國 立 交 通 大 學

運 輸 科 技 與 管 理 學 系

博 士 論 文

以貝氏方法構建與求解
路徑基礎之時變交通指派模式

Time-varying Path-based Traffic Assignment
Using Bayesian Approach

研究生： 藍健綸

指導教授： 卓訓榮 教授

周幼珍 副教授

中華民國九十七年十一月

以貝氏方法構建與求解路徑基礎之時變交通指派模式

學生：藍健綸　　　　　　　　　　　　　　指導教授：卓訓榮

周幼珍

國立交通大學運輸科技與管理學系 博士班

摘 要

現今之交通運輸系統不再是以「大規模的建設」來解決所有的運輸問題，而是朝向一個更細緻化、對環境更友善且能夠永續經營的方式。智慧型運輸系統，整合了電子、通訊、資訊處理等技術，希望能夠透過有效之管理方式來減少交通擁擠狀況。於智慧型運輸系統當中的先進交通管理系統，則需要即時交通狀況之資訊，才能據以進行分析與控制。因此，本研究發展一路徑基礎之動態交通量指派模式，並據此推估路網現況。

大多數動態交通量指派之研究，均專注於探討使用者均衡或系統最佳狀態。但路網現況是否滿足使用者均衡，仍有許多學者存疑。是故，本研究不以使用者均衡指觀點描述交通量指派問題，而是透過線性動態系統構建路徑基礎之動態交通量指派模式。由最基礎之非時變且不考慮旅行時間之模式，逐漸放鬆成為時變且考量旅行時間之模式。過去針對此種動態系統所構建之模式，大多需要歷史之起迄流量資訊、路徑選擇矩陣或狀態轉移矩陣；但在現實環境中，這些資訊不一定能夠順利取得。本研究透過貝氏方法以及卡門濾波兩者之結合，放鬆上述之假設條件，並且提出一整合型演算法求解此模式。為確認此演算法之收斂，本研究亦提出一平行數列收斂性確認方式，作為此演算法之收斂停止條件。由於此演算法需要大量之計算，為增進計算效率，本研究將此演算法以通訊量最小化的目標進行平行化，並執行於平行電腦上。透過真實路網流量資訊，本研究得以驗證此模式之估計與執行效率。

Time-varying Path-based Traffic Assignment Using Bayesian Approach

Student: Chien-Lun Lan Advisors: Hsun-Jung Cho

Yow-Jen Jou

Department of Transportation Technology and Management

National Chiao Tung University

ABSTRACT

Transportation system nowadays is no longer "extraordinary construction" but becomes more elegant, environmental-friendly, and sustainable. Intelligent Transportation System (ITS) integrates the telecommunications, automation, electronics, and information processing system, is considered possessing the potential to solve the traffic congestion problem. Advanced Traffic Management Systems (ATMS), requires real-time traffic condition, is one of the key issues of ITS. Therefore, we suggest a path-based traffic assignment model to describe the network flow status.

Most existing research works on dynamic traffic assignment focus on the user-equilibrium or system optimal. Nevertheless, the existence of such assumption in real world network is questionable to many researchers. An approach without these assumptions while keeping the basic traffic relationship might be useful facing the disequilibria issue. Therefore, we model the dynamic traffic assignment problem with dynamic system approach. Existing researches with this approach usually assume the prior information of O-D matrix, link-proportion matrix, or state transition matrix. In this paper, we relax such assumption by combining Gibbs sampler and Kalman filter in a state space model. A solution algorithm with parallel chain convergence control is proposed and implemented. To enhance its efficiency, a parallel structure is suggested with efficiency and speedup demonstrated using PC-cluster.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Transportation systems play an important role in the development of a country. It provides a source of mobility by the extensive network of freeways, expressways, and streets. In recent years, the mobility has been greatly obstructed by delays due to the congestion on overloaded transportation networks. The economy suffers a great loss owing to the worsening of traffic congestion. Dynamic traffic management is an efficient tool to control traffic congestion, and this tool requires information of traffic states. Traditional static methods are not suitable for this dynamic traffic management task. Therefore, a new assignment model that gives proper information to fulfill the need is worth to develop. In the introduction, we will discuss motivations, objectives, overview, and contributions of this research.

1.1 Research Motivations

Transportation system nowadays is no longer "extraordinary construction" but becomes more elegant, environmental-friendly, and sustainable. Building new highways is no longer a suitable option for every situation due to the prohibitively high costs involved, as well as social, political and environmental concerns. Intelligent Transportation System (ITS) integrates the telecommunications, automation, electronics, and information processing system, is considered possessing the potential to solve the traffic congestion problem. Advanced Traffic Management Systems (ATMS) is one of the key issues of ITS. Proper information, i.e. link flows, and origin-destination (O-D) demands, is a requirement to management the traffic system effectively. Information of can be obtained by comprehensive deployment of surveillance system, but the costs of extensive installation is nearly unbearable. Hence, estimate these precious information with reasonable amount of detectors is a research that worth some attention.

1.2 Research Objectives

The fundamental objective of this dissertation is to address a dynamic stochastic path-based assignment model. Utilizing link traffic counts, this model can give an overview of the network. Existing research works concerning this problem usually assume the existence of user-equilibrium or system optimal condition. Nevertheless, the existence of such assumption in real world network is questionable (Friesz, Bernstein, Mehta, Tobin and Ganhalizadeh, 1994; Friesz and Shah, 2001). An approach without these assumptions while keeping the basic traffic relationship (i.e., path-link incidence matrix) might be useful facing the disequilibria issue. A statistical approach will be introduced in this dissertation to relax such behavioral assumption. Issues concerning the objective include:

1. Development of a path-base assignment model without user behavior assumptions.
2. Modeling the path-base assignment problem with linear time-varying coefficient dynamic system, that addresses the time-dependent path flows.
3. Present a solution algorithm with parallel computing capability to improve computing efficiency.

4. Demonstrate the path-base assignment model with real network data.

1.3 Research Overview

This dissertation is organized as follows. First, the introduction chapter gives an overview of the motivation, research objectives, and overview of this dissertation. Second, a literature review of related researches in the relevant areas. The literature review chapter concerns about topics include: i) Static Trip distribution models and ii) dynamic trip assignment models.

A path-base assignment model with time-invariant coefficient dynamic system is proposed in chapter 3. Essential algorithms for solving this model, including Kalman Filter and Gibbs Sampler, are discussed. Macroscopic traffic flow model that represent the link dynamics is introduced along with finite difference scheme that solve it numerically. In chapter 4, a time-varying coefficient dynamic model with rolling horizon structure addressing the transition matrix is introduced.

Gibbs sampler, a powerful simulation method, draws values of a random variable from a sequence of distribution that converges to a desired target distribution. If used naively, it might give a misleading answer. Therefore, convergence assessments of the Gibbs Sampler with single chain method and multiple chain method are discussed in chapter 5. Gibbs sampler requires tremendous iteration during computation. To make the model more suitable for real-time use, parallel computing is also introduced in chapter 5 to increase the performance. Numerical examples with real data are also discussed in chapter 6. The last chapter presents the conclusions and perspectives of this study. The overview of the dissertation are illustrated in Figure 1.1.

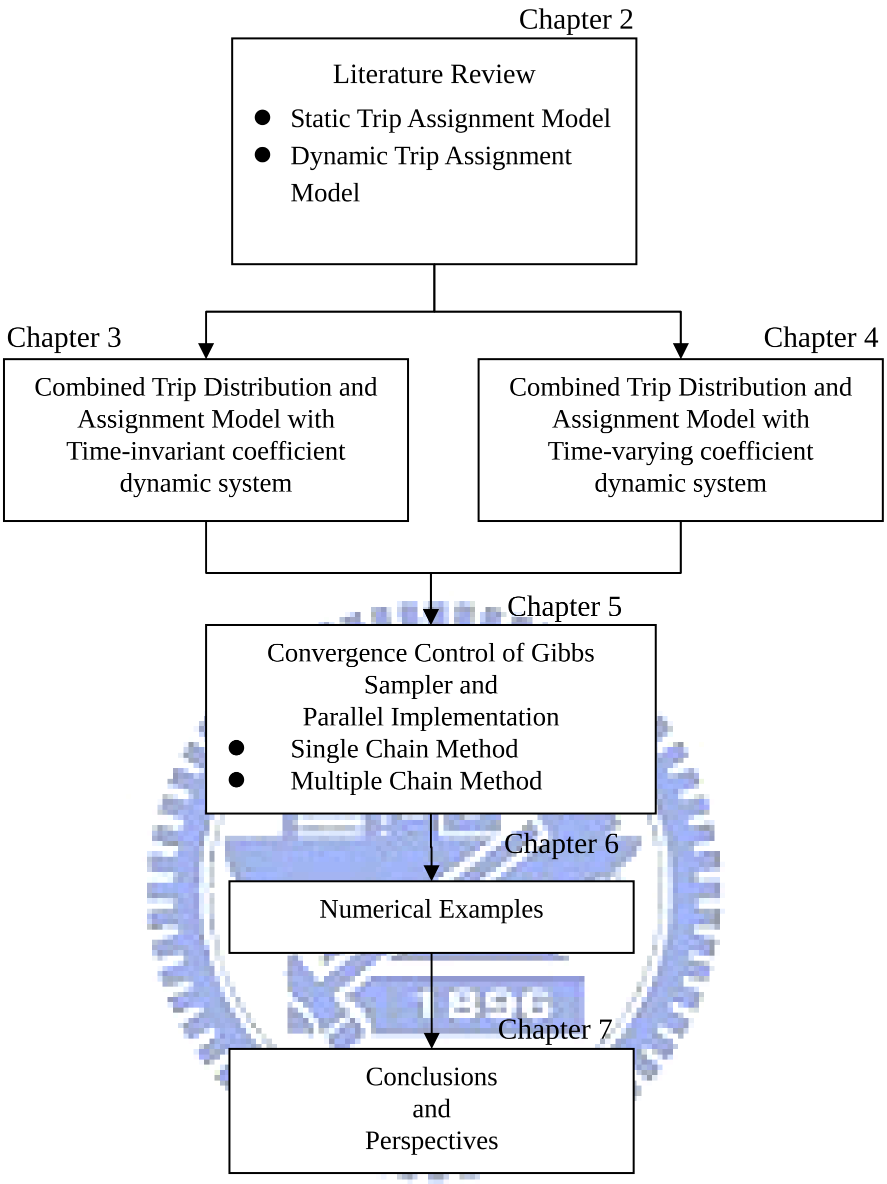


Figure 1.1 Overview of the dissertation.

1.4 Research Contributions

The principal contributions of this study are as follows:

1. Development of a dynamic path-base assignment model with no user behavior assumptions.

2. Existing research works on time-dependent origin-destination (O-D) estimation focus on the surveillance data and usually assume the prior information of the O-D matrix (or transition matrix) is know (or at least partially known). In this paper, we relax such assumption by combining Gibbs sampler and Kalman filter in a state space model.

3. A solution algorithm with parallel structure is proposed and implemented. With its efficiency and speedup demonstrated using real network data.

Chapter 2

Literature Review

This chapter provides literature reviews relevant to the formulation and solution algorithm of trip assignment problem. The following sections are organized as (i) static trip assignment models, (ii) dynamic trip assignment models, and (iii) summary and discussion.

Traditional four-step travel demand modeling decomposes the demand prediction problem in order to deal with its multi-dimensional character. These steps include trip generation, trip distribution, mode split, and trip assignment. It is a way to simplify the models for estimation and forecasting. Trip generation is the prediction of the number of trips produced by and attracted to each zone, which means, the number of trip ends "generated" within the area. That is, the trip generation phase of the analysis predicts total flows into and out of each zone in the study area, but it does not predict where these flows are coming from or going to. Trip ends are classified as being either a production or attraction. The variables are usually based on mathematical relationships between trip ends and socioeconomic or activity characteristics of the land use generating or attracting the trips.

In the traditional four stage aggregate approach, traffic assignment may not be viewed as strictly a demand model. It is the last stage of the model in which pre-determined origin-destination flows are assigned to links in the network. Various methods have been devised for assigning trips to network link, but have significant limitations.

