Inferring origin–destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data

Noelia Caceres*, Luis M. Romero and Francisco G. Benitez

Transportation Engineering, Faculty of Engineering, University of Seville, Camino de los Descubrimientos s/n, 41092, Seville, Spain

SUMMARY

The origin–destination matrix is an important source of information describing transport demand in a region. Most commonly used methods for matrix estimation use link volumes collected on a subset of links in order to update an existing matrix. Traditional volume data collection methods have significant shortcomings because of the high costs involved and the fact that detectors only provide status information at specified locations in the network. Better matrix estimates can be obtained when information is available about the overall distribution of traffic through time and space. Other existing technologies are not used in matrix estimation methods because they collect volume data aggregated on groups of links, rather than on single links. That is the case of mobile systems. Mobile phones sometimes cannot provide location accuracy for estimating flows on single links but do so on groups of links; in contrast, data can be acquired over a wider coverage without additional costs. This paper presents a methodology adapted to the concept of volume aggregated on groups of links in order to use any available volume data source in traditional matrix estimation methodologies. To calculate volume data, we have used a model that has had promising results in transforming phone call data into traffic movement data. The proposed methodology using vehicle volumes obtained by such a model is applied over a large real network as a case study. The experimental results reveal the efficiency and consistency of the solution proposed, making the alternative attractive for practical applications. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS: transport demand; O–D matrix estimation; traffic flow; mobile phone data

1. INTRODUCTION

The origin–destination (O–D) trip matrix is an essential source of information about traffic demand used by administrative authorities for proper strategic planning and management of road infrastructure networks. An O–D matrix is difficult and often costly to obtain by direct measurements/interviews or surveys, although one may obtain a 'reasonable' estimate in a short term by using traffic counts and a prior O–D matrix. The prior O–D matrix is typically assumed to come from a sample survey and may be an old (probably outdated) matrix. Travel surveys are obtained through costly and laborious processes, which, from the initial data gathering to the result exploitation, are lengthy and may take years. Traffic counts are one type of information that can be collected automatically on a subset of links by fixed sensors embedded in the road (e.g. inductive loop, radio frequency identification (RFID), infrared, and camera). In most commonly used methods for matrix estimation, the prior O–D matrix is iteratively 'adjusted' or 'changed' to reproduce those observed

---

*Correspondence to: Noelia Caceres, Transportation Engineering, Faculty of Engineering, University of Seville, Camino de los Descubrimientos s/n, 41092, Seville (Spain). E-mail: noeliacs@esi.us.es

link volumes by means of modelling approaches. The reliability of sensor data is therefore critical for O–D matrix estimation. Combining loop data with other systems that also collect volume data might be of crucial importance for improving the efficiency and reliability of the estimated matrix.

Other existing systems based on collecting volume data aggregated for groups of links are not used in matrix estimation methods. One example of these systems is those providing measures in standard 'average daily traffic' format, aggregated in both directions (two links). Other technology that may also collect aggregate volumes is the mobile system. In most cases, the road network is denser than the cell distribution, so multiple roads may be covered by the same cell. The location accuracy provided without terminal and network modifications does not allow detection of movement of a handset along a particular road; it can only be asserted that it moves along one of the group of roads covered by the cell. Although mobile systems do not measure volumes as accurately as fixed sensors, volume data can be acquired on a wider coverage without additional costs arising. Thus, mobile phones can be regarded as a complementary solution to fixed sensors in order to enhance the available information for mobility monitoring purposes. The rationale behind it is the importance of the huge amount of data available at the mobile operators' database to be used as a rich source of information to diminish the uncertainty of mobility O–D matrices. Evidently, this type of data is complementary to other sources such as link traffic flow counts, licence plate recognition, audio video interleave data capturing techniques and RFID systems among others.

The paper presents a methodology for matrix estimation using aggregate volume on groups of links. Traditional modelling approaches for matrix estimation are formulated to use volume data measured on single links, so the concept of aggregate volumes may be very useful to include any available volume data in the estimation process. It is evident that an aggregate scheme may also handle individual (disaggregate) links as a special group composed of a single link. Therefore, the approach may combine aggregate and disaggregate data from different sources, but for the sake of clearness, the attention is focused on just the aggregate case using volume inferred from mobile phone data. These systems have become widespread in collecting traffic data because they provide great advantages over traditional methods of traffic measurement. Anyway, this approach may also be applied using aggregated data from any other technology that collects aggregate volumes.

This paper is organized as follows. In Section 2 a short review of matrix estimation approaches is provided, including related works about utilizing mobile phone data in transportation. Section 3 describes the model formulation and the solution method to estimate O–D matrices by means of aggregate volumes of groups of links. In Section 4, after introducing the mobility management strategy used in mobile systems, a method of inferring volume data from mobile phone counts is presented. A case study on a real network to demonstrate and verify the proposed method is summarized in Section 5. The last section concludes the paper with major findings and future extensions.

2. STATE OF THE ART

2.1. Origin–destination matrix estimation using classical modelling approaches

The O–D matrix is important in describing transport demand within a given region. An origin-destination matrix is difficult to obtain by direct measurements; observation of the population of interest in its entirety within a study area is not economically (or perhaps even technically) feasible. Then, traditionally O–D matrices have been estimated using different methodologies, such as carrying out surveys on a representative sample of individuals, applying a trip distribution model, or using traffic counts as measurements of link flows in a network model in order to update an existing matrix. The third alternative is the one that has mostly been used over the past 25 years, and a considerable amount of work has been documented in the literature [1]; a set of traffic counts and a prior O–D matrix are prerequisites. The prior matrix is typically assumed to come from a survey using a finite data set (instead of using the whole population). The survey data need to be corrected, expanded, and validated in order to achieve a representative and reliable prior matrix to be used in matrix estimation methods [2]. Therefore, this prior matrix can be regarded as an observation (a good approximation) of the 'true'

O–D matrix to be estimated. The way a prior matrix is obtained is profusely documented in specialized survey books and praxis protocols, and it is out of the scope of this paper to focus attention on this matter. However, the fact that a prior matrix is regarded as a 'good approximation' of the 'true' O–D matrix does not imply that it can be used directly as a result. Travel surveys are undertaken every 5–10 years because they are costly and laborious processes; hence, prior O–D matrices are probably outdated. Furthermore, a few errors may arise during the processes of building, calibrating, and forecasting the prior matrix with models, such as sampling errors, measurement errors, transfer errors, or aggregation errors. The prior matrix is therefore updated using traffic flows, which are one type of information that can be collected automatically on a subset of links in a network, not on all links (this would be impossible nowadays because of economic budgetary restrictions). In methods based on this third alternative, the prior O–D matrix is iteratively 'adjusted' or 'changed' to reproduce the observed traffic counts when assigned to the transportation network. The aforementioned errors can also be mitigated by adjusting the prior O–D matrix to satisfy the traffic counts. In this manner, one may obtain a 'reasonable' estimate of the O–D matrix; hence, this alternative is most widely used in practical applications.

2.2. Origin–destination matrix estimation using mobile phone data

The incorporation of new technologies plays a key role in improving transportation applications. The idea of using mobile phones to acquire traffic information is more and more widespread. Various review studies relating to this topic have been published in recent years [3,4]. With the specific goal of O–D matrix calculations, a few simulation works have been carried out to evaluate the potential of using mobile data for estimation of travel demand [5,6], with a view to its practical applicability. Regarding matrices obtained from network operators' data, we can refer to works using call data in order to extract mobility information of individuals [7] and to identify home or work locations [8] or even studies using information retrieved from turned-on phones for generating trip distribution tables [9] or time–space trajectories [10,11]. According to the reviewed literature, the major drawback in the use of phone data for matrix estimation is the composition of the sample, which may be constituted by active phones only (in communication—call, SMS, MMS, etc.) or by idle phones as well (turned on and not in communication). Thus, the phone status when samples are collected drastically changes the size or the location accuracy of the data. A sample from only active phones would guarantee higher location accuracy because the mobile system always knows the server cell to which the active phone is connected during a communication. At present, cell identification (cell ID) is the basic geographic unit provided without terminal or network modifications. However, the sample size is much smaller than those from idle phones because the size depends on the number of calls. In contrast, a sample from idle phones offers less accurate location data. Whenever the phone has idle status, the system does not need to know the cell where a phone is located but needs to know a set of candidate cells for becoming the serving cell. This set is called the location area, which consists of a group of adjacent cells.

