



Title	Travel Time Reliability Estimation Model Using Observed Link Flows in a Road Network
Author(s)	Uchida, Kenetsu
Citation	Computer-aided civil and infrastructure engineering, 30(6), 449-463 <a href="https://doi.org/10.1111/mice.12109">https://doi.org/10.1111/mice.12109</a>
Issue Date	2015-06
Doc URL	<a href="http://hdl.handle.net/2115/61969">http://hdl.handle.net/2115/61969</a>
Rights	This is the peer reviewed version of the following article: Computer-Aided Civil and Infrastructure Engineering, 30(6) 413–506 June 2015, which has been published in final form at <a href="http://onlinelibrary.wiley.com/doi/10.1111/mice.12109/full">http://onlinelibrary.wiley.com/doi/10.1111/mice.12109/full</a> . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Type	article (author version)
File Information	manuscript Uchida-R4.pdf



[Instructions for use](#)

# Travel time reliability estimation model using observed link flows in a road network

Kenetsu Uchida\*

*Hokkaido University, North 13 West 8 Kita-ku Sapporo, Japan*

**Abstract:** *This paper formulates a model that calculates travel time reliability in a road network. The sources of uncertainty addressed by the model are traffic capacity and travel demand. Since travel time reliability estimation problem, in general, is formulated as a path based problem, path enumeration can be required in advance. Thus, the path travel time reliability can depend on a path set enumerated in advance. For the purpose of determining the unique travel time reliability, a model that estimates stochastic path flows by using observed link flows is then presented. This method does not require predefined path set. However this method is similar to the standard maximum likelihood (ML) estimation method, the method presented in this study is easier to be solved since the number of unknown parameters is much smaller than that of the standard ML estimation method. Numerical experiments using two networks are carried out to demonstrate the model presented in this paper.*

## 1 INTRODUCTION

Conventional frameworks for analyzing and modeling transportation systems have been confined to average representation of the network state (e.g., average link flow or average travel demand). For instance, in the traditional traffic assignment model one can obtain a deterministic prediction of a future flow on a certain link in the network based on average origin-destination (O-D) demands, link capacities, and a form of proportional path choice model (either user equilibrium (UE) or stochastic user equilibrium (SUE)). This represents the deterministic view of the environment and the modeler's postulation that the variability or uncertainty in the system is not influential in system design and evaluation.

There has been a growing concern over the uncertainty of travel time in transport systems and its effect on the reliability of its services (SACTRA 1999). Research into

network reliability has emerged toward addressing this problem (Asakura, 1999; Bell, 2000; Chen et al., 2002; Sumalee and Kurauchi, 2006). From the traveler's perspective, the issue of travel time reliability has been a major concern. Travelers may experience excessive variability of their journey times from day to day on the same trips (Clark and Watling, 2005; Liu et al., 2007; Noland and Polak, 2002). From the planner's perspective, a key challenge is to design and operate the transportation system such that it copes with this variability, toward maintaining an acceptable level of service (Du and Nicholson, 1997; Lo and Tung, 2003; Sumalee et al., 2007).

Uncertainties in a network can be categorized into three main ones: those of supply, demand and behavior. It is possible to subcategorize these into recurrent and sporadic cases. Supply uncertainty relates directly to variability in the road or service capacity, which may be caused by recurrent events (e.g., traffic accidents, road work and parking) or sporadic causes (e.g., natural or man-made disasters). Regarding the demand side, evidence suggests that there exists a significant day-to-day variability in travel demand. Behavioral uncertainty is more related to the predictability of the model. In the context of path choice, one can define the probability (in contrast to the concept of proportion) of a certain path being chosen by an individual (Watling, 2002). The model is, thus, only capable of predicting the *likelihood* of each driver choosing each path, and not the actual choice made. This model provides only the likelihoods of the different states of the network.

In transport modeling, some advances have been made into incorporating these uncertainties into the network modeling framework, i.e., developing a stochastic network model. Watling (2002) proposed a second-order network equilibrium model that explicitly considers random path choice behavior. His model, in contrast to the conventional SUE model, uses path choice probability, as predicted by a SUE model, to define stochastic path flows that follow a multinomial distribution. The path flows derived using the traditional SUE model are, in fact, the expected flows of this multinomial distribution. In this model, the drivers

choose their paths such as to minimize their perceived long-run expected travel costs. With a non-linear travel cost function, this long-run expected travel cost will differ from the equilibrium cost computed by the conventional SUE model (Cascetta, 1989).

Clark and Watling (2005) extended this stochastic network model to the case with a Poisson distribution of O-D flows. Similarly, Nakayama and Takayama (2003) proposed a stochastic network model with random path choice behavior but using a binomial distribution of O-D flows. These models fully represent stochastic/uncertain path choice behavior with uncertain demands. However, the models do not consider supply uncertainties. Lo and Tung (2003) proposed a rather different formulation of the stochastic network model focusing on capacity uncertainty. The road capacity is assumed to follow a uniform distribution. Sumalee et al. (2011) address stochastic demand and stochastic capacity in a multi-modal network. Uchida et al. (2014) address network design problem in a stochastic multi-modal network model. Szeto and Solayappan (2010) proposed a nonlinear complementarity problem formulation for the risk-averse stochastic transit assignment problem. Lo et al. (2006) extended their original model to consider the concept of travel time budget in path choice decision making. Shao et al. (2006) adopted a similar postulation of the central limit theorem to derive the normal distribution of the path travel time but with O-D demand distribution (normal distribution). Szeto and Sumalee (2011) presented a cell-based model for multi-class stochastic dynamic traffic assignment. Ng and Waller (2012) presented a dynamic traffic assignment model based on the cell transmission model that accounts for travel time reliability. Uchida (2014) proposed a model which simultaneously estimates the value of travel time and of travel time reliability based on the risk-averse driver's route choice behavior. Castillo et al. (2014) dealt with the probabilistic and physical consistency of traffic-related random variables and models.

In calculating travel time reliability, as is often the case where a stochastic network model cannot be applied due to the lack of network data. In general, in calculating travel time reliability without applying a stochastic network model, stochastic properties of traffic flows need to be estimated. Hazelton (2000, 2001) presented methods to estimate O-D (or path) flows from observed link flows by assuming the O-D flows followed Poisson distributions. Duthie et al. (2011) presented a method for evaluating future travel demand uncertainty and finding an efficient technique for generating multiple realizations of demand. Ghosh et al. (2010) proposed a random process traffic volume model that enables estimation and prediction of traffic volume at sites where large and continuous data sets of traffic condition are unavailable.

Various transportation analysis models that adopted a frequency analysis techniques or/and a soft computing

techniques have been proposed. Karim and Adeli (2003) proposed a case-based reasoning model for freeway work zone traffic management. Jiang and Adeli (2003) proposed a similar freeway work zone traffic management model which took into account the effect of seasonal variation travel demand. Wavelet and neural network have been applied to various fields of transportation research (Adeli and Jiang, 2009; Adeli and Karim, 2005; Dharia and Adeli, 2003; Jiang and Adeli, 2004; Jiang and Adeli, 2005; Karim and Adeli, 2002).

This paper starts by formulating a model that calculates travel time reliability considering uncertainties in demand and supply. The model basically follows the framework of Clark and Watling (2005), which assumes a stochastic O-D demand. Further, multivariate normal distribution of traffic capacities (Uchida and Munehiro, 2010) is introduced to the framework of Clark and Watling (2005). The link flows calculated by the stochastic network model follow a multivariate normal distribution (Hazelton 2000, 2001; Nakayama and Takayama, 2003; Clark and Watling, 2005; Shao et al., 2006; Uchida and Munehiro, 2010). For the purpose of calculating travel time reliability in the network, path flows are required. Therefore, path flows calculated by a stochastic network model can depend on a path set enumerated. Unrealistic flow pattern can be obtained even though an implicit route enumeration algorithm, e.g., Papola and Marzano (2013), is applied. Beside that, there may be a case where a stochastic network model cannot be applied due to the lack of network data. Therefore, a path flow estimation model which does not require path enumeration is also proposed in this paper.

The paper has five more sections. The next section formulates the link flows and link capacities in the network. The third section presents a travel time reliability estimation model that expresses uncertainties in the demand and supply sides. Section four presents the path flow estimation model by using observed link flows. Section five carries out the numerical experiments for demonstrating the model presented in this study. The final section concludes the paper and discusses future research needs.

## 2 LINK FLOW AND LINK CAPACITY

### 2.1. Notation

The notations below are frequently used in this paper.

