     | 47      | 14 19 31      | 0.34   |
|         | 48      | 15 29 31      | 0.36   |
| 6-8     | 49      | 14 18         | 8.82   |
| 6-1     | 50      | 16 34         | 6.85   |
|         | 51      | 13 9          | 17.40  |
| 6-4     | 52      | 13 10         | 8.71   |
|         | 53      | 15 28 23      | 3.00   |
| 6-12    | 54      | 16            | 2.02   |

TABLE 6: Continued.

| OD pair | Path ID | Links          | $f_q$  |
|---------|---------|----------------|--------|
| 8-1     | 55      | 22 34          | 221.73 |
|         | 56      | 21 17 13 9     | 62.48  |
|         | 57      | 21 17 16 34    | 18.29  |
| 8-4     | 58      | 21 19 33 28 23 | 68.35  |
|         | 59      | 21 17 13 10    | 136.37 |
|         | 60      | 21 17 15 28 23 | 46.93  |
| 8-6     | 61      | 21 17          | 11.14  |
| 8-12    | 62      | 22             | 20.67  |
|         | 63      | 21 17 16       | 1.71   |
| 9-1     | 79      | 24 9           | 9.62   |
| 9-2     | 64      | 25 29 30       | 7.32   |
| 9-3     | 65      | 26 37          | 3.47   |
|         | 66      | 25 29 31       | 1.05   |
| 9-4     | 80      | 23             | 5.88   |
| 9-6     | 67      | 25 27          | 0.35   |
| 9-8     | 68      | 25 29 32 18    | 4.63   |
|         | 69      | 25 27 14 18    | 3.68   |
| 12-2    | 70      | 35 15 29 30    | 0.72   |
|         | 71      | 35 14 19 30    | 0.68   |
|         | 72      | 35 14 18 20    | 1.92   |
|         | 73      | 36 20          | 62.88  |
|         | 74      | 35 15 29 31    | 57.03  |
| 12-3    | 75      | 35 14 19 31    | 54.23  |
| 12-6    | 76      | 35             | 7.38   |
| 12-8    | 77      | 35 14 18       | 2.50   |
|         | 78      | 36             | 81.76  |

reasonable choice since, in terms of error in the estimation of flows, it offers similar results to models with a greater number of routes per OD pair, such as  $k = 6$  or  $7$ .

5.1.2. *Influence of Threshold Flow ( $F_{thres}$ ).* The degree of network simplification implemented in Step 1 of the proposed algorithm depends on the demand and cost threshold values. The definition of these values will determine the total number of routes in the set  $\mathcal{Q}$ , obtained from the simplification of the reference set  $\mathcal{R}$ .

In the case in which the reference route set is not simplified, i.e., with a null threshold flow, the algorithm operates with no simplified network and, therefore, with the same number of routes as that in the reference set. When the threshold flow increases, however, the network will be simplified according to the defined flow value. For the network model studied in this paper, different values of threshold flow have been considered. Following the conclusions drawn in the previous section, this analysis was carried out using  $k = 4$  (i.e., the number of routes in  $\mathcal{R}$  is 162) and  $B = 16$  again.

The effects of the values of  $F_{thres}$  considered in this paper on the size of the route set are reflected in the third column of Table 5. If higher values of  $F_{thres}$  are considered, the number of routes with respect to the reference is reduced; hence, the number of routes in set  $\mathcal{Q}$  is smaller, and once the

TABLE 7: OD matrix resulting after the application of Step 1 according to  $F_{\text{thres}} = 30$ .