2.1 Static Trip Assignment Models

Static assignment models assume that link flows and link trip times remain constant over the planning horizon of interest. Hence, a static origin-destination (O-D) matrix is given and assigned to the network links, results a link flow pattern that is intended to replicate the actual flow based on some behavior assumption. The static equilibrium assignment models are adequate for long-term planning analysis. Studies have shown that these formulations fail to capture the essential features of traffic congestion, including queue, departure time shift, traffic propagations, and etc. (Herman and Lam, 1974; Lisco, 1983; Hendrickson and Planck, 1984). Early attempts ignored congestion and the equilibrium issue, assigning all trips between any given O-D pairs to the shortest travel cost path (all-or-nothing assignment). Refined approaches resulted from recognition of the need to incorporate congestion effects (Sheffi, 1985; Matsoukis, 1986)

User Equilibrium Concept

The first mathematical programming formulation for the static user equilibrium (UE) problem with fixed demand as an equivalent optimization problem is introduced by Beckmann et al. (1956). This formulation allows the derivation of existence and uniqueness properties of the solution, satisfying the Wardropian UE condition. While Wardrop's UE condition indicates no user can improve his/her travel time/cost by unilaterally switching routes (Wardrop, 1952). The static UE flow pattern is obtained by solving the Beckmann equivalent optimization problem, stated as the following mathematical program:

$$\min \quad Z(\mathbf{x}) = \sum_a \int_0^{x_a} t_a(\omega) d\omega \tag{2.2a}$$

subject to

$$\sum_k f_k^{rs} = q_{rs} \quad \forall r, s \tag{2.2b}$$

$$f_k^{rs} \geq 0 \quad \forall k, r, s \tag{2.2c}$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \quad \forall a \tag{2.2d}$$

where $\mathbf{x}$ is the vector of link flows, x_a represents the flow on link *a*, and $t_a(\bullet)$ is the link performance function for link *a* that specifies the link travel time as a function of the flow on the link. The link performance function, often refer to the BPR function, is a positive, increasing, and convex curve, as illustrated in Figure 2.1. The typical link performance function does not consider queued vehicles in the traffic stream nor the propagation of the traffic flow.

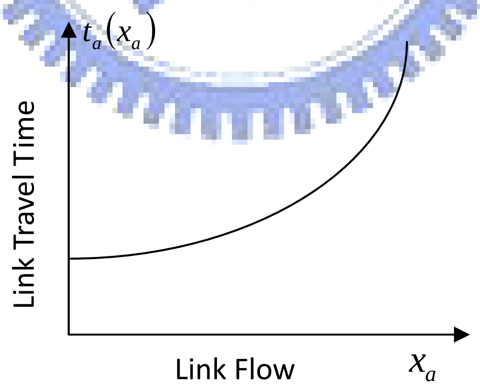


Figure 2.1 Typical Link Performance Function $t_a(x_a)$

The O-D demand between origin *r* and destination *s* is denoted by q_{rs} and the flow for O-D pair *r*-*s* assigned to path *k* is represented by f_k^{rs}. The static link-path incidence matrix relating path flows to link flows (equation 2.2d) are defined using link-path incidence variable $\delta_{a,k}^{rs}$ as follows:

$$\delta_{a,k}^{rs} = \begin{cases} 1, \text{if link } a \text{ is on path } k \text{ between O-D pair } r-s \\ 0, \text{otherwise} \end{cases} \tag{2.2e}$$

The objective function $Z(x)$, which is the sum over all arcs of the integrals of the link performance functions, does not have an intuitive economic or behavioral interpretation and is viewed strictly as a mathematical construct to solve equilibrium problems. Equation 2.2b indicates the set of flow conservation constraint which imply that all O-D demand have to be assigned to the network. The non-negativity conditions (equation 2.2c) ensure the solution of the program is physically meaningful. The network structure enters the formulation through the link-path incidence relationship (equation 2.2d) that relates the link-based objective function to the path-based constraint set.

Sheffi gives a comprehensive treatment of the static UE problem, addressing the conceptual, mathematical, algorithmic and computational aspects of the problem (Sheffi, 1985). A more difficult problem with asymmetric link interactions is addressed by Dafermos, and Fisk and Boyce by using variational inequality (VI) techniques (Dafermos, 1980, 1982; Fisk and Boyce, 1983). Nagurney, Mahmassani and Mouskos, and Patricksson address computational issues related to the VI problem (Nagurney, 1984,1986; Mahmassani and Mouskos, 1988, 1989).

System Optimal Concept

The other major class of assignment is system optimal (SO) formulation. The SO seeks a flow pattern that achieves some system-wide objectives. The static SO assignment problem can be formulated as follows:

$$\min \quad Z(x) = \sum_a x_a t_a(x_a) \qquad (2.3a)$$

subject to

$$\sum_k f_k^{rs} = q_{rs} \quad \forall r,s \qquad (2.3b)$$

$$f_k^{rs} \geq 0 \quad \forall k,r,s \qquad (2.3c)$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \quad \forall a \qquad (2.3d)$$

The objective function is the only difference from UE formulation with the interpretation of total system travel cost. The SO flow pattern does not always represent an equilibrium solution, as individual travelers may reduce their travel time by switching their routes. Hence, the SO flow pattern is not expected to hold without some control strategy such as road pricing or restriction. Consequently, the SO flow pattern is not an appropriate descriptive model of actual user behavior. It can be treated as a performance index of the network. The solution procedures for SO are identical to those of UE except that they differ in the specification of link cost functions (average cost function in UE; marginal cost function in SO).

Ben-Akiva enumerates the shortcomings of using static models in modeling congestions (Ben-Akiva, 1985). The static assignment has a major shortcoming of inadequately link congestion model. As discussed earlier in the thesis, the congestion is presented by a link performance function which gives the average trip time as a function of the average link flow. The average link flow can even exceed the actual capacity of the road section which is unrealistic for the control purpose. Another major problem lies on the description of traffic propagation. The static volume-delay curve cannot describe the propagation of traffic flow, especially in high flow levels. Thereby, static assignment models are inappropriate for real-time traffic control application, especially for congested networks.

2.2 Dynamic Trip Assignment Models

Dynamic network assignment is under intensive research, for both user equilibrium and system optimal problems. One common feature of these researches is that they differ from the standard static assignment assumptions to deal with time-varying flows. Another feature shared by these researches is that none presently provides a universal solution for general networks.

The first attempt to formulate the DTA problem as a mathematical program is introduced by Merchant and Nemhauser (1978a, 1978b). The model (referred to as the M-N model) is limited to the fixed-demand, single-destination, deterministic, system optimal scenario. A link exit function is utilized to propagate traffic and a static link performance function is introduced to present the travel cost as a function of link flow. It results a flow-based, discrete time, non-convex non-linear programming formulation. The global solution can be derived by solving a piecewise linear version of the model.

System Optimal Concept

Carey (1987) reformulates the M-N model as a convex nonlinear program by the manipulation of exit function, which gives mathematical advantage over the original M-N model. The formulation differs from M-N model mainly in the consideration of multiple destinations, and the exit function $g_a(\bullet)$.

$$\min Z(x) = \sum_t \sum_a h_a^t\left(x_a^t\right) x_a^t \qquad (2.4a)$$

subject to

$$g_a\left(x_a^t\right) \geq b_a^t \quad \forall a,t \qquad (2.4b)$$

$$b_a^t = x_a^t - x_a^{t+1} + d_a^t \quad \forall a,t \qquad (2.4c)$$

$$\sum_b d_b^t = F_k^t + \sum_c b_c^t \quad \forall a,t,b \in B(k), c \in C(k) \qquad (2.4d)$$

$$x_a^0 = E_a \quad \forall a \qquad (2.4e)$$

$$b_a^t, d_a^t, x_a^t \geq 0 \quad \forall a,t \qquad (2.4f)$$

where x_a^t represents the number of vehicles on link *a* at the beginning of interval *t*. $h_a^t\left(x_a^t\right)$ represents the travel cost incurred by the volume x_a^t and assumed to be continuous, convex, nondecreasing and nonnegative. The variable b_a^t and d_a^t denote the number of vehicle exiting and entering link *a* in interval *t*, respectively. F_k^t is referred to the exogenous demand at node *k* in period *t*. E_a is the initial volume on arc *a*. $B(k)$ and $C(k)$ respectively represent the set of links incident from and to node *k*. $g_a(\bullet)$ the exit function define the maximum number of vehicles that can exit from link *a* and is a function of traffic conditions on the link; it is assumed to be a continuous, non-negative, non-decreasing, and concave function. Figure 2.2 illustrates a possible shape of such link exit function.

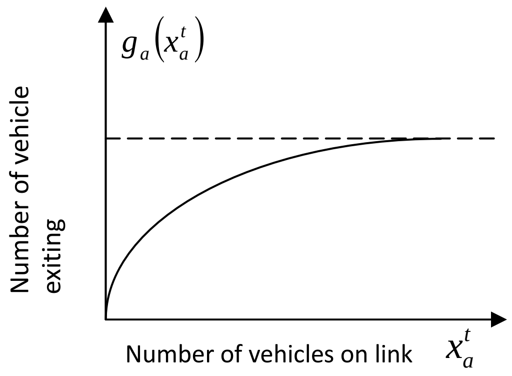


Figure 2.2 A possible shape of link exit function

Although after the manipulation of exit function, the formulation is convex, but it remains problematic by the non-convexity issues arising from first-in first-out (FIFO) requirement. The FIFO violation implies some traffic physically jumps over another to reduce system cost which is inconsistent with traffic realism. The FIFO requirement is easily satisfied in single destination formulation. While facing general networks, the FIFO requirement would introduce additional constraints that yield a non-convex constraint set and increasing the computational burden severely (Carey, 1992). As SO flow patterns, it may often be advantageous to favor certain traffic movement over others to minimize system-wide travel cost. For example, traffic at minor approach of an intersection may be holding back in favor of the major approach. That means vehicles may be artificially delayed for a time that might be considered as *unfair* or *unreasonable*; and the flow pattern may not be acceptable for real-world operation. Ziliaskopoulos introduces a linear programming formulation for the single destination system optimal DTA problem based on the cell transmission model (Ziliaskopoulos, 2000; Daganzo, 1994). This model circumvents the need for link performance function as the flow propagates according to the cell transmission model, hence is more sensitive to traffic realities.

User Equilibrium Concept

The user equilibrium formulation is generalized from the Wardrop condition for the static problem; it becomes the equilibration of the experienced path travel times of users. Janson represents one of the earliest attempts at modeling the UE dynamic traffic assignment problem as a mathematical program (Janson, 1991). The link-based UE model formulated as a mathematical program as follows:

$$\min Z(x) = \sum_t \sum_a \int_0^{x_a^t} \lambda_a^t(\omega) d\omega \tag{2.5}$$

subject to

Equation (2.4b) – (2.4f)

The λ_a^t represents the cost of traveling link *a* at the beginning of interval *t* when there exist x_a^t vehicles on the link.

Birge and Ho extend the M-N model to the stochastic case by relaxing the assumption that O-D demands are known for the entire planning horizon (Birge and Ho, 1993). It assumes a finite number of scenarios, defined as a possible combination of past O-D demands in every time interval, while assignment decisions are independent of future O-D demands.

Another approach is modeling dynamic traffic assignment problem in a continuous manner. The O-D demands are assumed to be known continuous functions of time; link flows are treated as continuous functions of time. Constraints of optimal control formulations are analogous to those of the mathematical programming formulations, but they are defined in a *continuous-time manner*. Friesz et al. discuss link-based optimal control formulation for both SO and UE objectives for the single destination case (Friesz, Luque, Tobin, and Wie, 1989). The model assumes that changes from one system state to another may occur concurrently as the network conditions change; that implies the routing decision are made based on current network conditions, and can be continuously modified as conditions change. The SO model is represented as follows:

$$\min Z(x) = \sum_a \int_0^T C_a(x_a)dt \tag{2.6a}$$

subject to

$$\frac{dx_a(t)}{dt} \equiv \dot{x}_a = u_a(t) - g_a[x_a(t)] \quad \forall a \in A, t \in [0,T] \tag{2.6b}$$

$$S_k(t) = \sum_{a \in A(k)} u_a(t) - \sum_{a \in B(k)} g_a[x_a(t)] \quad \forall k \in M, t \in [0,T] \tag{2.6c}$$

$$x_a(0) = x_a^0 \geq 0 \quad \forall a \in A \tag{2.6d}$$

$$u_a(t) \geq 0 \quad \forall a \in A, t \in [0,T] \tag{2.6e}$$

Equation 2.5b describe the rate of change of traffic volume with respect to time for link *a* will be considered as the difference of the flow entering link *a*, $u_a(t)$, and the flow exiting link *a*, $g_a[x_a(t)]$. In equation 2.6c, $S_k(t)$ represents the traffic flow generated at node *k*, which is assumed to be a nonnegative and continuous function of time. $A(k)$ and $B(k)$ respectively represent the set of links incident to and from node *k*. The nonnegative constraints of both traffic volume on the link and traffic volume entering the link are indicated in equation (2.6d) and (2.6e). As for the UE case, the objective function is illustrate as follows,

$$\min Z(x) = \sum_a \int_0^T \int_0^{x_a} c_a(\omega_a) g'_a(\omega_a) d\omega_a dt \tag{2.7}$$

subject to

equation (2.6b)-(2.6e).