When cells or location areas are used as home or work zones for building an O–D matrix, mobility patterns may be different from when using traditional traffic zones. The spatial extent of zones is of crucial importance for the understanding of trips that are generated and attracted within the zone. In conventional transport studies, a zoning system is defined according to transportation planning criteria, such as homogeneous land uses and/or population composition using census and socioeconomic data [2]. In mobile networks, generally base stations (cells) are placed to provide adequate coverage for communication rather than the aforesaid criteria. Moreover, cell size varies according to urban or rural areas and may cover areas with different land uses (e.g. home and work zones fit into the same cell); hence, the fitting problem is worse for location areas. However, even using cells (active phones sample), a typical traffic zone and a cell do not overlap exactly in both the size and the features of the area. Either poor network representations or too coarse zoning systems may invalidate the results of even the most theoretically appealing model [2]. Anyway, it is difficult to exactly detect the starting point and the destination of a trip by means of communication records; some users only have 1–2 records per day, and the points where users make calls may not coincide with a trip origin and destination.

There are other drawbacks to estimating O–D matrices directly from phone data, such as the uncertainty about the number of phones carried in the vehicles, or the fact that not every phone is under contract with the network operator that delivers the data. Therefore, phone trips do not match vehicle trips, and the matrix must be corrected by methods based on other vehicle-monitoring systems. Despite these drawbacks, an O–D matrix generated using phone data directly must be regarded as valuable prior information about the transport demand, and it can help to obtain more accurate results in future. Until now, the aforementioned modelling approaches based on traffic counts and a prior matrix remain the most reliable ones for matrix estimation.

The preceding arguments provide the rationale for not drawing up O–D matrices directly from phone data but using the inferred flows between two adjacent/distant cells (inferred from call data) to update an existing prior matrix by solving an optimization problem.

2.3. Traffic flow estimation using mobile phone data

The most used sensor to collect traffic flow data is the loop detector. Data provided by these sensors are available with little effort, but they are limited in terms of cost and coverage. Their measurements depend on the existing infrastructure, which tends to be expensive to extend or modify because of the cost of devices, installation, or maintenance. Although loops generally produce reasonable data all the time, they may occasionally provide missing values or erroneous measurements due to various malfunctions. There exist two main types of errors [12]: first, the detectors tend to undercount vehicles; second, a large percentage of detectors tend to count vehicles in neighbouring lanes in addition. By considering these negative aspects, other alternatives for collecting traffic flow data should be included for use in matrix estimation approaches. According to the literature, a fair number of works relating to traffic flow estimations using phone data have shown good quality results in order to be regarded as a feasible alternative for mobility detection purposes. Most of them use data generated by active phones. As already mentioned, a sample from only active phones guarantees higher location accuracy based on cell ID. Certainly, having a sample from idle phones increases the sample size, but location data are less accurate at the level of the location area, which consists of a group of adjacent cells. Then, the problem of matching users' movements to roads is more difficult to solve than using location data at cell level. This is the main reason why most studies so far have used datasets from active phones to obtain traffic flows, even though the sample size is smaller.

Regarding active phones, the most common event for detecting the user's movement between two cells is the handover. A handover record is inserted into the system databases when a phone with a call changes from one base station/cell to another. The analysis of these records provides a measurement of the number of active phones moving from one cell to another in a given period, that is to say a boundary-crossing rate. Two works [13,14] found that the phones' flow from handover calls is closely related to the vehicles' flow measured by loop detectors, having similarities for most hours of an ordinary day. Typical flow peaks associated with times both in the morning and in the afternoon appear in both cases. However, those studies concluded that accurate vehicle flows cannot be obtained directly from phone data because of sample size problems. Flow data derived from active phones yield information on only a statistical sample of all the travelling vehicles; some vehicles may carry more than one mobile phone, either phones of other operators or switched-off phones; even no calls may occur per vehicle. For these last situations, phones in movement are not detected. These aspects imply special treatment to correlate phone and vehicle counts. A calibration process is therefore required that uses vehicle volume measures obtained, for example, from loop detectors located spatially in the same monitored section. In this regard, different works using loop detector data gave accurate results by means of procedures to bring phone counts to vehicle counts based on empiric transfer functions [15], correction factors [10], or Newtonian relaxation techniques [16]. In other recent work [17,18], a set of models containing additional information about calling users' behaviour was proposed to infer vehicle volumes from call data. The models have had promising results in transforming call data into traffic movement data, which makes it possible to calculate accurate traffic flows on any road in any period.

Considering the findings in the field of O–D matrices and traffic flows, a methodology for matrix estimation that combines the traditional modelling approaches with traffic flows inferred from mobile phones is proposed. A mixed approach makes phone data attractive for matrix estimation applications because of the fact that they involve relatively low costs and wider coverage. However, the methodology continues to use a matrix derived from a sample survey as a prior matrix. Survey data offer the possibility of obtaining more useful trip data (trip purpose, transportation mode, trip length, etc.). At present, these pieces of information cannot be derived accurately from raw phone data, and they are necessary for the other stages in travel demand modelling.

3. METHOD FOR UPDATING ORIGIN–DESTINATION MATRICES

3.1. Introduction

As already commented, most methods for matrix estimation use volumes on a subset of links and a prior O–D matrix so that the prior matrix is iteratively updated to reproduce the observed volumes. The matrix estimation problem can then be interpreted as pursuing two types of objectives, of which one is to fit the traffic counts as exactly as possible and the other is to search for a solution that is as close as possible to the prior matrix. Better matrix estimates can be obtained when information is available about the overall distribution of traffic through time and space. Other existing technologies are not used in matrix estimation methods because they collect volume data aggregated for groups of links, rather than on single links. The new feature of this contribution with respect to the traditional modelling approaches is the use of aggregate volumes on groups of links. Some modifications in the model formulation to use aggregate volumes must therefore be developed. Anyway, it is evident that an aggregate scheme may handle individual (disaggregate) links as one group composed of a single link. The proposed methodology does not exclude the use of information coming from either link volume data or other sources.

3.2. Mathematical formulation

Consider a study area that has been partitioned into traffic zones with trips from any origin to all destinations. Each traffic zone is represented by a node called a *centroid*, where trips originate or terminate. An O–D trip matrix is denoted by $T = [T_{ij}]$, its (i,j) element being the number of trips from origin i to destination j during a certain period. A road network corresponding to the study area is abstracted into a graph model consisting of a set of *regular nodes* and a set of *directed links*. The service level associated with the links is given by link performance functions $s_a(v_a)$, which relate the travel time on each link to the flow across the link. Finally, the assignment of the matrix to the network model in order to obtain flow and travel time on each link is considered to be a deterministic (or stochastic) user-equilibrium procedure, whose behavioural principles are described by the two conditions usually attributed to J.G. Wardrop [19]. Regarding the adjustment problem, we have reformulated it to use aggregate volumes on groups of links. The formulation proposed to adjust the prior matrix includes a quadratic term to control deviations with respect to the observed volume data; the distance between prior and estimated matrices is controlled by a set of variable bounds and functional constraints, which define admissible ranges for individual O–D pairs, zone productions and attractions, and total number of trips. Therefore, it must be noted that we are not really minimizing the distance between both matrices but limiting the variations involved to certain feasible intervals. Next, the necessary mathematical conventions to formulate the matrix adjustment bi-level approach are summarized.

