Path Flow and Trip Matrix Estimation Using Link Flow Density

Louis de Grange and Felipe González
School of Industrial Engineering, Diego Portales University, Santiago de Chile.
Phone: 56-2-22130469 ; e-mail: **louis.degrange@udp.cl , felipe.gonzalezr@udp.cl**

Shlomo Bekhor
Faculty of Civil and Environmental Engineering Technion - Israel Institute of Technology Haifa 32000, Israel.
Phone: 972-4-8292460; e-mail: **sbekhor@technion.ac.il**

ABSTRACT

A macroscopic model is presented that simultaneously estimates route flows and trip matrices for congested road networks using data on link densities instead of link flows. The advantage of this approach is that it avoids errors that may occur in the flow-cost relationship for each link when congestion is heavy. The proposed methodology allows the estimation of trip demand matrices and route flows, using an image of the network such as an aerial photograph identifying the number of vehicles on each link. The model is formulated as a maximum entropy optimization problem, subject to linear constraints given by vehicle densities on the links. The model is validated using analytic examples and traffic microsimulations. The results demonstrate the superiority of the link-density approach over the traditional flow-based method.

Keywords: trip matrix estimation; path flow; link flow density; maximum entropy, macroscopic traffic model.

1. INTRODUCTION

This article proposes a macroscopic-level static equilibrium model of a road network that simultaneously generates route flow and trip matrix estimates under conditions of heavy congestion. In addition to link cost data, the model uses data on each network link's vehicle density. This latter information is obtained directly for a given instant from a static image of the network (e.g., a photograph from which vehicle counts can be made for each link) on the assumption the fundamental traffic equation relating speed, flow and density is known for each link. The main advantage of this use of density data is that it captures the relationship between flow and cost (trip time or speed) on each link more accurately. Classical models for estimating trip matrices from traffic flow counts under congestion (see references, Section 2) normally assume this relationship is increasing and convex. In reality, however, it behaves quite differently in highly congested situations, as is apparent from the fundamental traffic equation. This implies that using density directly instead of flow may enable us to generate better estimates of link costs, a conjecture we will attempt to demonstrate.

Another advantage of our proposed approach is that it obviates the need for counting vehicle flows (veh/h) on each network link. All we need to know is the number of vehicles on each link at a given instant—in other words, the density (veh/km)—and the relationship between density, speed and flow for each link. The information for the latter is easily obtained with "loop detectors" installed in various links in the road network. Our model is based on the specification of a maximum entropy optimization problem subject to linear constraints defined by the link flow data for a situation in which each link's density and cost are in equilibrium. The model is validated using microsimulations of traffic on a small network and obtains good results, the proposed formulation proving capable of reproducing the simulation data for the model variables. For comparison purposes, we also estimate the model using link flow data, the traditional approach in the specialized literature. As will be seen, our proposed approach using link densities instead of link flows delivers better results. The remainder of this paper is divided into six sections. Section 2 presents a brief literature survey dealing with the principal aspects of our model; Section 3 introduces our approach for estimating route flows using link densities; Section 4 presents a microsimulation for validating our methodology; Section 5 compares the results of the simulation with those obtained using the classic link-flow approach; Section 6 develops an extension for the simultaneous of route flows and trip matrices; and finally, Section 7 sets out our main conclusions.

2. LITERATURE SURVEY

This literature survey focuses on three areas of particular relevance to our study: the estimation of trip demand matrices from traffic flow counts, the estimation of route flows, and the flow-speed-density relationship at the macroscopic level. The first two areas are closely related and will be treated together in Section 2.1 while the third is taken up in Section 2.2.

2.1 Estimation of trip matrices and route flows

There are two traditional approaches to estimate origin-destination matrices based on traffic counts (a good survey of these methods may be found in Abrahamsson (1998) and Bera and Rao (2011)). The first approach involves maximum entropy or minimum information models, which use a given information criterion (e.g., maximum entropy, as per Wilson (1970)) to estimate the most likely trip matrix consistent with the observed flows (Van Zuylen and Willumsen, 1980). This class of formulations consists of mathematical programming models that attempt to determine the most suitable trip matrix consistent with the information contained in the network link traffic volume data while maximizing entropy or minimizing information from a prior trip matrix using an objective function. Since this approach does not consider the effect of congestion and trip times on the estimated matrix, there will be a linear proportionality between matrix trips and flows, which may not be correct. For this reason, the optimization problem is usually formulated subject to road network flow conservation constraints. Under this paradigm, Lam and Lo (1991) compare the maximum entropy and minimum information methods. Regardless of the modelling criterion used, models for estimating trip matrices from link flow information generally employ the following equation relating interzonal trips and link flows:

$$f_a = \sum_{ij} T_{ij} p_{ij}^a, \quad 0 \le p_{ij}^a \le 1 \tag{1}$$

where the variable f_a is the flow on link a and must equal the sum of all trips between the different origin-destination pairs (i, j) using that link, variable T_{ij} represents the trips between the zone containing origin i and the zone containing origin j, and variable p_{ij}^a is the proportion of trips between pair (i, j) that uses a. The value of p_{ij}^a can be estimated from various traffic assignment criteria (Farhangian and LeBlanc, 1982; Iida et al., 1994).