The model proves that at the optimal solution, the instantaneous flow marginal costs on the used paths for an O-D pair are identical and less than or equal to the ones on the unused paths. As for the UE case, the model is in the form of equilibration of instantaneous user path costs.

Ran et al. use the optimal control approach to obtain a convex model for the instantaneous UE dynamic traffic assignment problem by defining link inflows and outflows to be control variables (Ran, Boyce, and LeBlanc, 1993) They recognize the inability of the usual cost functions to account for dynamic queuing and congestion costs, and propose splitting the link travel cost into moving and queuing parts. Boyce et al. proposed a methodology to solve the above model using Frank-Wolfe algorithm, but no implementations are illustrated (Boyce, Ran, and LeBlanc, 1995).

Because of the limitations of obtaining analytic mathematical properties, researchers focused on analytical DTA models have gradually migrated toward the variational inequality (VI) formulations. Variational inequality provides a general formulation platform for several different problems. Variational inequality approach is first introduced to the static traffic equilibrium by Dafermos (1980). Friesz et al. introduce the VI formulation into network design problem and suggest a sensitivity analysis based heuristic algorithm (Friesz, Tobin, Cho, Mehta, 1990; Friesz, Cho, Mehta, Tobin, Anandalingam, 1992). Friesz et al. is the first to show there is a variational inequality formulation of dynamic user equilibrium with simultaneous route choice and departure time decisions (Friesz, Bernstein, Smith, Tobin, Wie, 1993). Wie et al. formulate the dynamic network user equilibrium problem as a variational inequality problem in discrete time in terms of unit path cost functions (Wie, Tobin, Friesz, Bernstein, 1995). They also demonstrate that, assuming certain regularity conditions hold, discrete time dynamic network user equilibrium is guaranteed to exist. They define the time varying flow pattern $h*$ and associated minimum cost $\mu*$ is a discrete time dynamic network user equilibrium if the following conditions are satisfied:

$$h_p^*(t)\left[c_p\left(t,h^*\right)-\mu_{ij}\left(h^*\right)\right]=0 \quad \forall p\in P_{ij}, i\in I, j\in J, t=0,1,\ldots,T \tag{2.8a}$$

$$c_p\left(t,h^*\right)-\mu_{ij}\left(h^*\right)\geq 0 \quad \forall p\in P_{ij}, i\in I, j\in J, t=0,1,\ldots,T \tag{2.8b}$$

$$\sum_{p\in P_{ij}}\sum_{t=0}^{T} h_p^*(t)=Q_{ij} \quad \forall i\in I, j\in J \tag{2.8c}$$

$$h_p^*(t)\geq 0 \quad \forall p\in P_{ij}, i\in I, j\in J, t=0,1,\ldots,T \tag{2.8d}$$

In equation 2.8, *i* and *j* are the origin and the destination node respectively. P_{ij} denotes the set of all possible paths between origin *i* and destination *j*. $h_p(t)$ is the number of vehicles entering the first link on path *p* in period *t*, and $h=\left[h_p(t): p\in P, t=0,1,\ldots,T\right]$. $c_p(t,h)$ is the nonnegative unit travel cost incurred by travelers departing their origin in period *t* and choosing path *p* to their destination. $\mu(h)=\left[\mu_{ij}(h): i\in I, j\in J\right]$ is the vector of minimum unit travel costs. Q_{ij} is the total fixed O-D demand between *i* and *j* during time interval $0\leq t\leq T$. Since a complete path enumeration is required, an efficient method to identify a possible path set should be introduced to relief the computation burden.

Ran and Boyce propose a link-based discretized VI formulation SO DTA model with fixed departure time (Ran and Boyce, 1996). They equilibrate the experience travel time, the same as Friesz (1993). A queuing delay component is introduced in the model, but the capacity and oversaturation constraints increase the computation complexity significantly. Chen and Hsueh propose a link-based VI formulation UE DTA model and a nested diagonalization solution algorithm (Chen and Hsueh, 1998). The constraint set of the model is nonlinear and nonconvex, and multiple local solutions might exist. The variational inequality approach gives greater analytical flexibility and convenience than other analytical approaches. Although it brings mathematical advantages, the variational inequality approach is much more computationally intensive, especially facing the complete path enumeration for path-based formulation.

Simulation based models are mostly based on the mathematical programming models, but the critical constraints that describe the traffic flow propagation (i.e. flow conservation, vehicular movement) are addressed through simulation instead of analytical representation. With the traffic simulation, these models can address the traffic flow more realistic. However, the theoretical insights cannot derive analytically, which is the key issue of simulation-based models. A deterministic DTA model, with both SO and UE solutions, is proposed by Mahmassani and Peeta (1993). A meso-scopic traffic simulator is used as part of an iterative algorithm, with complete priori information of O-D demands for the entire planning horizon. Ghali and Smith propose a deterministic SO DTA model with congestion arises exclusively at specified bottlenecks modeled as deterministic queues (Ghali and Smith, 1995). Simulation based models can describe the traffic flow more realistic, which is troublesome in analytical formulations. However, the limitation lies on the inability to derive the associated mathematical properties.

More recently, Ashok and Ben-Akiva introduced stochasticity to map the assignment matrix between time-dependent O-D flows and link volumes both in off-line and real-time application (Ashok and Ben-Akiva, 2002). Ben-Akiva et al. propose DynaMIT, a meso-scopic simulator, as a dynamic traffic assignment system to estimate and predict current and future traffic conditions (Ben-Akive, Koutsopoulos, Mishalani, Yang, 1997). The model considers both historical information and drivers' response to information, supply and demand simulators work together to generate UE route guidance. Nie and Zhang proposed a relaxation approach for estimating static O-D matrix that minimizes a distance metric between measured and estimated traffic condition while the condition satisfies user equilibrium (Nie and Zhang 2008).

2.3 Summary and Discussion

To design and manage a transportation system, there is a need for efficient analyzing tool to describe the usage of the system. The traditional static method, four-stage planning process, includes trip generation, trip distribution, modal split, and trip assignment. Traditional sequential four-stage method is not suitable in the operation perspective. Since the travel costs used in the trip distribution stage are functions of the trip assignment outcomes; that is, the stages have to be repeated. From the literature review, estimate the network flows directly from time-series of link flows seems to be a reasonable candidate for real-time traffic operation aspect.

This chapter has reviewed several topics relevant to the trip assignment problems. Most existing research works assumed the existence of user-equilibrium or system-optimal conditions. However, the existence of equilibrium states in real traffic networks is questionable; an alternative approach to relax the assumption of user-equilibrium or system optimal might worth be established.

Chapter 3

A Path-base Assignment Model with Time-invariant Coefficient Dynamic System

In this chapter, some essential concepts of the path-base assignment model with time-invariant coefficient dynamic systems are discussed. Begin with the brief introduction to dynamic system models, the model assumption and notation is addressed in section 3.1. A state space approach that modeled the path estimation is illustrated in section 3.2. Link dynamics that describe the propagation of traffic flows is introduced, in section 3.3, to modify the proposed model.

3.1 Assumption and Notation

Existing research works on traffic assignment usually assume the existence of user equilibrium status or pursuit some system optimal situation. However, it is questionable whether such equilibrium state really exists or not. In this study, we relax such assumption by some statistical approaches.

The basic assumption in this research is the existence of unknown relationship between consecutive path-flows; the assumption is similar to that of Okutani. While Okutani assumed a time invariant relation between time-series O-D flow exists; and this relationship can be estimated by some prior information. In this study, we do not estimate the relationship by prior information; the relationship is estimated simultaneously with path flows.

For convenience, we use the following notations in this chapter:

Table 3.1 Notations of Time-invariant State Space Model

Notation	**Descriptions**
F	a $p \times p$ path flow transition matrix
ψ_t	a $p \times 1$ network path flows on time t
y_t	a $q \times 1$ link traffic observation vector on time t
H	a $q \times p$ zero-one matrix which denotes the path-link incidence matrix
p	number of elements in path set P
q	number of observations in the target network
σ_t, γ_t	independently and identically distributed Gaussian noise terms
^{T}	The transport of a matrix is mark by superscript T.

3.2 Estimation of Path Flow by State Space Model without Prior Information

Isaac Newton introduced the differential equations in the 17th century and provided mathematical models for many dynamic systems. Given a finite number of initial conditions, one can uniquely determine the system status for all time. The finite dimensional representation of a problem is the basic idea for the state-space approach to the representation of dynamic systems. The dependent variables of the equations are the state variable of the dynamic systems. The principal dynamic system models are listed in Table 3.2 below.

Table 3.2 Principal dynamic system models

Model	Continuous	Discrete
Time-invariant		
Linear	$\dot{\psi}(t) = \Phi\,\psi(t) + Cu(t)$	$\psi_{t+1} = F\psi_t + \Gamma u_t$
General	$\dot{\psi}(t) = f\left(\psi(t), u(t)\right)$	$\psi_{t+1} = f\left(\psi_t, u_t\right)$
Time-varying		
Linear	$\dot{\psi}(t) = \Phi(t)\psi(t) + C(t)u(t)$	$\psi_{t+1} = F_t\psi_t + \Gamma_t u_t$
General	$\dot{\psi}(t) = f\left(t, \psi(t), u(t)\right)$	$\psi_{t+1} = f\left(t, \psi_t, u_t\right)$

Reference: [Grewal & Andrews, 1993]

We focused on the discrete dynamic systems in this dissertation. In the dynamic model, F is a $n \times n$ dynamic coefficient matrix. The matrix $\psi_t = [\psi_1(t) \quad \psi_2(t) \quad \psi_3(t) \quad \cdots \quad \psi_n(t)]^T$ is called the state vector, where $\psi_n(t)$ denotes the n^{th} state variable in time *t*. The n-dimensional domain of the state vector is called the state space of the dynamic system. The state variables are related to the system outputs by a system of linear equations that can be represented in vector form, as follows.

$$y_t = H\psi_t + \gamma_t$$

where

$$y_t = [y_1(t) \quad y_2(t) \quad y_3(t) \quad \cdots \quad y_l(t)]^T$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n} \\ h_{31} & h_{32} & h_{33} & \cdots & h_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{l1} & h_{l2} & h_{l3} & \cdots & h_{ln} \end{bmatrix}$$

The y_t is a *l*-vector called the measurement vector (also called observation vector) of the system. The matrix H is a measurement sensitivity matrix with measurement sensitivities h_{ln} measures the scale of l^{th} output to n^{th} state variable.

Real world problems tend to have some kind of unpredictability in behaviors, due to some unknown exogenous inputs. Furthermore, output-measuring process with physical sensors will always introduce some amount of sensor noise, which will cause errors in the estimation process. While facing real world problems, using a statistical approach that taking uncertainties into account instead of deterministic would be a better approach. These dynamic systems with uncertainties are characterized by statistical parameters such as means, correlations, and covariances.

3.2.1 Modeling

In this section, only the most basic modeling is introduced. The transition matrix that describe the relationship among path-flows in different time period is assumed to be fixed. The link travel time, which is an important issue, is not considered. Although the above strict assumption are made in this section, but they will be relaxed in the following section.

State space model is introduced to estimate path flows from link traffic counts. The state space model is coupled with two parts: transition equations and observation equations. First, the state equation which assumed that the path flows at time $t+1$ can be related to the path flows at time t by the following autoregressive form,

$$\psi_t = F\psi_{t-1} + \sigma_t, \quad t = 1,2,3,\ldots,n \tag{3.1}$$

where ψ_t is the state vector which is unobservable, F is a random transition matrix, $\sigma_t \sim N_p(0,\Sigma)$ is independently and identically distributed noise term, where

N_p denotes the *p*-dimensional normal distribution, Σ is the corresponding covariance matrix. ψ, the $p \times 1$ state vector, is defined to be the path flows belonging to O-D pairs.

Next, the observation equation,

$$y_t = H\psi_t + \gamma_t, \quad t = 1,2,3,...n \tag{3.2}$$

where y_t is the $q \times 1$ observation vector which means there are q detectors on the network. The number of paths is denoted by p. H is a $q \times p$ zero-one matrix, which denotes the path-observation incidence matrix. The path-observation is pre-determined by generating possible path set and given travel time on each link. γ_t is also a noise term that $\gamma_t \sim N_q(0, \Gamma)$. Both ψ and F are unobservable, thus Kalman filter is not suitable to directly estimate and forecast the state vector. Hence, Gibbs sampler is used to tackle the problem of simultaneous estimation of F and ψ by available information.