| Node        | 1       | 2      | 3      | 4      | 6     | 8      | 12    | Gen. trips |
|-------------|---------|--------|--------|--------|-------|--------|-------|------------|
| 1           | 0.00    | 206.88 | 342.37 | 0.00   | 23.04 | 263.31 | 0.00  | 835.60     |
| 2           | 277.28  | 0.00   | 0.00   | 230.68 | 5.30  | 0.00   | 20.52 | 533.78     |
| 3           | 419.89  | 0.00   | 0.00   | 98.72  | 1.36  | 0.00   | 31.55 | 551.52     |
| 4           | 0.00    | 209.82 | 98.14  | 0.00   | 12.88 | 267.04 | 0.00  | 587.88     |
| 6           | 24.25   | 3.87   | 0.70   | 11.71  | 0.00  | 8.82   | 2.02  | 51.37      |
| 8           | 302.50  | 0.00   | 0.00   | 251.65 | 11.14 | 0.00   | 22.38 | 587.67     |
| 9           | 9.62    | 7.32   | 4.52   | 5.88   | 0.35  | 8.31   | 0.00  | 36.00      |
| 12          | 0.00    | 66.20  | 111.26 | 0.00   | 7.38  | 84.26  | 0.00  | 269.10     |
| Attr. trips | 1033.54 | 494.09 | 556.99 | 598.64 | 61.45 | 631.74 | 76.47 |            |

cameras are installed on the network, a bigger number of new routes in  $\mathcal{C}$  will appear (from 9 added routes with  $F_{\text{thres}} = 0$  to 20 with  $F_{\text{thres}} = 60$ ). Having a lower number of routes in set  $\mathcal{Q}$  has important advantages in terms of computational cost solving location problems (10)–(15). Note also that due to the OD matrix configuration (see Figure 5), the simplification process leads to the same final network for  $F_{\text{thres}} = 40$  than 50. For illustration purposes, Tables 6 and 7 show the simplified set of routes and the simplified OD matrix for  $F_{\text{thres}} = 30$ .

Figure 8 shows  $\text{RMARE}^{\text{best}}$  obtained for different cases shown in Table 5. The smallest quantified error in link flows corresponds to a case without simplification, as expected. The value of the error for the remaining scenarios increases with the value of the threshold flow used in the simplification step, and again, higher improvements in terms of error occur in the first iterations of the algorithm.

It is important to point out that, despite all the facts exposed above, the results of all the cases are similar in terms of error. This is because the simplification algorithm always keeps the routes with higher reference flow which are used for problems (10)–(15) to locate devices. Therefore, this leads to lower estimation errors and hence a better performance.

For the sake of comparison with existing models, Figure 8 also shows the evolution of the  $\text{RMARE}^{\text{best}}$ , keeping the set of routes in the estimation process as constant (i.e., setting  $\mathcal{C} = \mathcal{Q}$  for all the cases). In this case also, the proposed model outperforms the estimates given by the existing models.

**5.1.3. Influence of the Available Budget  $B$ .** In terms of traffic flow estimation, the number of installed devices on a network may be the most important factor to consider, even more if one evaluates the unusual possibility of obtaining full observability of the network.

Several authors have presented two versions of the scan device location problem: the full flow-observability problem and the partial flow-observability problem. In the first version, given a set of scanned links  $\mathcal{S}\mathcal{L}$ , the equation system is fully observable, and the coefficient matrix has full rank, and it is not necessary to estimate the flows of interest. In the second version, given a set of scanned links  $s$ , the equation system is not observable for all flow variables (Gentili and Mirchandani [34]), either due to limitations of the number of devices to be installed and hence not to have a

