ESTIMATING LINK-DEPENDENT ORIGIN-DESTINATION MATRICES FROM SAMPLE TRAJECTORIES AND TRAFFIC COUNTS

G. Michau[1,2,*], *P. Borgnat*[1], *N. Pustelnik*[1], *P. Abry*[1]

A. Nantes[2], *E. Chung*[2]

[1] Laboratoire de Physique, CNRS
ENS de Lyon, Univ. de Lyon, France

[2] Smart Transport Research Centre
QUT, Brisbane, Australia

ABSTRACT

In transport networks, Origin-Destination matrices (ODM) are classically estimated from road traffic counts whereas recent technologies grant also access to sample car trajectories. One example is the deployment in cities of Bluetooth scanners that measure the trajectories of Bluetooth equipped cars. Exploiting such sample trajectory information, the classical ODM estimation problem is here extended into a link-dependent ODM (LODM) one. This much larger size estimation problem is formulated here in a variational form as an inverse problem. We develop a convex optimization resolution algorithm that incorporates network constraints. We study the result of the proposed algorithm on simulated network traffic.

Index Terms— OD matrices, urban transport, inverse problem, convex optimization, inference on network

1. INTRODUCTION

Estimating Origin-Destination matrices (ODM) is of utmost importance in Internet traffic management (see e.g., [1, 2, 3] and references therein) as well as in transport networks, on which we focus here. ODM consist of trip volumes between origin and destination regions in the transport network, over some period of time. In transport, ODM are traditionally assessed either through surveys, which are effective yet expansive and still prone to errors due to subjectivity [4, 5], or through traffic count (e.g., by magnetic loops at traffic lights) based estimations, a fairly intricate issue. However, the ubiquity of ICT (Information and Communications Technology) devices makes it now possible to identify and track individual vehicles at reasonable costs. An operating example is provided by the Bluetooth scanners operated in Brisbane by the Brisbane City Council [6], whose study constitute our long-term target. Previous works have shown the benefits of such technology to construct car trajectories from Bluetooth enabled devices [7, 8]. In nature, such data are more informative than traffic counts (which ignore trajectories). However, they do not account for the full traffic as only Bluetooth equipped cars trajectories are obtained. The objective of the present work is to propose a method for recovering the ODM, by combining Bluetooth sample trajectories data and the traffic count data. In addition, we extend ODM estimation to the challenging task of estimating Link-dependent Origin-Destination matrices (LODM), which is possible only by the availability of Bluetooth data. These LODM provide an additional and useful layer of information as compared to ODM: the actually chosen path(s) used to go from an origin to a destination. On road networks, the number of OD pairs varies like the square of the number of nodes. When further multiplied by the number of available links, the problem dimensionality grows large, thus calling for efficient procedures. In the following, we will formulate LODM estimation problem as an inverse problem.

Related Work. ODM estimation has already been studied as an inverse problem, notably using a prior ODM obtained from surveys or estimates [9, 10, 11]. ODM is usually estimated by minimizing an objective function based on its difference to the prior ODM, and on the difference between the measured traffic counts and the traffic counts derived from the *assignment* of the ODM. Several optimization methods were used: Entropy maximisation [12, 13, 14], Bayesian inference [15, 16], Generalised least squares estimator [17, 18]. Still, these methods relies on a modelling approach, mostly the *four-step model* [12] – assignment being the fourth step. They thus inherit the drawbacks of the assignment which, either using strong assumptions or being done with black-box traffic simulators, is a major source of uncertainty in ODM estimation. A second limitation is the need to either use prior ODM estimates, or models for generation and distribution of traffic (the first two steps). The final and major limitation is that these methods have not been extended to LODM estimation in transport.

Objectives. We will study how the estimation of LODM (and ODM as a by-product) can be formulated as an inference problem on the transport networks, and then be solved by convex optimization techniques [19, 20]. The principle is to make full use of the available data (sample trajectories and traffic counts), using constraints induced by the topology of the transport network only – thus removing the relevance of the assignment step. After introducing notations, the direct problem is defined in Section 2 and the estimation as an inverse problem is formulated in Section 3. In Section 4, the proposed LODM estimation performance are explored using traffic simulations.