$N$ :	Set of nodes in the network
$A$ :	Set of links in the network
$I$ :	Set of O-D pairs
$J_i$ :	Set of paths between O-D pair $i$
$o_i$ :	Origin centroid node for O-D pair $i$
$d_i$ :	Destination centroid node for O-D pair $i$
$\delta_{aj}$ :	Variable that equals 1 if link $a$ is part of path $j$ , and equals 0 otherwise

- $L_n^{in}$ : Set of links whose destination node is  $n$   
 $L_n^{out}$ : Set of links whose origin node is  $n$   
 $n_o^a$ : Origin node of link  $a$   
 $n_d^a$ : Destination node of link  $a$   
 $v_a$ : Link flow of link  $a$   
 $V_a$ : Stochastic link flow of link  $a$   
 $\mathbf{V}$ : Vector of stochastic link flows  
 $\Sigma$ : Variance-covariance (var-cov) matrix of stochastic link flows  
 $C_a$ : Stochastic traffic capacity of link  $a$   
 $c_a$ : Traffic capacity of link  $a$   
 $F_{ij}$ : Stochastic flow of route  $j$  between O-D pair  $i$   
 $f_{ij}$ : Mean flow of route  $j$  between O-D pair  $i$   
 $Q_i$ : Stochastic traffic demand for O-D pair  $i$   
 $q_i$ : Mean traffic demand for O-D pair  $i$   
 $p_{ij}$ : Path choice probability of traffic demand for O-D pair  $i$  choosing path  $j$

## 2.2. Link flow and link capacity

We assume that link flows in a network follow multivariate normal distribution (MVN). This assumption is supported if O-D flows follow independent Normal distributions (e.g., Shao et al., 2006), independent Poisson distributions (e.g., Hazelton 2000, 2001; Clark and Watling, 2005; Uchida and Munehiro, 2010) or independent Binomial distributions (e.g., Nakayama and Takayama, 2003). Following Sumalee and Xu (2011), O-D travel demand,  $Q_i$ , is assumed to be an independent random variable with the mean of  $E[Q_i] = q_i$  and the variance of  $\text{var}[Q_i] = VMR_i \cdot q_i$ , where  $VMR_i$  is the variance to mean ratio of random travel demand. By applying the same assumption that the coefficient of variation of path flow is equal to that of O-D demand as employed in Chen et al. (2011), stochastic flow on path  $j \in J_i$ ,  $F_{ij}$ , is then given by

$$F_{ij} = p_{ij} \cdot Q_i \quad \forall i \in I, \forall j \in J_i$$

$F_{ij}$  is the random variable with the mean of  $E[F_{ij}] = p_{ij} \cdot E[Q_i]$  and the covariance of  $\text{cov}[F_{ij}, F_{ik}] = p_{ij} \cdot p_{ik} \cdot \text{var}[Q_i]$  where  $p_{ij} (j \in J_i)$  is path choice probability which can be determined by a path choice model, e.g. UE, SUE and so on. It may be helpful to present a relationship between  $p_{ij}$  and  $VMR_i$  when we take into account the effect of  $p_{ij}$  on travel time reliability which is given by

$$\text{var}[F_{ij}] = \text{var}[p_{ij} \cdot Q_i] = (p_{ij})^2 \cdot \text{var}[Q_i] = VMR_{ij} \cdot E[F_{ij}]$$

where  $VMR_{ij} = p_{ij} \cdot VMR_i$  is the variance to mean ratio of random path flow. The conservation of the path flow variance in relation to the O-D demand variance holds as follows

$$\begin{aligned} \text{var}[Q_i] &= \sum_j \text{var}[F_{ij}] + \sum_{j_1} \sum_{j_2} \text{cov}[F_{ij_1}, F_{ij_2}] \\ &= (p_{ij})^2 \cdot \text{var}[Q_i] + \sum_{j_1} \sum_{j_2 \neq j_1} p_{ij_1} \cdot p_{ij_2} \cdot \text{var}[Q_i] \\ &= \text{var}[Q_i] \cdot \left( \sum_j p_{ij} \right)^2 = VMR_i \cdot q_i \end{aligned}$$

Stochastic flow of link  $a$ ,  $V_a$ , is given by:

$$V_a = \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot F_{ij} \quad \forall a \in A$$

By applying the stochastic O-D flow assumed above, the link flows follow the multivariate normal distribution with the mean of:

$$\begin{aligned} E[V_a] &= \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot E[F_{ij}] \\ &= \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot p_{ij} \cdot E[Q_i] \quad \forall a \in A \end{aligned} \quad (1)$$

and the var-cov of:

$$\text{cov}[V_a, V_b] = \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot \delta_{bj} \cdot \text{var}[F_{ij}] \quad \forall a, b \in A \quad (2)$$

We assume that link capacities  $C_a (a \in A)$  follow a multivariate normal distribution with the mean capacities of  $E[C_a] (a \in A)$  and the var-cov of  $\text{cov}[C_a, C_b] (a, b \in A)$ . The link capacities can be estimated by measuring traffic flow data (Uchida and Munehiro, 2010).

## 3. TRAVEL TIME RELIABILITY

In this study, link travel time is represented by the following BPR function (Bureau of Public Roads, 1964):

$$t_a(v_a; c_a) = t_a^0 \cdot \left( 1 + \gamma \cdot \left( \frac{v_a}{c_a} \right)^\lambda \right) \quad (3)$$

where  $t_a^0$  is the free flow travel time of link  $a$ , and  $\gamma$  and  $\lambda$  are calibration parameters. By substituting  $v_a$  and  $c_a$  in Eq. (3) with  $V_a$  and  $C_a$ , respectively, we obtain:

$$t_a(V_a; C_a) = t_a^0 \cdot \left( 1 + \gamma \cdot (V_a)^\lambda \cdot (C_a^{-1})^\lambda \right) \quad (4)$$

By denoting  $D_a = V_a \cdot C_a^{-1}$ , Eq. (4) is then:

$$t_a(V_a; C_a) \Rightarrow \tilde{t}_a(D_a) = t_a^0 \cdot \left( 1 + \gamma \cdot (D_a)^\lambda \right) \quad (5)$$

By performing a  $m$ th ( $m \geq 1$ ) order Taylor expansion on Eq. (5) at  $D_a = E[D_a]$ , following Clark and Watling (2005), we obtain:

$$\tilde{t}_a(D_a) = \sum_{k=0}^m b_{ka} \cdot (D_a - d_a)^k$$

where  $d_a = E[D_a]$ , and  $b_{ka}$  is the coefficient of the  $k$ th term of the Taylor expansion which is given by:

$$b_{ka} = \frac{1}{k!} \cdot \left. \frac{\partial \tilde{t}_a^{(k)}(D_a)}{\partial D_a^{(k)}} \right|_{D_a=d_a} = \begin{cases} t_a(E[V_a]) \cdot E[C_a] & \text{if } k = 0 \\ \gamma \cdot t_a^0 \cdot \frac{\prod_{l=1}^k (\lambda - l + 1)}{k!} \cdot d_a^{\lambda-k} & \text{otherwise} \end{cases}$$

Mean link travel times are then calculated as:

$$E[\tilde{t}_a(D_a)] = \sum_{k=0}^m b_{ka} \cdot E[(D_a - d_a)^k] = \begin{cases} \sum_{k=0}^{m/2} b_{2k,a} \cdot (\text{var}[D_a])^k & m : \text{even number} \\ \sum_{k=0}^{(m-1)/2} b_{2k,a} \cdot (\text{var}[D_a])^k & m : \text{odd number} \end{cases} \quad (6)$$

The var-cov of the link travel times are:

$$\begin{aligned} & \text{cov}[\tilde{t}_a(D_a), \tilde{t}_b(D_b)] \\ &= \sum_{k=0}^m \sum_{l=0}^m b_{ka} \cdot b_{lb} \cdot E[(D_a - d_a)^k \cdot (D_b - d_b)^l] \\ & \quad - E[\tilde{t}_a(D_a)] \cdot E[\tilde{t}_b(D_b)] \\ &= \left( \sum_{k=1}^m \phi_{ab}^k \right) \cdot \delta_a \cdot \delta_b \quad \forall a, b \in A, \end{aligned} \quad (7)$$

where

$$\phi_{ab}^k = k \cdot \frac{\prod_{l=1}^k (\lambda - l + 1)}{k! \cdot d_a^k} \cdot \frac{\prod_{s=1}^k (\lambda - s + 1)}{k! \cdot d_b^k} \cdot (\text{cov}[D_a, D_b])^k \quad \forall a, b \in A \quad (8)$$

$$\delta_a = t_a(E[V_a]) \cdot E[C_a] - t_a^0 \quad \forall a \in A$$

Eq. (6) and Eq. (7) can be calculated by applying a method originally proposed by Isserlis (1918) given the moments of  $D_a$ . The calculation of  $\text{cov}[D_a, D_b]$  ( $a, b \in A$ ) in Eq. (8) is presented in Appendix 1. The travel times of path  $j$  which serves O-D pair  $i$  ( $\Xi_{ij}$ ) are given by:

$$\Xi_{ij} = \sum_{a \in A} \tilde{t}_a(D_a) \cdot \delta_{aj} \quad \forall i \in I, \forall j \in J_i$$

The mean path travel time and the var-cov of the path travel times are respectively given by:

$$E[\Xi_{ij}] = \sum_{a \in A} E[\tilde{t}_a(D_a)] \cdot \delta_{aj} \quad \forall i \in I, \forall j \in J_i \quad (9)$$

$$\begin{aligned} \text{cov}[\Xi_{i_1 j_1}, \Xi_{i_2 j_2}] &= \sum_{a_1 \in A} \sum_{a_2 \in A} \text{cov}[\tilde{t}_{a_1}(D_{a_1}) \cdot \delta_{a_1 j_1}, \tilde{t}_{a_2}(D_{a_2}) \cdot \delta_{a_2 j_2}] \\ &= \sum_{a_1 \in A} \sum_{a_2 \in A} \delta_{a_1 j_1} \cdot \delta_{a_2 j_2} \cdot \text{cov}[\tilde{t}_{a_1}(D_{a_1}), \tilde{t}_{a_2}(D_{a_2})] \\ & \quad \forall i_1, i_2 \in I, \forall j_1 \in J_{i_1}, \forall j_2 \in J_{i_2} \quad (10) \end{aligned}$$

For the calculation of the total travel time in the network, one can refer to Clark and Watling (2005). They calculate the moments of the stochastic total travel time by applying the method proposed by Isserlis (1918).

Risk-averse drivers may take into account the mean travel times as well as the variations of the travel times, such as standard deviations, in their path choice decisions. Such path choice problems are formulated as path-based problems due to the non-separable property of path travel time variance (or covariance) shown by Eq. (10). These problems can be solved by a stochastic network model. In general, path enumeration is required in advance for solving these path-based problems. For the purpose of calculating the unique travel time reliability in the network, unique path flows need to be provided. However, path flows are determined depending on a path set enumerated in advance. This means that travel time reliability in the network can change depending on the path set enumerated in advance. In addition, due to the lack of network data, there may be a case where a stochastic network model cannot be applied to calculate travel time reliability. Therefore, if several sets of observed link flows are available, it may be reasonable to estimate the path flow based on the observed link flows instead of solving the path choice problems. The next section presents a path flow estimation method based on the observed link flows. This model does not require path enumeration in advance.