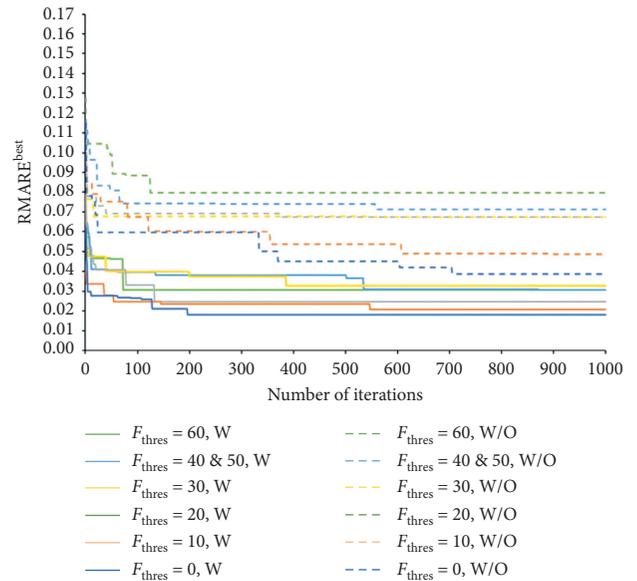


FIGURE 8: Evolution of the error  $\text{RMARE}^{\text{best}}$  with the number of iterations made, considering different values of  $F_{\text{thres}}$  with (W) or without (W/O) the proposed model.

full-rank coefficient matrix. In any case, the effects of partial knowledge of the routes analyzed in this paper have received very limited attention in the literature.

For the analysis, several values of  $B$  have been introduced into the location model so that the influence of the number of devices to be installed can be analyzed in terms of the quantification of the error in the flow estimation. In this case, we have developed the analysis assuming  $k=4$  and  $F_{\text{thres}} = 40$ .

Figure 9 compares the evolution of  $\text{RMARE}^{\text{best}}$  for different values of  $B$ . As expected, the cases with a higher budget value  $B$  show less error than the other cases, i.e., the error increases, while the value of the budget decreases. Here, a comparison with the results of the existing methods is also provided to show the improvements achieved using the proposed model.

Table 8 shows the number of added routes to set  $\mathcal{Q}$  depending on the available budget for constant simplification  $F_{\text{thres}}$  and  $k$ . It is interesting that, for higher number of installed devices, the number of added routes is also higher. This means that the transportation planner can perform severe simplification of the network (higher  $F_{\text{thres}}$ ) if the

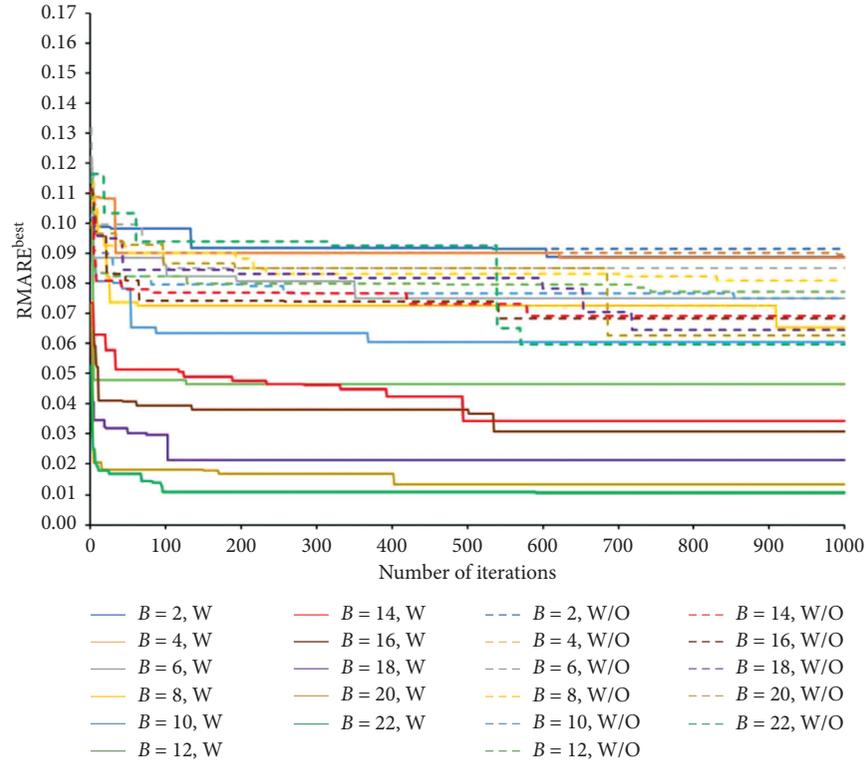


FIGURE 9: Evolution of  $\text{RMARE}^{\text{best}}$  with the number of iterations considering different values of  $B$  with (W) or without (W/O) the proposed model.