Indices and sets:

$i \in I$ Origin zones (n_o is the total number of origin zones)
$j \in J$ Destination zones (n_d is the total number of destination zones)
$a \in A$ Network links

$g \in G$ Groups of links observed

$k \in K_{ij}$ Routes or paths from origin i to destination j

Constants:

δ_{ak}	1 if link a belongs to path k, 0 otherwise
$\delta_{a|g}$	1 if link a belongs to group g, 0 otherwise
u_{ij}, l_{ij}	Upper and lower bounds for (i,j) O–D pair
u_i^O, l_i^O	Upper and lower bounds for trips generated by zone i
u_j^D, l_j^D	Upper and lower bounds for trips attracted by zone j
u, l	Upper and lower bounds for total network trips
$\bar{v}|_g$	Observed travel demand through the group of links g
$\alpha|_g$	(Optional) weight factor associated with the volume on group of links g

Functions:

$s_a(v_a)$ Performance (volume delay or cost) function of link a

Variables:

v_a	Volume on link a
h_k	Flow on path k
$p_k = \frac{h_k}{T_{ij}}$	Probability or proportion of path k
$P_{ij,a} = \sum_{k \in K_{ij}} \delta_{ak} p_k$	Proportion of trips from node i to node j through the link a
$v|_g$	Estimated travel demand through the group of links g
T_{ij}	Travel demand (trips) from origin i to destination j.

Then, the bi-level approach proposed in this investigation is formulated as follows:

<u>Upper Level</u>

$$\underset{T_{ij}}{\text{Minimize}} \quad f\left(T_{ij}\right) = \frac{1}{2} \cdot \sum_{g \in \hat{G}} \alpha\big|_g \left(v\big|_g - \bar{v}\big|_g\right)^2$$

s.t. $l_{ij} \le T_{ij} \le u_{ij}$ $\forall i \in I, j \in J$ (a)

     $l_i^O \le \sum_{j \in J} T_{ij} \le u_i^O$ $\forall i \in I$ (b)

     $l_j^D \le \sum_{i \in I} T_{ij} \le u_j^D$ $\forall j \in J$ (c)

     $l \le \sum_{i \in I} \sum_{j \in J} T_{ij} \le u$ (d)

<u>Lower Level</u>

$$\underset{v_a}{\text{Minimize}} \quad g(v_a) = \sum_{a \in A} \int_0^{v_a} s_a(v) dv$$

s.t. $v_a = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_{ij}} \delta_{ak} h_k, \quad \forall a \in A$

     $\sum_{k \in K_{ij}} h_k = T_{ij}, \quad \forall i \in I, j \in J$

     $h_k \ge 0 \quad\quad\quad \forall k \in K_{ij}, i \in I, j \in J$

(1)

The lower-level programme, known as Beckmann's transformation, is the basic model for obtaining those volumes v_a on all network links satisfying the *user-equilibrium* conditions for a given fixed demand T_{ij} [19]. In order to obtain the aggregate volume $v|_g$ on every group $g \in G$, it is necessary to carry out a *sub-problem* assignment that provides the standard results of the lower level (Equation (1)). The results can then be expressed in terms of path flows or path proportions as follows:

$$v\big|_g = \sum_{a \in g} v_a = \sum_{a \in A} \delta_{a|g} v_a = \sum_{a \in A} \delta_{a|g} \left(\sum_{i \in I} \sum_{j \in J} T_{ij} \sum_{k \in K_{ij}} \delta_{ak} p_k \right) \quad\quad (2)$$

$$= \sum_{a \in A} \delta_{a|g} \left(\sum_{i,j} T_{ij} P_{ij,a} \right) \forall \text{group} \ g \in G$$

Expression (2) shows that the aggregate volume on the group of links g is the sum of the volume on each link belonging to the group. The problem of estimating an O–D matrix from the aggregate

volume consists in finding the values of $n_o \times n_d$ variables (T_{ij}) from as many equations as there are independent observed volumes in the network. In most cases, the number of unknowns is greater than the total number of independent equations that can be found for the groups of links having the counts, so it will not be possible to determine a unique matrix, and the problem is underspecified. Thus, the objective of the estimation process becomes the estimation of a solution that is considered to be the best one according to some criteria. For the present approach, the criteria have already been set out and mathematically modelled in the upper level (Equation (1)) by means of the set of variable bounds (a) and functional constraints (b–d). Because of the existing uncertainty in the definition of the matrix, potential errors in the definition of the transportation network model, and noise associated with the vehicle counts, the rationale behind this makes reasonable that all the matrices in an adequate range can be considered equally suitable. In the formulation proposed, two different strategies manage the constraints, according to the type: the bounded constraints (restrictions on each element of the matrix) and the functional constraints (constraints on origin-row, on destination-column aggregations, and on total number of trips). For instance, for an outdated 5-year-old matrix to be brought up-to-date with the actual vehicle counts, one can use the outdated matrix as a prior matrix, letting the total target matrix trips vary into an interval ranging from the outdated total trips (it is not usual for mobility to decrease in 5 years) and the predicted percentage increment by the time series record (lower and upper constraints on total trips). The constraints provide a better physical perception to pick out the most suitable matrix (for example, constraints might be imposed restricting the admissible variations to ±20% in the trips production of a concrete traffic analysis zone, or accepting an increase greater than 15% in total mobility). These conditions would be part of the functional constraints and treated as terms of a new objective function, defined by the augmented Lagrangian function (ALF). Consequently, this type of constraint can be regarded as a function of some metrics (distance). The bounded constraints require a different treatment because of the large number of restrictions, their augmented Lagrangian would impose $2\,n_o\,n_d$ additional terms in the objective function; this makes the optimization problem intractable, an approach to encircle this issue is dealt in the following section.

3.3. Implementation

As mentioned earlier, the formulation using aggregate volume of groups of links is needed in order to use other available data from systems providing aggregate measurements. According to the literature, there are a large number of works that have already developed methodologies to estimate O–D matrices using link volumes. The approach followed in this research has selected one of these previously developed methods [20] as the starting point for the reformulation using volumes on groups of links. The reformulation and the validation of the resulting algorithm is a simple and unproblematic task that has never been used in matrix estimation to our knowledge, but it can be very useful for combining any available volume data in order to improve efficiency in matrix estimation.

The selected method addresses the problem of estimating and updating observed O–D matrices based upon available link-flow information, preserving the prior structure of a survey matrix. Although the modification could be applied to any approach using observed link volumes, we have selected the cited method because of the well-behaved results that the method has obtained in matrix estimation. This approach combines an augmented Lagrangian method, called the method of multipliers, with the linear approximation method credited to Frank and Wolfe. The idea of the augmented Lagrangian scheme [21] consists of adding a quadratic loss penalty term to the classical Lagrangian to form an unconstrained function whose minimum is a Kuhn–Tucker point of the original problem. However, note that the ALF proposed here will not be completely unconstrained because variable bounds will not be included. In the model, the Lagrangian function includes the functional restrictions (corresponding to trips generated, attracted, and total) present in Equation (1b)–(1d) but bound ones associated with trips between O–D pairs in Equation (1a), which are those dealt with directly by the Frank–Wolfe method. The use of these functional restrictions and variable bounds limits the possible changes in the estimated matrix regarding the information contained in the prior matrix. For the sake of brevity, the details of the mathematical optimization

scheme are not pursued further and can be found in [20]. Then, the ALF for aggregate volume data can now be written as follows:

$$
L_{\text{aug}}(T, \sigma, \tau) = \frac{1}{2} \sum_{g \in \hat{G}} \sum_{i \in I} \sum_{j \in J} \alpha|_g \left(\sum_{a \in g} \left(T_{ij} \sum_{k \in K_{ij}} \delta_{ak} p_k \right) - \bar{v}|_g \right)^2
$$