If the observed flow pattern is in user equilibrium, the author's extended entropy model (Fisk, 1988) will generate the same solution as a combined distribution and assignment model. This is demonstrated in Fisk (1989), where the author also concludes that with network equilibrium-based formulations, maximum entropy and combined distribution and assignment approaches can be expected to produce the same results for congested networks as long as the observed link traffic volumes are equilibrium flow patterns. A paper by Gothe et al. (1989) complements this work while Maher et al. (2001) offers an extension to stochastic user equilibrium instead of the deterministic equilibrium proposed by Fisk (1989). Since our focus is macroscopic, dynamic assignment models (Mahmassani, 2001; Ziliaskopoulos and Peeta, 2001; Watling and Hazelton, 2003b; Zhang and Nie, 2005) are beyond the scope of this paper.

The second traditional approach to the estimation of trip matrices uses statistical techniques such as generalized least squares or maximum likelihood. These techniques also estimate trip matrices from link flow data. Maximum likelihood seeks to identify the trip distribution most consistent with the observed flows for a given density function while the generalized least squares (GLS) approach attempts to estimate a matrix whose road network assignment minimizes the differences between observed and modelled flows, with each link having a different weight. The maximum likelihood method is applied by Spiess (1987) to small networks while generalized least squares is utilized by Cascetta (1984) and Bell (1991), but all under a proportional assignment approach that does not consider flow-delay relations in network links. Maher (1983) proposes a Bayesian approach. The first revisions of these approaches were published by Nguyen (1984) and Cascetta and Nguyen (1988). The latter work focuses on statistical inference techniques and the former mainly on minimum information approximations or maximum entropy. Cascetta and Nguyen provide no particular approach of their own but do offer a general structure for addressing the O-D matrix estimation problem using traffic counts. They also cover public as well as private transport while Nguyen (1984) only considers private vehicles (i.e., cars). Yang et al. (1992) and Yang (1995) extend the generalized least squares model by incorporating equilibrium traffic assignment into the model.

2.2 The flow-speed-density relationship (macroscopic traffic models)

Traffic flows are not normally uniform but rather vary both in space and time and thus are difficult to describe. Nevertheless, their behaviour has traditionally been explained by just three variables: flow volume, speed and density (or concentration).

Mean speed (v) is defined as the average speed of vehicles passing a given point or location; flow (or volume) f is defined as the number of vehicles crossing a given section or segment of road or highway within a given period of time, and density or concentration D is defined as the number of vehicles occupying a given section or segment. For multi-lane roads or highways which are generally characterized by mixed-traffic conditions, flow is expressed as a vehicles per unit of time per number of lanes (Arasan and Krishnamurthy, 2008). Density is also expressed in terms of the total width of the highway.

Macroscopic traffic models are generally modelled at the road level. Geroliminis and Daganzo (2008) have shown that traffic in large urban regions can be modelled dynamically at an aggregate level, if the neighbourhoods are uniformly congested. Using data from San Francisco (United States) and Yokohama (Japan), they have shown that urban neighborhoods approximately exhibit a ''macroscopic fundamental diagram'' that relates the number of vehicles (accumulation) in the neighbourhood to the neighbourhood's average speed (or flow). May and Keller (1967) proposed the following traffic flow speed-density relation:

$$v = v_{\min} + \left(v_{\max} - v_{\min}\right)\left(1 - \left(D/D_j\right)^{\alpha}\right)^{\gamma} \tag{2}$$

where $v_{\min}$ is the minimum speed, $v_{\max}$ is the maximum speed and α, γ are parameters to be calibrated. Ardekani and Ghandehari (2008) proposed a Greenberg-type model modified to include a minimum density D_0, thus assuming there are always some vehicles on the highway. In this formulation,

$$v = v_c \cdot \ln\left[\frac{D_j + D_0}{D + D_0}\right] \tag{3}$$

In the final analysis, which formulation to use will depend mainly on how well each one fits the available data. As will be described in the next section, the present authors opted for the model designed by May and Keller (1967).

