1 2	A Maximum Likelihood Estimation of Trip Tables for the Strategic User Equilibrium Model
3	
4	Tao Wen*
5	School of Civil and Environmental Engineering, University of New South Wales
6	and National ICT Australia (NICTA)
7	Sydney NSW 2052, Australia, t.wen@unsw.edu.au
8	
9	Chen Cai
10	National ICT Australia (NICTA)
11	Sydney NSW 2052, Australia, chen.cai@nicta.com.au
12	
13	Lauren Gardner
14	School of Civil and Environmental Engineering, University of New South Wales
15	Sydney NSW 2052, Australia, l.gardner@unsw.edu.au
16	
17	Vinayak Dixit
18	School of Civil and Environmental Engineering, University of New South Wales
19	Sydney NSW 2052, Australia, v.dixit@unsw.edu.au
20	
21	S. Iravis Waller
$\frac{22}{22}$	School of Civil and Environmental Engineering, University of New South wales
23 24	Sudney NSW 2052 Australia (NICTA)
24 25	Sydney NSW 2052, Australia, s.waller@ullsw.edu.au
25 26	Fang Chen
20	National ICT Australia (NICTA)
28	Sydney NSW 2052 Australia fang chen@nicta.com.au
29	by anoy 105 († 2052, Hustrana, Tang.onon e mou.com.au
30	
31	*Corresponding author
32	
33	
34	
35	
36	
37	A Paper Submitted for Presentation at the 94th Annual Meeting of the Transportation
38	Research Board and Publication in the Transportation Research Record
39	
40	
41	
42	Total words = $4587+2$ Tables and 5 Figures= $6337$ words
43	

### 1 ABSTRACT:

2 This paper proposes a novel framework to estimate trip tables for the strategic user 3 equilibrium traffic assignment model. The proposed framework uses a bi-level estimation 4 model, where the upper-level is a new maximum likelihood estimation method and the lower-5 level is the strategic user equilibrium assignment model which accounts for some aspects of 6 day-to-day volatility in traffic flow. The maximum likelihood method proposed in this paper 7 illustrates its ability to utilize information from day-to-day observed link flows in order to 8 provide a unique estimation of the total trip demand distribution. This is accomplished by 9 passing the total trip demand distribution to the strategic user equilibrium model to produce a 10 set of link flow distributions which can then be compared to the link level observations. The mathematical proof demonstrates the convexity of the model. In addition, a numerical 11 12 analysis is conducted on a test network to illustrate the efficiency of the proposed framework. 13

14 Keywords: strategic user equilibrium, O-D estimation, maximum likelihood

#### 1 1. INTRODUCTION

Enhanced origin-destination (O-D) matrix estimation methodologies would be extremely useful for transportation planning. Traditionally, the O-D matrix is obtained from plate surveys, household surveys or roadside surveys. Such survey activities may be financially expensive for large size networks at frequent intervals, and usually suffer from limited response and sampling coverage. As an alternative, O-D matrix estimation provides a statistical approach for estimating or calibrating an O-D matrix from observed link flows and some prior knowledge of the O-D demand.

9 In this paper, we propose a framework that combines the maximum likelihood method 10 for O-D matrix estimation and the strategic user equilibrium model (StrUE) for traffic 11 assignment(1). This framework is hereby referred as MLStrUE. The StrUE model is defined such that "at strategic user equilibrium all used paths have equal and minimal expected cost". 12 For each user present in a given demand scenario, their chosen route is followed regardless of 13 14 the realized travel demand on a given day. Therefore the link flows will not result in an 15 equilibrium state in any particular demand realization, but instead equilibrium exists stochastically over all demand realizations. The StrUE model was proposed to be able to 16 capture the impact of day to day demand volatility on reliability, and eventually route choice. 17 Therefore, apart from the O-D splits, a fundamental parameter that needs to be estimated is 18 19 the variance in the total trip demand distribution.