Notations. The following notations are used for problem formulation: $\underline{A}$, $\underline{\underline{A}}$ and $\underline{\underline{\underline{A}}}$ respectively refer to vectors, matrices and tensors. The Hadamard product (element-wise product) of $\underline{\underline{\underline{A}}}$ and $\underline{\underline{\underline{B}}}$ is denoted $\underline{\underline{\underline{A}}} \circ \underline{\underline{\underline{B}}}$. For LODM, the two first dimensions are labelled i and j, referring to origin and destination; the third dimension, labeled l, stands for the links in the network. The symbol $\bullet$ is used to denote the dimension that does not contribute to a sum: e.g., the sum over first and third dimensions is written $\sum_{i,\bullet,l} \underline{\underline{\underline{A}}}$.

2. ESTIMATING LINK-DEPENDENT ODM

2.1. Problem Presentation

The road network is represented as a graph $G = (V, L)$. The finite set of nodes V models the major intersections of the road network; each node is also a possible origin or destination. Each edge of the set L is directed and corresponds to a direct itinerary (or road) linking two nodes (*i.e.,* not going through another node in V). The

*Funding for this project was provided by a grant from Région Rhône-Alpes and the french ANR-12-SOIN-0001-02 grant.

structure of the graph is then given by two matrices $\underline{\underline{I}}$ and $\underline{\underline{E}}$ called the *incidence* and *excidence* matrices respectively. These matrices describe the relations between the nodes and the edges: for every $(v, l) \in \{1, \ldots, |V|\} \times \{1, \ldots, |L|\}$,

$$\underline{\underline{I}}_{vl} = \begin{cases} +1 & \text{if the edge } l \text{ is arriving to the node } v, \\ 0 & \text{otherwise,} \end{cases}$$

$$\underline{\underline{E}}_{vl} = \begin{cases} +1 & \text{if the edge } l \text{ is starting from the node } v, \\ 0 & \text{otherwise.} \end{cases}$$

Note that in graph theory, it is the difference $\underline{\underline{I}} - \underline{\underline{E}}$ that would be named as "incidence matrix".

The information assumed available on this graph are $\underline{\underline{\underline{B}}}$ and $\underline{q}$. The tensor $\underline{\underline{\underline{B}}}$, of dimension $|V|^2 \times |L|$ gathers information from Bluetooth trajectories. Each trajectory adds a count of 1 into the elements of $\underline{\underline{\underline{B}}}$ corresponding to the OD and the links it uses (*i.e.,* edges in G). The vector $\underline{q}$ of dimension $|L|$, is the traffic flow measured on each edge. These counts can be obtained by magnetic loops. Such measurements are subject to count errors modelled here by a noise $\underline{\epsilon}$.

The quantity to be estimated is the count of trajectories for all cars over the OD and links, denoted $\underline{\underline{\underline{Q}}}$, which has to satisfy:

$$\underline{q} = \sum_{i,j,\bullet} \underline{\underline{\underline{Q}}} + \underline{\epsilon}. \tag{1}$$

The traffic flow is proportional to the flow of Bluetooth equipped cars. We thus introduce a supplementary variable $\underline{\underline{\underline{\alpha}}}$ such that $\underline{\underline{\underline{Q}}} = \underline{\underline{\underline{\alpha}}} \circ \underline{\underline{\underline{B}}}$, hence:

$$\underline{q} = \sum_{i,j,\bullet} \underline{\underline{\underline{\alpha}}} \circ \underline{\underline{\underline{B}}} + \underline{\epsilon}. \tag{2}$$

The challenge in this work is to estimate $\underline{\underline{\underline{\alpha}}}$ (and thus $\underline{\underline{\underline{Q}}}$) by using $\underline{q}$ and $\underline{\underline{\underline{B}}}$, the only available information. This is a highly underdetermined inverse problem, admitting thus a large number of different solutions. For constraining this set of solutions, a variational approach is proposed, that consists in solving

$$\widehat{\underline{\underline{\underline{\alpha}}}} \in \underset{\underline{\underline{\underline{\alpha}}}}{\operatorname{Argmin}} \sum_{k=1}^{K} F_k(\underline{\underline{\underline{\alpha}}}) \tag{3}$$

where the functions $F_k \colon \mathbb{R}^{|V| \times |V| \times |L|} \to\ ]-\infty, +\infty]$, for every $k \in \{1, \ldots, K\}$, model several network properties.