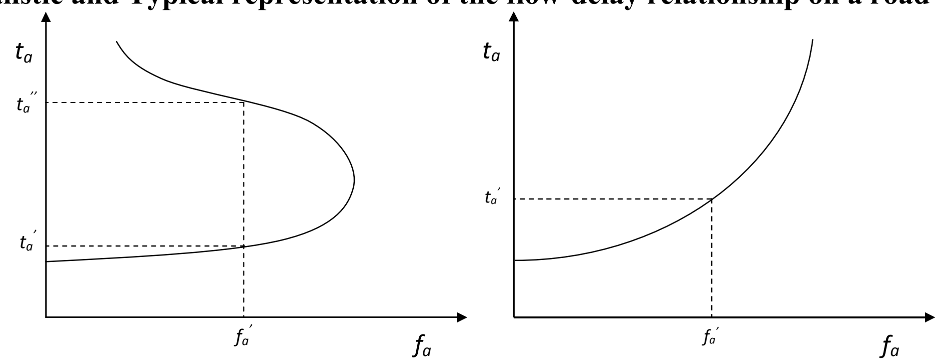
3. ROUTE FLOW ESTIMATION MODEL

Assume that we know, or can satisfactorily estimate, the vehicle density (d_a) on a given link $a \in A$ where A is the set of road network links. Density is defined as the number of vehicles on a link at a given instant. With a photograph of a city's road network we can then estimate, for each street, both its length and the number of vehicles on it and therefore the density of its traffic flow (number of vehicles per kilometre) at the moment the photograph was taken.

Alternatively, the density could be estimated with occupancy data from loop detectors. Thus, upon identifying the number of vehicles n_a on link a and the length of the link l_a, we estimate the density as $d_a = \left(n_a / l_a\right)$. From the fundamental traffic equation we know that the flow on link a $\left(f_a\right)$ is related to its speed $\left(v_a\right)$ and density by $f_a = v_a \cdot d_a$. This implies that upon estimating the density of link a, which we denote d_a^*, we can then directly estimate the corresponding speed $v_a^* = v_a\left(d_a^*\right)$ and therefore also the flow $f_a^* = v_a^* \cdot d_a^*$.

An advantage of using the expression $v_a^* = v_a\left(d_a^*\right)$ is that it allows us to define the cost of the link as a function of the density rather than the flow, thus avoiding the typical problem illustrated in Figure 1 where the presence of congestion generates two alternative trip times or costs for the same flow level and the usual relationship posited in matrix estimation models incorporating congestion. Clearly, this limits the usefulness of such models for high-congestion situations.

Figure 1
Realistic and Typical representation of the flow-delay relationship on a road link



If we define the speed on link a as $v_a = \frac{l_a}{t_a}$, where t_a is the trip cost (or time), then cost for the link can be estimated as $t_a^* = \frac{l_a}{v_a^*}, \forall a \in A$. With the flow estimates $\left(f_a^*\right)$ and a definition based on some criterion (Bekhor et al., 2006; Prato, 2009) of the possible set of alternative routes (choice-set) used for the trips between the O-D pairs of the network (R_w), we can now solve the following entropy optimization problem (Rossi et al, 1989; Janson, 1993, Bar-Gera, 2006):

$$\min \;\; Z_1 = \sum_{r\in R_w}\sum_{w} h_w^r\left(\ln h_w^r - 1\right)$$
$$s.t.: \sum_{r\in R_w} h_w^r = T_w, \;\; \forall w \quad \left(\mu_w\right) \qquad \sum_{\substack{r\in R_w \\ r\supseteq a}} h_w^r = f_a^*, \;\; \forall a \quad \left(\beta_a\right) \qquad h_w^r \geq 0 \tag{4}$$

where h_w^r is the flow on route r between O-D pair w and T_w is the fixed and exogenous demand for trips between the pair. In this first model we are assuming the trip matrix is known and the trip times are exogenous; later, in our second model (Section 6), we will relax this assumption. The optimality conditions are:

$$h_w^r = T_w \frac{\exp\left(\sum_{a\subseteq R_w} \beta_a\right)}{\sum_{r\in R_w} \exp\left(\sum_{a\subseteq R_w} \beta_a\right)}, \quad \forall r, w \tag{5}$$

This expression is a logit model that allows us to obtain route flows consistent with the observed flows. We may also assume that the route choice is given by a logit multinomial model (Hazelton, 2003b) with the following form:

$$h_w^r = T_w \frac{\exp\left(\theta C_w^r\right)}{\sum_{r \in R_w} \exp\left(\theta C_w^r\right)}, \quad \forall r, w \tag{6}$$

where C_w^r is the composite cost of route r between pair w. This cost is given by $\theta C_w^r = \sum_k \theta_k X_{k,w}^r$, the sum of a series of attributes $X_{k,w}^r$ of the route (trip time, tolls, fuel, etc.). Since the value of the k-th attribute of the route is the sum of the individual k-th attribute values of the links making up the route, we define $X_{k,w}^r = \sum_a \delta_a^{r,w} x_{a,k}$, where $\delta_a^{r,w}$ is equal to 1 if link a belongs to the route and 0 otherwise, and $x_{a,k}$ represents the value of the link's k-th attribute. Since (5) is logit multinomial, then, if we assume the set of route alternatives between a given O-D pair (R_w) is known, we may posit the following equivalence between (5) and (6):