20 An important aspect of the StrUE model is that the total trip demand is assumed to 21 follow a certain statistical distribution; traditionally a lognormal distribution has been used (2, 22 3). Under the assumption of a log-normally distributed demand, this paper focuses on 23 estimating the demand distribution parameters. Note that other distributions can also be used 24 if they do not change the convexity of the objective function. Only the distribution of the total 25 trip demand needs to be estimated because for the StrUE model the O-D proportions are 26 assumed to be fixed. Furthermore, for the StrUE model a log-normally distributed total trip 27 demand allows for a closed form probability density function of the link flow, which can be 28 shown to follow a lognormal distribution as well. The direct relationship between the link 29 flow variables and the total demand in StrUE allows for the use of day-to-day observed link 30 flows (which in turn provide actual link flow distributions) to calibrate the total demand 31 distribution. In this study this calibration is accomplished by implementing the maximum 32 likelihood estimation method, in which we maximize the joint probability of observing all the 33 link flows within a time period.

34 A bi-level programming method is proposed to eliminate the impact of strongly 35 biased initial estimates, where the upper level provides the total demand distribution to the 36 StrUE model, and the StrUE model can provide link flow distributions to the upper level. A 37 benefit of the proposed modelling framework includes the incorporation of actual day-to-day 38 observed link flows and the corresponding distributions, instead of aggregated or averaged 39 values. Additionally, the performance of the MLStrUE approach can be assessed based on the 40 accuracy of its estimations for both expected link flows and link flow distributions, which are 41 a direct output of the StrUE model.

The remainder of this paper includes a literature review of previous relevant research, presented in Section 2. Section 3 defines the mathematical model and includes a derivation for the analytical solution to the total demand estimation. Numerical analysis is demonstrated in Section 4; conclusions and future research are presented in Section 5.

#### 46 **2. LITERATURE REVIEW:**

47 Although the traditional O-D matrix estimation mainly focused on statistical approaches 48 based on loop counts, a wide range of methods have been explored in previous studies, 49 including the generalized least square method(4, 5), the maximum likelihood(6), bi-level 1 programming approach(7) and maximum entropy(8). Generally the problem is to find an O-D 2 matrix to optimize an objective function subject to a set of constraints. However, the problem 3 is often challenging due to the number of observable links in a traffic network typically being 4 much smaller than the number of O-D pair demands; this means that it may not be possible to 5 obtain a unique solution from a single set of link counts alone. The problem was further extended to account for the stochastic nature of observed flows (9, 10). Recently, dynamic 6 7 approaches were introduced to account for the time dependent characteristics in the 8 network(11, 12). However, the application to large scale network and the computation 9 complexity still remains a problem.

10 Among the research, relatively little attention has been paid to the higher order of the 11 variables in a network, such as their variance and covariance that can potentially provide 12 more constraints to the optimization problem. Cremer and Keller demonstrated that aggregating or averaging link count data collected over a sequence of time period may lose 13 14 some important information.(13). Hazelton (14) proposed a weighted least squares method to 15 account for the covariance of links, and assumed a parameter to explain the circumstances when the variance exceeds the mean if a Poisson distribution is used. Bell (15) proposed a 16 maximum likelihood method and found the analytical solution to the covariance of O-D 17 18 matrix by using a Taylor approximation. However, these research contributions still have 19 some limitations in the assumptions. For example, the O-D demand was assumed to follow 20 the Poisson or multinomial distribution, which stipulates certain relationships between the 21 mean and variance of the O-D demand. In the MLStrUE, the O-D demand is assumed to 22 follow a lognormal distribution, which allows the mean and variance of total demand to be 23 independent of each other, and assures the non-negativity of the demand. In a well-24 constructed network, loop detectors can easily provide link counts on a day-to-day basis; therefore, it is important to consider the variation of link flows and the distribution of O-D 25 26 total demand as effective information to calibrate the O-D trip matrix. The proposed 27 MLStrUE framework estimates the distribution of the total O-D demand and thus 28 significantly reduces computation complexity.