2.2. Consistency of Flows on Edges and Nodes

According to (2), a first function has to be designed in order to ensure the consistency of the solution with the measured flows on the links, thus a constraint on the edges. A usual choice for such a function is:

$$F_1(\underline{\underline{\underline{\alpha}}}) = \|\underline{q} - \sum_{i,j,\bullet} \underline{\underline{\underline{\alpha}}} \circ \underline{\underline{\underline{B}}}\|^2. \tag{4}$$

A second constraint comes from the balance of the flows on each node. It can be written using the classical ODM, denoted $\underline{\underline{T}}$:

$$\underline{\underline{T}} = \sum_{\bullet,\bullet,l} \underline{\underline{\underline{I}}} \circ \underline{\underline{\underline{Q}}} = \sum_{\bullet,\bullet,l} \underline{\underline{\underline{E}}} \circ \underline{\underline{\underline{Q}}} \tag{5}$$

where $\underline{\underline{\underline{I}}}$ and $\underline{\underline{\underline{E}}}$ are respectively the $|V|$-replication of the previous incidence and excidence matrices, defined as follows:

$$(\forall k \in V) \quad \underline{\underline{\underline{I}}}_{kjl} = \underline{\underline{I}}_{jl} \quad \text{and} \quad \underline{\underline{\underline{E}}}_{ikl} = \underline{\underline{E}}_{il} \tag{6}$$

The balance requires that, at every node v, the flow having for destination v, computed as $\underline{D}_v = \sum_i \underline{\underline{T}}_{i,v}$, minus the flow originating from v, computed as $\underline{O}_v = \sum_j \underline{\underline{T}}_{v,j}$, should equal the flow going through the node v. This is written as:

$$\underline{D}_v - \underline{O}_v = \sum_{\bullet,l} (\underline{\underline{I}} - \underline{\underline{E}})_{v,l}\ \underline{q}_l. \tag{7}$$

Using variable $\underline{\underline{\underline{\alpha}}}$ and data $\underline{\underline{\underline{B}}}$ and $\underline{q}$ with eq. (5), it reads as

$$\sum_{i,\bullet,l} \underline{\underline{\underline{I}}} \circ \underline{\underline{\underline{\alpha}}} \circ \underline{\underline{\underline{B}}} - \sum_{\bullet,j,l} \underline{\underline{\underline{E}}} \circ \underline{\underline{\underline{\alpha}}} \circ \underline{\underline{\underline{B}}} = (\underline{\underline{I}} - \underline{\underline{E}})\underline{q}. \tag{8}$$

The function resulting from this constraint is

$$F_2(\underline{\underline{\underline{\alpha}}}) = \| \sum_{i,\bullet,l} \underline{\underline{\underline{I}}} \circ \underline{\underline{\underline{B}}} \circ \underline{\underline{\underline{\alpha}}} - \sum_{\bullet,j,l} \underline{\underline{\underline{E}}} \circ \underline{\underline{\underline{B}}} \circ \underline{\underline{\underline{\alpha}}} - (\underline{\underline{I}} - \underline{\underline{E}})\underline{q}\|^2.$$