$$\frac{\exp\left(\sum_{a \subseteq R_w} \beta_a\right)}{\sum_{r \in R_w} \exp\left(\sum_{a \subseteq R_w} \beta_a\right)} = \frac{\exp\left(\theta C_w^r\right)}{\sum_{r \in R_w} \exp\left(\theta C_w^r\right)}, \quad \forall r, w \tag{7}$$

$$\frac{\exp\left(\sum_{a \subseteq R_w} \beta_a\right)}{\sum_{r \in R_w} \exp\left(\sum_{a \subseteq R_w} \beta_a\right)} = \frac{\exp\left(\sum_k \theta_k \sum_a \delta_a^{r,w} x_{a,k}\right)}{\sum_{r' \in R_w} \exp\left(\sum_k \theta_k \sum_a \delta_a^{r',w} x_{a,k}\right)}, \quad \forall r, w \tag{8}$$

The consistency between (7) and (8) is more clearly illustrated by the following analytic example based on the road network with demand parameters shown in Figure 2. The routes and assigned trips for this example are described in Table 1.

We begin by assuming the following speed-density function (May and Keller, 1967) which we will later calibrate by microsimulation (see Section 4).

$$v_i = v_{\min} + \left(v_{\max} - v_{\min}\right)\left(1 - \left(D_i / D_j\right)^{\alpha}\right)^{\gamma} \tag{9}$$

where

v_t : average speed in period i.

$v_{\min}$: minimum speed; for calibration purposes, $V_{\min} = 0\, km/hr$.

$v_{\max}$: maximum speed; for calibration purposes, $V_{\max} = 70\, km/hr$.

D_t : average density in period *i*.

D_j : Jam density, that is, extreme traffic density associated with completely stopped traffic flow; for calibration purposes, $D_j = 150\, veh/km$.

α, γ : parameters to be calibrated.

The parameter values are $\alpha = 1.17; \gamma = 1.28$ (details in Section 4).

Figure 2
Modelled road network example

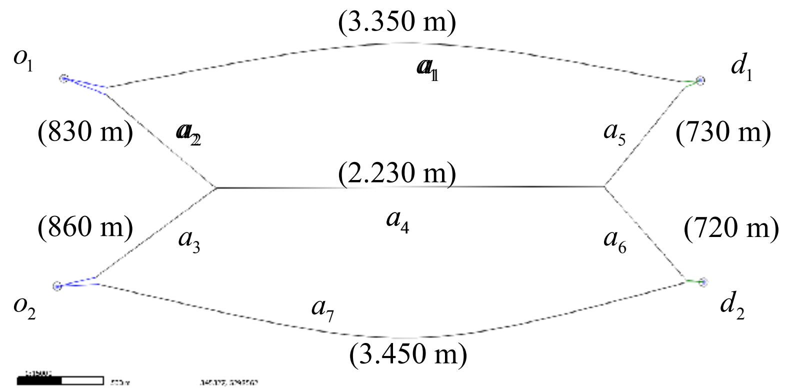


Table 1
O-D trips and routes for modelled example in Figure 3

O-D pair *w*	Trips	Route link sequences
$o_1 - d_1$	1,500	R1: a_1
$o_1 - d_1$		R2: $a_2 - a_4 - a_5$
$o_2 - d_2$	3,000	R3: $a_3 - a_4 - a_6$
$o_2 - d_2$		R4: a_7

Substituting these values into (9), we have

$$v_i = 0 + (70 - 0)\left(1 - \left(\frac{d_i}{150}\right)^{1.17}\right)^{1.28} = 70\left(1 - \left(\frac{d_i}{150}\right)^{1.17}\right)^{1.28} \qquad (10)$$

Also, the population θ parameter in the route choice model is defined as $\theta = -0.33$.

The link flow densities d_i must be consistent in each period *i* with speed-density function (10) but also with the route flows h_w^r and the link flows f_a. The densities that ensure these flow and speed consistencies are set out in Table 2 for the various links in the network (note that this is a one-period analysis).

Table 2
Equilibrium variable results

Variable	Arco						
	a_1	a_2	a_3	a_4	a_5	a_6	a_7
Density (veh/km)	13.27	9.68	22.77	38.70	9.68	22.77	28.54
Speed (km/hr)	64.66	66.29	60.09	51.94	66.29	60.09	57.19
Flow (veh/hr)	858	642	1368	2010	642	1368	1632
Length (km)	3.35	0.83	0.86	2.23	0.73	0.72	3.45
Trip time (min)	3.11	0.75	0.86	2.58	0.66	0.71	3.62

The traffic flow described in Table 2 is incorporated into the constraints in the model as set out in (4). Solving the model produces the following Lagrange multipliers:

Table 3
Lagrange multipliers, entropy optimization model (4)

Link	a_1	a_2	a_3	a_4	a_5	a_6	a_7
β_a multipliers	6.755	6.465	7.221	0.000	0.000	0.000	7.398