#### 4 PATH FLOW ESTIMATION

We assume that mean link flows as well as several sets of observed link flows in a network are given. The mean flow of a link in the network may be estimated by using the observed flows of the link, e.g., the mean link flow can be an average of the observed flows of the link. Stochastic path flow estimation model presented in this section does not require all link flows to be observed in the network as discussed later. Considering the evolutions of ITS technology, it may be possible to observe a large number of link flows in the road network.

Following Sumalee and Xu (2011), the  $VMR_{ij}$  of path flows are assumed to equal to those of the corresponding stochastic O-D flow, i.e.,  $VMR_i = VMR_{ij}$ . This relationship is obtained when we regard  $F_{ij}$  as an independent random variable that follows the same statistical distribution as O-D

flow. Based on this assumption, the conservation of the path flow variance in relation to the O-D demand variance still holds as follows

$$\sum_j VMR_{ij} \cdot E[F_{ij}] = \sum_j VMR_i \cdot (p_{ij}) \cdot E[Q_i] = VMR_i \cdot q_i$$

In addition, the  $VMR_i$  of O-D demand flows are assumed to be the same for all O-D pairs, i.e.,  $VMR_i = VMR \ \forall i \in I$ . These assumptions are required to derive the expression of covariance of two link flows by using only the mean path flows,  $f_{ij} = E[F_{ij}]$ , with given  $VMR$ , shown by

$$\text{cov}[V_a, V_b] = VMR \cdot \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot \delta_{bj} \cdot f_{ij} \quad \forall a, b \in A$$

Even though these two assumptions do not hold, a path flow estimation model presented in this section can work however the expression of  $\text{cov}[V_a, V_b]$  becomes a more complicated one. These two assumptions are provided to keep the path flow estimation model as simple as possible.

Based on the definition of  $\text{cov}[V_a, V_b]$  shown by Eq. (2),  $\text{cov}[V_a, V_b]$  is the summation of the mean path flows that pass through both links  $a$  and  $b$  multiplied by  $VMR$ . Since no cyclic path flow is assumed, we obtain:

$$\text{cov}[V_a, V_b] = 0 \quad \forall a \neq b \in L_n^{out}, \forall n \in N \quad (11)$$

$$\text{cov}[V_a, V_b] = 0 \quad \forall a \neq b \in L_n^{in}, \forall n \in N \quad (12)$$

$$\text{cov}[V_a, V_b] = 0 \quad \text{if } n_o^a = n_d^b \ \& \ n_d^a = n_o^b \quad \forall a (\neq b) \in A \quad (13)$$

Conservation of link flows, at node  $n$ , that pass through link  $a$  is given by:

$$\sum_{b \in L_n^{in}} \text{cov}[V_a, V_b] = \sum_{c \in L_n^{out}} \text{cov}[V_a, V_c] \quad \forall a \in A, \forall n \in N \quad (14)$$

Eq. (14) implies that the summation of link flows that pass through link  $a$  and then flow into node  $n$  via one of the links  $b \in L_n^{in}$  has to flow out from the node via one of the links  $c \in L_n^{out}$ .

We set the following four assumptions on the centroid nodes from/to which O-D demand is generated/attracted:

$$L_{o_i}^{in} = \emptyset \quad \forall i \in I$$

$$o_i \neq o_{i'} \quad \forall i, i' (i \neq i') \in I$$

$$L_{d_i}^{out} = \emptyset \quad \forall i \in I$$

$$d_i \neq d_{i'} \quad \forall i, i' (i \neq i') \in I$$

The above four assumptions imply that a unique origin centroid node and a unique destination centroid node are provided at each O-D pair, and that there is no path flow that passes through the centroid nodes. By putting the conservation of O-D flows together with these assumptions on the centroid nodes, we have the following three conservation laws with respect to O-D flows:

$$E[V_a] = VMR \cdot \sum_{b \in L_{d_i}^{in}} \text{cov}[V_a, V_b] \quad \forall i \in I, \forall a \in L_{o_i}^{out} \quad (15)$$

$$\sum_{a \in L_{o_i}^{out}} E[V_a] = VMR \cdot \sum_{a \in L_{o_i}^{out}} \sum_{b \in L_{d_i}^{in}} \text{cov}[V_a, V_b] = q_i \quad \forall i \in I \quad (16)$$

$$\sum_{c \in L_{o_i}^{out}} \text{cov}[V_a, V_c] = \sum_{b \in L_{d_i}^{in}} \text{cov}[V_a, V_b] \quad \forall i \in I, \forall a \in A \quad (17)$$

Eq. (15) implies that the path flow which is generated from the centroid node  $o_i$  and which flows into one of the links  $a \in L_{o_i}^{out}$  has to be attracted to the centroid node  $d_i$  via one of the links  $b \in L_{d_i}^{in}$ . Eq. (16) implies that the summation of Eq. (15) with respect to  $a \in L_{o_i}^{out}$  is equal to the mean O-D flow  $q_i$ . Eq. (17) implies that the summation of path flows that flow out from the links  $c \in L_{o_i}^{out}$  and then flow into link  $a$  have to be attracted to the centroid node  $d_i$  via one of the links  $b \in L_{d_i}^{in}$ . Since a pair of centroid nodes includes the unique origin node and the unique destination node, we obtain the following three conditions on the link flows whose origin node or destination node is a centroid node:

$$\text{cov}[V_a, V_b] = 0 \quad \forall a \in L_{o_i}^{out}, \forall b \in L_{d_{i_2}}^{in} \quad (\forall i_1 (\neq i_2) \in I) \quad (18)$$

$$\text{cov}[V_a, V_b] = 0 \quad \forall a \in L_{o_{i_1}}^{out}, \forall b \in L_{o_{i_2}}^{out} \quad (\forall i_1 (\neq i_2) \in I) \quad (19)$$

$$\text{cov}[V_a, V_b] = 0 \quad \forall a \in L_{d_{i_1}}^{in}, \forall b \in L_{d_{i_2}}^{in} \quad (\forall i_1 (\neq i_2) \in I) \quad (20)$$

As mentioned earlier, the vector of the mean link flows  $E[\mathbf{V}]$  and the diagonal elements of the var-cov matrix  $\Sigma$  are given. We have to obtain the off-diagonal elements of  $\Sigma$  that satisfy the conditions shown by Eqs. (11) - (20) for the purpose of calculating the path flows. We assume that there exist some observed vectors of mean link flows  $\mathbf{v}_k (k=1, \dots, K)$ . The off-diagonal elements of  $\Sigma$ , i.e.,  $\text{cov}[V_a, V_b] (a (\neq b) \in A)$ , can be then estimated by maximizing the following log-likelihood function for  $MVN(E[\mathbf{V}], \Sigma)$ :

$$\ln L = -\frac{K \cdot |A|}{2} \cdot \log(2\pi) - \frac{K}{2} \cdot \log|\Sigma| - \frac{1}{2} \cdot \sum_{k=1}^K (\mathbf{v}_k - E[\mathbf{V}])^T \Sigma^{-1} (\mathbf{v}_k - E[\mathbf{V}])$$

subject to Eqs. (11) - (20), where  $|A|$  is the number of links and the superscript T denotes a matrix or vector transposition. For each row of  $\Sigma$ , the number of unknown covariance is  $|A| - 1$ , since each row contains a diagonal element that is given. Since at each node but excluding centroid nodes we have Eq. (14), the total number of linear equations we obtain is  $(|N| - |N_c|)$ , where  $|N|$  is the number of nodes, and  $|N_c|$  is the number of centroid

nodes. Therefore, for each row, the number of unknown covariance is  $|A| - 1 - (|N| - |N_c|)$ . Thus, total number of unknown covariance in  $\Sigma$  seems to be  $|A| \cdot (|A| - 1 - (|N| - |N_c|))$ . However, at each link  $L_{o_i}^{out} (i \in I)$ , we have Eq. (15). In addition, the var-cov matrix is always a symmetric matrix, i.e.,  $\text{cov}[V_a, V_b] = \text{cov}[V_b, V_a] \forall a, b$ . Consequently, the number of unknown covariance in  $\Sigma$  is less than:

$$\frac{1}{2} \cdot \left( |A| \cdot (|A| - 1 - (|N| - |N_c|)) - \sum_{i \in I} |L_{o_i}^{out}| \right)$$

where  $|L_{o_i}^{out}| (i \in I)$  is the number of links whose origin node is  $o_i$ , since some of covariance in  $\Sigma$  are given by Eqs. (11) - (13) and (18) - (20).