$$
+ \frac{1}{2} \left(\sum_{i=1}^{n_o} \omega_i^O \left(\left\langle u_i^O - \sum_{j=1}^{n_d} T_{ij} + \sigma_i^O \right\rangle^2 - \left(\sigma_i^O \right)^2 \right) + \sum_{i=1}^{n_o} \omega_i^O \left(\left\langle \sum_{j=1}^{n_d} T_{ij} - l_i^O + \tau_i^O \right\rangle^2 - \left(\tau_i^O \right)^2 \right) \right)
$$

$$
+ \frac{1}{2} \left(\sum_{j=1}^{n_d} \omega_j^D \left(\left\langle u_j^D - \sum_{i=1}^{n_o} T_{ij} + \sigma_j^D \right\rangle^2 - \left(\sigma_j^D \right)^2 \right) + \sum_{j=1}^{n_d} \omega_j^D \left(\left\langle \sum_{i=1}^{n_o} T_{ij} - l_j^D + \tau_j^D \right\rangle^2 - \left(\tau_j^D \right)^2 \right) \right)
$$

$$
+ \frac{1}{2} \omega \left(\left(\left\langle u - \sum_{i=1}^{n_o} \sum_{j=1}^{n_d} T_{ij} + \sigma \right\rangle^2 - \sigma^2 \right) + \left(\left\langle \sum_{i=1}^{n_o} \sum_{j=1}^{n_d} T_{ij} - l + \tau \right\rangle^2 - \tau^2 \right) \right),
$$

$$
\text{where} \langle \alpha \rangle = \begin{cases} \alpha & \text{if} \ \alpha < 0 \\ 0 & \text{if} \ \alpha \geq 0 \end{cases}
$$

(3)

and $(\omega_i^O, \omega_j^D, \omega)$ are the weight factors associated with the production constraints of zone i, the attraction constraints of zone j, and the total number of trips constraint, respectively. Lastly, $\sigma = \{\sigma^O, \sigma^D\}$, $\tau = \{\tau^O, \tau^D\}$ stand for vectors of multipliers associated with the mentioned constraints. The updating rule for σ and τ takes the following form:

$$
\sigma_i^{(m+1)} = \left\langle u_i^O - \sum_{j=1}^{n_d} T_{ij}^{(m)} + \sigma_i^{O \ (m)} \right\rangle \tau_i^{(m+1)} = \left\langle \sum_{j=1}^{n_d} T_{ij}^{(m)} - l_i^O + \tau_i^{O \ (m)} \right\rangle, \text{for each } i \in I
$$

$$
\sigma_j^{(m+1)} = \left\langle u_j^D - \sum_{i=1}^{n_o} T_{ij}^{(m)} + \sigma_j^{D \ (m)} \right\rangle \tau_j^{(m+1)} = \left\langle \sum_{i=1}^{n_o} T_{ij}^{(m)} - l_j^D + \tau_j^{D \ (m)} \right\rangle, \text{for each } j \in J \quad (4)
$$

$$
\sigma^{(m+1)} = \left\langle u - \sum_{i=1}^{n_o} \sum_{j=1}^{n_d} T_{ij}^{(m)} + \sigma^{(m)} \right\rangle \tau^{(m+1)} = \left\langle \sum_{i=1}^{n_o} \sum_{j=1}^{n_d} T_{ij}^{(m)} - l + \tau^{(m)} \right\rangle
$$

Let m be the counter to identify the successive estimates obtained by the algorithm. Then, the updating rule of the successive estimates of the Frank–Wolfe algorithm can be written as $T_{ij}^{(m+1)} = T_{ij}^{(m)} + \lambda^{(m)} \cdot (y_{ij}^{(m)} - T_{ij}^{(m)}) = T_{ij}^{(m)} + \lambda^{(m)} \cdot d_{ij}^{(m)}$ for each $i \in I$, $j \in J$, in accordance with the O–D notation. Note that $d_{ij}^{(m)} = y_{ij}^{(m)} - T_{ij}^{(m)}$ is the Frank–Wolfe descent direction, $y_{ij}^{(m)}$ is the auxiliary extreme point, and $\lambda^{(m)}$ is the step length in the mth iteration. This step is subject to lying within the interval $\lambda \in [0,1]$ because the auxiliary point y is located in a vertex (and/or edge) of the feasible region defined by the linear problem. To obtain the step length, the ALF (Equation (3)) should be minimized by expressing it as a function of λ at the new $m+1$ estimate, that is, evaluating the function for $T_{ij}^{(m+1)} = T_{ij}^{(m)} + \lambda^{(m)} \cdot d_{ij}^{(m)}$ where $T_{ij}^{(m)}$ and $d_{ij}^{(m)}$ are known. Thus, a minimization of this one-dimensional function, where $\lambda \in [0,1]$, has to be performed. Because of the numerical oscillations of the objective function values observed in the cited optimization approach, additional novelties for determining the optimum step have been developed. In this regard, the presence of the bracket operator $\langle \cdot \rangle$ has motivated the selection of the Golden Section Search algorithm for minimizing a quasi-convex function over the interval [a,b]. This algorithm is based on finding the minimum of the unimodal function by successively narrowing the range of values of λ. In our problem, the optimum value of λ is found by applying this algorithm iteratively within the interval $\lambda \in [0,1]$ in the descent direction $d_{ij}^{(m)}$.

The termination criteria of the algorithm designed are based on the following: (i) the maximum number of sub-problems (or Lagrangian stages) and (ii) the maximum number of iterations per sub-problem.

Specific rules are used to choose these maximum numbers of steps, and so, as was the case of the upper and lower bounds in Equation (1), they depend on error analysis and statistical confidence intervals.

4. ESTIMATES OF TRAFFIC VOLUMES FROM MOBILE PHONE DATA

4.1. Introduction

In mobile networks, a service coverage area is divided into smaller areas referred to as cells. A cell is defined as the area in which a phone can communicate with a certain base station. A set of adjacent cells is grouped into one location area. The cell radius depends on many parameters, such as antenna type, power levels, or even topology and buildings around. The cell radius is smaller in urban areas, where population density is high and more antennae are necessary to provide good communication services, than in rural areas, varying from a few hundred metres to several kilometres. One of the main features of mobile systems is the ability of users to access communication services from any location within the telecommunications network. To support users' mobility, the mobile phone network has an inbuilt positioning system based on the cell of origin, which is used to keep track of the position of phones so as to deliver incoming calls, messages, or data packets to subscribers. When a call occurs, the system inserts a record into its databases including parameters related to the call such as start/end time, duration, caller phone number, or identification of the server cell (cell ID). During a call, the mobile system always knows the server cell to which the phone is connected in order to route the call to the antenna. Because subscribers are free to move within the network's service area, the mobility management in mobile networks includes processes that keep the databases updated with the user's position. When a phone with a call moves from one cell to another (it crosses the boundary of different cells), the call is routed to the new cell. The transference is executed by the handover process, which automatically updates the databases with the new cell ID and the time stamp. More details about mobile systems can be found in the literature [22].