2.3. Low Variability

A sound assumption is that the proportion, called penetration rate, of Bluetooth amongst the vehicles in use on the road network is not varying much. Measures have shown that in Brisbane the Bluetooth penetration rate varies between 22 and 30%. Thus, solutions with limited variability of the Bluetooth penetration rate are more likely. The variable $\underline{\underline{\underline{\alpha}}}$ is the inverse of the penetration rate depending on links and OD. Its variability can be quantified through the following function, which is all the more interesting as being strongly convex, it insures the uniqueness of the solution :

$$F_3(\underline{\underline{\underline{\alpha}}}) = \sum_{i,j,l} (\underline{\underline{\underline{\alpha}}}_{ijl} - \alpha^o)^2 \tag{9}$$

where α^o is the a priori average sampling ratio, computed as

$$\alpha^o = \frac{\sum_l \underline{q}_l}{\sum_{i,j,l} \underline{\underline{\underline{B}}}_{ijl}} \tag{10}$$

2.4. Range

As the total flow is at least greater or equal to the flow of Bluetooth enabled vehicles, it is further imposed that $\underline{\underline{\underline{\alpha}}}$ belongs to the following convex constraint set:

$$C = \left\{ \underline{\underline{\underline{\alpha}}} = \left(\underline{\underline{\underline{\alpha}}}_{ijl}\right)_{(ijl) \in V \times V \times L} \in \mathbb{R}^{|V| \times |V| \times |L|} \mid \underline{\underline{\underline{\alpha}}}_{ijl} \geq 1 \right\} \tag{11}$$

In the criterion, this constraint appears through an indicator function $F_4(\underline{\underline{\underline{\alpha}}}) = \iota_C(\underline{\underline{\underline{\alpha}}})$, equals to 0 if $\underline{\underline{\underline{\alpha}}} \in C$ and $+\infty$ otherwise.

3. ALGORITHM

The criterion to obtain a relevant *transport* solution, designed using the topology of the networks and the data available, then reads:

$$\widehat{\underline{\underline{\underline{\alpha}}}} \in \underset{\underline{\underline{\underline{\alpha}}}}{\operatorname{Argmin}}\ \gamma_1 F_1(\underline{\underline{\underline{\alpha}}}) + \gamma_2 F_2(\underline{\underline{\underline{\alpha}}}) + \gamma_3 F_3(\underline{\underline{\underline{\alpha}}}) + \iota_C(\underline{\underline{\underline{\alpha}}}) \tag{12}$$

with $\gamma_k \geq 0$ for every $k \in \{1, \ldots, 3\}$, the weight of each constraint.

The functions involved in criterion (12) are convex, lower semicontinuous and proper. Moreover, $\gamma_1 F_1 + \gamma_2 F_2 + \gamma_3 F_3$ is differentiable with a β-Lipschitz gradient where the value of β depends on the norm of the matrices involved in each function. The function $F_4 = \iota_C$ is non-differentiable but it has a closed form expression for its projection [21]. To find $\widehat{\underline{\underline{\alpha}}}$, we used the forward-backward algorithm, adapted from [22, 20, 23], described as follow.

Algorithm 1 Forward-backward algorithm

Set $\tau = 1.99\beta^{-1}$
For $n = 0, 1, \ldots$ until convergence
$$\underline{\underline{\alpha}}^{[n+\frac{1}{2}]} = \underline{\underline{\alpha}}^{[n]} - \tau\left(\gamma_1 \nabla F_1 + \gamma_2 \nabla F_2 + \gamma_3 \nabla F_3\right)\left(\underline{\underline{\alpha}}^{[n]}\right)$$
$$\underline{\underline{\alpha}}^{[n+1]} = \max\left\{\underline{\underline{\alpha}}^{[n+\frac{1}{2}]}, 1\right\}$$

The initial condition $\underline{\underline{\alpha}}^{[0]}$ is always taken set to zero. According to [23], the sequence $(\alpha^{[n]})_{n\in\mathbb{N}}$ converges to $\widehat{\alpha}$. Moreover, its convergence rate has been described in [24]. In practice, we consider the convergence is achieved when the relative error between two iterates is such that $\frac{\|\alpha^{[n]}-\alpha^{[n-1]}\|^2}{\|\alpha^{[n]}\|^2} \leq 10^{-6}$.