Due to the conditions of flow conservation shown by Eqs. (14) and (15), there may be a case in which the log-likelihood function cannot be defined since the var-cov matrix  $\Sigma$  becomes singular, e.g.,  $\text{cov}[V_a, V_c] (a \in A)$  for a link  $c$  from the set  $L_n^{in} \cup L_n^{out}$  is determined by using  $\text{cov}[V_a, V_b] (b \neq c) \in L_n^{in} \cup L_n^{out}, a \in A$ . For the purpose of ensuring that the var-cov matrix is regular, we consider the problem where one of the links  $a \in L_n^{in} \cup L_n^{out}$  is reduced at each node  $n \in N$ , and where one of the links  $a \in L_{o_i}^{in} \cup L_{d_i}^{out}$  is reduced at each O-D pair  $i \in I$ , from  $\mathbf{V}$ ,  $\mathbf{v}_k$  and  $\Sigma$ . Note that without loss of generality, the covariance that relate to the reduced links can be duplicated by using the covariance of the other links. Let  $\hat{\mathbf{V}}$ ,  $\hat{\mathbf{v}}_k$  and  $\hat{\Sigma}$  denote the reduced versions of  $\mathbf{V}$ ,  $\mathbf{v}_k$  and  $\Sigma$ , respectively. The log-likelihood function for  $MVN(E[\hat{\mathbf{V}}], \hat{\Sigma})$  is then:

$$\begin{aligned} \ln \hat{L} = & -\frac{K \cdot |\hat{A}|}{2} \cdot \log(2\pi) - \frac{K}{2} \cdot \log|\hat{\Sigma}| \\ & - \frac{1}{2} \cdot \sum_{k=1}^K (\hat{\mathbf{v}}_k - E[\hat{\mathbf{V}}])^T \hat{\Sigma}^{-1} (\hat{\mathbf{v}}_k - E[\hat{\mathbf{V}}]) \end{aligned} \quad (21)$$

where  $|\hat{A}|$  is the number of links in the reduced problem. By maximizing Eq. (21) subject to Eqs. (11) - (20), we obtain the unknown off-diagonal elements of  $\hat{\Sigma}$ . This discussion means that the model proposed in this study does not require all link flows to be observed in the network considering the conservation of all link flows connected to one node. In addition, even though some of link flows are not observed, such unobserved link flows can be estimated by applying a link flow estimation model with incomplete data, e.g., Tanyimboh and Templeman (1993).

It is expected that there may still be a case in which  $\hat{\Sigma}$  becomes singular when some elements of  $\hat{\Sigma}$  take specific values. Note that even though  $\hat{\Sigma}$  becomes singular, such  $\hat{\Sigma}$  do exist as the results of the driver's path choice. In this case, it may be reasonable to apply the log-likelihood function for singular MVN distribution (e.g., Srivastava and Rosen, 2002) given by:

$$\begin{aligned} \ln \tilde{L} = & -\frac{K \cdot \rho}{2} \cdot \log(2\pi) - \frac{K}{2} \cdot \log \left( \prod_{i=1}^{\rho} \lambda_i \right) \\ & - \frac{1}{2} \cdot \sum_{k=1}^K (\hat{\mathbf{v}}_k - E[\hat{\mathbf{V}}])^T \hat{\Sigma}^{-} (\hat{\mathbf{v}}_k - E[\hat{\mathbf{V}}]) \end{aligned} \quad (22)$$

where  $\lambda_i (i=1, \dots, \rho)$  are the nonzero eigenroots of  $\hat{\Sigma}$ , and  $\hat{\Sigma}^{-}$  is any generalized inverse of  $\hat{\Sigma}$  that satisfies  $\hat{\Sigma} \hat{\Sigma}^{-} \hat{\Sigma} = \hat{\Sigma}$ . If a Moore-Penrose inverse is applied, the generalized inverse is then:

$$\hat{\Sigma}^{-} = \Gamma \Lambda^{-1} \Gamma^T$$

where  $\Gamma$  is the  $|\hat{A}| \times \rho$  matrix of the eigenvectors of  $\hat{\Sigma}$ ; and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_\rho)$  is  $\rho \times \rho$  diagonal matrix. Note that if  $\hat{\Sigma}$  is not singular, then the Moore-Penrose inverse is equivalent to  $\hat{\Sigma}^{-1}$  ( $\hat{\Sigma}^{-} = \hat{\Sigma}^{-1}$ ) and we obtain a resulting relationship that implies that Eq. (21) is equal to Eq. (22) if  $\hat{\Sigma}$  is not singular. Thus, without loss of generality, Eq. (22) can be used for estimating the covariance of  $\hat{\Sigma}$ . As is often the case with the log-likelihood function denoted by Eq. (22) in which the var-cov matrix can change from non-singular to singular or vice versa, at a singular point the log-likelihood function may not be continuous. Furthermore, if the singular point maximizes the log-likelihood function, we have to examine the stability of the solution, which is the subject of the next section.

It should be noted that the mean link flows need to hold Eqs. (11) - (20). When the mean link flows do not hold these conditions, we can still estimate the mean link flows that hold the conditions by applying a link flow estimation model with incomplete data, e.g., a model presented in Tanyimboh and Templeman (1993). On the other hand, the observed link flows do not need to hold the conditions since the mean and var-cov of stochastic link flow can be defined in the path flow estimation model. It may be reasonable to apply a maximum likelihood (ML) estimation method with lower unknown parameters rather than the one with larger unknown parameters when the number of observations is small. In this regard, the model presented in this section does not require a number of observations compared to the standard ML estimation method in which the number of unknown parameters is  $|A| + (1/2) \cdot |A| \cdot (|A| + 1)$ . However, similar to the standard ML estimation model, the variance of an estimated parameter becomes larger when the number of observations is small. Also, it should be noted that the

proposed model can work for uncongested networks rather than congested networks since the model does not consider the constraint from static models.

## 5 NUMERICAL EXPERIMENTS

In this section, two numerical experiments are carried out to demonstrate the model proposed in this study. In the first experiment, a network with an O-D pair is addressed whereas a network with two O-D pairs is addressed in the second experiment. In the second experiment, stochastic path flows as well as path travel time distributions are estimated. In the numerical experiments, we adopt the variance to mean ratio of random travel demand of one, i.e.,  $VMR = 1$ . It seems that only mean traffic demand(s) is/are addressed in the numerical experiments. However, we take into account the effect of uncertainty in the traffic demand on the path flow by introducing several sets of the observed link flows. Note that, we assume that a set of observed link flows is sampled from a multivariate normal distribution.

### 5.1 An O-D pair

Figure 1 shows test network that consists of three nodes and four links with an O-D pair. The mean O-D demand for this network is 100. We set the following vector of mean link flows:

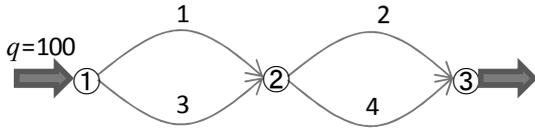


Figure 1 Test network 1

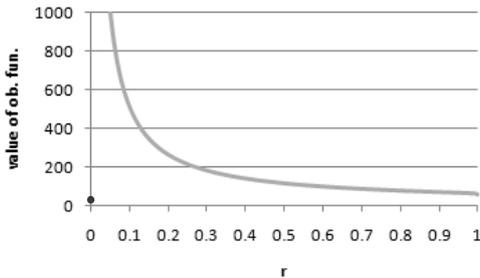


Figure 2 Inverted objective function for the first case

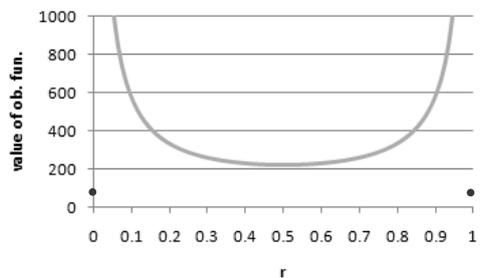


Figure 3 Inverted objective function for the second case

$$E[\mathbf{v}_1] = (50 \ 50 \ 50 \ 50)^T$$

According to Eq. (11), we obtain:

$$\text{cov}[V_1, V_3] (= \text{cov}[V_3, V_1]) = 0, \text{cov}[V_2, V_4] (= \text{cov}[V_4, V_2]) = 0$$

At node 2, Eqs. (14) and (16) are equivalent, and this gives us:

$$\text{cov}[V_a, V_1] + \text{cov}[V_a, V_3] = \text{cov}[V_a, V_2] + \text{cov}[V_a, V_4] \forall a \in A$$

We denote  $\text{cov}[V_1, V_2] = r \cdot \text{var}[V_1]$  by using the unknown parameter  $r$  ( $0 \leq r \leq 1$ ). The resulting var-cov matrix ( $\Sigma_1$ ) is then:

$$\Sigma_1 = \begin{pmatrix} \text{var}[V_1] & \text{cov}[V_1, V_2] & \text{cov}[V_1, V_3] & \text{cov}[V_1, V_4] \\ & \text{var}[V_2] & \text{cov}[V_2, V_3] & \text{cov}[V_2, V_4] \\ & & \text{var}[V_3] & \text{cov}[V_3, V_4] \\ & & & \text{var}[V_4] \end{pmatrix} = \begin{pmatrix} 50 & 50 \cdot r & 0 & 50 \cdot (1-r) \\ & 50 & 50 \cdot (1-r) & 0 \\ & & 50 & 50 \cdot r \\ & & & 50 \end{pmatrix}$$

Note that we can reduce the number of unknown parameters for this problem to one. Next, we set the following two observed vectors  $\mathbf{v}_k$  ( $k=1,2$ ) which are regarded as the link flows observed on day  $k$  given by:

$$\mathbf{v}_1 = (100 \ 100 \ 0 \ 0)^T, \mathbf{v}_2 = (50 \ 50 \ 50 \ 50)^T$$

We prepared two sets of observed link flows for the purpose of showing a clear change in the shape of the objective function after adding one more observation. Of course in this example, since the number of observations is small, the accuracy of an estimated parameter is low.