The information about phones crossing inter-cell boundaries is related to the estimation of users moving from one area to another. Therefore, a mobile system can be regarded as a kind of detector monitoring the movement of phones with calls through inter-cell boundaries. The topology of a road network is complex, and several roads may connect two cells. By analysing the road network topology and cell distribution, the routes connecting two cells are identified; in Figure 1a, there are two routes connecting cell 1 and cell 3, but only one route connects cell 1 and cell 2. By representing the road network by nodes and links, an inter-cell boundary comprises the group of links whose starting node is in one cell and whose ending node is in another cell (Figure 1b). The ideal situation is to have a large set of individual links where the movement of phones can be analysed independently. Unfortunately (indeed fortunately), in most cases, the road network is denser than the cell distribution, so multiple links connect two cells. In this case, there is no certainty that a particular handset is moving along a particular link; it can only be asserted that it is moving along one of the roadways connecting two cells. In this case, an aggregate data approach is more valuable because it allows the number of observed links to be increased. If only volume data inferred from mobile networks at the individual link level
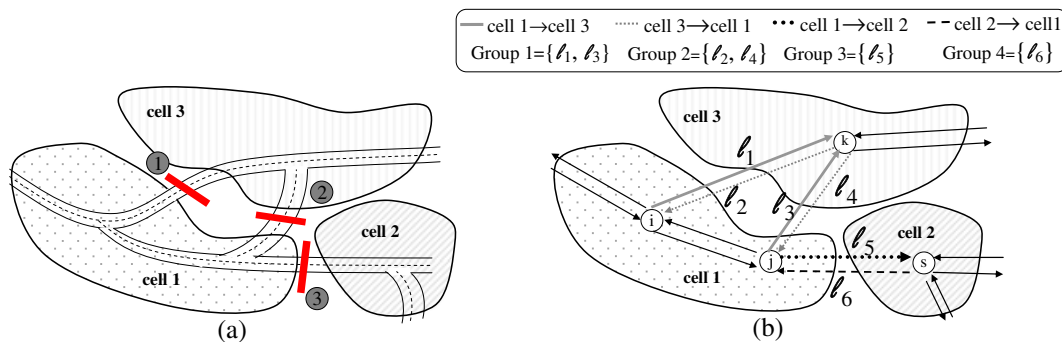

Figure 1. Groups of (a) roads and (b) links in a transport network.

were used, a significant number of links would be discarded. The situation in which a unique roadway connects two cells does not occur frequently in mobile networks. Because the adjustment procedure will become more efficient when the number of observed links taken into account increases, an approach using aggregate data on groups of links is essential for incorporating volume data inferred from mobile networks. An aggregate volume scheme must therefore be formulated in order to use this alternative in the field of O–D matrix estimation. Additionally, it is evident that an aggregate scheme may also handle individual (disaggregate) links, collected by other systems (e.g. loop detectors). Thus, an additional advantage of the presented approach is that it can combine heterogeneous data sources. Besides the use of aggregate data, not only volume inferred from phones but from other systems such as traffic count data given in the standard 'average daily traffic' format (traffic measures aggregated for both directions) are pieces of information to be considered.

It is necessary to bear in mind that the cell borders are not static, but are to some extent dynamic. Each cell coverage area has an effective radius defined by means of signal strength measurements, where a phone can communicate with the server base station. This radius may vary according to certain random factors (weather, call load, etc.) although the fluctuation is controlled by control measurements to ensure the desired cell edge reliability. These issues relating to cell borders do not drastically impact on traffic flow estimations. The approach estimates the number of vehicles moving from one cell to another, that is, the vehicles moving along the roads that connect the two cells. These roads will remain the same even if cell borders change. The only change will be reflected in the road section where the area of intersection between the inter-cell boundary and the road is located. In addition, the cells overlap at the edges to prevent holes in coverage when designing the cell layout for a mobile system. The cell overlapping area is defined as the overlap between adjacent cells with regard to the primary coverage. To avoid this problem, only roads connecting the effective radius of the cells should be considered to comprise valid group; that is to say, links whose starting and ending nodes are inside the effective radius of the corresponding cells.

4.2. Model for estimating vehicle volume using phone call data

The motion of phones while being involved in calls can be regarded as an easily accessible and low-cost alternative to infer the number of vehicles moving from one cell to another, in particular, vehicles moving along the roads that connect the two cells. The approach for inferring traffic volume data uses call data records, available without additional cost, provided by a mobile operator. From these calls, only those made by phones in motion are of interest so that the collected data must be filtered before delivery to any estimation volume procedure. In this research, we have used one of the models for estimating vehicle volumes developed by the authors [17,18]. The models transform the filtered call data into traffic movement data to calculate traffic flows in any network link in any given time. The dataset used in that study consisted of all 'outgoing calls' (calls initiated by the user, not received calls) recorded in the study area, which were provided by the operator Vodafone. For every outgoing call, the dataset included the exact time of the call, the encrypted ID number of the phone, duration, and the cell ID to which the phone was connected during the call. All the cell IDs that the phone is connected to at call time were available. Furthermore, an additional parameter is included in the collected call data, which is related to the reason for the call drop in a cell, being one of them the handover event. By using this parameter, the measure of handovers from the data was derived. Because of privacy issues, the encrypted ID for every phone was a unique and randomized number based on the original phone ID. The models were defined, calibrated, and evaluated in a field test over cells covering roads with different traffic backgrounds and characteristics. In this case study, 12 inter-cell boundaries were analysed. Only a road connected each pair of studied cells. Traffic volumes measured by detectors located next to boundaries between each pair of cells were also used. These two types of data were collected for 18 regular working days, composing a historical dataset. This dataset was divided into two parts: a calibrating set for estimating model parameters and a testing set for model assessment.

The cited work previously performed data filtering in order to detect situations of phones' movement from call data. According to the literature, the handover is the most used event to identify such situations. However, that work did not use only handover events. If the call ends before the phone enters the new cell, the handover is not performed, and the phone movement is not detected. Hence, an additional

event was considered on the basis of the realization of two consecutive calls in different cells within a short period. In particular, the model selected for use in the current study is based on the approximation of the physical phenomenon of 'in-motion calls'. A phone makes an 'in-motion call' either when the user has an active call and moves from one cell to another, which is the case called the handover (Figure 2a), or when the user makes a call in each of the two cells composing the boundary within a short period (Figure 2b). The count of in-motion calls transforms call data into traffic movement data so that the number of in-motion calls is used as input in the model to calculate traffic flows.

In order to formulate a theoretical model that reliably represents the relation between the number of in-motion calls and the number of vehicles, an expression is defined to model each of the situations in which an in-motion call may occur. The fact that the studied cells are away from urban environments allows us to suppose that practically all of the in-motion calls are made by users on board vehicles travelling along roads that connect pairs of cells. This number of in-motion calls depends on the proportion of all the moving vehicles whose occupants have made such calls. This proportion is founded on the probability of making an in-motion call on board a vehicle, using the monitored operator, in an hour period j and in a pair of cells (or inter-cell boundary) k, $P(j,k)$. The estimation model is defined by means of the hypothesis $n_{imc}(j,k) \approx P(j,k) \cdot n_{veh}(j,k)$, considering $n_{imc}(j,k)$ and $n_{veh}(j,k)$ as the number of in-motion calls counted and the number of vehicles travelling along the road that connects the pair of cells k in period j, respectively. The expression for the probability of making an in-motion call is therefore required. Regarding in-motion calls, it is necessary to comment that they are made by only a sample of all the phones on board vehicles. Some vehicles may carry more than one phone, either from other operators or switched off, although either various calls or none may be made from the same vehicle. These features are already included in the calculation of $P(j,k)$ because its expression is based on the ratio of calls to vehicles, not phones to vehicles; therefore, $P(j,k)$ stands for the probability that a phone makes an in-motion call (not any call), which may be of type 1 (associated with handover) or type 2 (associated with consecutive calls in different cells within a short time). After modelling the situations involving in-motion calls, this probability is estimated by adding the success probability of each type of in-motion calls, that is, $P(j,k) = P_j \cdot Q(j,k) + P_j^2$, whose terms are related to the types of in-motion calls. With regard to the first type (Figure 2a), the probability of making a call on board a vehicle and this call executing a handover is given in a simplified manner by $P_j \cdot Q(j,k)$, where P_j is the probability of 'making a call on board a vehicle, using the monitored operator' in the hour period j, and $Q(j,k)$ stands for the probability of handover given by

$$Q(j,k) = \frac{1}{\alpha(j,k)} \left[1 - e^{-\alpha(j,k)} \right], \quad \text{with} \quad \alpha(j,k) = \frac{L_k}{V_k \cdot T_j} \tag{5}$$