29 Estimation of the O-D trip matrix also requires a proper assignment model. When applying the assignment model to a large network, realism and computational complexity are 30 31 both critical and must be equally considered to determine a model's practical applicability. 32 Further, a major complication in transportation modelling is the ability to properly account for the inherent uncertainties regarding demand (16, 17) and capacity levels (18, 19). 33 34 Additionally, as has been noted, uncertainty on these variables directly affects route choice 35 behaviour (20). It is, therefore, necessary to incorporate these stochastic elements into models 36 to ensure robust planning capabilities, but to do so in a manner that maintains computational 37 tractability. The strategic user equilibrium (1) effectively accounts for the impact of demand 38 uncertainty; the model relies on users minimizing their expected travel time based on the previous trip experiences in which they have gathered knowledge on demand (daily trips). 39 40 The user's knowledge of each can be represented by a given distribution, with a known 41 expected value and variance. Based on these known distributions, each user selects a travel 42 route to minimize their expected travel time subject to Wardrop's UE conditions.

43 44

45

46

The contribution of this study can be highlighted as the following:

- 1) By assuming that the total demand follows a lognormal distribution, we exploit its positiveness. Additionally, unlike the Poisson distribution, it allows the variance and the mean of the total demand to be different.
- 47
  48
  2) We consider the day-to-day link flow variations and use a maximum likelihood method to relate this information to the total demand distribution.
- 49 3) We also apply the strategic user equilibrium model to account for the impact of50 variation in demand.

4) The StrUE model is used to provide link flows according to the estimated total demand distribution; multiple sets of link flows, thus the link flow distribution, is generated by sampling the total demand.

# 4 **3. PROBLEM FORMULATION:**

5 This section defines the mathematical concept of the MLStrUE framework. Table 1 explains

- 6 the notations used in this paper.
- 7

1

2

3

TABLE 1 Summary of notations.

Ν	Link (index) set
$K_{RS}$	Path set
Α	Node (index) set
$p_n$	Proportion of total demand on link $n$ ; $P = (p_1,, p_n)$
$t_n$	Travel time on link $n$ ; $\mathbf{t} = (, t_n,)$
$t_{nf}$	Free flow travel time on link <i>n</i>
$c_k^{rs}$	Travel time on path k connecting O-D pair r-s; $\mathbf{c}^{rs} = (, c_k^{rs})$ ; $\mathbf{c} = (, c^{rs},)$
$q_{rs}$	Fraction of total trips that are between O-D pair r-s; $\sum_{\forall rs} q_{rs} = 1$
Т	Random variable for total trips with probability distribution $g(T)$
g(T)	Lognormal probability density function of the total demand
x <sub>ni</sub>	Observed flow on link <i>n</i> , for day <i>i</i> .
$l_n$	Flow on link <i>n</i> .
$v_k^{rs}$	Proportion of flow on path k, connecting O-D pair r-s
$C_n$	The capacity on link <i>n</i>
$\delta^{rs}_{n,k}$	Indicator variable $\delta_{a,k}^{rs} = \begin{cases} 1 \text{ if nis included in path } k \\ 0 \text{ otherwise} \end{cases} (\Delta^{rs})_{n,k} = \delta_{n,k}^{rs};$
	$\Delta = (\dots, \Delta^{rs}, \dots)$
и	Mean of the corresponding normal distribution, also called the location
<i>r</i> -	parameter for the lognormal distribution.
$\sigma^2$	Variance (= the second central moment) of the corresponding normal
	distribution, also called the scale parameter for the lognormal distribution.
σ	Standard deviation of the corresponding normal distribution
$m_T$	The expected total demand
$v_T$	Variance of the lognormal distribution
$m_n$	The expected link flow on link <i>n</i> .
$v_n$	The variance of link flow on link <i>n</i> .
$S_T$	The standard deviation of the total demand
<i>s</i> <sub>n</sub>	The standard deviation of link flow on link <i>n</i> .
соv	Coefficient of variation, defined as the ratio of the standard deviation to the mean of a variable.