For the four routes we thus obtain

Table 4
Route-level multipliers and trip times

Route link sequence	$\sum_{a\subseteq P_w} \beta_a$	$\sum_{a\subseteq P_w} t_a^*$
R1: a_1	6.755	3.112
R2: $a_2 - a_4 - a_5$	6.465	3.983
R3: $a_3 - a_4 - a_6$	7.221	4.148
R4: a_4	7.398	3.619

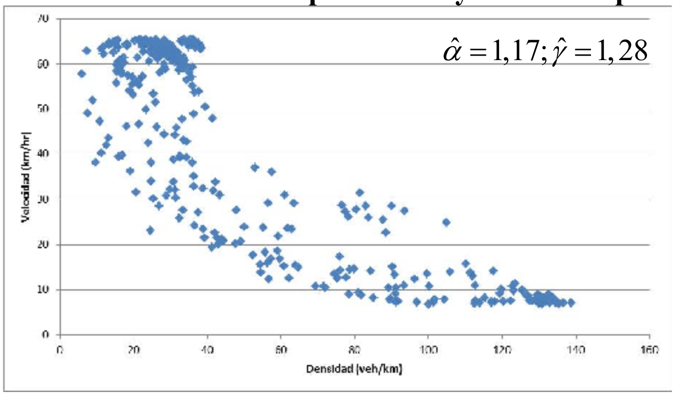
Given the values in Table 1 and the fixed parameter value $\theta = -0.33$, the equality conditions defined in (7) are satisfied. The validity of the proposed model and the equivalence posited in (7) are thus confirmed. The above example was, of course, constructed to reproduce the proposed model. In what follows we present a validation using traffic microsimulation in order to generate variability in the estimates.

4. VALIDATION OF THE PROPOSED MODEL USING MICROSIMULATION

The modelled example network in Figure 2 was replicated in the Aimsun traffic microsimulator. The microsimulator was used for two specific purposes: first, to calibrate the relation between density and speed defined in (9); and second, to generate the densities for different simulated periods that were then used to implement the methodology described in Section 3 for estimating the θ parameters (one for each period).

For calibrating (9) the microsimulations created a network of 7 links and 4 centroids, equivalent to the example network in Figure 3 but now with a range of demand levels (from 100 to 4,000 vehicles per hour, uniformly distributed and increasing in units of 300 with each successive simulation) in order to obtain greater variability in the results. Ten one-hour replications were performed for each demand level and every 30 minutes the average speed and density data for each network link were stored. Thus, the total number of simulations executed was 10×13×2 = 260. The α and γ parameters in (9) were estimated using non-linear least squares. A scatter graph showing the dispersion between speed and density together with the calibrated parameter values is shown in Figure 4.

Figure 4
Microsimulation of speed-density relationship



To generate the input data for the maximum entropy optimization problem (4), forty 30-minute network replications were executed. Two O-D pairs were simulated with the following trip distribution: 1,500 trips between $o_1 - d_1$ and 3,000 trips between $o_2 - d_2$. The model used in the simulations to assign the trips to the network was a multinomial route choice formulation with the following specification:

$$h_w^r = T_w \frac{\exp\left(\theta t_w^r\right)}{\sum_{r \in R_w} \exp\left(\theta t_w^r\right)}, \quad \forall r, w \tag{11}$$

where t_w^r is trip time (in minutes) for route *r* between O-D pair *w* and $\theta = -0.33$. The routes and assigned trips were the same as those defined in Table 1.

The 40 simulations generated 40 average density values $\left(d_{a,i}\right)$, where *a* indicates the link and *i* the simulation period, for each of the 7 network links. The calibrated speed-density function (12) was then utilized to find the average speed $\left(v_{a,i}\right)$ corresponding to each density value for each link. The estimated link flows, which were incorporated into (4) as constraints, were obtained from the formula $f_{a,i} = d_{a,i} v_{a,i}$ and the trip times for the links were estimated as $t_{a,i} = l_{a,i} / v_{a,i}$. Using these link flow data for each microsimulation period, the parameters $\beta_{a,i}$ were then obtained by solving the optimization problem (4) forty times with different values of $f_{a,i}$. One possible criterion for estimating the θ parameters in (7) is as follows:

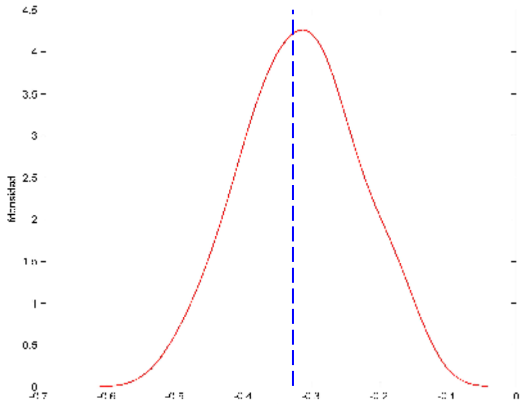
$$\min_{\theta_k} \sum_i \sum_{r,w} \left[\frac{\exp\left(\sum_{a\subseteq R_w} \beta_{a,i}\right)}{\sum_{r\in R_w} \exp\left(\sum_{a\subseteq R_w} \beta_{a,i}\right)} - \frac{\exp\left(\sum_k \theta_k \sum_a \delta_a^{r,w} x_{a,i,k}\right)}{\sum_{r'\in R_w} \exp\left(\sum_k \theta_k \sum_a \delta_a^{r',w} x_{a,i,k}\right)} \right]^2 \tag{13}$$