The expression (5) depends on the hour period j and the boundary k due to the term $\alpha(j,k)$. Through this last parameter, the dependence on the distance, which a phone must travel within a cell until it enters another cell (crosses the boundary), is captured. The smaller this distance, the easier is that the call must execute a handover. This length depends on the road that passes through the origin cell of the boundary k. The term $Q(j,k)$ introduces characteristics of the roads in the cell because it depends on the mean velocity at boundary k, V_k; the length travelled in the cell before leaving it, L_k; and the
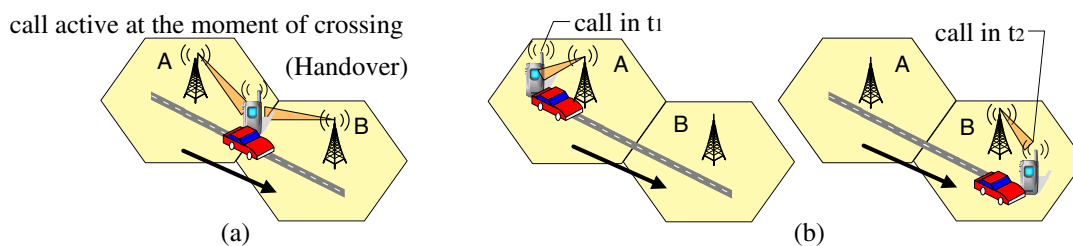

Figure 2. In-motion calls: (a) handover and (b) two calls in t_1 and t_2 such that $t_2 - t_1 \leq T$.

mean call duration in interval j, T_j. The expression is derived from a simple mobility model after introducing randomness into the trajectories that users describe across the network [23]. With regard to the second type of in-motion calls (Figure 2b), the success probability is modelled as P_j^2. This simplification is founded on the making of two calls during the hour period j taking into account that P_j is defined as the probability of 'making a call on board a vehicle, using the monitored operator' in the period j. One should bear in mind that the process of making a call is not independent of the fact of having made a previous call recently. Moreover, the sample does not deal with the whole population of phone users but with a particular group of users: those travelling on board a vehicle. The calling activity of this sample of users may be considerably different from those of the entire population. In order to simplify the modelling of in-motion calls, the hypothesis that the two calls are uncorrelated in spite of a bias may arise with this assumption. The values of P_j are determined using an empirical approach based on the ratio of successfully knowing the number of in-motion calls and the number of moving vehicles.

In order to investigate the dependence of the call probability P_j with space (location), the empirical approach based on the proportion of times in which a favourable or successful event occurs with regard to the total number of possible results was applied. In this case, the event of interest was 'making a call on board a vehicle supported by the operator providing call data'. The successful event considered takes place when an in-motion call occurs, being the number of in-motion calls directly extracted from the collected call data. Furthermore, the total number of possible results, which is related to the number of vehicles moving along each inter-cell boundary, is known by means of counting stations that were near such a boundary. Then the empirical approach for determining the call probability was computed separately for each pair of cells in order to consider the possible dependencies on the time and on the space. The outcome shows that the estimated values of the probability for each location remained within similar ranges for the same period. This coincidence was attributed to the similarity of calling activity of users who travel along the associated roadways. The cells mainly support freeway traffic, away from sites such as residential areas, shopping areas, commuter hubs, and so on. In these last cases, the behaviour of making a call drastically changes from site to site because each of them serves users with considerably different call patterns. In contrast, no significant factor influences the calling activity of users over the studied cells, showing a regular, stable behaviour. In these cells, the calling behaviour is not affected by the site activity so that it seems reasonable to assume that the locations may be aggregated. Then, the process was repeated but aggregating all locations into a single sample. Logically, the aggregate scheme provokes masked errors, but this alternative was selected for the transferability of findings to other locations with similar traffic features and for increasing the sample size.

Regarding the two types of events generating in-motion calls, it is worth to note that the proportion of occurrence of each of them varies in the range 45%–55%. This percentage distribution remains within the same range over time and space so that no significant changes were remarkable to be showed in the results.

The functional formula of the model is presented in Equation (6), after introducing a set of parameters $\varphi = \{a, b_1, b_2, c, d\}$ and substituting the expression of $Q(j,k)$ given in Equation (5). The detailed model derivation can be found in [17,18]. The model uses the number of in-motion calls counted in period j and inter-cell boundary k, $n_{\mathrm{imc}}(j,k)$, as input data in order to estimate the corresponding number of moving vehicles, $\hat{n}_{\mathrm{veh}}(j,k)$.

$$n_{\mathrm{veh}}(j,k) \approx \frac{n_{\mathrm{imc}}(j,k)}{P_j^2 + P_j \cdot Q(j,k)} \quad \xrightarrow[\{a,\, b_1,\, b_2,\, c,\, d\}]{} \quad \hat{n}_{\mathrm{veh}}(j,k) = \frac{a \cdot n_{\mathrm{imc}}(j,k)}{P_j^2 + P_j \cdot \frac{b_1}{\alpha(j,k)}\left[1 - e^{-b_2 \cdot \alpha(j,k)}\right] + c} + d \quad (6)$$

The model parameters were estimated using a calibration stage where the criterion is based on the minimization of the sum of the absolute relative error between the observed and the modelled values. The evaluation of this proposed model was carried out using criteria such as error measures in absolute values, percentiles, and linear and rank correlation between the observed volumes and the estimates; some of the results are showed in Table 1. Figure 3 shows a comparison between vehicle flows observed and estimated at two inter-cell boundaries. The plot reveals that the estimates follow

Table I. Error measurements (aggregate for all periods and boundaries).

MAE	MARE	MedARE	Rank correlation	Linear correlation
203.5960	0.2000	0.1646	0.5662	0.5333

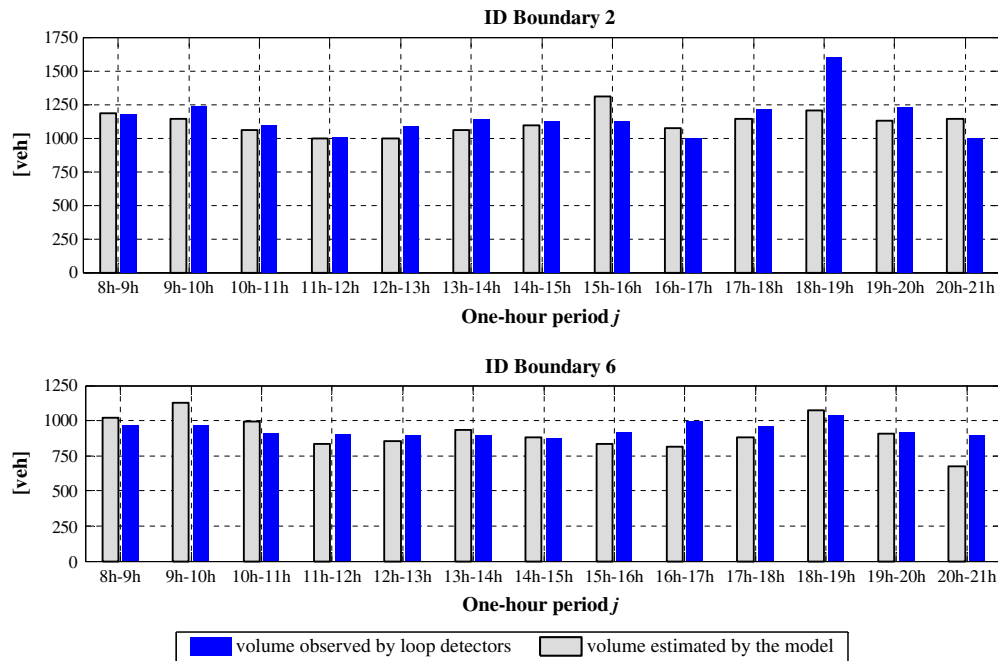MAE, mean absolute error; MARE, mean absolute relative error; MedARE, median absolute relative error.