8 It is important to realize that  $\mu$  and  $\sigma^2$ , which appear in the equations of the log-9 normal distribution, do not denote the mean and the variance of the log-normal distribution, 10 but of the corresponding parameters of the normal distribution. The mean and the variance of 11 the log-normal distribution are indicated in the following discussion by *m* and *v*.

The assignment map in the StrUE model is a vector noted as the link proportions,
which is the proportion of the link flow to the total demand:

15 
$$l_n = p_n T$$
  $p_n \in P, n \in N$ 

1 The link proportions are assumed fixed in the O-D matrix estimation problem; hence 2 each link flow also follows a lognormal distribution if the total demand follows a lognormal 3 distribution. It has been validated that numerical convolution of lognormal distributions is a 4 distribution which follows the lognormal law with a fair approximation(21). The link flow 5 distribution is related to the total demand distribution by:

$$7 m_n = p_n m_T [2]8 v_n = p_n^2 v_T [3]$$

The parameters for the link flow distribution can be obtained by the definition:

11 
$$\mu_n = \ln m_n - \frac{1}{2} \ln (1 + \frac{v_n}{m_n^2})$$
 [4]

12 
$$\sigma_n^2 = ln(1 + \frac{\nu_n}{m_n^2})$$
 [5]

Substitute Eq.2 and Eq.3 into Eq.4 and Eq.5, we have the transformation of the total
demand distribution to link flow distribution:

$$\begin{array}{ll}
16 & \sigma_n = \sigma_T \\
17 & \mu_n = lnp_n + \mu_T \\
18 & & & [6]
\end{array}$$

19 Since each link flow follows a lognormal distribution, the probability of observing  $x_n$ 20 trips on link *n* is: 21

22 
$$f(x_n) = \frac{1}{x_n \sigma_n \sqrt{2\pi}} e^{-\frac{(\ln x_n - \mu_n)^2}{2\sigma_n^2}} n \in N$$
 [8]

23  $x_n$ -Observed flow on link n

The joint probability of observing a set of link flows can be obtained by the product of the probability density functions:

27

28

29

24

6

9 10

$$j(x_n) = \prod_{1}^{n} \frac{1}{x_n \sigma_n \sqrt{2\pi}} e^{-\frac{(\ln x_n - \mu_n)^2}{2\sigma_n^2}} \quad n \in N$$

Furthermore, we may collect more than one set of loop counts, namely the observed day-to-day link flows. It is therefore necessary to maximize the joint probability of observing all sets of link flows, in order to reduce the effect of noise and observation failure. Here the observed link flows are indicated by a *n*-by-*i* matrix, *n* is the number of links and *i* is the number of observations:

35 36  $x_{ni} = \begin{bmatrix} x_{11} & \cdots & x_{1i} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{ni} \end{bmatrix}$  37 [10]

The maximum likelihood method here is to maximize the joint probability of
 observing all sets of link flows, which is given by the following equation:

41 
$$j(x_n) = \prod_{1}^{i} \prod_{1}^{n} \frac{1}{x_{ni}\sigma_n\sqrt{2\pi}} e^{-\frac{(\ln x_{ni}-\mu_n)^2}{2\sigma_n^2}}$$
 [11]  
42

[9]

1 Conventionally, we maximize the logarithm of the joint probability, because taking 2 log of the function won't change its convexity. By plugging in Eq.6 and Eq.7 into Eq.11 and 3 changing the signs, the objective function becomes: 4

5 min: 
$$J(t_j^*) = \sum_{1}^{i} \sum_{1}^{n} \ln(x_{ni}\sigma_T \sqrt{2\pi}) + \frac{(\ln \frac{x_{ni}}{p_n} - \mu_T)^2}{2\sigma_T^2}$$
 [12]  
6 Subject to:  $\sigma_T > 0$ 

To prove the convexity of the objective function, we only need to show that for an arbitrary x, the function below is convex:

7 8

11 
$$f(\mu_T, \sigma_T) = \sum_{1}^{n} \ln(x \sigma_T \sqrt{2\pi}) + \frac{(\ln x - \mu_T)^2}{2\sigma_T^2}$$
 [13]