TABLE 8: Number of routes that appear with the application of the proposed algorithm according to the  $B$  value.

| $B$ | Number of routes in set $\mathfrak{R}$ | Number of routes in set $\mathcal{Q}$ | Added routes with $\mathcal{S}\mathcal{L}$ | Number of routes in $\mathcal{C}$ |
|-----|--|---------------------------------------|--|-----------------------------------|
| 2   | 162                                    | 72                                    | —  | 72                                |
| 4   | 162                                    | 72                                    | 2  | 73                                |
| 6   | 162                                    | 72                                    | 3  | 75                                |
| 8   | 162                                    | 72                                    | 3  | 75                                |
| 10  | 162                                    | 72                                    | 5  | 77                                |
| 12  | 162                                    | 72                                    | 6  | 78                                |
| 14  | 162                                    | 72                                    | 18   | 90                                |
| 16  | 162                                    | 72                                    | 20   | 92                                |
| 18  | 162                                    | 72                                    | 37   | 109                               |
| 20  | 162                                    | 72                                    | 70   | 142                               |
| 22  | 162                                    | 72                                    | 75   | 147                               |

number of devices to install is higher. This leads to a better performance of the algorithm while the expected estimation results will also be satisfactory.

**5.2. The Ciudad Real Network.** In this section, we illustrate the application of the proposed method to a real-size network. In particular, we have adapted the Ciudad Real network used in Castillo et al. [1] and in Owais et al. [45]. This network has 218 links and 105 nodes and has been divided into a total of 20 traffic zones, originating a matrix of 380 OD pairs (see Figure 10). After applying the method exposed in

Section 3, the extended OD matrix is composed by 10,374 node-based OD pairs. After that, the  $k$ -shortest path was carried out, assuming  $k=3$ , discarding routes with more than 1.1 times the shortest path, resulting in a total of 18,630 routes in set  $\mathfrak{R}$ . Then, the proposed algorithm has been applied assuming  $B=50$  and  $F_{\text{thres}} = 100, 150, 200,$  and  $250$ , and the number of resulting routes in  $\mathcal{Q}$  is shown in Table 9. Note that the number of resulting routes in  $\mathcal{C}$  after applying the algorithm is on the same order of magnitude which again leads to good results in all cases. Figure 11 shows the evolution of  $\text{RMARE}^{\text{best}}$  showing the same trend as in the Nguyen-Dupuis network.

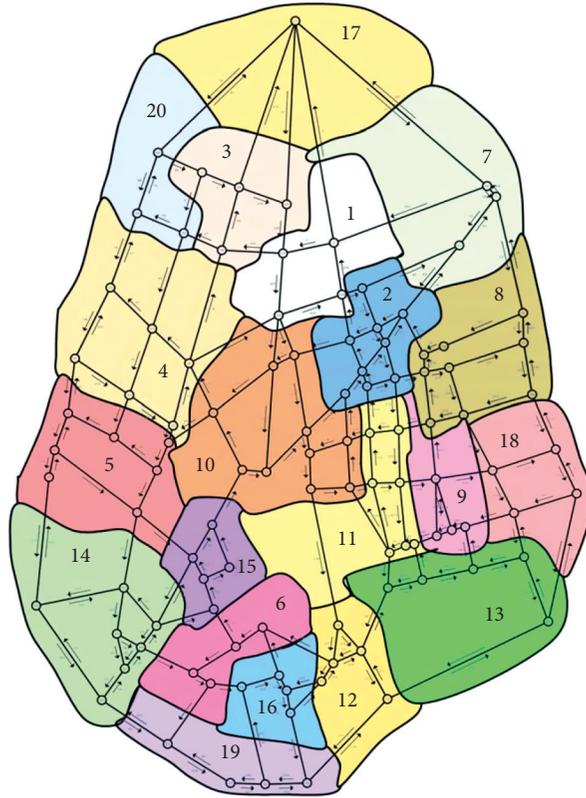


FIGURE 10: The Ciudad Real network (adapted from Castillo et al. [1]).