Figure 3. Vehicle volumes observed and estimated at boundaries 2 and 6 in each 1-hour period.

the peaks and valleys of the observed volumes within an admissible error level (MARE < 20%; MedARE < 16.5%, aggregate for all periods and boundaries). Higher error levels are obtained from late afternoon/early evening. This result might be explained by several factors, the beginning of cheaper tariffs being one of them. This factor has a strong effect on calling behaviour—duration and number of calls—which fluctuates greatly in these periods from day to day; hence, error levels increase. The vehicle occupancy is another important factor affecting the accuracy of the estimation model. There are significant differences in vehicle occupancy by time of day; the morning peak period has lower average vehicle occupancy than the mid-day period and the evening peak period. The vehicle occupancy for going to work during the morning peak period is regular on a daily basis. From evening hours, aside from returning home, there are numerous other activities in which people engage on a less-than-daily basis, such as visiting friends or relatives, shopping, entertainment, fitness, and so on. People may engage in these activities alone or accompanied, and no stable trends exist in vehicle occupancy for these periods. Because it is difficult to exactly detect the vehicle occupancy using only call data, the model has been developed using an average vehicle occupancy rate. The model accuracy may change when vehicle occupancy is drastically different from this rate. By contrast, these error levels decrease to 6%–8% in periods when the trends in making calls and vehicle occupancy are more stable (9:00–14:00).

A larger and more comprehensive comparison between different derived models is found in a work by Caceres [18]. In that work, several models are tested, some of them are based on pure mathematical inference without inspiration from the physical phenomenon, by combining exponential and polynomial functions with other terms related to the users' calling behaviour due to the existence of a relationship between typical call activity and traffic mobility; others are formulated based on an approximation

of the physical phenomenon of in-motion calls. All models are evaluated and compared for selecting those that produced the best results. This research concluded that the 'physical' model was one of those providing better balance between all the error measurements computed.

After comparing volumes estimated by the model with volumes collected by detectors, it seems reasonable to assert that such estimates can also be utilized as observed data in methodologies to estimate O–D matrices using traffic counts. Traditionally, these approaches use volume data collected by loop detectors, in which it is known there are two main types of errors [12]. First, the detectors tend to count a few less vehicles than there actually are. In most cases, this error is less than 10% of the real volume. Secondly, detectors tend to count vehicles in neighbouring lanes in addition. In some cases, the share of the additionally counted vehicles has a 15% of divergence. The standards defined are that the total traffic volume should not vary from reality by more than 20% [12]. Then, the error levels obtained using the estimation model are within the limits for fulfilling the standards. Moreover, demand matrices are studied for periods representing 1 or 2 hours, usually morning peaks, and the achieved error levels are lower during these morning hours.

5. EMPIRICAL SETTING FOR ORIGIN–DESTINATION MATRIX ESTIMATION

The present paper has presented a methodology for estimating matrices using volume data aggregated on groups of links and a prior matrix. The approach may combine aggregate and disaggregate data from different sources, but for the sake of clearness, the attention is focused on just the aggregate case using volume inferred from phone data.

5.1. Data source

The test network is based on the transport network model corresponding to the region of Madrid (Spain), whose basic dimensions are 1079 centroids (based on conventional traffic zones defined from criteria such as land uses and socioeconomic data), 9457 regular nodes, and 21 713 directional links. The model contains topological information such as connectivity relationships between nodes and links, directions of links, link costs, and the clustering of nodes into cells in the mobile system covering the transport network. In order to estimate the final O–D matrix by applying the methodology, a prior O–D matrix and volume data on a subset of links are required. For this purpose, the prior O–D trip matrix used is taken from a previous study developed over this network, which had an available morning peak hour (8:00–9:00) O–D trip matrix corresponding to the Madrid Mobility Survey 2004, containing 1079×1079 O–D pairs.

With regard to volume data, a large amount of observed links are required; but the field test developed in [17,18] has estimates available over only 12 roads. In order to validate the methodology, the volumes on groups of links are obtained according to a distortion procedure. This procedure is based on the statistical error distribution of the estimation model explained in Section 4. This distribution is determined by comparing real vehicle volumes and estimates using the model, with the error levels being around 6%–8% for periods associated with morning peak hours. The aggregate volumes of every group of links (inter-cell boundary) are therefore derived after applying a random value following the aforementioned error distribution to available volumes. Previously, the groups of links for every boundary are identified, according to the cell layout, by clustering links whose starting node and ending node are in different and adjacent cells (Figure 1b). The cell layout used has been 'modified' in accordance with an arbitrary criterion (in order not to reveal proprietary data from the operator) using 350 cells. The study area has around 900 inter-cell boundaries. However, not all possible boundaries are 'valid' because of their characteristics (e.g. cells near to densely populated areas where pedestrians highly distort the volume inference). As a first step in the study, a processing of the initial set of boundaries is executed to filter non-valid links and to get the groups of links to be used in the adjustment process. In this test network, there are 408 groups of links related to inter-cell boundaries, involving 1582 links. In a previous study for the adjustment of the same prior matrix using loop detectors over the same network, 1420 links were observed. The criterion for selecting observed links based on cell distribution gets a similar number of links as traditional approaches based on counter stations.

5.2. Results

For the sake of conciseness, the values of the (optional) weight factors $\alpha|_g$ for all groups are set to 1, but different values can be used in order to give more importance to a group than to others. For the weight factors $(\omega_i^O, \omega_j^D, \omega)$ associated with the functional constraints integrated into the augmented Lagrangian (3), a common weight factor W is used to penalize the whole term corresponding to the functional constraints. A suitable and simple rule used in similar matrix estimation approaches [20] consists of choosing values of W of the same order of magnitude as the initial value of the objective function. For the upper level problem (1), the values of the bounds [upper, lower] limiting the range of distortion are, in the case of each O–D pair, [+25%, −25%] of the prior O–D matrix $T_{ij}^{(0)}$; for zonal productions, [+15%, −15%] of $O_i^{(0)} = \sum_{j \in J} T_{ij}^{(0)}$; for zonal attractions, [+15%, −15%] of $D_j^{(0)} = \sum_{i \in I} T_{ij}^{(0)}$; and for the total trips, [+10%, 0%] of $T^{(0)} = \sum_{i \in I} \sum_{j \in J} T_{ij}^{(0)}$.

Figure 4a shows aggregate volumes of groups of links that cross every inter-cell boundary, obtained by assigning the prior O–D matrix to the test network, versus observed volumes on those groups of links, which would be derived from the model using phone data. Figure 4b is the same comparison but showing aggregate volumes modelled by assigning the adjusted matrix. An acceptable correlation between adjusted and observed aggregate volumes is found ($R^2 = 0.76$). In a previous study developed to update the same prior matrix but using volumes collected by loop detectors, the correlation achieved was better ($R^2 = 0.80$). No significant differences in correlations are found by comparing the two cases, but phone data provide a cost-effective source compared with the high costs involved in installation or maintenance of detectors.

As already commented, the prior matrix contains valuable information about the 'true' matrix, so the adjustment method pays careful attention to the distortion of the information contained in it. By comparing the number of trips of each O–D pair in the prior matrix and the adjusted one using the proposed methodology, we have an obvious high correlation value ($R^2 = 0.98$) between them because of the constraints used for limiting the distortion range regarding the prior O–D matrix.