There are four possible paths: those that consist of links 1 and 2, of links 1 and 4, of links 2 and 3, and of links 3 and 4. We examine the path flow estimation problem where the covariance relating to link 1 are reduced by applying Eq. (14). For this problem, the var-cov matrix  $\hat{\Sigma}_1$  becomes singular at  $r = 0$  or  $1$ . Therefore, the log-likelihood function shown in Eq. (22) must be used for estimating the covariance of  $\hat{\Sigma}_1$ . The resulting problem is to maximize Eq. (22), with respect to  $r$  ( $0 \leq r \leq 1$ ), and subject to:

$$E[\hat{\mathbf{v}}_1] = (E[V_2] = 50 \ E[V_3] = 50 \ E[V_4] = 50)^T$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 50 & 50 \cdot (1-r) & 0 \\ & 50 & 50 \cdot r \\ & & 50 \end{pmatrix}$$

$$\hat{\mathbf{v}}_1 = (100 \ 0 \ 0)^T, \hat{\mathbf{v}}_2 = (50 \ 50 \ 50)^T$$

The value of the objective function multiplied by minus one, which will be referred to as the inverted objective function, for this problem is shown in Figure 2. Thus, the optimal solution is the minimum value of the function that is 32.9 at  $r = 0$ . As predicted, at the singular point of  $r = 0$ ,

the objective function is not continuous, i.e., the value at the singular point is  $-\ln \tilde{L}(r=0)=32.9$  whereas  $-\lim_{r \rightarrow +0} \tilde{L}(r) = \inf$ . Obviously, this solution is not stable in terms of perturbation stability, although stability theory cannot be applied since even the *least convex set* (Filippov, 1960) is not defined at the singular point due to the gradient of  $-\inf$ , i.e.,  $-\partial \lim_{r \rightarrow +0} \tilde{L}(r)/\partial r = -\inf$ . If the *least convex set* is well defined, then practical stability for discontinuous systems (Michel and Porter, 1972), which we do not explain further in this paper, may be applicable for checking the stability of the solution. The solution of  $r=0$  is stable starting only from  $r=0$ . Otherwise, if a small perturbation is made in  $r=0$ , i.e.,  $r=r+\delta r$  where  $\delta r$  is a small feasible positive value, then the other solution of  $r=1$  which is also singular point will be obtained. Further, since at this singular point the objective function is continuous, i.e.,  $-\ln \tilde{L}(r=1) = -\lim_{r \rightarrow 1} \tilde{L}(r) = 86.9$ , this solution is stable in terms of perturbation. The solution which is stable in terms of perturbation for this problem seems to be  $r=1$ , and that gives us:

$$\Sigma_1 = \begin{pmatrix} 50 & 50 & 0 & 0 \\ & 50 & 0 & 0 \\ & & 50 & 50 \\ & & & 50 \end{pmatrix}$$

The solution says that there are two paths: path 1, which consists of the links 1 and 2 with the mean flow of 50; and path 2, which consist of the links 3 and 4 with the mean flow of 50. Due to  $\mathbf{v}_1$ , the two paths that consist of the links 1 and 4 and of the links 2 and 3 are not selected as the used paths.

We examine then the other problem, where  $\mathbf{v}_3 = (0 \ 100 \ 100 \ 0)^T$  is further added to the sets of observed link flows. Figure 3 shows the inverted objective function for this case. From Figure 3, the minimum value of the function is 86.9 at  $r=0$  or 1, which is the singular point. At each singular point, the function is discontinuous. By applying the same discussion provided in the previous paragraph to this problem, the stable solution seems to be  $r=0.5$ , at which the function takes the value of 224.8, which implies that there are two paths with the mean flow of 50: paths that consist of the links 1 and 4, and of the links 2 and 3.

The discussion in this section clarifies that:

- i) at the singular point, the objective function for the path flow estimation problem may not be continuous;
- ii) at the singular point, the objective function can be optimized (maximized); and
- iii) even though the value of the objective function is optimized at the singular point, the solution is not

stable when the objective function is discontinuous at the singular point.

Thus, for the purpose of obtaining a stable solution, the continuity of the objective function at the singular point must be checked. Continuity check of the objective function can be numerically done as follows. For a given singular point  $\tilde{\mathbf{r}}$ , i.e., the vector of unknown parameters at which  $\hat{\Sigma}$  becomes singular, if  $\tilde{L}(\tilde{\mathbf{r}}) = \lim_{\delta \tilde{\mathbf{r}} \rightarrow 0} \tilde{L}(\tilde{\mathbf{r}} + \delta \tilde{\mathbf{r}})$  then the objective function is continuous at the singular point ( $\tilde{\mathbf{r}}$  is a feasible solution), otherwise it is discontinuous ( $\tilde{\mathbf{r}}$  is not a feasible solution), where  $\delta \tilde{\mathbf{r}}$  is the feasible small perturbation vector with nonzero elements.

## 5.2 Multiple O-D pairs

We next consider an example different from the previous one in terms of the number of O-D pairs and of sets of observed link flows. Figure 4 shows a test network of nine nodes and fourteen links with two O-D pairs. The mean demand for O-D pair 1, whose origin centroid node is 1 and whose destination centroid node is 9, is  $q_1=1000$ . The other O-D demand is  $q_2=1000$  for the O-D pair whose origin centroid node is 7 and whose destination centroid node is 3. Considering both the network topology and the locations of the centroid nodes for each O-D pair, two mean flows of links  $i$  and  $ii$  are always same, for example the case of O-D pair 1, and this gives:

$$\text{cov}[i, a] = \text{cov}[ii, a] \quad \forall a \in A$$

This is true for the three pairs of links of (iii, iv), (v, vi) and (vii, viii). Thus, we can get the simplified network shown by Figure 5, where the pair of two links that always carry the same mean link flow is represented by a link. In the figure, link 1 represents links  $i$  and  $ii$  from Figure 4, link 2 represents links  $iii$  and  $iv$ , link 8 represents links  $v$  and  $vi$ , and link 9 represents links  $7$  and  $8$ . If Eq. (14) is applied to the simplified network at node 5, one of the links among 1, 2, 5, 6, 8 and 9 must be reduced. The same discussion can be applicable at node 4, in which case one of the links among 3, 5 and 7 must be reduced, and at node 6, in which case one of the links among 4, 6 and 10 must be reduced. If Eq. (17) is applied to O-D pair 1, one of the links among 1, 3, 9 and 10 must be reduced. However, we cannot further reduce one of links among 2, 4, 7 and 8 by applying Eq. (17) to O-D pair 2, since link 8 has already been reduced. Thus, we consider the problem where links 1, 5, 6, and 8, for example, are reduced from the simplified network. These reduced links are depicted by dashed arrows in Figure 5. Without loss of generality, all of the covariance of link flows for test network 3 can be duplicated by using the covariance relating the links of 2, 3, 4, 7, 9, and 10 in Figure 5 and by applying Eqs. (11) - (20).

We set the observed link flows  $\mathbf{v}_k$  ( $k=1, \dots, 19$ ), the vector of link flows observed on day  $k$ , as shown in Table 1, and the following vector of the mean link flows given by:

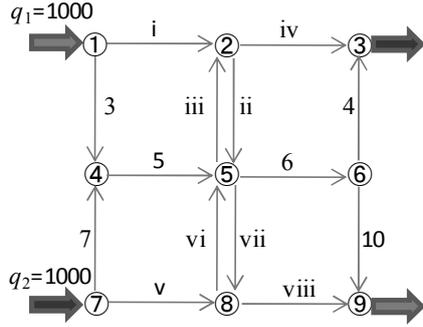


Figure 4 Test network 2

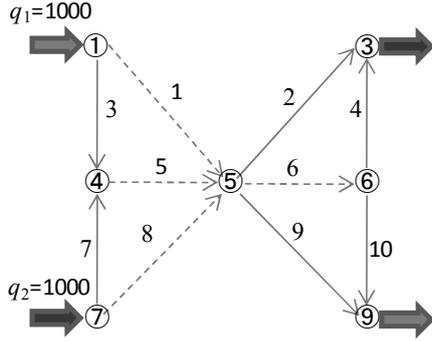


Figure 5 Simplified network

 Table 1 Link flows observed on day  $k$ 

link flow/ day	$E[V_1]$	$E[V_2]$	$E[V_3]$	$E[V_4]$	$E[V_5]$	$E[V_6]$	$E[V_7]$	$E[V_8]$	$E[V_9]$	$E[V_{10}]$
1	1000	1000	0	0	0	0	0	1000	1000	0
2	500	500	500	500	1000	1000	500	500	500	500
3	667	667	333	333	667	667	333	667	667	333
4	500	500	500	500	1000	1000	500	500	500	500
5	600	600	400	400	800	800	400	600	600	400
6	667	667	333	333	667	667	333	667	667	333
7	571	571	429	429	857	857	429	571	571	429
8	625	625	375	375	750	750	375	625	625	375
9	667	667	333	333	667	667	333	667	667	333
10	600	600	400	400	800	800	400	600	600	400
11	636	636	364	364	727	727	364	636	636	364
12	583	583	417	417	833	833	417	583	583	417
13	615	615	385	385	769	769	385	615	615	385
14	643	643	357	357	714	714	357	643	643	357
15	600	600	400	400	800	800	400	600	600	400
16	625	625	375	375	750	750	375	625	625	375
17	647	647	353	353	765	765	412	588	588	412
18	667	667	333	333	722	722	389	611	611	389
19	632	632	368	368	737	737	368	632	632	368

$$E[V_2] = (632 \ 632 \ 368 \ 368 \ 737 \ 737 \ 368 \ 632 \ 632 \ 368)^T$$

The var-cov matrix in which  $\text{cov}[V_a, V_b]$  is simply denoted as  $\text{cov}[a, b]$  is then:

$$\Sigma_2 = \begin{pmatrix} 632 & 0 & 0 & 0 & 0 & \text{cov}[1,6] & 0 & 0 & \text{cov}[1,9] & \text{cov}[1,10] \\ & 632 & 0 & 0 & \text{cov}[2,5] & 0 & \text{cov}[2,7] & \text{cov}[2,8] & 0 & 0 \\ & & 368 & 0 & \text{cov}[3,5] & \text{cov}[3,6] & 0 & 0 & \text{cov}[3,9] & \text{cov}[3,10] \\ & & & 368 & \text{cov}[4,5] & \text{cov}[4,6] & \text{cov}[4,7] & \text{cov}[4,8] & 0 & 0 \\ & & & & 736 & \text{cov}[5,6] & \text{cov}[5,7] & 0 & \text{cov}[5,9] & \text{cov}[5,10] \\ & & & & & 736 & \text{cov}[6,7] & \text{cov}[6,8] & 0 & \text{cov}[6,10] \\ & & & & & & 368 & 0 & 0 & 0 \\ & & & & & & & 632 & 0 & 0 \\ & & & & & & & & 632 & 0 \\ & & & & & & & & & 368 \end{pmatrix}$$