TABLE 9: Sets of routes that appear with the application of the proposed algorithm for the Ciudad Real network case.

| Threshold flow, $F_{\text{thres}}$ | Number of routes in set $\mathfrak{R}$ | Number of routes in set $\mathcal{Q}$ | Added routes with $\mathcal{SL}$ | Number of routes in $\mathcal{C}$ |
|------------------------------------|--|---------------------------------------|----------------------------------|-----------------------------------|
| 100                                | 18,630                                 | 2987                                  | 6151                             | 9138                              |
| 150                                | 18,630                                 | 1389                                  | 9122                             | 10,511                            |
| 200                                | 18,630                                 | 361                                   | 11,417                           | 11,778                            |
| 250                                | 18,630                                 | 138                                   | 11,687                           | 11,825                            |

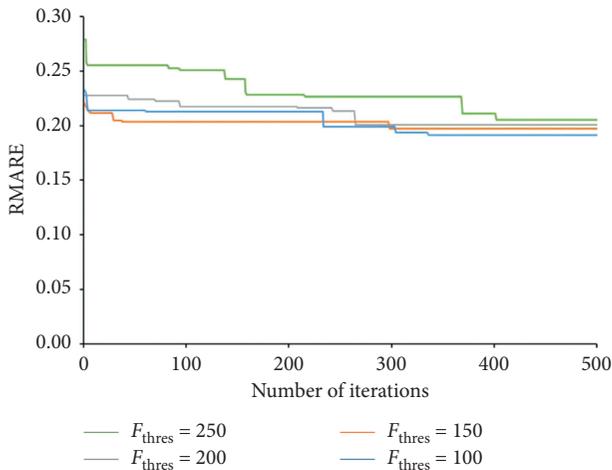


FIGURE 11: Evolution of  $\text{RMARE}^{\text{best}}$  for the Ciudad Real case.

## 6. Conclusions

This paper proposes a two-step methodology that may have important advantages from a practical point of view compared with the existing methodologies that estimate traffic flows using plate scanning data.

In the first step, a new methodology for traffic network modeling that does not use centroids and connectors is presented. The problems derived from the use of these tools are well known and have been analyzed in several studies and projects. Instead, the proposed methodology is an important step forward from a practical point of view since it defines a network with more detail so that the displacements between nodes (i.e., the network routes) have a better definition and also do not lead to artificial congestion in links. Both considerations make this method compatible with the data obtained from plate scanning as a particular case of AVI readers.

In the second step, a heuristic algorithm is proposed to face the uncertain or partial knowledge of routes which is essential for the correct application of the plate scanning technique or other AVI-based methods. For this, an iterative process has been formulated to obtain the device location to expect the best possible result in terms of link flow estimation. The proposed algorithm improves the expected flow estimation quality for the same value of the objective function used in other papers found in the literature. In addition, this reduction is achieved in the first 300–500 iterations of the algorithm. One of the reasons for this improvement is the incorporation of new routes to the estimation model once the field data have been collected. This fills a gap of the existing methods since they did not mention what to do with those scanned vehicles whose scanned patterns do not match with the modeled routes.

Finally, to evaluate the influence of different parameters of the algorithm on the final solution, we have performed a sensibility analysis using the well-known Nguyen-Dupuis network. In particular,