6. CONCLUDING REMARKS

An O–D matrix is crucial for any transport demand modelling. Better matrix estimates can be obtained when information about the overall distribution of traffic through time and space is available, not only at specified locations in the network. Mobile systems, without terminal and network modifications, provide a complementary alternative to detect phones in motion in terms of movements from one cell
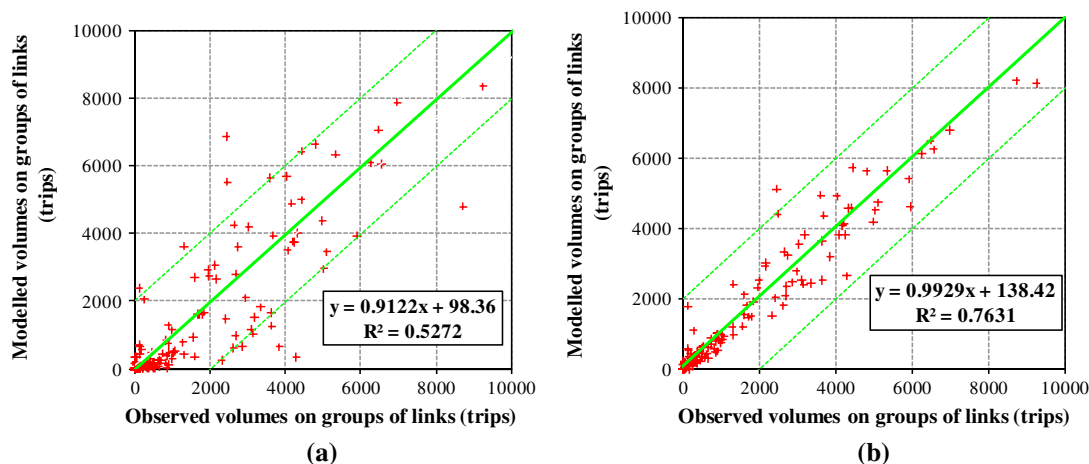

Figure 4. Correlation between observed and modelled aggregate volumes of groups of links crossing inter-cell boundaries by assigning (a) the prior O–D matrix and (b) the adjusted one.

to another. Utilizing existing mobile systems has a major advantage compared with other traffic flow capturing techniques because it avoids cost and coverage limitations associated with infrastructure-based solutions. The relatively low costs and wider-coverage data make mobile systems attractive for use as a complementary solution for collecting traffic volumes when detector data are unavailable or insufficient. Mobile phones sometimes cannot provide location accuracy for estimating flows on single links but can do so on groups of roadways (links) connecting two either adjacent or distant cells.

After reviewing the state of the art, this paper presents a methodology adapted to the concept of volume aggregated on groups of links in order to use any available volume data source. To calculate volume data, we have used a model that transforms phone call data into traffic movement data. In comparison with measurements provided by counting stations, volumes inferred from call data achieve reasonable accuracy for use in matrix estimation. By applying the methodology for matrix estimation to vehicle volume obtained by the model, the experimental results reveal the efficiency and consistency of the solution proposed in comparison with a similar study using loop data. The most important objectives that have been reached are as follows: (i) mobile phone technology is proposed as a feasible system to infer aggregate volume on groups of links to be used as observed data for O–D matrix estimation; an aggregate scheme has never been used in matrix estimation to our knowledge, and it has great potential for combining any available data source and obtaining more accurate matrices; (ii) reformulation of the objective function taking aggregate volume notation into account; and (iii) numerical validation of the proposed methodology. Other minor contributions are as follows: (iv) analytical expressions for the gradient of the new objective function; and (v) methodological changes to correct oscillations observed in applying the optimization method, using the golden section algorithm instead of a bisection search.

Future works are oriented towards developing mixed formulations using O–D matrices and traffic flows derived from phone data, as well as from traditional techniques (e.g. field surveys, loop detectors, and cameras).

REFERENCES

1. Abrahamsson T. Estimation of origin–destination matrices using traffic counts—a literature survey. Report IR-98021, Int. Institute for Applied Systems Analysis, 1998.
2. Ortuzar J de D, Willumsen LG. *Modelling Transport*. Third edn. John Wiley and Sons, Ltd.: West Sussex, England, 2001.
3. Caceres N, JP Wideberg, Benitez FG. Review of traffic data estimations extracted from cellular networks. *IET Intelligent Transport Systems* 2008; **2**(3):179–192.
4. Qiu Z, Jin J, Cheng P, Ran B. State of the art and practice: cellular probe technology applied in ATIS. *Transportation Research Board, 86th Annual Meeting*, 2007; paper no. 07-0223.
5. Sohn K, D Kim. Dynamic origin–destination flow estimation using cellular communication system. *IEEE Transaction on Vehicular Technology* 2008; **57**(5):2703–2713.
6. Zhang Y, Qin X, Dong S, Ran B. Daily O–D matrix estimation using cellular probe data. *Transportation Research Board, 89th Annual Meeting*, 2010; paper no. 10–2472.
7. White J, Wells I. Extracting origin destination information from mobile phone data. *11th International Conf. on Road Transport Information and Control*, 2002; pp. 30–34.
8. Ahas R, Silm S, Saluveer E, Järv O. Modelling home and work locations of populations using passive mobile positioning data. In *Location Based Services and TeleCartography II*, Gartner G, Rehr K (Eds). Springer: Berlin, 2009; 301–315.
9. Pan C, Lu J, Di S, Ran B. Cellular-based data-extracting method for trip distribution. *Transportation Research Record* 1945, Journal of the Transportation Research Board, National Research Council, Washington, D.C., 2006; 33–39.
10. Friedrich M, Immisch K, Jehlicka P, Otterstätter T, Schlaich J. Generating OD matrices from mobile phone trajectories. *Transportation Research Board*, 89th Annual Meeting, 2010; paper no. 10-1972.
11. Schlaich J, Otterstätter T, Friedrich M. Generating trajectories from mobile phone data. *Transportation Research Board, 89th Annual Meeting*, 2010; paper no. 10-0374.
12. Lehnhoff N. Quality of automatic data collection with loop detectors. *Proc. of the 2nd International Symposium "Networks for Mobility"*, 2004; Stuttgart, Germany.

13. Höpfner M, Lemmer K, Ehrenpfordt I. Cellular data for traffic management—first results of a field test. *ITS Europe Conference*, 2007; paper no. 2407.
14. Thiessenhusen KU, Schäfer RP, Lang T. Traffic data from cell phones: a comparison with loops and probe vehicle data. *ITS World Congress*, 2006; paper no. 1550.
15. Bekhor S, Hirsh M, Nimre S, Feldman I. Identifying spatial and temporal congestion characteristics using passive mobile phone data. *Transportation Research Board, 87th Annual Meeting*, 2008; paper no. 08-1534.
16. Herrera J, Bayen A. Traffic flow reconstruction using mobile sensors and loop detector data. *Transportation Research Board, 87th Annual Meeting*, 2008; paper no. 08-1868.
17. Caceres N, Romero L, Benitez FG, del Castillo JM. Traffic flow estimates inferred from mobile phone networks. *12th World Conference on Transport Research*, 2010; paper no. 2720.
18. Caceres N. Mobility matrix estimate by using mobile phone data (in Spanish). Ph.D. Thesis Dissertation, Univ. of Seville, Spain. 2010; http://www.esi2.us.es/GT/docs/TesisNCS.pdf.
19. Sheffi Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1985.
20. Doblas FJ, Benitez FG. An approach to estimating and updating origin–destination matrices based upon traffic counts preserving the prior structure of a survey matrix. *Transportation Research Part B*, 2005; **39**(7):565–591.
21. Reklaitis GV, Ravindran A, Ragsdell KM. *Engineering Optimization: Methods and Applications*. John Wiley & Sons, Inc.: New York, 1983.
22. Wakefield T, McNally D, Bowler D, Mayne A. *Introduction to Mobile Communications: Technology, Services, Markets*. Auerbach Publications: Boca Raton, FL, 2007
23. Sallent O, Perez-Romero J, Umbert A, Agusti R. Teaching wireless telecommunications: experiencing the network to stimulate the learning process. *International Technology Education and Development Conference*. 2008; 1–5.