where the given covariance that are obtained from  $E[V_2]$ , Eqs. (11) - (14) and (18) - (20) are denoted by constants. By using two parameters  $0 \leq r_1, r_2 \leq 1$ ,  $\text{cov}[1,10]$  and  $\text{cov}[8,4]$ , which are arbitrarily selected from the unknown covariance, can be respectively denoted as:

$$\text{cov}[1,10] = r_1 \cdot \text{cov}[1,1] = 632 \cdot r_1$$

$$\text{cov}[8,4] = r_2 \cdot \text{cov}[8,8] = 632 \cdot r_2$$

By applying Eqs. (11) - (20), the other unknown covariance can be denoted by using both the parameters and the given covariance (Appendix 2). Thus, the number of unknown parameters in the path flow estimation problem is only two. The mean equilibrium path flows and the var-cov matrix for the reduced problem are respectively given by:

$$E[\hat{V}_2] = (E[V_2] = 632 \ E[V_3] = 368 \ E[V_4] = 368 \ E[V_7] = 368 \ E[V_9] = 632 \ E[V_{10}] = 368)^T$$

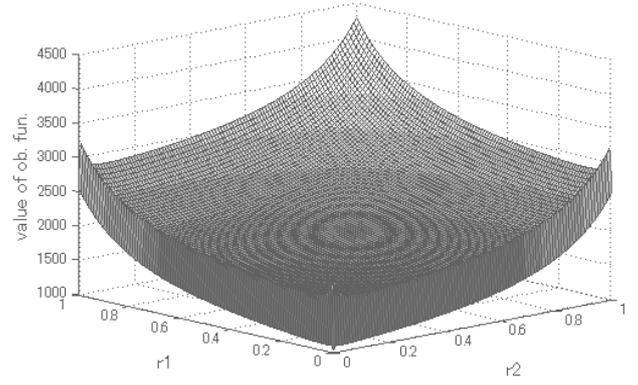
$$\hat{\Sigma}_2 = \begin{pmatrix} 632 & 0 & 0 & \text{cov}[2,7] & 0 & 0 \\ & 368 & 0 & 0 & \text{cov}[3,9] & \text{cov}[3,10] \\ & & 368 & \text{cov}[4,7] & 0 & 0 \\ & & & 368 & 0 & 0 \\ & & & & 632 & 0 \\ & & & & & 368 \end{pmatrix} = \begin{pmatrix} 632 & 0 & 0 & 632 \cdot r_2 & 0 & 0 \\ & 368 & 0 & 0 & 632 \cdot r_1 & 368 - 632 \cdot r_1 \\ & & 368 & 368 - 632 \cdot r_2 & 0 & 0 \\ & & & 368 & 0 & 0 \\ & & & & 632 & 0 \\ & & & & & 368 \end{pmatrix}$$

Figure 6 and Figure 7 show the surface of the inverted objective function and the contour of the function shown in Figure 6, respectively. The surface of the inverted function is calculated by using the iteration relationship:  $r_i^{j+1} = r_i^j + \delta r$  for  $i = 1$  and  $2$  in which the increment  $\delta r$  equals 0.01, which starts from  $r_i^0 = 0$  and ends in  $r_i^{101} = 1.0$ .  $\hat{\Sigma}_2$  becomes singular at  $r_1$  (or  $r_2$ ) = 0 or 1. At the singular point of  $(r_1, r_2) = (0, -)$  or  $(-, 0)$  where the sign - denotes any real numbers between 0 and 1, the objective function is discontinuous (since

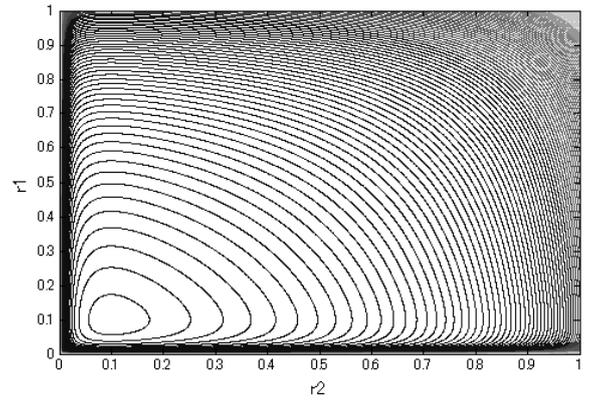
$\lim_{r_1 \rightarrow +0} -\tilde{L}(r_1, -) = \lim_{r_2 \rightarrow +0} -\tilde{L}(-, r_2) = \inf$ ). On the other hand, at the singular point of  $(r_1, r_2) = (1, -)$  or  $(-, 1)$  the objective function is continuous. Thus, the feasible solution region for this problem is  $0 < r_1 \leq 1$  and  $0 < r_2 \leq 1$ .

Path flows are then estimated by maximizing the objective function shown by Eq. (21) and by excluding the unfeasible solution region which is  $(r_1, r_2) = (0, -)$  or  $(-, 0)$ . We obtain the internal solution of  $r_1 = r_2 = 0.1$ . At this solution, the function takes a value of 1419.7. This solution implies that there are four paths for each O-D pair. The first path for the O-D pair 1 (denoted as 1-1) consists of the links i, ii, vii and viii in Figure 4, with the mean path flow of 568.8. The second path for the O-D pair (denoted as 1-2) consists of the links i, ii, 6 and 10, with the mean flow of 63.2. The third path for the O-D pair (denoted as 1-3) consists of the links 3, 5, 6 and 10, with the mean flow of 304.8. The fourth path for the O-D pair (denoted as 1-4) consists of the links 3, 5, vii and viii, with the mean flow of 63.2. Since a path serving O-D pair 2 has the same topology as one of the paths serving O-D pair 1 and carries the same mean path flow as the corresponding path, the paths serving O-D pair 2 are not mentioned further. By using Eq. (9), the mean path travel times for the O-D pair 1  $E[\Xi_{1j}] (j=1, \dots, 4)$  are all calculated as 53; however, their respective variances  $E[\Xi_{1j}] (j=1, \dots, 4)$  calculated by using Eq. (10) are  $4.8^2$ ,  $4.5^2$ ,  $4.8^2$ , and  $4.5^2$ .

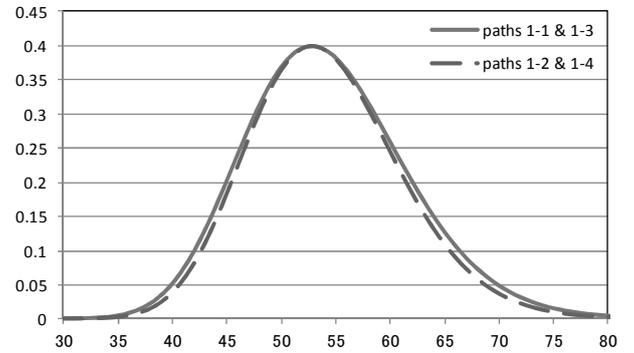
Next, consider one of the other feasible solutions to  $r_1 = r_2 = 0$ . This solution implies that there are two paths for each O-D pair. The first path for the O-D pair 1 (denoted as 1'-1) consists of the links i, ii, vii and viii, with the mean path flow of 632. The second path for the O-D pair (denoted as 1'-2) consists of the links 3, 5, 6 and 10, with the mean path flow of 368. The mean path travel times for the O-D pair 1  $E[\Xi_{1j}] (j=1, 2)$  are all 53, and their variances  $\text{var}[\Xi_{1j}] (j=1, 2)$  are all  $4.8^2$  which is the same as paths 1-1 and 1-3. Thus, depending on the path flows, the distribution of path travel times changes although the mean path travel times do not change.



**Figure 6** Surface of the inverted objective function



**Figure 7** Contour of the function shown in Figure 6



**Figure 8** Distribution of the path travel time

Following Clark & Watling (2005), we carried out curve fitting, i.e., we used a flexible family of probability densities known as *Johnson curves* (Johnson 1949), by adapting the techniques described in Hill et al. (1976) and Hill (1976) for the purpose of obtaining the probability density function (PDF) of path travel time. In calculating travel time distributions, methods described in section 3 were applied. We assume that all links in the network have the same link performance functions in which we set  $t_a^0 = 10$ ,  $\gamma = 0.5$  and  $\lambda = 4.0$ . For the traffic capacities, we

assume that  $E[C_a] = 700, \sqrt{\text{var}[C_a]}/E[C_a] = 0.1$  and  $\text{cov}[C_a, C_b] = 2450 (a, b \in A)$  suggesting a 0.5 coefficient of correlation between the two capacities. A lognormal system was selected in which the parameters presented in Clark & Watling (2005) are estimated as  $\hat{\delta} = 17.9$   $\hat{\gamma} = -30.8$  for the paths 1-1 and 1-3, and  $\hat{\delta} = 16.8$   $\hat{\gamma} = -28.9$  for the paths 1-2 and 1-4 under the assumption on the shift parameter of  $\hat{\xi} = 0$ . Figure 8 shows the estimated PDF of the travel time distribution.