4. SIMULATION AND STUDY OF THE RESULT

4.1. Simulation setting

A simulation is developed to produce ground truth data. First, a schematic road network is built by locating a set of nodes randomly on a grid. The nodes are first linked by a minimum spanning tree (computed by the Kruskal's algorithm [25]). Then, links are randomly added to connect the nodes with lower degree (sum of in and out edges) provided that the added links do not cross an existing one. This is stopped when the average total degree becomes 6 per node, a value consistent with that of real road networks (notably Brisbane transport network). In practice for this simulation, the number of nodes is $|V| = 50$. This choice is driven by the tractability of the experiment and the possibility of testing varied setups easily.

Then trajectories are drawn with random origin and destination with uniform law, and use the shortest path connecting the two. Their number is set proportional to the number of links (and thus to the number of nodes).

For future comparison to Brisbane's network, the number of vehicles is set to 500 per links. Measures show that few hundreds vehicles per link are detected on average by the scanners per 15 minutes (a duration deemed relevant in transport to estimate ODM). With a penetration rate of Bluetooth devices estimated at around 30%, 500 vehicles per link is a reasonable value.

The penetration rate is drawn for each OD pair from a Gaussian distribution of mean 30% and standard deviation of 10% (and truncated to be between 0 and 1). This choice accounts for the variability of the ownership distribution of Bluetooth devices (which is not known) from one node to another depending, as an example, on the wealth of the neighbourhoods of the node. Finally, for each trajectory, it is drawn with a probability equal to the penetration rate on its OD whether it is a sample Bluetooth trajectory, or not. This allows us to have data $\underline{\underline{B}}$ while the full set of trajectories gives $\underline{\underline{Q}}$ for ground truth. The traffic flow per link $\underline{q}$ is obtained from $\underline{\underline{Q}}$ by adding a noise ϵ drawn from a Gaussian $\mathcal{N}(0, 0.1)$ distribution.

Figure 1 displays a simulated transport network with traffic counts $\underline{q}$ and Bluetooth proportion $\underline{\underline{\alpha}}$ represented on the edges.

4.2. Examining the role of the functions in the criterion

First, we examine the effect of the various functions that are involved in the criterion of (12). For that, (12) is run several times with different weights γ_1, γ_2 and γ_3 representing the importance of each term. To study their role separately, we first use only one of the term together with the range constraint $\iota_C(\alpha)$, thus setting the other two weights to zero. The results are reported in Table 1, in the second block of lines. As expected, if only one term is active among F_1 to F_3, the other two depart significantly from the minimum value that can be achieved separately. Moreover, taking F_3 only is leads to a solution extremely close to α^o from (10) as shows a comparison of lines 1 and 4 of the Table 1 .

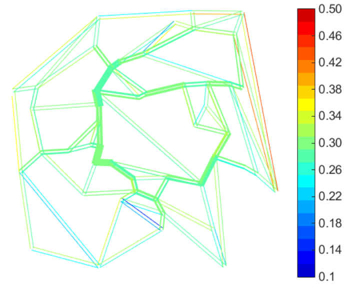


Fig. 1. A realisation of the simulation of a transport network, with volumes per link shown by the width of the edge (maximum width is for counts of 10,000) and Bluetooth proportion on the links coded in grey levels (see level bar)).

We then check whether the minimisation of one term is induced by the minimisation of the others and could be thus dispensed with. Clearly, using one term only leads to far from optimal values for the other two terms.

Using two terms without optimizing the weights (whose values are taken as 1), one sees in the second block in Table 1 that the third function always takes a large value as compared from the minimum achievable.

Finally, if using the three terms in the criterion, one obtains a fully coherent solution that respects the properties described in 2.2 and 2.3.

4.3. Optimal solution

The quality of the solutions can be checked by looking at the relative distance D_Q between the simulated link-dependent OD matrix $\underline{\underline{Q}}$ and the estimated one. This is computed as a ℓ_2 norm of the difference divided by the norm of the actual LODM. Others relative metrics are computed the same way: D_α and D_T for, respectively, the inverse penetration rates $\underline{\underline{\alpha}}$ and the OD matrix $\underline{\underline{T}}$.