There are two major elements to be incorporated in the solution method, 1) filtering states by observations, and 2) sampling scheme of transition matrix, F, and state vector, ψ. Since the observations, y_t, are not used in the conditional distribution, the Kalman filter and the Gibbs sampler must be combined. After the estimation of state vector, O-D flows can be calculated by the summation of path flows.

3.2.2 Kalman Filter

Kalman filter is an estimator for the linear-quadratic-Gaussian problem, which estimates the instantaneous state of a linear dynamic system perturbed by Gaussian white noise. It utilizes measurements, corrupted by Gaussian white noise, linearly related to the state and gives a statistically optimal estimator with respect to quadratic function of estimation error. The Kalman filter does not only provide a means for inferring missing information from indirect and noisy measurements, but also used for predicting the likely future trends of dynamic systems. The structure of the filter can be derived in a Bayesian framework as follows.

The first stage (i.e. t =1), there's no observation exists, thus the state vector ψ_0 must be generated by a prior distribution that $\psi_0 \sim N_p(\mu_0, V_0)$, where μ_0 is the mean and V_0 is the covariance matrix. By using equation (3.1), the distribution for the state vector in the first stage will be normal with parameters

$$E[\psi_t \mid y_{t-1}] = \mu_{t|t-1} = F\mu_{t-1} \tag{3.3}$$

$$Var[\psi_t \mid y_{t-1}] = V_{t|t-1} = FV_{t-1}F^T + \Sigma \tag{3.4}$$

where $\mu_{t|t-1}$ denotes the expect value of ψ_t and $V_{t|t-1}$ denotes the variance of ψ_t when y_{t-1} is observed. By the above information, the forecast observation would be normal distribution with parameters

$$E[y_t \mid y_{t-1}] = \hat{y}_t = H\mu_{t|t-1} \tag{3.5}$$

$$Var[y_t \mid y_{t-1}] = M_t = HV_{t|t-1}H^T + \Gamma \tag{3.6}$$

The above equation (3.3)-(3.6) holds for any t.

As the new observation y_t become available, the parameter vector would be updated according to Baye's rule,

$$p(\psi_t \mid y_t) \propto p(y_t \mid \psi_t) p(\psi_t \mid y_{t-1}).$$

By using Bayes' rule and standard Bayesian theory, the posterior distribution will be normal with parameters

$$\mu_{t|t} = \mu_{t|t-1} + V_{t|t-1} H^T M_t^{-1} \left(y_t - H\mu_{t|t-1} \right) \tag{3.7}$$

$$V_{t|t} = V_{t|t-1} - V_{t|t-1} H^T M_t^{-1} H V_{t|t-1} \tag{3.8}$$

The algorithm of Kalman filter is illustrated as follows,

Algorithm Kalman Filter

Input: $\mu_0, V_0, \quad y_t := observation\ sequence$

Output: Filtered μ and V

Begin

FOR each time step of the observation sequence

Generate prediction of the new observation by

$$\hat{y}_t = H \cdot F \cdot \mu_{t-1}$$

$$M_t = H\left(F \cdot V_{t|t-1} \cdot F^T + \Sigma\right)H^T + \Gamma$$

Update the parameter by

$$\mu_{t|t} = \mu_{t|t-1} + V_{t|t-1} H^T M_t^{-1} \left(y_t - H\mu_{t|t-1} \right)$$

$$V_{t|t} = V_{t|t-1} - V_{t|t-1} H^T M_t^{-1} H V_{t|t-1}$$

END FOR

END

3.2.3 Gibbs Sampler

Gibbs sampler is a technique for generating random variables from a distribution indirectly, without having to calculate the density. In this paper, we make the following assumptions, (i) The initial $\psi_0 \sim N(\mu_0, V_0)$, (ii) The covariance matrix Σ and Γ are known, and (iii) Given *F*, the distribution ψ_t is Gaussian.

The state equation can be written

$$\psi_t^T = \psi_{t-1}^T F^T + \sigma_t^T, \quad t = 1,2,3,\cdots,n \tag{3.9}$$

that is

$$\begin{bmatrix} \psi_1^T \\ \vdots \\ \psi_n^T \end{bmatrix} = \begin{bmatrix} \psi_0^T \\ \vdots \\ \psi_{n-1}^T \end{bmatrix} F^T + \begin{bmatrix} \sigma_1^T \\ \vdots \\ \sigma_n^T \end{bmatrix}$$

The following notation is used for simplification.

$$\boldsymbol{\psi}_n = \begin{bmatrix} \psi_1^T \\ \vdots \\ \psi_n^T \end{bmatrix}, \ \boldsymbol{\psi}_{n-1} = \begin{bmatrix} \psi_0^T \\ \vdots \\ \psi_{n-1}^T \end{bmatrix}, \ \boldsymbol{\sigma}_n = \begin{bmatrix} \sigma_1^T \\ \vdots \\ \sigma_n^T \end{bmatrix}, \ F^T = \begin{bmatrix} F_1^T & \cdots & F_i^T & \cdots & F_p^T \end{bmatrix}$$

where F_i^T denotes the i^{th} column vector of F^T. Then the equation (3.9) can be re-written as

$$\boldsymbol{\psi}_n = \boldsymbol{\psi}_{n-1} F^T + \boldsymbol{\sigma}_n$$

Consider the element of S, the $p \times p$ covariance matrix being used to estimate the variance-covariance matrix of F,

$$S(F^T) = \{S_{ij}(F_i^T, F_j^T)\}$$

where S_{ij} denotes the (i, j) element of the matrix. S_{ij} can be calculated by the following equation.

$$\begin{aligned} S_{ij} &= \left(\boldsymbol{\psi}^T_{n(i)} - \boldsymbol{\psi}^T_{n-1}F_i^T\right)^T \left(\boldsymbol{\psi}^T_{n(j)} - \boldsymbol{\psi}^T_{n-1}\hat{F}_j^T\right) \\ &= \left(\boldsymbol{\psi}^T_{n(i)} - \boldsymbol{\psi}^T_{n-1}\hat{F}_i^T\right)\left(\boldsymbol{\psi}^T_{n(j)} - \boldsymbol{\psi}^T_{n-1}\hat{F}_j^T\right) + \left(F_i^T - \hat{F}_i^T\right)\boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}\left(F_j^T - \hat{F}_j^T\right) \end{aligned} \tag{3.10}$$

where $\hat{F}_i^T = \left(\boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}\right)^{-1}\boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n(i)}$ is the least square estimate of F_i^T, and $\boldsymbol{\psi}_{n(i)}$ is the i^{th} column vector of $\boldsymbol{\psi}_n$. Consequently,

$$S\left(F^T\right) = A + \left(F^T - \hat{F}^T\right)^T \boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}\left(F^T - \hat{F}^T\right) \tag{3.11}$$

where $\boldsymbol{A}$ is a $p \times p$ matrix. $A = \{a_{ij}\}$, where a_{ij} is the (i, j) elements of $\boldsymbol{A}$, with

$$a_{ij} = \left(\boldsymbol{\psi}^T_{n(i)} - \boldsymbol{\psi}^T_{n-1}\hat{F}_i^T\right)^T \left(\boldsymbol{\psi}^T_{n(j)} - \boldsymbol{\psi}^T_{n-1}\hat{F}_j^T\right). \tag{3.12}$$

That means $\boldsymbol{A}$ is proportional to the sample covariance matrix. From the general result in the Gaussian model, the posterior distribution of F^T is then

$$\begin{aligned} p\left(F^T \mid \boldsymbol{\psi}\right) &\propto \left|S\left(F^T\right)\right|^{-\frac{n}{2}}, -\infty < F^T < \infty \\ &= \left|A + \left(F^T - \hat{F}^T\right)^T \boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}\left(F^T - \hat{F}^T\right)\right|^{-n/2} \end{aligned} \tag{3.13}$$

The distribution in equation (3.13) is a matrix-variate generalization of the *t*-distribution. The following sampler for generating F^T and $\boldsymbol{\psi}$ is then proposed. The sampling scheme generate from the conditional distributions

a. $\psi_t \mid F, \psi_{t-1}, \Sigma \sim N\left(F\psi_{t-1}, \Sigma\right)$

b. $F^T \mid \boldsymbol{\psi}, \Sigma \sim \left[k(n, p, p)\right]^{-1} |A|^{(n-p)/2} \left|\boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}\right|^{p/2} \left|A + F\boldsymbol{\psi}_{n-1}\boldsymbol{\psi}^T_{n-1}F^T\right|^{-n/2}$

The above sampling scheme would be the key component of the Gibbs sampler.

The Gibbs sampler is a Markovian updating scheme that proceeds as follows. Given an arbitrary starting set of values $\{Z_1^{(0)}, Z_2^{(0)}, Z_3^{(0)}, ..., Z_k^{(0)}\}$, and then draw

$Z_1^{(1)} \sim [Z_1 | Z_2^{(0)}, Z_3^{(0)}, ... Z_k^{(0)}]$,

$Z_2^{(1)} \sim [Z_2 | Z_1^{(0)}, Z_3^{(0)}, ... Z_k^{(0)}]$,

$Z_3^{(1)} \sim [Z_3 | Z_1^{(0)}, Z_2^{(0)}, ... Z_k^{(0)}]$, …

$Z_k^{(1)} \sim [Z_k | Z_1^{(0)}, Z_2^{(0)}, ... Z_{k-1}^{(0)}]$.

Each variable is visited in the natural order and a cycle requires k random variate generations. After i iterations we have $(Z_1^{(i)}, Z_2^{(i)}, Z_3^{(i)}, ..., Z_k^{(i)})$. Under mild conditions, the following results hold (Geman and Geman, 1998)

Result 1: Convergence

As the iteration continue, $(Z_1^{(i)}, Z_2^{(i)}, Z_3^{(i)}, ..., Z_k^{(i)}) \rightarrow [Z_1, Z_2, Z_3, ..., Z_k]$. Hence, for each sequence s, $Z_s^{(i)} \rightarrow [Z_s]$ as $i \rightarrow \infty$.

Result 2: Rate

Using the sup norm, the joint density of $(Z_1^{(i)}, Z_2^{(i)}, Z_3^{(i)}, ..., Z_k^{(i)})$ converges to the true density at a geometric rate, under visiting in the natural order.

Result 3: Ergodic theorem

For any measurable function T of $Z_1, Z_2, Z_3, ..., Z_k$ whose expectation exists,

$$\lim_{i \rightarrow \infty} \frac{1}{i} \sum_{l=1}^{i} T(Z_1^{(l)}, Z_2^{(l)}, Z_3^{(l)}, ..., Z_k^{(l)}) \rightarrow E(T(Z_1, Z_2, Z_3, ... Z_k)).$$

Then for every function T on the possible configurations of the system and for every starting configuration holds with probability one.

Analytical convergence rates for Gibbs sampler applied to state space models are further discussed by Pitt and Schephard in 1996 and Robert (Pitt and Schephard, 1996; Robert, 1998)

As Gibbs sampling through m replications of the aforementioned i iterations produces i independent and identically distributed k tuples $Z_{1j}^{(i)}, Z_{2j}^{(i)}, Z_{3j}^{(i)}, ..., Z_{kj}^{(i)}$, $j = 1,2,3,...,m$, which the proposed density estimate for $[Z_s]$ having form

$$[\hat{Z}_s] = \frac{1}{m}\sum_{j=1}^{m}[Z_s | Z_r^{(j)}, r \neq s].$$

The above Gibbs sampling scheme on a random transition matrix and state vector forms the center part of the algorithm. In the process of generate state vectors, Kalman filtering mechanism is added. While a simple monitoring of the chain (Z_s) can only expose strong non-stationarities, it is more relevant to consider the cumulated sums, since they need to stabilize for convergence to be achieved. The convergence control of Gibbs sampler is discussed in chapter 5.