## 6 CONCLUSIONS

This study proposes a model that calculates the travel time reliability in a road network. The sources of uncertainty are O-D demand and traffic capacity. In general, path enumeration is required in advance to estimate travel time reliability by applying a stochastic network model due to non-separable property of travel time reliability. For the purpose of calculating unique travel time reliability, the stochastic network model require unique path flows which are inevitable to depend on a predefined paths set. In addition, due to the lack of network data, there may be a case where we cannot apply a stochastic network model to calculate travel time reliability. Therefore, this study developed a model that estimates path flows without requiring path enumeration. This model is developed based on the standard ML estimation method. Numerical experiments were then carried out to demonstrate the proposed model.

Travel time reliability evaluation in a large network is a practical concern from the viewpoint of road policy development. One approach may be the development of an efficient algorithm for sophisticated but complicated models that are difficult to be solved when targeting a large network problem. The other approach may be the simplification of the models themselves into a model for which efficient algorithms have already been developed. If a stochastic network model cannot be applied due to the lack of network data, we have to apply different approaches for travel time reliability evaluation. An approach is the application of the path flow estimation model presented in this study to the travel time reliability evaluation. The path flow estimation problems addressed in this study have the two advantages in terms of computational load over a problem to which the standard ML estimation method needs to be applied. The first one is that a vector of averages which needs to be estimated by the standard ML estimation method is given. The second one is that diagonal elements of a var-cov matrix which need to be estimated by the standard ML estimation method are also given. Constraints of off-diagonal elements of the var-cov matrix are obtained by Eqs. (11) - (20). It is likely that the var-cov matrix can be obtained even though this process may be time

consuming for targeting a large network. Therefore, it can be concluded that the path flow estimation problems addressed in this study are easier to be solved than a problem to which the standard ML estimation method needs to be applied. Thus, we speculate that the model for path flow estimation presented in the study can be applied to a moderately-large network problem through the use of the property of the var-cov matrix of link flows, i.e., a sparse and symmetric matrix. The time-consuming procedure in path flow estimation is the calculation of the inverse matrix. Efficient calculation methods for the inverse of the sparse and symmetric matrix have been developed (e.g., George and Liu, 1981). As shown by the var-cov matrices used in the numerical experiments, a number of the elements in the var-cov matrix are zero; thus, the number of unknown parameters is equal to the number of non-zero elements in the var-cov matrix. The number of unknown parameters becomes smaller by considering additionally the constraints on flow conservation compared to the number of non-zero elements. By adopting an efficient method for the inverse of the sparse matrix, and by utilizing the property of the path flow estimation problem that the number of unknown parameters can decrease, we can make the path flow estimation problems easier to solve. However, the path flow estimation by applying the model to a large network, e.g., a network in which there are more than one thousand links, is difficult. An improvement of the model is required to estimate the path flow in a large network. That is our future task.

## REFERENCES

- Adeli, H. and Jiang, X., *Intelligent Infrastructure – Neural Networks, Wavelets, and Chaos Theory for Intelligent Transportation Systems and Smart Structures*, CRC Press, Taylor & Francis, Boca Raton, Florida, 2009.
- Adeli, H. and Karim, A., *Wavelets in Intelligent Transportation Systems*, John Wiley and Sons, West Sussex, United Kingdom, 2005.
- Asakura, Y. (1999), Evaluation of network reliability using stochastic user equilibrium. *Journal of Advanced Transportation*, 33, pp. 147-199.
- Bureau of Public Roads (1964), *Traffic Assignment Manual*, U.S. Department of Commerce, Urban Planning Division, Washington D.C.
- Bras, R. L., and Georgakakos, K. P. (1989), Real time nonlinear filtering techniques in streamflow forecasting: a statistical linearization approach, In *Proceedings of the Third International Symposium on Stochastic Hydraulics*, pp. 95-105.
- Cascetta, E. (1989), A stochastic process approach to the analysis of temporal dynamics in transportation networks, *Transportation Research*, 23B(1), pp. 1-17.

- Castillo, E, Calviño, A, Nogal, M., Lo., H. K., On the probabilistic and physical consistency of traffic random variables and models, *Computer-Aided Civil and Infrastructure Engineering*, 29, pp. 496–517.
- Chen, A., Yang, H., Lo, H. K., and Tang, W. H. (2002), Capacity reliability of a road network: an assessment methodology and numerical results, *Transportation Research*, 36B(3), pp. 225-252.
- Clark, S. D., and Watling, D. P. (2005), Modelling network travel time reliability under stochastic demand. *Transportation Research*, 39B(2), pp. 119-140.
- Dharia, A. and Adeli, H. (2003), Neural Network Model for Rapid Forecasting of Freeway Link Travel Time, *Engineering Applications of Artificial Intelligence*, Vol. 16, Nos. 7-8, pp. 607-613.
- Du, Z.-P., and Nicholson, A. (1997), Degradable transportation systems: Sensitivity and reliability analysis, *Transportation Research*, 31B(3), pp. 225-237.
- Duthie, J.C., Unnikrishnan, A., and Waller, S.T. (2011), Influence of demand uncertainty and correlations on traffic predictions and decisions, *Computer-Aided Civil and Infrastructure Engineering*, 26 (1), pp. 16-29.
- Filippov, A. G. (1960), Application of the theory of differential equations with discontinuous right-hand sides to non-linear problems in automatic control, In: *Proceedings of the 1<sup>st</sup> IFAC*, pp. 923-927.
- George, A., and Liu, J.W.H. (1981), *Computer solution of large sparse positive definite systems*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Ghosh, B., Basu, B., and O'Mahony, M. (2010), Random process model for traffic flow using a wavelet - bayesian hierarchical technique, *Computer-Aided Civil and Infrastructure Engineering*, 25:8, pp. 613-624.
- Hazelton, M. (2000), Estimation of origin-destination matrices from link flows in uncongested networks, *Transportation Research Part B*, 34, pp. 549-566.
- Hazelton, M. (2001), Inference for origin-destination matrices: estimation, prediction and reconstruction, *Transportation Research Part B*, 35, pp. 667-676.
- Hill I. D. (1976), Algorithm AS100: normal-Johnson and Johnson-normal transformations, *Applied Statistics*, 25 (2), pp. 190-192.
- Hill I. D., Hill, R., and Holder, R. L. (1976), Algorithm AS99: Fitting Johnson curves by moments, *Applied Statistics* 25 (2), pp. 180-189.
- Isserlis, L. (1918), On a formula for the product-moment coefficient of any order of normal frequency distribution in any number of variables, *Biometrika*, 12 (1-2), pp. 134-139.
- Jiang, X. and Adeli, H. (2003), Freeway Work Zone Traffic Delay and Cost Optimization Model, *Journal of Transportation Engineering*, ASCE, Vol. 129, No. 3, pp. 230-241.
- Jiang, X. and Adeli, H. (2004), Wavelet packet-autocorrelation function method for traffic flow pattern analysis, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19, No. 6 2004, pp. 324-337.
- Jiang, X. and Adeli, H. (2005), Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting, *Journal of Transportation Engineering*, Vol. 131, No. 10, pp. 771-779.
- Karim, A. and Adeli, H. (2002), Incident Detection Algorithm Using Wavelet Energy Representation of Traffic Patterns, *Journal of Transportation Engineering*, ASCE, Vol. 128, No. 3, pp. 232-242.
- Karim, A. and Adeli, H. (2003), CBR Model for Freeway Work Zone Traffic Management, *Journal of Transportation Engineering*, ASCE, Vol. 129, No. 2, pp. 134-145.
- Liu, H. X., He, X., and Recker, W. (2007), Estimation of the time-dependency of values of travel time and its reliability from loop detector data, *Transportation Research*, 41B(4), pp. 448-461.
- Lo, H. K., Luo, X. W., and Siu, B. W. Y. (2006), Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion, *Transportation Research*, 40B(9), pp. 792-806.
- Lo, H. K., and Tung, Y.-K. (2003), Network with degradable links: capacity analysis and design, *Transportation Research*, 37B (4), pp. 345-363.
- Johnson, N. L. (1949), Systems of frequency curves generated by methods of translation, *Biometrika*, 16 (1/2), pp. 149-176.
- Michel A. N., and Porter, D. W. (1972), Practical stability and finite-time stability of discontinuous systems, *IEEE Transactions on Circuit Theory*, 19-2, pp. 123-129.
- Nakayama, S., and Takayama, J. (2003), Traffic network equilibrium model for uncertain demands, *The 82nd Annual Meeting of the Transportation Research Board*, Washington, D.C.
- Ng, M.W. and Waller, S.T. (2012), A dynamic route choice model considering uncertain capacities, *Computer-Aided Civil and Infrastructure Engineering*, 27:4, pp. 231-243.
- Noland, R. B., and Polak, J. W. (2002), Travel time variability: A review of theoretical and empirical issues, *Transport Review*, 22(1), pp. 39-54.
- Papola, A., Marzano, V. (2013), A network generalized extreme value model for route choice allowing implicit route enumeration, *Computer-Aided Civil and Infrastructure Engineering*, 28, pp.560–580.
- SACTRA. (1999) *Transport and the economy*. HMSO, London.
- Shao, H., Lam, W. H. K., and Tam, M. L. (2006), A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand, *Networks and Spatial Economics*, 6(3-4), pp. 173-204.