For our example, the estimates of θ were obtained by solving the following problem:

$$\min_{\theta_k} \sum_i \sum_{r,w} \left[\frac{\exp\left(\sum_{a\subseteq R_w} \beta_{a,i}\right)}{\sum_{r\in R_w} \exp\left(\sum_{a\subseteq R_w} \beta_{a,i}\right)} - \frac{\exp\left(\theta \sum_a \delta_a^{r,w} t_a\right)}{\sum_{r'\in R_w} \exp\left(\theta \sum_a \delta_a^{r',w} t_a\right)} \right]^2 \tag{14}$$

The solution was derived using the data generated with the 40 density values obtained every 30 minutes from the simulated network ($i = 1, 2, \ldots, 40$). The resulting distribution for the θ parameter is shown in Figure 5.

Figure 5
Probability density of the θ parameter estimates



As can be seen in the figure, the mean of the forty θ estimates appears very close to the defined population parameter value $\theta = -0.33$, its exact value being -0.317.

To determine whether the difference between these two is significant, we performed a hypothesis test using the *t*-statistic in which the null hypothesis was that the values were significantly different. Thus,

$$\left.\begin{aligned} H_0 &: \theta = -0.33 \\ H_1 &: \theta \neq -0.33 \end{aligned}\right\} \rightarrow \left| \frac{-0.33 - (-0.317)}{0.082} \right| = 0.161 < t_{39;0.95} = 2.02 \tag{15}$$

The null hypothesis was therefore rejected, meaning that the estimator of θ cannot be said to be different from the population value.

5. LINK FLOWS VERSUS LINK DENSITIES: A COMPARATIVE ANALYSIS

As noted in Section 2, estimates of both trip matrices and route flows are traditionally based on link flow rather than link density data. The two data types often lead to different results, mainly in conditions of high congestion (see Figure 1). In this section we compare and contrast the results generated using our density approach in the previous section with a new set of estimates based on the classic flow data approach. Using the data generated by the traffic microsimulations we can obtain the average flows and trip times for the various links in our simulated example network (Figure 2) and then estimate the typical flow-delay functions employed in deterministic static assignment models (BPR, exponential, conical etc.). For the example network, the flow-delay functions that provide the best fit are the exponential type:

$$t_{a,i} = t_a^0 \exp\left(\lambda_a \left(f_{a,i}/k_a\right)\right) \tag{16}$$

where $t_{a,i}$ is the average travel time on link a in simulation period i, t_a^0 is the free flow time on the link, $f_{a,i}$ is the link's flow and k_a its capacity. Using our microsimulation values for $t_{a,i}$, $f_{a,i}$ and k_a (half-hourly averages), we take the natural log of both sides of (16) and estimate the λ_a parameter directly using ordinary least squares. The resulting λ_a estimates for each link and their respective *t*-test statistical significance values are summarized in Table 5.

Table 5
Flow-delay function parameter estimates

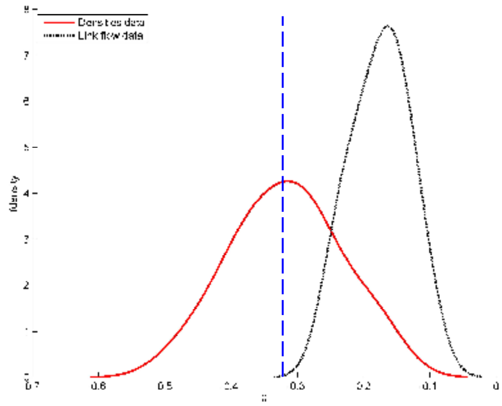
Parameter	Estimated value	*t*-test
λ_1 , λ_7	0.257	12.99
λ_2 , λ_3 , λ_5 , λ_6	0.289	13.31
λ_4	0.268	13.52

Having estimated the flow-delay functions, we use them to obtain the theoretical relationship between flow and trip time. Assuming we know the link flow levels, (16) allows us to estimate the corresponding trip time (or cost) which we then introduce into (14) in order to estimate the θ parameter. The distribution of the θ estimates derived from the link flow data (instead of densities) is shown in Figure 6.