Algorithm Gibbs Sampler

Input: $H := path - observation\ incidence\ matrix$, $y_t := observation\ sequence$

Output: $\hat{\psi}, \hat{F}$

Begin

Initialize

$F^{(0)} := I_p$, $\Sigma := I_p$, $\Gamma := I_p$

$\psi_{store} = \{\phi\}$, $F_{store} = \{\phi\}$

SET GibbsCount (g) to 0

WHILE not Converge

Generate $\psi^{(g)} \sim N(\mu, V)$

Append $\psi^{(g)}$ to ψ_{store}

CALL Kalman Filter with μ, V, and observation sequence

Generate $F^{T(g)}$ by

$$A^{(g)} = \{a_{ij}^{(g)}\}, \quad a_{ij} = \left(\psi_{n(i)}^{T(g)} - \psi_{n-1}^{T(g)} \hat{F}_i^{T(g)}\right)^T \left(\psi_{n(j)}^{T(g)} - \psi_{n-1}^{T(g)} \hat{F}_j^{T(g)}\right)$$

Generate $w \sim Wishart\left(\psi_{n-1}^{(g)} X_{n-1}^{T(g)}, n-p\right)$

Generate $Z = \left(z_1^T, z_2^T, z_3^T, \ldots, z_p^T\right), z_k \overset{iid}{\sim} N_p\left(0, A^{(g)}\right)$

COMPUTE $F^{T(g)} = \left(\left(w^{\frac{1}{2}}\right)^T\right)^{-1} Z$

APPEND $F^{T(g)}$ to F_{store}

INCREMENT GibbsCount

END WHILE

READ last k items from ψ_{store} and put in ψ_n

COMPUTE $\hat{\psi} = \frac{1}{k}\sum \psi_n$

READ last k items from F_{store} and put in F_n

COMPUTE $\hat{F} = \frac{1}{k}\sum F_n$

END

3.2.4 Solution Framework

In this section, the solution framework of applying Gibbs sampler and Kalman filter to the state space model is illustrated. The solution framework that combines the algorithm in both section 3.2.2 and 3.2.3 is demonstrated in figure 3.1. The solution framework is an iterative estimation of state vectors and transition matrix; the iteration counts are denoted as g. It first filters the state vector by given transition matrix, path-observation incidence matrix (some may refer as mapping matrix), and observation vector; and then turns to estimate the transition matrix by filtered state vector. During the Kalman filter stage of the framework, transition matrix and mapping matrix are fixed; while in the transition matrix estimating stage, the state vector is fixed. As the algorithm reach convergence, the transition matrix, $\hat{F}$, and state vector, $\hat{\psi}$, can be estimated.

After deriving transition matrix and state vector, the prediction of state vector can be represent as,

$$\hat{\psi}_{t+1} = F \cdot \psi_t + \sigma_{t+1}. \tag{3.14}$$

And the estimation of $\hat{\psi}_{t+h}$ is

$$\begin{aligned} \hat{\psi}_{t+h} &= F \cdot \hat{\psi}_{t+h-1} + \sigma_{t+h} \\ &= F \cdot \left(F \cdot \hat{\psi}_{t+h-2} + \sigma_{t+h-1}\right) + \sigma_{t+h} \\ &= F \cdot \left(F \cdot \left(F \cdot \left(F \cdots \left(F \cdot \psi_t + \sigma_{t+1}\right) \cdots\right) + \sigma_{t+h-2}\right) + \sigma_{t+h-1}\right) + \sigma_{t+h} \end{aligned} \tag{3.15}$$

Since the expectation value of σ is zero. The estimate of $\hat{\psi}_{t+h}$ can be written as

$$\hat{\psi}_{t+h} = F^h \psi_t. \tag{3.16}$$

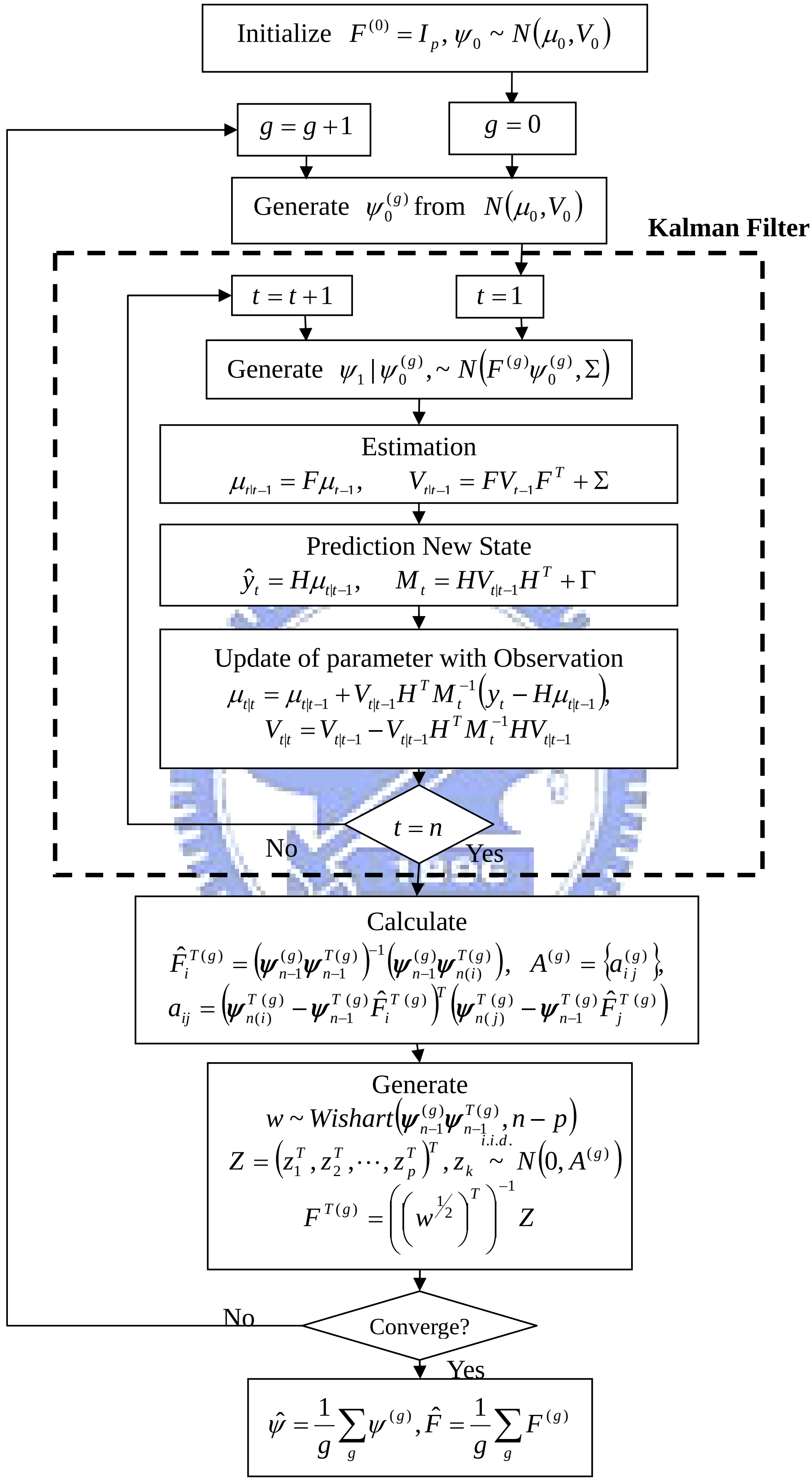


Figure 3.1 The solution framework of state space model

3.3 Estimation of Path Flow Considering Link Travel Time

Travel time is an important issue while facing transportation problems. It is no exception in path flow estimation. We relax the assumption of zero travel time in this section. To describe the flow propagation from its origin to destination, travel time on each links should be considered. A macroscopic traffic continuum model, the PW model, is introduced to describe the link dynamics. Notations increased or changed in the section is shown in table 3.3.

Table 3.3 Notations of Time-invariant State Space Model considering Link Travel Time

Notation	Descriptions
ψ_t^{lag}	a $p(MaxLag+1)\times 1$ matrix composed of series of path flows
$\psi_t(i)$	i^{th} component of the vector ψ_t
H_t	a $q\times p(MaxLag+1)$ zero-one matrix which denotes the path-link incidence matrix at time t
Δt	uniform width of time mesh for link dynamic model
k_j^n	density at time density at time $t_0+n\Delta t$ and space $j\Delta x$
u_j^n	speed at time density at time $t_0+n\Delta t$ and space $j\Delta x$
g_j^n	net entering or leaving rate of vehicle at time density at time $t_0+n\Delta t$ and space $j\Delta x$

3.3.1 Modeling

The model in this section had been modified from the previous section (section 3.2) to consider travel time effect. The state transition equation is the same as in previous section, but the observation equation had been modified as follows.

$$\psi_t = F\psi_{t-1} + \sigma_t, \quad t = 1,2,3,\ldots,n \tag{3.17}$$

$$y_t = H_t \psi_t^{lag} + \gamma_t, \quad t = 1,2,3,\ldots n \tag{3.18}$$

The state variable, ψ_t^{lag} , is modified here to take the travel time into account. It is a $p(MaxLag+1)\times 1$ matrix composed of series of state vectors,

$$\psi_t^{lag} = \left\{ \psi_t^T \quad \psi_{t-1}^T \quad \psi_{t-2}^T \quad \cdots \quad \psi_{t-MaxLag}^T \right\}^T.$$

Let the element in ψ_t, a $p\times 1$ vector, be represent as $\begin{bmatrix} \psi_t(1) \\ \psi_t(2) \\ \vdots \\ \psi_t(p) \end{bmatrix}_{p\times 1}$. Then,

$$\psi_t^T = \left[\psi_t(1) \quad \psi_t(2) \quad \cdots \quad \psi_t(p)\right]_{1\times p}, \text{ and}$$

$$\psi_t^{lag} = \left\{ \left[\psi_t(1) \quad \cdots \quad \psi_t(p)\right]\left[\psi_{t-1}(1) \quad \cdots \quad \psi_{t-1}(p)\right] \quad \cdots \right.$$
$$\left. \left[\psi_{t-MaxLag+1}(1) \quad \cdots \quad \psi_{t-MaxLag+1}(p)\right]\left[\psi_{t-MaxLag}(1) \quad \cdots \quad \psi_{t-MaxLag}(p)\right] \right\}^T_{1\times p(MaxLag+1)}.$$

The term $MaxLag$ denotes the maximum time lag that path flow can be observed on observation sites at time t. The *path-observation* incidence matrix, H_t, is a zero-one $q\times p(MaxLag+1)$ matrix at time t. If the path flow of a certain time can be observed at detector q, then the corresponding element in H_t is one; else, it is zero. The incidence matrix can be generated by the interaction with link dynamics that is described in the next section.

3.3.2 Link Dynamics

Traffic simulation techniques appeared in the early 1950s in the field of transportation science. Computer-based traffic simulation tools, mostly developed in the past few decades, exist as a cost-effective assisting tool for researchers and practitioners to verify and evaluate traffic management strategies. Traffic simulation models can be characterized as microscopic, mesoscopic or macroscopic. The microscopic models simulate every vehicle in the network; mainly include three behaviors, accelerating, decelerating and lane changing. This kind of models, try to describe the actions and reactions of the vehicle that make up the traffic as accurately as possible, are the so-called car-following model. In order to achieve accuracy in modeling traffic, it leads to a simulation model with high degree of parameters (50 parameters is common). The simulation time heavily depends on the number of vehicle that exist simultaneously in the simulated network, that make it hard to meet the requirements of simulating large-scale congested traffic networks for ITS applications, especially at real-time level (Yang and Koutsopoulos, 1996).

The macroscopic approach, based on an analogy between traffic flow and a real fluid flow, is also called continuum traffic-flow model. These models mainly based on traffic density, volume and speed have been widely analyzed in the past (Lighthill and Witham, 1955; Payne, 1979; Leo and Pretty, 1990; Helbing, 1995). Macroscopic models usually involve partial differential equations defined on appropriate domains with boundary conditions describing traffic phenomena. The models present a higher level of abstraction than the microscopic model and lead to some computing advantages. The computing time required for a macroscopic model do not increase with the number of existing vehicle on the simulation network, and this advantage makes it easier to implement on a large-scale network.

Macroscopic traffic continuum models had been classified into first-order continuum models, such as Lighthill, Whitham and Richards' well-known flow conservation model (LWR model), and high-order continuum models, such as Payne's momentum conservation models (PW model). The models are composed of one or several partial differential equations (PDEs) defined on appropriate domains with initial and boundary conditions.. The LWR model consists of the fundamental conservation principle in the form of a PDE,

$$\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(k)}{\partial x}=g(x,t) \tag{3.19}$$

together with standard definition of flux function

$$q(k)=k(x,t)\times u(k) \tag{3.20}$$

It is assumed that the empirical *u-k* relationships follow the Greenshields traffic stream model

$$u(k)=u_f\times\left(1-k/k_b\right) \tag{3.21}$$

where u_f denote the free flow speed and k_b the density with vehicles bumper to bumper.