From Table 1, these distances have minimal values when all the three terms F_1 to F_3 are active at the same time.

The optimal solution that can be recovered with (12) is sought by exploring the space of possible weights γ_k (varying them systematically). The optimal solution retained is when D_Q is minimum. The optimal combination of the weights, and the characteristics of the corresponding solution, are on the second to last line of Table 1. It turns out that, for the same weights, D_α and D_T are close to the smallest obtained value; this is not surprising and it tells that the formulation of the inverse problem for the variable $\underline{\underline{\alpha}}$ is sound for estimation of the LODM $\underline{\underline{Q}}$ or the ODM $\underline{\underline{T}}$. We have also checked that the results are not too sensitive to some variations of the γ_k around the best weights; for instance, having unity constant weights provides a result close to the optimal solution. Finally, let us emphasize that the value D_Q is close to its minimum for several solutions. It is stemming from a well-known issue of this estimation prob-

γ_1	γ_2	γ_3	F_1	F_2	F_3	D_Q	D_α	D_T
α^o			533	267	**0**	0.384	0.593	0.382
1	0	0	**0.072**	592	0.0034	0.617	0.729	0.610
0	**1**	0	10^5	$\mathbf{10^{-6}}$	0.0034	0.702	0.842	0.692
0	0	**1**	533	267	$\mathbf{10^{-11}}$	0.384	0.593	0.382
1	0	**1**	**14**	185	$\mathbf{10^{-5}}$	**0.382**	0.593	0.379
0	**1**	**1**	466	**2.7**	$\mathbf{10^{-5}}$	0.383	0.593	0.377
1	**1**	0	**0.08**	$\mathbf{10^{-6}}$	0.0034	0.615	0.729	0.605
1	**1**	**1**	**14**	**2.4**	$\mathbf{10^{-5}}$	**0.382**	**0.592**	**0.376**
0.6	0.4	0.8	**21**	**7.7**	$\mathbf{10^{-5}}$	**0.382**	**0.592**	**0.376**
$\widetilde{\alpha}$			10^{-26}	183	10^{-4}	**0.382**	**0.592**	0.378

Table 1. Comparison of the values of the objectives functions F_1 to F_3 and of the quality metrics D_Q, D_α and D_T, for different values of γ_k. When $\gamma_k = 0$, the associated function is not involved in the criterion yet still evaluated to understand how far from optimising each term the retrieved solution is. As F_k have high values when not involved in the algorithm, they all have to be used. The values in boldface are the one associated to the active terms in columns γ_1 to F_3, and the ones with the smallest values for columns D_Q, D_α and D_T. The second to last line shows the best solution, and the last line refers to solution (13) in 4.4.

lem: When not properly constraint, it is highly under-determined and therefore many solutions can perform with a given criteria (D_Q here). However, the solutions involving all the constraints F_k have also small values for these functions. This is an important results as minimising the function F_k does not require any ground truth. In addition, it discriminates between a set of acceptable solutions (with respect to D_Q) by ensuring that the properties of the retained solution are consistent with the measured data and with having continuity of the flows at the nodes. Others solutions do not have all these consistencies.

Figure 2 is a graphical representation of the results, aggregated on the edges of the network. It shows the counts and the estimated penetration rates for 2 solutions: (a) with constant penetration rate where $\underline{\underline{\alpha}} = \alpha^o$, and (b) the optimal solution from (12). It appears that the links with lowest traffic volumes are the hardest to recover: they have the highest relative errors both in terms of recovered volumes and of inverse penetration rates. For the solution of the inverse problem, the results are closer to the ground truth values than with solution (a) which is consistent with the obtained values of F_1 (bigger for solution (a)) and highlight the importance of these criteria.