As can be seen in the Figure 6, the mean value of these estimates is 0.175. That this is significantly less in statistical terms than the population value of 0.33 is confirmed by the following hypothesis test:

$$\left.\begin{array}{l} H_0 : \theta = -0.33 \\ H_1 : \theta \neq -0.33 \end{array}\right\} \rightarrow \left|\frac{-0.33-(-0.175)}{0.043}\right| = 11.78 > t_{39;0.95} = 2.02 \tag{17}$$

Figure 6
Probability density of the θ parameter estimates



In Table 6 these model estimates, generated directly using link flows (Figure 6), are compared to those obtained in the previous section using flow densities (Figure 5).

Table 6
Comparison of route choice model estimates: densities vs flows

Estimator	Using densities	Using flows
θ parameter	0.317	0.175
MSE	1.94	2.21
MAE	17.72	20.80

The MSE statistic in Table 6 is the mean square error, defined as follows:

$$MSE = \sum_{i,r,w} \frac{\left(h_{r,i}^{w} - \hat{h}_{r,i}^{w}\right)^2}{n} \quad (18)$$

where $h_{r,i}^{w}$ is the flow for the period i microsimulation on route r between O-D pair w and $\hat{h}_{r,i}^{w}$ is the corresponding flow estimated with the route choice model (using link density and link flow input data). The MAE statistic is the mean absolute error, defined as follows:

$$MAE = \sum_{i,r,w} \frac{\left|h_{r,i}^{w} - \hat{h}_{r,i}^{w}\right|}{n} \quad (19)$$

The Table 6 results clearly demonstrate that the value of the θ parameter estimator based on link density data is practically the same as the population value but significantly different from the value derived from link flow data. As for the ECM and EAM values, they are lower for the density-based estimates.

6. EXTENSION: SIMULTANEOUS TRIP MATRIX AND ROUTE FLOW ESTIMATION MODEL

In optimization problem (4) it was assumed the T_w values for trip demand between each O-D pair w on a network were already known.

In reality, however, this information is usually difficult to estimate if not simply non-existent. By contrast, estimating total trips (T) at a given instant is less complicated, or to put it another way, can be done more accurately. With an estimate of (T) we can formulate a second optimization problem that in addition to route flows will enable us to estimate the most likely trip matrix. This problem is expressed as follows:

$$\min \quad Z_2 = \sum_w T_w\left(\ln T_w - 1\right) + \frac{1}{\gamma}\left(\sum_{r\in R_w}\sum_w h_w^r\left(\ln h_w^r - 1\right) - \sum_w T_w\left(\ln T_w - 1\right)\right)$$

$$s.t.: \sum_{r\in R_w} h_w^r = T_w, \ \forall w \quad (\mu_w) \qquad \sum_w T_w = T, \qquad (\Phi) \tag{20}$$

$$\sum_{\substack{r\in R_w \\ r\supseteq a}} h_w^r = f_a^*, \ \forall a \quad (\beta_a) \qquad h_w^r \geq 0$$

where T_w is the demand between O-D pair w and is now a design variable of the problem (in problem (4) the T_w variables were exogenous whereas here in (20) only our total trips estimate T is). The formula used to estimate T is applicable to a steady-state situation at a given instant and can be simply written as $T = \varphi \sum_{a\in A} n_a$, where n_a is known previously (from a satellite photograph, for example) and φ is a positive parameter. The optimality conditions are

$$T_w = T\frac{\exp\left(L_w\right)}{\sum_w \exp\left(L_w\right)}, \quad \forall w, \quad L_w = \frac{1}{\gamma}\ln\sum_{r\in R_w}\exp\left(\gamma\sum_{a\subseteq R_w}\beta_a\right) \tag{21}$$

With these expressions we can derive a trip matrix that is consistent with the densities estimated for each network link.

7. CONCLUSIONS

A model was presented that estimates the route flows and trip matrices for a road network from link density data instead of link flow data, the traditional method.

This proposed approach has the advantage of incorporating into the analysis the fundamental traffic equation relating flow, speed and density rather than the flow-delay relationship, which may be incorrect when congestion is heavy. The proposed formulation is a maximum entropy model with flow conservation constraints. The constraint flows were estimated first from observed densities and then from a macroscopic relationship between (i) density and speed and (ii) flow, density and speed. The model was validated with traffic microsimulations.

Our first conclusion is that using link densities (veh/km)—obtainable, for example, from aerial photographs—in place of link flows (veh/h) yields satisfactory estimates of route choice and route flow models.

Indeed, the density-based approach performed better, producing an unbiased estimate of the population parameter defined in the route choice model used with the microsimulation whereas the flow-based approach generated an estimate that was biased downward.

Our second conclusion is that the predictive ability of the model estimates using densities was superior to that of the model estimated directly with flow data. Our third conclusion is that the proposed approach is easily extended to the estimation of trip matrices from link densities. Validation of these estimates can be performed using the same microsimulation methods employed for validating the other models in this study.