Although the LWR model is widely cited in researches, it is also known to have some deficiencies. The steady state velocity assumption, which means that velocity changes instantaneously as density change is certainly not valid in traffic flow. Payne used a motion equation to obtain a more complex equation to describe speed dynamics (Payne, 1979),

$$\frac{\partial u}{\partial t}+u(\nabla u)=-\frac{1}{k}\nabla(P_e(k))+\frac{1}{\tau}(u_e(k)-u) \tag{3.22}$$

where $u_e(k)$ is an equilibrium speed-density relation and τ is the relaxation time. Since it is difficult to find the analytical solution of the traffic continuum model, numerical methods, such as Lax or Upwind method, had been used by researchers to simulate the numerical solutions for traffic continuum model. Lax-F finite difference method is used in solution scheme.

Lax-F scheme transfer PDEs to finite difference equations by using centered difference skill. The Lax-F difference equation or PW continuum model can be written as,

$$k_j^{n+1} = \frac{k_{j+1}^n + k_{j-1}^n}{2} - \frac{\Delta t}{\Delta x}\frac{q_{j+1}^n + q_{j-1}^n}{2} + \frac{\Delta t}{2}\left(g_{j+1}^n + g_{j-1}^n\right), \tag{3.23}$$

and

$$u_j^{n+1} = u_j^n - \Delta t\left\{u_j^n \frac{u_j^n - u_{j-1}^n}{\Delta x} + \frac{1}{\tau}\left[u_j^n - u_e\left(k_j^n\right) + \frac{v}{k_j^n}\frac{k_{j+1}^n - k_j^n}{\Delta x}\right]\right\}. \tag{3.24}$$

After the computation of the Lax-F scheme, densities on each mesh can be obtained. In this dissertation, a link travel time estimation method for the continuum traffic models proposed by Hwang and Cho is used (Hwang and Cho, 2006).

3.3.3 Solution Framework

A solution framework that combines the link dynamics to address travel time issue of the state space model is suggested. It first set the travel time on each link to be the free-flow travel time and generate a initial path-link incidence matrix, H_t. The path-link incidence matrix is then used in the path flow estimation. After the estimation of path flows, the path flows are transformed into link flows and the link dynamics are incorporated to give a travel time estimation of certain link. The link travel time is then used to generate the path-link incidence matrix for the next iteration. The algorithm terminates while the difference of path flows in consecutive iterations are less than ε. The solution framework is illustrated in figure 3.2.

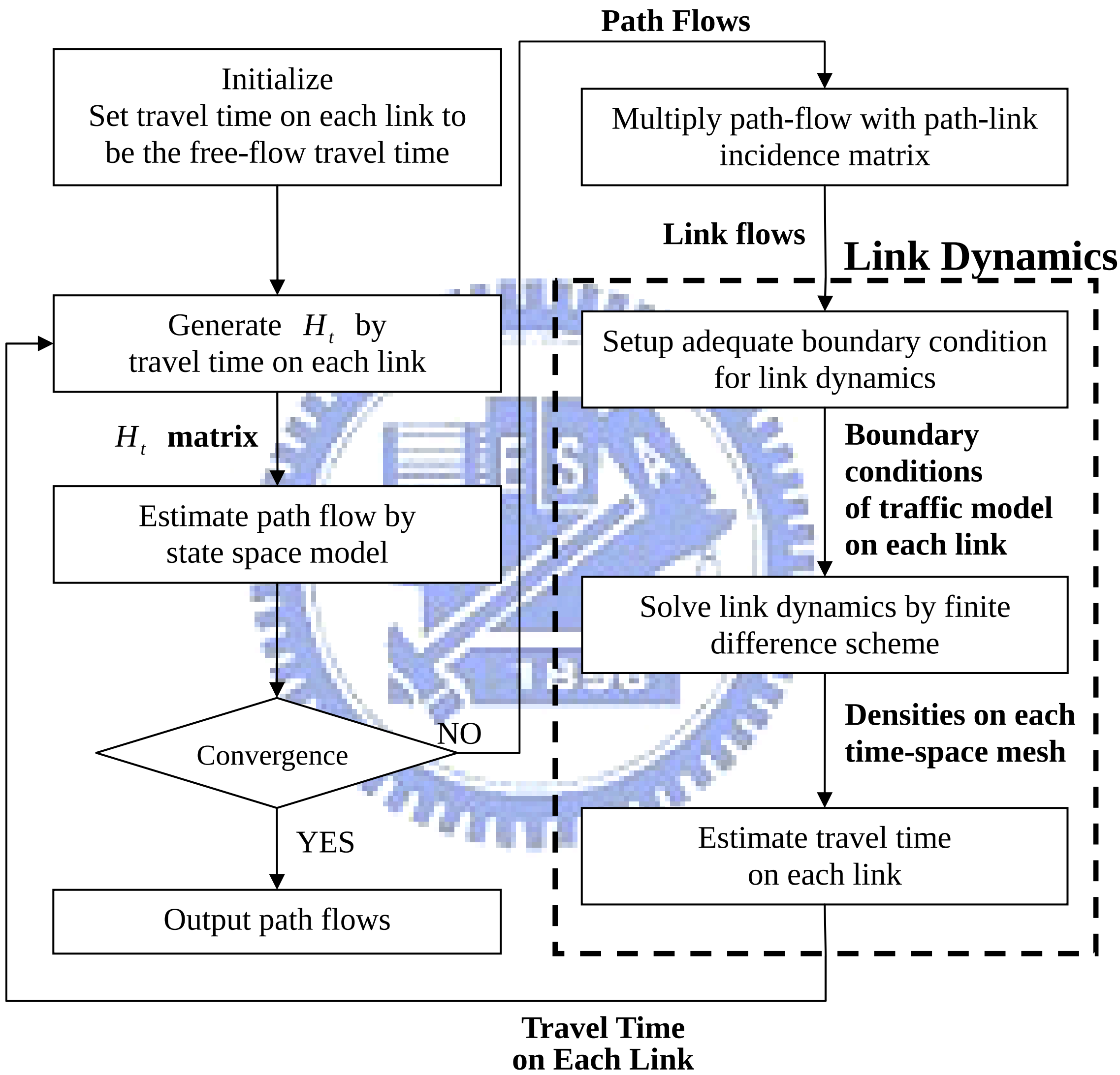


Figure 3.2 The solution framework of time-invariant coefficient state space model considering link travel time

Chapter 4

A Path-base Assignment Model with Time-varying Coefficient Dynamic System

Chapter 3 introduced the path-base assignment model with time-invariant coefficient dynamic systems. That model assumes a time-invariant relationship among path flows in different times. In reality, the characteristic of traffic is different from time to time, i.e., peak hours and non-peak hours, trips heading to work or going home. This chapter tries to relax such time-invariant assumption by introduce a time-varying coefficient state space model.

4.1 Assumptions

In chapter3, the transition matrix assumed to be time-invariant. We relax such assumption by introducing a rolling horizon structure and Wiener process to address the transition matrix. Wiener process (often called Brownian motion) is a stochastic process plays an important role in both pure and applied mathematics. It is very useful in modeling a trend with random walk. The basic assumption in this chapter is the change of time-varying transition matrix can be represented by Wiener process. Notations increased or changed in the section is shown in table 4.1.

Table 4.1 Notation of time-varying state space model

Notation	Descriptions
W	a $p \times p$ drift matrix that describe the relation between F_t and F_{t+1}
F_t	a $p \times p$ path flow transition matrix on time t
ρ_t	Independently and identically distributed Gaussian noise terms with variance-covariance matrix of $\boldsymbol{\rho}$
Ψ_t	is a $p \times (m+1)$ matrix composed of $m+1$ state vectors at time t
$\boldsymbol{\sigma}_t$	Is a $p \times (m+1)$ matrix composed of $m+1$ random term σ_t

4.2 Estimation of Path Flow by Time-varying Coefficient State Space Model

In this section, we estimate the path flows by time-varying coefficient state space model. The modeling is introduced in section 4.2.1 with a rolling horizon structure. Section 4.2.2 briefly introduces the Wiener process that addresses the time-varying character of transition matrix. The solution algorithm is discussed in section 4.2.3.

4.2.1 Modeling

Time-varying coefficient state space model is introduced to estimate path flow from link traffic counts. In this model, transition matrix at time t can be related to transition matrix at time $t-1$ with an additive form,

$$F_t = W + F_{t-1} + \rho_t \tag{4.1}$$

where W is the drift term and ρ_t is a noise term.

The transition equation is represented as equation 4.2 with a rolling horizon manner.

$$\Psi_t = F_{t-1}\Psi_{t-1} + \boldsymbol{\sigma}_t, \quad t = m, m+1, m+2, \ldots, n \tag{4.2}$$

where $\Psi_t = [\psi_{t-m}, \psi_{t-m+1}, \cdots, \psi_t]$ is a $p \times (m+1)$ matrix composed of $m+1$ state vectors. Let the element in ψ_t, a $p \times 1$ vector, be represent as $\begin{bmatrix} \psi_t(1) \\ \psi_t(2) \\ \vdots \\ \psi_t(p) \end{bmatrix}_{p\times 1}$. Then,

$$\Psi_t = \begin{bmatrix} \psi_{t-m}(1) & \psi_{t-m+1}(1) & \cdots & \psi_t(1) \\ \psi_{t-m}(2) & \psi_{t-m+1}(2) & \cdots & \psi_t(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{t-m}(p) & \psi_{t-m+1}(p) & \cdots & \psi_t(p) \end{bmatrix}_{p\times(m+1)}.$$

The term m is the number of periods in a rolling stage with $m > p-1$ due to the degree of freedom requirement in the computation stage. The $\boldsymbol{\sigma}_t = [\sigma_{t-m}, \sigma_{t-m+1}, \cdots, \sigma_t]$ is composed of $m+1$ noise vectors.

The third equation is the observation equation,

$$y_t = H\psi_t + \gamma_t, \quad t = 1,2,3,\ldots n. \tag{4.3}$$

This observation equation, describe the relationship between path flow and observation, is identical to eq. 3.2 in section 3.2.

The rolling horizon structure can be expressed as follows,

$$\begin{array}{l}
\underbrace{\psi_1 \quad \psi_2 \quad \cdots \quad \psi_m \quad \psi_{m+1}}_{\text{Transition with } F_{m+1}} \quad \psi_{m+2} \quad \cdots \quad \psi_{n-m-1} \quad \psi_{n-m} \quad \cdots \quad \psi_{n-1} \quad \psi_n \\
\psi_1 \quad \underbrace{\psi_2 \quad \cdots \quad \psi_m \quad \psi_{m+1} \quad \psi_{m+2}}_{\text{Transition with } F_{m+2}} \quad \cdots \quad \psi_{n-m-1} \quad \psi_{n-m} \quad \cdots \quad \psi_{n-1} \quad \psi_n \\
\vdots \\
\psi_1 \quad \psi_2 \quad \cdots \quad \psi_m \quad \psi_{m+1} \quad \psi_{m+2} \quad \cdots \quad \underbrace{\psi_{n-m-1} \quad \psi_{n-m} \quad \cdots \quad \psi_{n-1}}_{\text{Transition with } F_{n-1}} \quad \psi_n \\
\psi_1 \quad \psi_2 \quad \cdots \quad \psi_m \quad \psi_{m+1} \quad \psi_{m+2} \quad \cdots \quad \psi_{n-m-1} \quad \underbrace{\psi_{n-m} \quad \cdots \quad \psi_{n-1} \quad \psi_n}_{\text{Transition with } F_n}
\end{array} . \quad (4.4)$$

Solution algorithms, including Kalman Filter and Gibbs Sampler, mentioned in section 3.2 are applied to each rolling stage. After the estimation, state vectors of the same time in each stage will be averaged to give an estimated value.

4.2.2 Wiener Process

Wiener process is a continuous-time stochastic process, often called Brownian motion. A Wiener process Z_t is characterized by

1. $Z_0 = 0$.

2. Z_t is almost surely continuous.

3. Z_t has independent increments with distribution $Z_t - Z_s \sim N(0, t-s)$.

   As in discrete form, $\Delta Z = Z_{t_k} - Z_{t_{k-1}}$, $\Delta t = t_k - t_{k-1}$, $\Delta Z \sim N\left(0, \left(\sqrt{\Delta t}\right)^2\right)$.

The generalized Wiener process $\Omega_t = At + \rho Z_t$ is called a Wiener process with drift A and variance ρ^2. In this case, $\Omega_t \sim N\left(A \cdot \Delta t, \left(b\sqrt{\Delta t}\right)^2\right)$.

From equation 4.1, the expectation value of $E[F_t - F_{t-1}] = E[W + \rho_t] = W$. The drift term W can be estimated by least square method, that

$$\hat{W} = \frac{1}{n}\sum_{t=1}^{n}(F_t - F_{t-1}) = \frac{1}{n}(F_n - F_1). \tag{4.5}$$

The ρ can be estimated by the sample covariance matrix of F_t, where its variance-covariance matrix is

$$\rho = \frac{1}{n}\sum_{t=1}^{n}(F_{i\,t} - \overline{F}_i)(F_{i\,t} - \overline{F}_i)^T \tag{4.6}$$

where $F_{i\,t}$ is the i^{th} row vector of F_t.