4.4. Comparing to a closed-form solution using link information

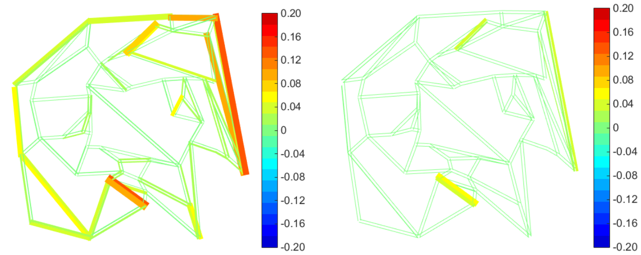
Interestingly, the solution obtained by the proposed LODM estimation procedure shows proximity with a closed-form expression that could be obtained differently. The Bluetooth trajectory samples can be used to count the number of Bluetooth enabled cars come over each link. A model $\widetilde{\alpha}$ for $\underline{\underline{\alpha}}$ is to assume a constant the penetration rate over each link:

$$\forall i \in V, \forall j \in V, \quad \widetilde{\alpha}_l = \frac{q_l}{\sum_{i,j,\bullet} \underline{\underline{B}}_{ijl}} \tag{13}$$

Using equation (13) leads to a closed-form estimate $\widetilde{\underline{\underline{Q}}}$ of $\underline{\underline{Q}}$:

$$\widetilde{\underline{\underline{Q}}}_{ijl} = \frac{\underline{\underline{B}}_{ijl}}{\sum_{i,j,\bullet} \underline{\underline{B}}_{ijl}} \cdot q_l \tag{14}$$

This estimate is in fact computed as a fraction of the flow on the link, where the fraction is set to the proportion of OD pairs that use link



(a) (b)

Fig. 2. Representation on the network of the error on recovered volumes per link (width) and on recovered Bluetooth penetration rate (color). (a) For the solution $\underline{\underline{\alpha}} = \alpha^o$. The maximum width correspond to a 35% error. (b) For the optimal solution of (12).

l in the Bluetooth sample trajectories. It appears that this solution is often close to the solution from the algorithm (around 2.5% difference for the example shown). This is due to the fact that $\widetilde{\alpha}$ written in (13) appears to lead already to low value for F_1 and F_3 (cf. last line of table1). The difference with the optimal solution for (12) is coming from the term F_2 which is large for solution $\widetilde{\alpha}$.

It shows *a posteriori* that such deterministic solution is relevant, regarding the minimisation of (12). However, it is not consistent with our knowledge of the problem to assume that the penetration rate is constant for each link independently of the OD (as we simulated with a penetration rate depending on the OD). Moreover, the variational approach is better with respect to all the consistency requirements, especially on the nodes, emphasising once more the importance of the functions F_k. Also, for future work, we intend to add other constraints and regularizing priors that would relate itineraries information to links or involving smoothness in the network. The closed-form solution is not malleable like that.

5. CONCLUSION

For transport networks, an approach to estimate origin-destination matrices in a link-dependent way is developed. It uses as input the traffic counts on the edges of the networks and a set of sample trajectories like the ones that can be obtained through the use of Bluetooth scanners. The first result is that a simple forward-backward algorithm allows us to recover a good enough solution when formulating the inverse problem as (12); a second result is that all the terms in the criterion have a role to play and are useful to discriminate, amongst a set of solutions with correct quality metric, one that is better consistent with the measured data and the constraints on the networks. These results are an encouragement to implement further constraints (*e.g.,* , continuity of the flow along the paths) and it would be relevant to introduce regularizing priors in the criterion (possibly using then a different suitable algorithm). For instance we could postulate that the inverse penetration rate is varying smoothly for its variables (i, j) describing OD on regions adjacent in the network, and one way to take that into account would be by introducing, as defined for instance in the recent work [26], a norm of smoothness of $\underline{\underline{\alpha}}$ along the graph, be it a ℓ_2 norm or a total variation one to allow for some jumps. This is under current inspection.

6. REFERENCES

[1] M Coates, A. O Hero, R Nowak, and B Yu, "Internet tomography," *Signal Processing Magazine, IEEE*, vol. 19, no. 3, pp. 47–65, 2002.

[2] A Girard, B Sansò, and F Vazquez-Abad, *Performance Evaluation and Planning Methods for the Next Generation Internet*, vol. 6, Springer, 2006.