13 The Hessian matrix of  $f(\mu_T, \sigma_T)$  can be found by taking second partial derivatives 14 with respect to  $\mu_j$  and  $\sigma_j$ :

16 
$$H = \begin{bmatrix} \sigma_T^{-2} & 2\sigma_T^{-3}(\ln x - \mu_T) \\ 2\sigma_T^{-3}(\ln x - \mu_T) & \sigma_T^{-2} + 3\sigma_T^{-4}(\ln x - \mu_T)^2 \end{bmatrix} > 0$$
[14]

18 The Hessian is positive definite, hence the function is strictly convex. The sum of 19 several convex functions is also a convex function, therefore we have proved that our 20 objective function is strictly convex, the unique optimal solution is assured. The optimal 21 solutions can be found by taking the first derivative with respect to mean and variance of total 22 demand:

23  
24 
$$\mu_T = \frac{\sum_{i=1}^{i} \sum_{j=1}^{n} \ln \frac{x_{ni}}{p_n}}{\prod_{i=1}^{ni} x_{ii}}$$
 [15]

25 
$$\sigma_T^2 = \frac{\sum_{1}^{r} \sum_{1}^{n} (\ln \frac{m}{p_n} - \mu_T)^2}{ni}$$
 [16]

Assuming that the StrUE model represents the route choice behaviour, we can then formulate a bi-level programming problem, where the upper level is the maximum likelihood demand estimation problem; the lower level is the StrUE model, which has the objective function:

32 
$$Minimize: z(f) = \sum_{n \in \mathbb{N}} \int_0^{f_n} \int_0^\infty \int_0^\infty t_n(p_n T) g(T) dT df$$
[17]

33 Subject to:

$$34 \quad \sum_{k} v_k^{rs} = q_{rs} \quad \forall k, r, s$$
[18]

$$35 \quad v_k^{rs} \ge 0 \qquad \quad \forall k, r, s \tag{19}$$

$$36 \quad p_n = \sum_r \sum_s \sum_k v_k^{rs} \delta_{n,k}^{rs} \qquad \forall k, r, s$$
[20]

37

The fraction of the total demand between O-D pair *r*-*s*, namely  $q_{rs}$ , can be obtained from the prior estimates, i.e. household survey data, or field experiments. The link travel time function for the StrUE model is defined by the U.S. Bureau of Public Roads (22) cost function due to its widespread use in transport planning models:

$$t_n(flow) = t_{nf} \left[ 1 + \alpha t_{nf} \left( \frac{f_n T}{C_n} \right)^{\beta} \right]$$
[21]

3

2

4 where  $\alpha$  and  $\beta$  are the parameters for the BPR function.

5 The objective functions of the upper and the lower levels are both strictly convex, 6 therefore the model always has feasible solutions. A solution algorithm has been proposed to 7 the bi-level programming:

8 Step 1: (Initialization) k=0. Start from the prior O-D matrix; obtain the fraction of 9 total trips  $q_{rs}$  and initial values for the mean and the variance of the total demand. Produce a 10 set of link proportions from the StrUE model. Note that  $q_{rs}$  will be kept invariant over the bi-11 level iterations while  $\mu_T^k$  and  $\sigma_T^k$  will be calibrated.

12 Step 2: (Optimization) Substituting the link-flow proportion matrix  $P_k$ , solve the 13 upper-level to obtain  $\mu_T^k$  and  $\sigma_T^k$  of the total demand.

14 Step 3:(Simulation) Using  $\mu_T^k$  and  $\sigma_T^k$ , apply the StrUE model to produce a new set of 15 link flow proportions  $P_{k+1}$ .

16 Step 4: (Convergence test) Calculate the deviation between simulated and observed 17 link flows, and the deviation between estimated and target O-D matrices. If stopping 18 criterion is met, stop. After enough iteration, the results will always converge.