4.2.3 Solution Framework

A rolling horizon structure is incorporated in the solution framework. The solution procedure first takes observations from $t = 1 \sim m+1$ into account, and utilize the solution framework presented in section 3.2 to estimate the state vectors, Ψ_{m+1}, and transition matrix, F_{m+2}. Second, the procedure continues to use the observations from $t = 2 \sim m+2$ to estimate their corresponding variables. The solution procedure will keep rolling until it reaches its final stage. State vectors of the same time in each rolling stage, i.e. ψ_2 would exists in stage 1 and 2, will be averaged to give an estimated value.

After deriving F_t by rolling horizon method. The estimation of F_{t+h} can be computed by

$$\begin{aligned} \hat{F}_{t+h} &= W + \hat{F}_{t+h-1} + \rho_{t+h} \\ &= W + \left(W + \hat{F}_{t+h-2} + \rho_{t+h-1}\right) + \rho_{t+h} \\ &= hW + F_t + \sum_{i=1}^{h} \rho_{t+i} \end{aligned} \tag{4.7}$$

Since the expectation value of ρ_t equals to zero, the estimation of F_{t+h} is

$$\hat{F}_{t+h} = hW + F_t. \tag{4.8}$$

The prediction of state vector at $t+1$ can be represent as,

$$\hat{\psi}_{t+1} = \hat{F}_t \cdot \psi_t = (W + F_t + \rho_t) \cdot \psi_t = (W + F_t) \cdot \psi_t. \tag{4.9}$$

And the prediction of state vector at $t+2$ can be represent as,

$$\begin{aligned}\hat{\psi}_{t+2} &= \hat{F}_{t+1} \cdot \hat{\psi}_{t+1} = (W + \hat{F}_{t+1} + \rho_{t+1}) \cdot \hat{\psi}_{t+1} = (W + \hat{F}_{t+1}) \cdot \hat{\psi}_{t+1} \\ &= (W + \hat{F}_{t+1}) \cdot [(W + F_t) \cdot \psi_t] \\ &= [W + (W + F_t)] \cdot [(W + F_t) \cdot \psi_t] \\ &= (2W + F_t) \cdot (W + F_t) \cdot \psi_t\end{aligned} \tag{4.10}$$

The prediction of state vector at $t+h$ can be generalized as,

$$\begin{aligned}\hat{\psi}_{t+h} &= \hat{F}_{t+h-1} \cdot \hat{\psi}_{t+h-1} \\ &= (hW + F_t) \cdot [((h-1)W + F_t) \cdot ((h-2)W + F_t) \cdots (W + F_t) \cdot \psi_t] \\ &= \left[\prod_{i=1}^{h} (iW + F_t)\right] \cdot \psi_t\end{aligned} \tag{4.11}$$

4.3 Estimation of Path Flow Considering Link Travel Time

In the previous section (section 4.2), we relax the assumption of fixed transition matrix by introducing time-varying coefficient dynamic system. However, travel time is not addressed in that section. The travel time issue can be easily taking into consideration in this section. Notations used in this section is the same as that of section 3.2, 3.3, and 4.2.

The time-varying coefficient state space model considering travel time effect is shown as follows.

$$F_t = W + F_{t-1} + \rho_t \tag{4.12}$$

$$\Psi_t = F_{t-1}\Psi_{t-1} + \sigma_t, \quad t = m, m+1, m+2, ..., n \tag{4.13}$$

$$y_t = H_t \psi_t^{lag} + \gamma_t, \quad t = 1,2,3,...n \tag{4.14}$$

Equation 4.12 that describe the relationship of time-dependent coefficient is the same as equation 4.2 in the previous section (section 4.2). The transition equation 4.13 is also the same as equation 4.3 in the previous; while equation 4.14 the describe the link observation considering time lag effect is introduced in section 3.3.

The model proposed in this section is developed from models proposed in previous chapters. Therefore, solution framework can be straight forward derived by combining solution frameworks discussed in the previous chapters. The solution framework is illustrated in figure 4.1.

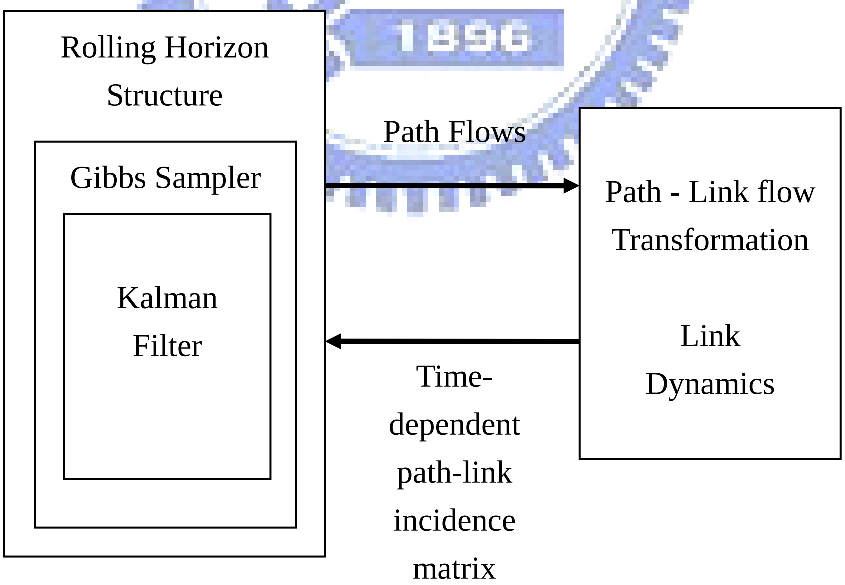


Figure 4.1 The Solution framework of Time-varying coefficient state space model considering travel time effect.

Chapter 5

Convergence Control of Gibbs Sampler and Parallel Implementation

In this chapter, the convergence control method of Gibbs Sampler and parallel implementation of solution framework are discussed. Begin with the convergence assessment for single chain in section 5.1, parallel chain method is then illustrated in section 5.2. Finally, Parallel computing technique and its implementation is addressed in section 5.3.

5.1 Single Chain Convergence Assessments

Gibbs sampler, an example of a Markov Chain Monte Carlo method, is an algorithm that generate sequences of samples from a joint probability distribution. It has been used to tackle a wide variety of statistical problems. If the method had been used naively without convergence control, it might result a misleading answer. Properties of applying Gibbs Sampler to state space model can be referred to Frühwirth-Schnatter (1994), Carter and Kohn (1994), Gamerman (1998), and Jungbacker and Koopman (2007); while analytic convergence rates for Gibbs sampler applied to uni-variate state space model can be referred to Pitt and Shephard (1996).

Theoretical guarantee of convergence is different from the practical requirement. It is thus necessary to develop a tool that determines whether the chain is converged or not. Monitoring the chain that Gibbs sampler produce is a reasonable method.

However, monitoring the elements in a chain would only expose strong non-stationarities. Therefore, it is more relevant to consider the cumulated averages; since they need to stabilize for convergence to be achieved. Let the number of Gibbs iterations be denoted as g, and the corresponding element $x^{(g)}$. The difference between cumulated averages should be less than ε,

$$\left| \frac{1}{g}\sum_{g} x^{(g)} - \frac{1}{g-1}\sum_{g-1} x^{(g)} \right| \le \varepsilon . \tag{5.1}$$

Possible alternatives to empirical averages would be importance sampling, conditional expectations, or Riemann approximation techniques. However, the average may only correspond to the exploration of a single mode of the distribution by the chain; the single chain methods can never guarantee that the whole support of target distribution has been explored (Robert and Cellier, 1998).

5.2 Multiple Chain Convergence Assessments

The main idea of Gibbs sampler, an iterative simulation method, is to draw values of a random variable *x* from a sequence of distributions that converge to a desired target distribution of *x*. Single chain (sequence) methods hardly bring information on the regions of the space it does not visit. Parallel chain methods try to overcome such defect by generating parallel chains, aiming at eliminating the dependence on initial conditions. The convergence control is most often based on the comparison of the estimations of different quantities for the parallel chains. More precisely, the criterion is based on the difference between a weighted estimator of variance for each chain and the variance of the estimators on the different chains. The estimation method composed of two steps. First, create an estimate of the target distribution, centered about its mode (or modes), and over-dispersed in the sense of being more variable than the target distribution. The approximate distribution is then used to start several independent chains of the iterative simulation. The second step is to analyze the multiple chains to form a distributional estimate of the target random variable.

The monitor of convergence of the iterative simulation is to estimate the factor by which the scale of the current distribution for the target distribution might be reduced if the simulations were continued to the limit $n \to \infty$. This potential scale reduction can be estimated by

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}}{W}\frac{df}{df-2}} = \sqrt{\left(\frac{n+1}{n} + \frac{M+1}{Mn}\frac{B}{W}\right)\frac{df}{df-2}} \tag{5.2}$$

where

$$df = \frac{\left(\sum_{m=1}^{M}\frac{S_m^2}{n}\right)^2}{\frac{\sum_{m=1}^{M}S_m^4}{n^2(n-1)}} = \frac{\left(\sum_{m=1}^{M}S_m^2\right)^2(n-1)}{\sum_{m=1}^{M}S_m^4} \tag{5.3}$$

$\hat{R}$ declines to 1 as $n \to \infty$. $\hat{R}$ is the ratio of the current variance estimate, $\hat{V}$, to the within–chain variance, W. n is the count of iterations, S_m are the standard deviation of each chain, B is the between-chain variance and the M is the number of parallel chains. The between- and within-chain variances, B and W, are defined as follows.

$$B = \frac{1}{M}\sum_{m=1}^{M}\left(\bar{\xi}_m - \bar{\xi}\right)^2, \tag{5.4}$$

$$W = \frac{1}{M}\sum_{m=1}^{M}S_m^2 = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}\left(\xi_m^{(i)} - \bar{\xi}_m\right)^2, \tag{5.5}$$

with

$$\bar{\xi}_m = \frac{1}{n}\sum_{i=1}^{n}\xi_m^{(i)}, \qquad \bar{\xi} = \frac{1}{M}\sum_{m=1}^{M}\bar{\xi}_m \tag{5.6}$$

where $\xi_m^{(i)}$ is the *i*-th element in chain *m*.

Once $\hat{R}$ is near 1, it is conclude that each set of the simulated values is close to the target distribution. In practice, the potential scale reduction is chosen to be 10% ($\hat{R} \le 1.1$). After it converges, we can calculate the desired sample value (the state vector) based on the empirical distribution of the *n* simulated iterates for each simulated chain.

Some Gibbs Sampler estimation examples are demonstrated in this section. In the following examples, four parallel chains with different random seeds are illustrated. Example in Figure 5.1 converged ($\hat{R} \le 1.1$) after 1.92×10^5 iterations, and the estimated value is 23.75; example in Figure 5.2 converged after 4.12×10^5 iteration, and the estimated value is 0.11.

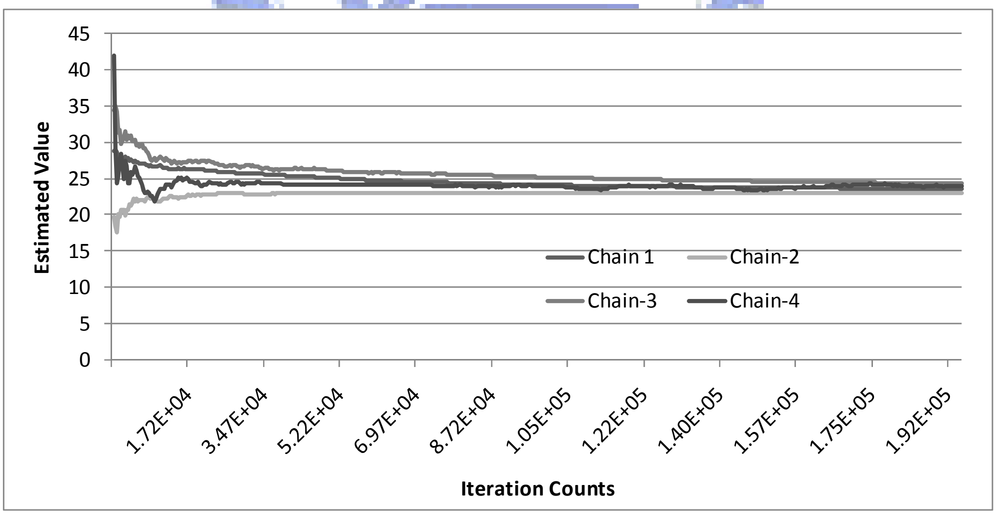


Figure 5.1 Example 1 of four parallel chains.