- Srivastava, M., and Rosen, D. (2002), Regression models with unknown singular covariance matrix. *Linear Algebra and Its Applications*, 354, pp. 255-273.
- Sumalee, A., and Kurauchi, F. (2006), Network capacity reliability analysis considering traffic regulation after a major disaster, *Networks and Spatial Economics*, 6(3-4), pp. 205-219.
- Sumalee, A., Uchida, K. and Lam, W. H. K. (2011), Stochastic multi-modal transport network under demand uncertainties and adverse weather condition, *Transport. Res. Part C*, Vo. 19, pp. 338-350.
- Sumalee, A., Watling, D. P., and Nakayama, S. (2007), Reliable network design problem: the case with uncertain demand and total travel time reliability, *Transportation Research Record*, 1964, pp. 81-90.
- Sumalee, A., Xu, W. (2011), First-best marginal cost toll for a traffic network with stochastic demand, *Transportation Research Part B*, 45(1), pp. 41-59.
- Szeto, W.Y. and Sumalee, A. (2011), A cell-based model for multi-class doubly stochastic dynamic traffic assignment, *Computer-Aided Civil and Infrastructure Engineering*, 26:8, pp. 595-611.
- Szeto, W.Y. and Solayappan, M. (2010), Reliability-based transit assignment for congested stochastic transit networks, *Computer-Aided Civil and Infrastructure Engineering*, 26:4, pp. 311-326.
- Uchida, K., and Munehiro, K. (2010), Impact of stochastic traffic capacity on travel time in road network. *The 89th Annual Meeting of the Transportation Research Board*, Washington, D.C.
- Uchida, K. (2009), User equilibrium assignment models under stochastic demand and supply with perception error. *Doboku Gakkai Ronbunshuu D*, 65 (3), pp. 386-398 (in Japanese).
- Uchida, K. (2014), Estimating the value of travel time and of travel time reliability in road networks, *Transportation Research Part B*, 66, pp.129-147.
- Uchida, K., Sumalee, A., Ho, H.W. (2014), A stochastic multi-modal reliable network design problem under adverse weather conditions, *Journal of Advanced Transportation*, DOI: 10.1002/atr.1266.
- Tanyimboh, T.T. and Templeman, A.B. (1993), Calculating Maximum Entropy Flows in Networks. *The Journal of the Operational Research Society*, 44 (4), pp.383-396.
- Watling, D. P. (2002), A second order stochastic network equilibrium model I: Theoretical foundation, *Transportation Science*, 36(2), 149-166.

## APPENDIX 1

Bras and Georgakakos (1980) proposed an analytical method that approximates the  $m$ th power of  $X$  by applying a linear function of  $X$  and  $E[X]$ , where  $X$

follows a normal distribution and  $m$  is a real number.

According to the method,  $X^m$  can be approximated as:

$$X^m \approx \hat{X}^m = \alpha^{(X,m)} \cdot E[X] + \beta^{(X,m)} \cdot (X - E[X]) \quad (a1)$$

where  $\alpha^{(X,m)}$  and  $\beta^{(X,m)}$  are the calibration parameters of the  $r$ th approximation that are respectively given by:

$$\alpha^{(X,m)} = E[X]^{m-1} \cdot \left( 1 + \sum_{l=1}^r \frac{\prod_{k=0}^{2l-1} (m-k) \cdot \prod_{s=1}^l (2s-1)}{(2l)!} \cdot (cv)^{2l} \right)$$

$$\beta^{(X,m)} = E[X]^{m-1} \cdot \left( \sum_{l=1}^r \frac{\prod_{k=0}^{2l-2} (m-k) \cdot \prod_{s=1}^l (2s-1)}{(2l-1)!} \cdot (cv)^{2l-2} \right)$$

where  $cv$  is the coefficient of the variation of  $X$ . These two parameters can be analytically derived by applying the least mean squared error criterion to  $E[(X^m - \hat{X}^m)^2]$ . By using

this analytical approximation method, the mean and the var-cov of the stochastic variable  $C_a^{-1}$  ( $a \in A$ ) are respectively given by:

$$E[C_a^{-1}] = \alpha_a \cdot E[C_a] \quad \forall a \in A$$

$$\text{cov}[C_a^{-1}, C_b^{-1}] = E[C_a^{-1} \cdot C_b^{-1}] - E[C_a^{-1}] \cdot E[C_b^{-1}]$$

$$= \beta_a \cdot \beta_b \cdot \text{cov}[C_a, C_b] \quad \forall a, b \in A$$

where  $\alpha_a$  and  $\beta_a$  ( $a \in A$ ) are the calibration parameters that are obtained by assuming  $X = C_a$  and  $m = -1$  in Eq. (a1). If the coefficient of variation of  $C_a$  is less than 0.2, it is reported that the approximation method gives sufficient precision (Uchida, 2009). The mean of  $D_a$  are then:

$$E[D_a] = E[V_a \cdot C_a^{-1}]$$

$$= E[(V_a - E[V_a]) - E[V_a]] \cdot ((C_a^{-1} - E[C_a^{-1}]) - E[C_a^{-1}])$$

$$= E[V_a] \cdot E[C_a^{-1}] \quad \forall a \in A.$$

The var-cov between  $D_a$  and  $D_b$  are given by:

$$\text{cov}[D_a, D_b]$$

$$= E[D_a \cdot D_b] - E[D_a] \cdot E[D_b]$$

$$= E\left[ \left( (V_a - E[V_a]) + E[V_a] \right) \cdot \left( (V_b - E[V_b]) + E[V_b] \right) \cdot \left( (C_a^{-1} - E[C_a^{-1}]) + E[C_a^{-1}] \right) \cdot \left( (C_b^{-1} - E[C_b^{-1}]) + E[C_b^{-1}] \right) \right]$$

$$- E[V_a] \cdot E[V_b] \cdot E[C_a^{-1}] \cdot E[C_b^{-1}]$$

$$= \text{cov}[V_a, V_b] \cdot \text{cov}[C_a^{-1}, C_b^{-1}] + E[C_a^{-1}] \cdot E[C_b^{-1}] \cdot \text{cov}[V_a, V_b]$$

$$+ E[V_a] \cdot E[V_b] \cdot \text{cov}[C_a^{-1}, C_b^{-1}]$$

$$\forall a, b \in A$$

## APPENDIX 2

By applying Eq. (15) to all the O-D pairs, we obtain:

$$\text{cov}[8,2] = \text{cov}[8,8] - \text{cov}[8,4] = 632 \cdot (1 - r_2)$$

$$\text{cov}[1,9] = \text{cov}[1,1] - \text{cov}[1,10] = 632 \cdot (1 - r_1)$$

$$\text{cov}[3,10] = \text{cov}[10,10] - \text{cov}[1,10] = 368 - 632 \cdot r_1$$

$$\text{cov}[3,9] = \text{cov}[9,9] - \text{cov}[1,9] = 632 \cdot r_1$$

$$\text{cov}[7,2] = \text{cov}[2,2] - \text{cov}[8,2] = 632 \cdot r_2$$

$$\text{cov}[7,4] = \text{cov}[4,4] - \text{cov}[8,4] = 368 - 632 \cdot r_2$$

By applying Eq. (14) at node 6 and at link 7, we obtain:

$$\text{cov}[7,6] = \text{cov}[7,4] + \text{cov}[7,10] \ \& \ \text{cov}[7,10] = 0$$

$$\Leftrightarrow \text{cov}[7,6] = \text{cov}[7,4]$$

In the same manner, we obtain:

$$\text{cov}[8,6] = \text{cov}[8,4] = 632 \cdot r_2$$

$$\text{cov}[1,6] = \text{cov}[1,10] = 632 \cdot r_1$$

$$\text{cov}[3,6] = \text{cov}[3,10] = 368 - 632 \cdot r_1$$

$$\text{cov}[4,6] = \text{cov}[4,4] - \text{cov}[4,10] \ \& \ \text{cov}[4,10] = 0$$

$$\Leftrightarrow \text{cov}[4,6] = \text{cov}[4,4] = 368$$

$$\begin{aligned} \text{cov}[5,10] &= \text{cov}[5,6] - \text{cov}[5,4] \\ &= 736 - 632 \cdot (r_1 + r_2) - (368 - 632 \cdot r_2) \\ &= 368 - 632 \cdot r_1 \end{aligned}$$

$$\text{cov}[6,10] = \text{cov}[6,6] - \text{cov}[6,4] = 736 - 368 = 368$$

By applying Eq. (14) at node 5 and at link 2, we obtain:

$$\text{cov}[5,2] = \text{cov}[2,2] - (\text{cov}[1,2] + \text{cov}[8,2]) \ \& \ \text{cov}[1,2] = 0$$

$$\Leftrightarrow \text{cov}[5,2] = \text{cov}[2,2] - \text{cov}[8,2] = 632 \cdot r_2$$

In the same manner, we obtain:

$$\text{cov}[5,9] = \text{cov}[9,9] - (\text{cov}[1,9] + \text{cov}[8,9]) \ \& \ \text{cov}[8,9] = 0$$

$$\Leftrightarrow \text{cov}[5,9] = \text{cov}[9,9] - \text{cov}[1,9] = 632 \cdot r_1$$

$$\text{cov}[5,6] = \text{cov}[5,5] + \text{cov}[5,1] + \text{cov}[5,8] - (\text{cov}[5,2] + \text{cov}[5,9])$$

$$\ \& \ \text{cov}[5,1] = \text{cov}[5,8] = 0$$

$$\Leftrightarrow \text{cov}[5,6] = \text{cov}[5,5] - (\text{cov}[5,2] + \text{cov}[5,9])$$

$$= 736 - 632 \cdot (r_1 + r_2)$$

By applying Eq. (14) at node 4 and at link 3, we obtain:

$$\text{cov}[3,5] = \text{cov}[3,3] + \text{cov}[3,7] \ \& \ \text{cov}[3,7] = 0$$

$$\Leftrightarrow \text{cov}[3,5] = \text{cov}[3,3] = 368$$

In the same manner, we obtain:

$$\text{cov}[4,5] = \text{cov}[4,3] + \text{cov}[4,7] \ \& \ \text{cov}[4,3] = 0$$

$$\Leftrightarrow \text{cov}[4,5] = \text{cov}[4,7] = 368 - 632 \cdot r_2$$

$$\text{cov}[5,7] = \text{cov}[5,5] - \text{cov}[5,3] = 736 - 368 = 368$$