#### 19 4. NUMERICAL RESULTS AND ANALYSIS:

20 The objective of the analysis is to test if the MLStrUE can effectively estimate the total 21 demand distribution from day-to-day observed link flows. The estimated total demand 22 distribution should closely approximate the actual total demand distribution; the link flow 23 distribution produced by the StrUE model should also closely match the observed link flows. 24 The main idea is to artificially determine the total demand distribution and generate random 25 link flow samples accordingly. The MLStrUE will reproduce the desired total demand 26 distribution from the random samples with perturbed prior estimates. The test should also 27 reflect the scalability of the MLStrUE to networks of substantial complexity.

Numerical tests are conducted on the Sioux Falls network (24 nodes and 76 links). The network properties are pre-defined in (23) (see Fig.1). The notations used in this section are defined in Table 1. The O-D demand is specified as proportions of the total network demand, therefore the demand for a given O-D pair is the O-D proportion multiplied by the total demand. The BPR parameters  $\alpha$  and  $\beta$  are equal to 0.15 and 4, respectively.

- 33 The observed link flows are generated by the following way:
- 34 Step 1: The true  $\mu_T$ ,  $\sigma_T$  are determined for the total demand.

Step 2: We implement the StrUE based on the total demand distribution and obtain a
 set of link proportions.

37 Step 3: We generate 100 samples of the total demand from the lognormal distribution 38 using  $\mu_T$ ,  $\sigma_T$  as parameters and each sample total demand is assigned to the network using the 39 pre-calculated link proportions.



FIGURE 1 The Sioux Falls network.

2 3

The actual expected total demand of the Sioux Falls network is  $m_A = 360600$ , and 4 5 the coefficient of variation cov is equal to 0.2, i.e. the standard deviation is 20% of the 6 expected total demand. In Table 2, each scenario represents a different initial estimate of the

7 total demand dis	stribution.

Scenario	Scenario description	$m_T$	S <sub>T</sub>
1	$m_T = 0.8 m_A$ and $cov = 0.1$	288480	28848
2	$m_T = 0.8 m_A$ and $cov = 0.3$	288480	86544
3	$m_T = 1.2m_A \ and \ cov = 0.1$	432720	43272
4	$m_T = 1.2 m_A \text{ and } cov = 0.3$	432720	129816
5	$m_T = 1.5 m_A \ and \ cov = 0.1$	540900	54090
6	$m_T = 1.5 m_A \text{ and } cov = 0.3$	540900	162270

8 **TABLE 2** Different scenarios of initial estimation of O-D matrix

9 In Fig.2 and Fig.3, the x-axis represents the number of iterations of the bi-level programming. In Fig.2, the y-axis represents the estimated expected total demand; In Fig.3, 10 the y-axis represents the estimated standard deviation of the total demand. Both figures show 11 that the estimated results converge to the actual ones in less than 3 iterations. This indicates 12 13 that the MLStrUE's robust performance against biased initial estimates, and demonstrates the 14 efficiency in arriving at convergence. In scenario 1 and 2, the estimated results of the first 15 iteration in both figures are very different from the actual ones. This is mainly because the 16 link proportions of the first iteration are obtained based on the initial estimates, and the initial 17 estimates in scenarios 1 and 2 are very biased, and therefore the results of the first iteration will be inaccurate. If the initial estimates of total demand distribution are guite different from 18 19 the actual distribution, the estimation of both expected total demand and standard deviation of 20 total demand will be very inaccurate, therefore we have shown that applying the bi-level 21 programming can reduce the impact of biased initial estimates.



FIGURE 2 Estimated expected total demand under different scenarios of initial estimation; results of 10 bi-level iterations are presented.





3

6

FIGURE 3 Estimated standard deviation of total demand under different scenarios of initial estimation; results of 10 bi-level iterations are presented.

Fig.4 demonstrates the performance of the MLStrUE at the link level; the *x*-axis indicates the actual expected link flow while the *y*-axis represents the estimated expected link flow. The estimated link flows are analytically produced by the StrUE model based on the total demand distribution after the convergence criterion has been met. The estimated expected link flows and the corresponding actual expected link flows are sorted from the smallest to the largest. The R squared value of the results is equal to 0.9837, which is very



1 close to 1. This indicates that the estimated results closely approximate the actual expected 2 link flows.