- (i) We have performed an analysis by varying the parameter  $k$  in the process of the enumeration of the  $k$ -shortest path between the nodes of the designed network. It has been observed that a high value of  $k$  allows for better estimates of flows in terms of a smaller RMARE. However, a high value for  $k$  would entail working with a network with a high number of routes, which would have some computational cost mainly in the location optimization problem. Small values of this parameter would avoid this problem, and we have proved that the quality of the solution is similar to the solutions obtained with a higher number of routes. This is because of the incorporation of routes to the estimation model once the field data are collected. The sensitivity analysis performed confirms that, for the plate scanning technique, a value of  $k$  between 3 and 4 is an acceptable value (endorsed by several authors) for this type of model.
- (ii) The analysis carried out to test the implications of the simplification of the network has been done by means of the elimination of those OD-pairs whose demand is lower than a defined threshold flow value  $F_{\text{thres}}$ . Again, the cases studied with different values of  $F_{\text{thres}}$  have obtained almost the same solution in terms of RMARE. This means that a medium-low degree of simplification leads to a good network model in terms of the final estimation process as well as the performance of the algorithm.
- (iii) It has been verified that the budget  $B$  or number of AVI devices to be installed on the network has a great influence on the estimation results. For all the cases studied, the number of devices has a substantial effect on the number of new routes and sets  $\mathcal{O}\mathcal{S}\mathcal{E}\mathcal{S}\mathcal{L}$  and therefore on the quality of the estimation errors. As expected, increasing the number of installed scanning devices on the network will

yield better observability and better estimation of flows, and hence more information about routes (both included in set  $\mathcal{Q}$  or not), which does not occur when working with a limited number of devices.

Finally, the methodology has also been applied to a real-size network. Despite, it has been observed that higher improvements on the solutions occur during the first 300–500 iterations, leading to few improvements when the algorithm has carried out a high number of iterations. It is of interest to find a method that allows to obtain the best solution for the scanned link set  $\mathcal{S}\mathcal{L}$  in earlier instances or iterations to avoid the computation costs arising from iterations that may be unnecessary. This observation deserves to be investigated and worked out in greater detail in future research, for which the use of advanced tools for heuristic optimization or machine learning is proposed to increase the efficiency of finding solutions in a shorter operating time.

## Data Availability

All the data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was funded by the Spanish Ministry of Economy and Competitiveness in relation to project TRA2016-80721-R (AEI/FEDER, UE). Similarly, the authors acknowledge Prof. Miguel Carrión (University of Castilla-La Mancha) and the university's technical staff for providing computer resources.

## References

- [1] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambronero, "Traffic estimation and optimal counting location without path enumeration using Bayesian networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 3, pp. 189–207, 2008.
- [2] E. Castillo, J. M. Menéndez, S. Sánchez-Cambronero, A. Calviño, and J. M. Sarabia, "A hierarchical optimization problem: estimating traffic flow using gamma random variables in a Bayesian context," *Computers & Operations Research*, vol. 41, pp. 240–251, 2014.
- [3] K. Perrakis, D. Karlis, M. Cools, D. Janssens, K. Vanhoof, and G. Wets, "A Bayesian approach for modeling origin-destination matrices," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 1, pp. 200–212, 2012.
- [4] Z. Huang, X. Ling, P. Wang et al., "Modeling real-time human mobility based on mobile phone and transportation data fusion," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 254–269, 2018.
- [5] S. Ibarra-Espinosa, R. Ynoue, M. Giannotti, K. Ropkins, and E. D. de Freitas, "Generating traffic flow and speed regional