[3] M Roughan, Y Zhang, W Willinger, and L Qiu, "Spatio-Temporal compressive sensing and internet traffic matrices (extended version)," *Networking, IEEE/ACM Transactions on*, vol. 20, pp. 662–676, 2012.

[4] S Fujii and T Gärling, "Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis," *Transportation Research Part A: Policy and Practice*, vol. 37, pp. 389–402, 2003.

[5] T Veitch, M Paech, and J Eaton, "What's missing from australian household travel surveys? – Off-Peak travel!," in *Australasian Transport Research Forum 2013*, Brisbane, Oct. 2013.

[6] A Bhaskar and E Chung, "Fundamental understanding on the use of bluetooth scanner as a complementary transport data," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 42–72, Dec. 2013.

[7] G Michau, A Nantes, E Chung, P Abry, and P Borgnat, "Retrieving dynamic Origin-Destination matrices from bluetooth data," in *Transportation Research Board, 93rd Annual Meeting*, Washington DC, Jan. 2014.

[8] G Michau, A Nantes, E Chung, P Abry, and P Borgnat, "Routes recovery from wireless detectors data: The case of bluetooth detectors in brisbane," in *Submitted for the 94rd Annual Meeting of the Transportation Research Board*, Washington DC, Jan. 2015.

[9] D. P Watling and M. J Maher, "A statistical procedure for estimating a mean Origin-Destination matrix from a partial registration plate survey," *Transportation Research Part B: Methodological*, vol. 26, pp. 171–193, 1992, 3.

[10] N. J Van Der Zijpp, "Dynamic Origin-Destination matrix estimation from traffic counts and automated vehicle identification data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1607, pp. 87–94, 1997.

[11] Y Asakura, E Hato, and M Kashiwadani, "Origin-Destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network," *Transportation*, vol. 27, no. 4, pp. 419–438, 2000.

[12] L. G Willumsen, "Estimation of an OD matrix from traffic Counts–A review," *Institute of Transport Studies, Universities of Leed*, 1978.

[13] A Wilson, "Entropy in urban and regional modelling: Retrospect and prospect.," *Geographical Analysis*, vol. 42, no. 4, pp. 364–394, 2010.

[14] W. H. K Lam and H. P Lo, "Estimation of origin-destination matrix from traffic counts: a comparison of entropy maximizing and information minimizing models," *Transportation Planning and Technology*, vol. 16, no. 2, pp. 85–104, 1991.

[15] M Maher, "Inferences on trip matrices from observations on link volumes: A bayesian statistical approach," *Transportation Research Part B: Methodological*, vol. 17, no. 6, pp. 435–447, Dec. 1983.

[16] 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, Jan. 2014.

[17] E Cascetta, "Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator," *Transportation Research Part B*, vol. 18, no. 4-5, pp. 289–299, 1984.

[18] M. G Bell, "The estimation of Origin-Destination matrices by constrained generalised least squares," *Transportation Research Part B: Methodological*, vol. 25, pp. 13–22, 1991.

[19] S Boyd and L Vandenberghe, *Convex Optimization*, Cambridge university press, 2009.

[20] H. H Bauschke and P. L Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics Ser. Springer, Apr. 2011.

[21] S Theodoridis, K Slavakis, and I Yamada, "Adaptive learning in a world of projections," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 97–123, 2011.

[22] P. L Combettes and V. R Wajs, "Signal recovery by proximal Forward-Backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[23] P. L Combettes and J. C Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, p. 185–212. Springer, 2011.

[24] G. H Chen and R. T Rockafellar, "Convergence rates in Forward–Backward splitting," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 24, May 1997.

[25] T. H Cormen, C. E Leiserson, R. L Rivest, and C Stein, "The algorithms of kruskal and prim," *Introduction to Algorithms*, p. 631–638, 2009.

[26] V Kalofolias, X Bresson, M Bronstein, and P Vandergheynst, "Matrix completion on graphs," *arXiv preprint arXiv:1408.1717*, vol. abs/1408.1717, 2014.