FIGURE 4 The estimated and actual expected link flow comparison, estimated results
are produced by the StrUE model based on estimated demand distribution.

6 One of the strengths of the StrUE model is that it can produce the link flow variation. 7 Since the total demand distribution is calibrated based on day-to-day observed link flows, it is 8 therefore necessary to compare the estimated standard deviation of link flow to the actual one. 9 In Fig.5, the estimated standard deviation of link flow is produced by the StrUE model based 10 on the total demand distribution after the bi-level convergence criterion has been met. The x-11 axis denotes the actual standard deviation of link flow while the y-axis indicates the estimated 12 one. It is illustrated in the figure that despite the fact that the R squared value is smaller than that of the expected link flow; the MLStrUE still provides relatively reliable estimation, 13 14 however, if the standard deviation of link flow is very high, the estimated results may be 15 more than 20% different from the actual ones.



estimated results are produced by the StrUE model based on estimated demand

distribution.

1 2

3

4

# 5 5. CONCLUSION:

6 This paper proposes a methodological framework (MLStrUE) to estimate the travel demand 7 distribution (trip table) based on day-to-day observed link flows. The estimated total demand 8 distribution maximizes the joint probability of observing all link flows. A bi-level 9 programming method is also included to reduce the impact of biased initial estimates of the 10 total demand distribution. A numerical analysis is conducted on a test network, and results for 11 both the system level and the link level demonstrated robust performance of the MLStrUE 12 framework. In the numerical experiment, the estimated mean and standard deviation of the total demand converged to the desired values regardless of the initial estimates after 2 or 3 13 14 iterations. Similarly, the link level analysis produced R squared values of 0.9837 and 0.942, 15 for the expected value and standard deviation of link flows, respectively. Based on the results, 16 the estimated link flow distribution closely approximates the actual link flow distribution, 17 suggesting that the MLStrUE can calibrate the total demand effectively and efficiently.

18 One limitation of the MLStrUE is the assumption of perfect traffic loop count 19 information. In this model, we generate loop counts by sampling from the results of the 20 assignment model based on actual demand distribution, which may not reflect real world 21 condition. In reality, the loop counts of some minor roads, or smaller regional roads might be 22 missing in practice, and the failure of the loop detectors may also have impact on the results. 23 The error may be reduced via statistical approaches such as outlier detection or noise analysis. 24 Another limitation is that the prior estimates of demand proportions may influence the results. 25 A real-world data set may be used in the future to validate the framework proposed here.

Future research will investigate the use of the covariance of loop counts, that is, if we have a large enough sample size, then the covariance matrix of link flows can be generated. This can potentially provide much more information than only mean and variance of link flows. Furthermore, the O-D demand may be assumed to follow a multivariate lognormal distribution, in this way the O-D demand is no longer aggregated as was the case with univariate lognormal distribution, possibly providing the covariance matrix of the link flows. 1 Since the OD estimation problem is a combination of a statistical optimization model and a 2 traffic assignment model, an improvement in either process warrants further research.

# **3 6. ACKNOWLEDGEMENTS:**

4 This research was in part made possible by the generous support of NICTA. NICTA is 5 funded by the Australian Government through the Department of Communications and the