- model data using internet GPS vehicle records,” *MethodsX*, vol. 6, pp. 2065–2075, 2019.
- [6] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, “Time-evolving O-D matrix estimation using high-speed GPS data streams,” *Expert Systems with Applications*, vol. 44, pp. 275–288, 2016.
  - [7] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, “The path most traveled: travel demand estimation using big data resources,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 162–177, 2015.
  - [8] T. A. Zin, K. Kyaing, K. K. Lwin, and Y. Sekimoto, “Estimation of originating-destination trips in Yangon by using big data source,” *Journal of Disaster Research*, vol. 13, no. 1, pp. 6–13, 2018.
  - [9] E. Castillo, I. Gallego, J. M. Menéndez, and A. Rivas, “Optimal use of plate-scanning resources for route flow estimation in traffic networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 380–391, 2010.
  - [10] C. Fu, N. Zhu, and S. Ma, “A stochastic program approach for path reconstruction oriented sensor location model,” *Transportation Research Part B: Methodological*, vol. 102, pp. 210–237, 2017.
  - [11] A. Federov, K. Nikolskaia, S. Ivanov, V. Shepelev, and A. Minbaleev, “Traffic flow estimation with data from a video surveillance camera,” *Journal of Big Data*, vol. 6, no. 1, p. 73, 2019.
  - [12] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ljaz, “Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models,” *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
  - [13] S. Sánchez-Cambronero, E. Castillo, J. M. Menéndez, and P. Jiménez, “Dealing with error recovery in traffic flow prediction using Bayesian networks based on license plate scanning data,” *Journal of Transportation Engineering*, vol. 137, no. 9, pp. 615–629, 2010.
  - [14] J. Bai and Y. Chen, “A deep neural network based on classification of traffic volume for short-term forecasting,” *Mathematical Problems in Engineering*, vol. 2019, Article ID 6318094, 10 pages, 2019.
  - [15] Z. Liu, Z. Li, K. Wu, and M. Li, “Urban traffic prediction from mobility data using deep learning,” *IEEE Network*, vol. 32, no. 4, pp. 40–46, 2018.
  - [16] M. Nigro, A. Abdelfatah, E. Cipriani, C. Colombaroni, G. Fusco, and A. Gemma, “Dynamic O-D demand estimation: application of SPSA AD-PI method in conjunction with different assignment strategies,” *Journal of Advanced Transportation*, vol. 2018, Article ID 2085625, 18 pages, 2018.
  - [17] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, Engelwood Cliffs, NJ, USA, 1985.
  - [18] H. Spies, “Conical volume-delay functions,” *Transportation Science*, vol. 24, no. 2, pp. 153–158, 1990.
  - [19] L. Huntsinger and N. Rouphail, “Bottleneck and queuing analysis: calibrating volume-delay functions of travel demand models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2255, no. 1, pp. 117–124, 2011.
  - [20] E. T. Mtoi and R. Moses, “Calibration and evaluation of link congestion functions: applying intrinsic sensitivity of link speed as a practical consideration to heterogeneous facility types within urban network,” *Journal of Transportation Technologies*, vol. 4, no. 2, pp. 141–149, 2014.
  - [21] W. W. Mann, “B-node model: new subarea traffic assignment model & application,” in *Proceedings of the Eighth TRB Conference on the Application of Transportation Planning Methods Transportation Research Board*, Corpus Christi, Washington, DC, USA, July 2002.
  - [22] M. Friedrich and M. Galster, “Methods for generating connectors in transport planning models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2132, no. 1, pp. 133–142, 2009.
  - [23] Z. S. Quian and H. M. Zhang, “On centroid connectors in static traffic assignment: their effects on flow patterns and how to optimize their selections,” *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1489–1503, 2012.
  - [24] E. Jafari, M. D. Gemar, N. R. Juri, and J. Duthie, “Investigation of centroid connector placement for advanced traffic assignment models with added network detail,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2498, no. 1, pp. 19–26, 2015.
  - [25] S. Hao and L. Yang, “Traffic network modeling and extended max-pressure traffic control strategy based on granular computing theory,” *Mathematical Problems in Engineering*, vol. 2019, Article ID 2752763, 11 pages, 2019.
  - [26] E. Castillo, J. M. Menéndez, and P. Jiménez, “Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations,” *Transportation Research Part B: Methodological*, vol. 42, no. 5, pp. 455–481, 2008.
  - [27] R. Mínguez, S. Sánchez-Cambronero, E. Castillo, and P. Jiménez, “Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks,” *Transportation Research Part B: Methodological*, vol. 44, no. 2, pp. 282–298, 2010.
  - [28] X. Zhou and H. S. Mahmassani, “Dynamic origin-destination demand estimation using automatic vehicle identification data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 105–114, 2006.
  - [29] J. Liu, F. Zheng, H. J. Van Zuylen, and J. Li, “A Dynamic OD prediction approach for urban networks based on automatic number plate recognition data,” *Transportation Research Procedia*, vol. 47, pp. 601–608, 2020.
  - [30] J. Li, H. v. Zuylen, Y. Deng, and Y. Zhou, “Urban travel time data cleaning and analysis for automatic number plate recognition,” *Transportation Research Procedia*, vol. 47, pp. 712–719, 2020.
  - [31] S. Sánchez-Cambronero, A. Rivas, R. M. Barba, L. Ruiz-Ripoll, I. Gallego, and J. M. Menéndez, “A methodology to model a traffic network which have their field data obtained by plate scanning technique,” *Transportation Research Procedia*, vol. 18, pp. 341–348, 2016.
  - [32] H. Yang and J. Zhou, “Optimal traffic counting locations for origin-destination matrix estimation,” *Transportation Research Part B: Methodological*, vol. 32, no. 2, pp. 209–126, 1998.
  - [33] A. Ehlert, M. G. H. Bell, and S. Grosso, “The optimisation of traffic count locations in road networks,” *Transportation Research Part B: Methodological*, vol. 40, no. 6, pp. 460–479, 2006.
  - [34] M. Gentili and P. B. Mirchandani, “Locating sensors on traffic networks: models, challenges and research opportunities,” *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 227–255, 2012.
  - [35] M. Salari, L. Kattan, W. H. K. Lam, H. P. Lo, and M. A. Esfeh, “Optimization of traffic sensor location for complete link flow observability in traffic network considering sensor failure,” *Transportation Research Part B: Methodological*, vol. 121, pp. 216–251, 2019.
  - [36] E. Castillo, M. Nogal, A. Rivas, and S. Sánchez-Cambronero, “Observability of traffic networks. Optimal location of