The formulation presented in this paper assumed a simple multinomial logit function for the route choice model. Since the focus of this paper is on the relationship between link and flow densities, the simple logit function allows for easier interpretation of the results. The optimization formulation can be extended to account for route overlapping, by suitably including an additional entropy function in the objective function (e.g. Bekhor and Prashker, 2001). Finally, an interesting extension for future research would be to apply it to large networks.

ACKNOWLEDGEMENTS

The authors are grateful for funding provided by the Centro de Desarrollo Urbano Sustentable (CEDEUS), Conicyt/Fondap/15110020.

REFERENCES

Abrahamsson, T. (1998). Estimation of Origin-Destination Matrices Using Traffic Counts – A Literature Survey. International Report, IIASA-98.

Ardekani, S. and Ghandehari M. (2008). A Modified Greenberg Speed-flow Traffic Model. Techn. Report N° 357, Dept. of Math, The University of Texas at Arlington, USA, 2008.

Bar-Gera, H. (2006). Primal method for determining the most likely route flows in large road networks. Transportation Science, 40, 269–286.

Bekhor, S., Ben-Akiva, M. E. and Ramming, S. (2006). Evaluation of choice set generation algorithms. Annals of Operations Research, 144, 235-247.

Bekhor, S., and Prashker, J. N. (2001). Stochastic user equilibrium formulation for generalized nested logit model. Transportation Research Record: Journal of the Transportation Research Board, 1752(1), 84-90.

Bell, M.G.H. (1991). The estimation of origin–destination matrices by constrained generalized least squares. Transportation Research, 25B, 13–22.

Farhangian, K. and LeBlanc, L. (1982). Selection of a Trip Table Which Reproduces Observed Link Flows," Transportation Research, 16B, 83-88.

Fisk, C. (1988). On Combining Maximum Entropy Trip Matrix Estimation with User-Optimal Assignment. Transportation Research, 22B, 69-79.

Fisk, C. (1989). Trip matrix estimation from link traffic counts: The congested network case. Transportation Research, 23B, 331–336.

Geroliminis, N. and Daganzo, C. F. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological, 42(9), 771-781.

Gothe, M.B., Jornsten, K.O. and Lundgren, J.T. (1989). Estimation of origin destination matrices from traffic counts using multi-objective programming formulation. Transportation Research, 23B, 257-269.

Hazelton, M. (2003a). Some comments on origin–destination matrix estimation. Transportation Research, 37A, 811–822.

Hazelton, M.L. (2003b). Total travel cost in stochastic assignment models. Networks and Spatial Economics, 3, 457-466.

Iida, Y.; Sasaki, T. and Yang, H. (1994). The Equilibrium-Based Origin-Destination Matrix Estimation Problem," Transportation Research, 28B, 23-33.

Janson, B. (1993). Most likely origin-destination link uses from equilibrium assignment Transportation Research, 27B, 333–350.

Lam, H.K.W. and Lo H.P. (1991). Estimation of Origin Destination Matrix from Traffic Counts: A Comparison of Entropy Maximizing and Information Minimizing Models. Transportation Planning and Technology, 16, 166-171.

Maher, M. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transportation Research, 20B, 435–447.

Maher, M.J., Zhang, X. and D.Van Vliet, D. (2001). A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows. Transportation Research, 35B, 23-40.

May, A. D. and Keller, H. E. (1967). A deterministic queueing model. Transportation Research, 1, 117-128.

Nguyen, S. (1984), Transportation Planning Models, Edited by M. Florian, Elsevier Science Publishers, Amsterdam, chapter Estimating origin-destination matrices from observed flows, 363-380.

Prato, C.G. (2009). Route choice modeling: past, present and future research directions. Journal of Choice Modelling, 2, 65–100.

Rossi, T.; McNeil, S. and Hendrickson, C. (1989). Entropy model for consistent impact-fee assessment. Journal of Urban Planning and Development, 115, 51–63.

Spiess, H. (1987). A Maximum Likelihood Model for Estimating Origin-Destination Matrices. Transportation Research, 21B, 395-412.

Van Zuylen, J. H. and Willumsen L. G. (1980). The most likely trip matrix estimated from traffic counts. Transportation Research, 14B, 281-293.

Watling, D. and Hazelton, M. L. (2003). The dynamics and equilibria of day-to-day assignment models. Networks and Spatial Economics, 3, 349-370.

Wilson A. G. (1970). Entropy in Urban and Regional Modeling, Pion, London.

Yang, H., Sasaki, T., Iida, Y. and Asakura, Y. (1992). Estimation of O–D matrices from link traffic counts on congested networks. Transportation Research, 26B, 417–434.

Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transportation Research, 29B, 231-242.

Zhang, H.M. and Nie, X. (2005). Some consistency conditions for dynamic traffic assignment problems. Networks and Spatial Economics 5, (1), 71–87.

Ziliaskopoulos, A. K. and Peeta, S. (2001). Review of Dynamic Traffic Assignment Models. Networks and Spatial Economics, 1, 233–267.