6 Australian Research Council through the ICT Centre of Excellence Program.

# 7 7. REFERENCES:

8 9 Dixit, V., L.M. Gardner, and S.T. Waller. Strategic User Equilibrium Assignment Under Trip (1)Variability. in Transportation Research Board 92nd Annual Meeting. 2013. 10 (2)Wen, T., et al. A Strategic User Equilibrium Model Incorporating Both Demand and Capacity 2 11 Uncertainty 3. in Transportation Research Board 93rd Annual Meeting. 2014. 12 (3) Duell, M., et al. Evaluation of a strategic road pricing scheme accounting for day-to-day and long term 13 14 demand uncertainty. in Transportation Research Board 93rd Annual Meeting. 2014. (4) Cascetta, E., Estimation of trip matrices from traffic counts and survey data: A generalized least 15 16 squares estimator. Transportation Research Part B: Methodological, 1984. 18(4-5): p. 289-299. (5) Bell, M.G.H., The estimation of origin-destination matrices by constrained generalised least squares. 17 Transportation Research Part B: Methodological, 1991. 25(1): p. 13-22. 18 (6) Spiess, H., A maximum likelihood model for estimating origin-destination matrices. Transportation 19 Research Part B: Methodological, 1987. 21(5): p. 395-412.  $\begin{array}{c} 20\\ 21\\ 22\\ 23\\ 24\\ 25\\ 26\\ 27\\ 28\\ 29\\ 30\\ 31\\ 32\\ 33\\ 34\\ 35\\ 36 \end{array}$ (7)Yang, H., et al., Estimation of origin-destination matrices from link traffic counts on congested networks. Transportation Research Part B: Methodological, 1992. 26(6): p. 417-434. (8) Fisk, C.S., On combining maximum entropy trip matrix estimation with user optimal assignment. Transportation Research Part B: Methodological, 1988. 22(1): p. 69-73. (9) Lo, H., N. Zhang, and W.H. Lam, Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. Transportation Research Part B: Methodological, 1996. 30(4): p. 309-324. (10) Lo, H., N. Zhang, and W. Lam, Decomposition algorithm for statistical estimation of OD matrix with random link choice proportions from traffic counts. Transportation Research Part B: Methodological, 1999. **33**(5): p. 369-385. (11)Frederix, R., F. Viti, and C.M.J. Tampère, Dynamic origin-destination estimation in congested networks: theoretical findings and implications in practice. Transportmetrica A: Transport Science, 2011. 9(6): p. 494-513. Bierlaire, M. and F. Crittin, An efficient algorithm for real-time estimation and prediction of dynamic (12) OD tables. Operations Research, 2004. 52(1): p. 116-127. (13) Cremer, M. and H. Keller, A new class of dynamic methods for the identification of origin-destination flows. Transportation Research Part B: Methodological, 1987. 21(2): p. 117-132. 37 (14)Hazelton, M.L., Some comments on origin-destination matrix estimation. Transportation Research Part 38 A: Policy and Practice, 2003. 37(10): p. 811-822. 39 (15)Bell, M.G.H., The Estimation of an Origin-Destination Matrix from Traffic Counts. Transportation 40 Science, 1983. 17(2): p. 198-217. 41 (16)Kim, J., F. Kurauchi, and N. Uno, Analysis of variation in demand and performance of urban 42 43 44 expressways using dynamic path flow estimation. Transportmetrica, 2009. 7(1): p. 63-84. (17)Bellei, G., et al., A demand model with departure time choice for within-day dynamic traffic assignment. European Journal of Operational Research, 2006. 175(3): p. 1557-1576. 45 (18)Brilon, W., J. Geistefeldt, and M. Regler. Reliability of freeway traffic flow: a stochastic concept of 46 capacity. in Proceedings of the 16th International symposium on transportation and traffic theory. 47 2005. 48 (19)Wu, X., P. Michalopoulos, and H.X. Liu, Stochasticity of freeway operational capacity and chance-49 constrained ramp metering. Transportation Research Part C: Emerging Technologies, 2010. 18(5): p. 50 741-756. 51 52 53 (20)Uchida, T. and Y. Iida. Risk assignment. A new traffic assignment model considering the risk of travel time variation. in Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation, Jul 21 - 23 1993. 1993. Berkeley, CA, United states: Publ by Elsevier Science 54 Publishers B.V. 55 (21)Fenton, L., The sum of log-normal probability distributions in scatter transmission systems. 56 Communications Systems, IRE Transactions on, 1960. 8(1): p. 57-67.

- (22) U.S, Department of Commerce. Bureau of Public Roads. Traffic Assignment Manual Urban Planning
- Division, US Department of Commerce, Washington, DC. 1964. Bar-Gera, H., Transportation test problems of Sioux falls network. website http://www.bgu.ac.il/~bargera/tntp/, 2012. (23)