- counting and scanning devices,” *Transportmetrica B: Transport Dynamics*, vol. 1, no. 1, pp. 68–102, 2013.
- [37] J. Yang and J. Sun, “Vehicle path reconstruction using automatic vehicle identification data: an integrated particle filter and path flow estimator,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 107–126, 2015.
- [38] S. Sánchez-Cambronero, P. Jiménez, A. Rivas, and I. Gallego, “Plate scanning tools to obtain travel times in traffic networks,” *Journal of Intelligent Transportation Systems*, vol. 21, no. 5, pp. 390–408, 2017.
- [39] M. Gentili and P. B. Mirchandani, “Review of optimal sensor location models for travel time estimation,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 74–96, 2018.
- [40] C. Cerrone, R. Cerulli, and M. Gentili, “Vehicle-id sensor location for route flow recognition: models and algorithms,” *European Journal of Operational Research*, vol. 247, no. 2, pp. 618–629, 2015.
- [41] N. Levy and I. Benenson, “GIS-based method for assessing city parking patterns,” *Journal of Transport Geography*, vol. 46, pp. 220–231, 2015.
- [42] S. Sánchez-Cambronero, E. Castillo, and J. Menéndez, *Traffic Prediction Models Using Bayesian Networks and Other Tools*, LAP Lambert Academic Publishing, Riga, Latvia, 2011.
- [43] E. Castillo, Z. Grande, A. Calviño, W. Szeto, and H. Lo, “A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks,” *Journal of Sensors*, vol. 2015, Article ID 903563, 26 pages, 2015.
- [44] J. Yen, “Finding the  $K$  shortest loopless paths in a network,” *Management Science*, vol. 17, no. 11, pp. 661–786, 1971.
- [45] M. Owais, G. Moussa, and K. Hussain, “Sensor location model for O/D estimation: multi-criteria meta-heuristics,” *Operations Research Perspectives*, vol. 6, pp. 1–12, 2019.
- [46] P. Bonsall, P. Firmin, M. Anderson, I. Palmer, and P. Balmforth, “Validating the results of a route choice simulator,” *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 6, pp. 371–387, 1997.
- [47] M. L. Hazelton, “Estimation of origin-destination matrices from link flows on uncongested networks,” *Transportation Research Part B: Methodological*, vol. 34, no. 7, pp. 549–566, 2000.