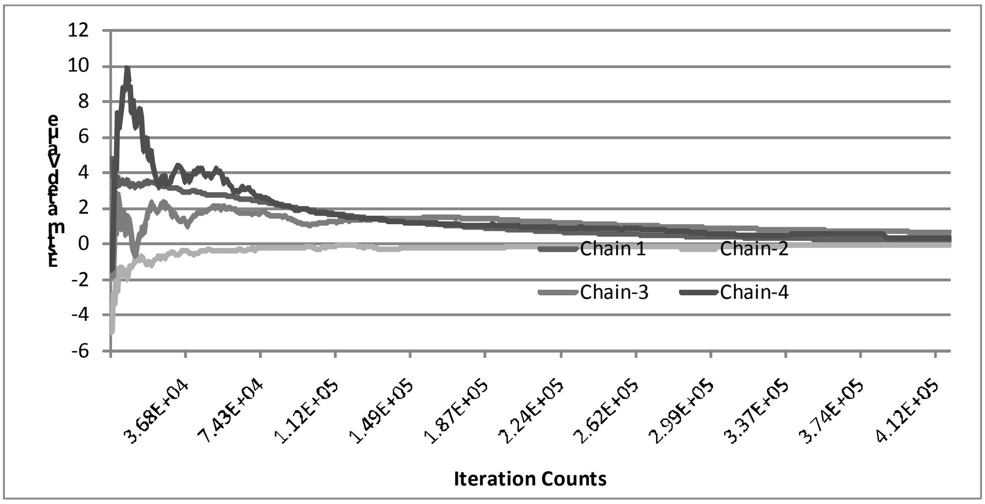


Figure 5.2 Example 2 of four parallel chains.

Another example of eight parallel chains is shown in Figure 5.3.

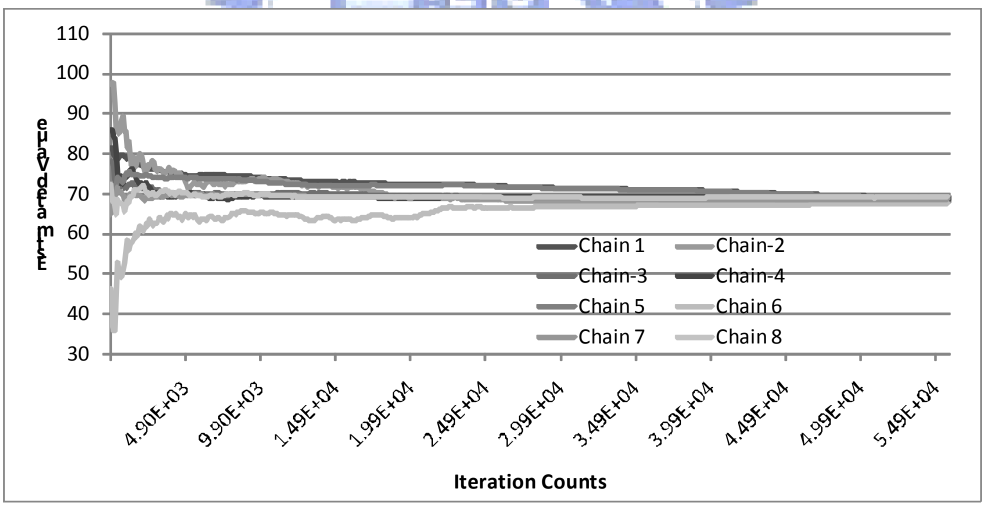


Figure 5.3 Example of eight parallel chains.

These chains converged ($\hat{R} \le 1.1$) after 5.5×10^4 iterations, and the estimated value is 68.57.

5.3 Parallel Implementation

Gibbs sampler, a particular type of Markov Chain Monte Carlo method, requires tremendous iterations during computation; normally, there would be tens of thousands iterations in the proposed algorithm. To achieve real-time information requirement, parallel computing technique is introduced to increase the performance. The solution algorithm should be modified to adopt the parallel implementation. Consider the Gibbs sampler algorithm in section 3.3.2, the algorithm is separated into several computing parts at the WHILE-LOOP. With different random seed, each computing part will lead to a different solution chain. The chain in each computing part will then be gathered to check the convergence. In this situation, communication between computing nodes is minimum, and computing power can be easily increased without communication bandwidth limitation. Figure 5.4 describes the parallel architecture. The parallel environment of this PC-cluster consists of 16 computing nodes; each contains 2 processors equivalent to Intel XEON 3.2 GHz and 1GB memory. Nodes are connected with a Gigabits Ethernet switch for MPI protocol and a 100 Mbits PCI fast Ethernet switch for Network File System (NFS) and Network Information System (NIS).

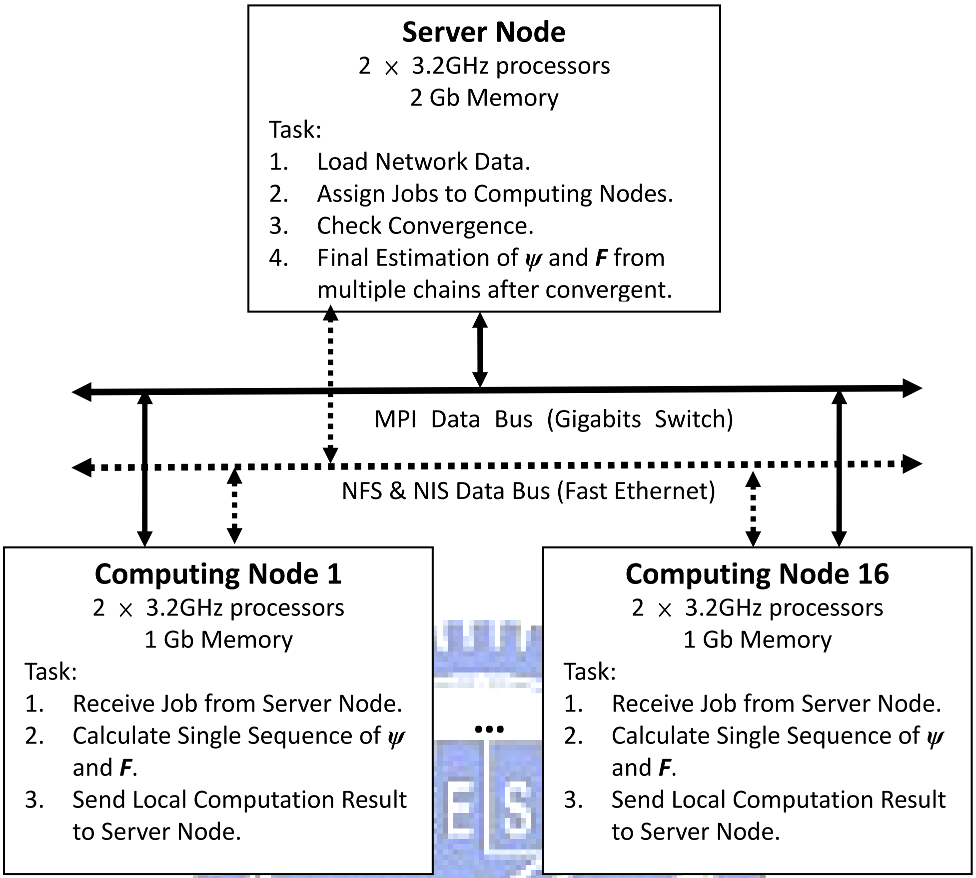


Figure 5.4 The Parallel computation structure.

In the server node, parameters used in our algorithm are initialized, so does the necessary input data. When assign jobs, these input data are been sent to computing nodes existing in the cluster through TCP/IP base intranet with Message Passing Interface (MPI) Library through gigabit switch. The computational procedure for the parallel process consists of:

1. Load input data and parameters. Initialize MPI environment.
2. Count the computing nodes exists in the cluster environment. Send data to each computing nodes.
3. Each computing nodes generate its own $\hat{\psi}$ and $\hat{F}'$ sequences by given input data. These results were sending to the server for convergence check.

4. The server check the convergence of each $\hat{\psi}$ and $\hat{F}'$. If $\hat{R} > 1.1$, the computing nodes will continue step 3. As $\hat{R} \leq 1.1$, the server estimate the global $\hat{\psi}$ and $\hat{F}'$ by sequences generated by each computing nodes.
5. Stop MPI environment. Output data.

Figure 5.5 shows the speedups and efficiencies of the proposed algorithm, where the speedups is the ratio of the code execution time on a single processor to that on multiple processors and efficiency is defined as the speedup divided by the number of processors. As shown in Figure 5.5, the parallel scheme has efficiency of 72.5% in a 32 processors environment.

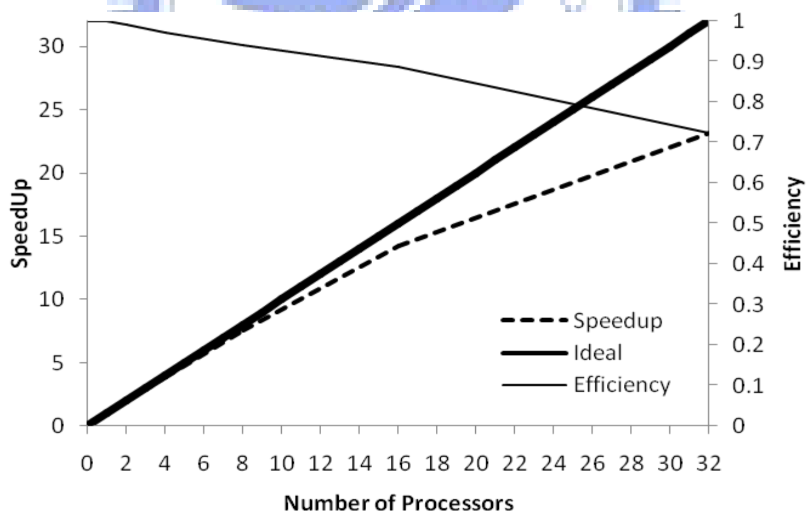


Figure 5.5 Speedups and efficiencies for the parallel computing.

Chapter 6

Numerical Examples

In this section, the proposed models are tested with real networks to discuss their performance. These networks include i) Taipei Mass Rapid Transit Network in section 6.1 and ii) Taiwan Freeway Network with Electronic Toll Collection information in section 6.2.

6.1 Examples 1 (Taipei Mass Rapid Transit Network)

A real network of Taipei MRT is tested in this section. The test network consists of 9 stations in Nangkang line, the topology of test network is demonstrated in figure 6.1. Real path-flow data is not likely to derive on a real road network, but it is possible in the MRT network. Therefore, MRT information is used here to give a measurement of model performance.

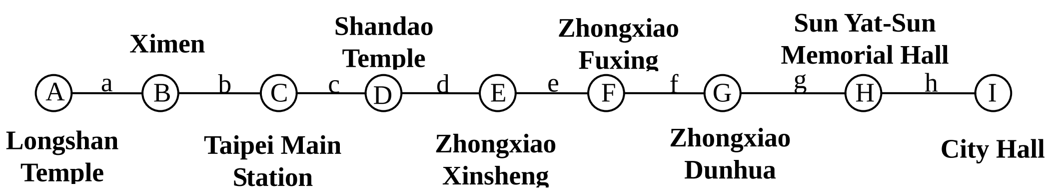


Figure 6.1 The MRT test network.

Real path-flow data are the numbers of passengers traverse between stations from 18:05~20:00 in five-minute interval; for example, there are 30 persons travel from Ximen to Taipei Main Station in the time-period of 18:00~18:35. Eight paths are included in this example defined as travelers toward Taipei Main Station from the rest of stations. These paths are demonstrated in Table 6.1.

Table 6.1 Path set of example MRT network.

Path	**Origin-Destination**	**Link Set**
$\psi(1)$	A → C	a , b
$\psi(2)$	B → C	b
$\psi(3)$	D → C	c
$\psi(4)$	E → C	d, c
$\psi(5)$	F → C	e, d , c
$\psi(6)$	G → C	f, e , d , c
$\psi(7)$	H → C	g, f , e , d ,c
$\psi(8)$	I → C	h, g , f , e , d , c

In the MRT test, travel time is not considered. Therefore, link flows can be easily obtained by the summation path flows on them, i.e., flow on link c is equal to $\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(7)+\psi(8)$, and the flow on link e is equal to $\psi(5)+\psi(6)+\psi(7)+\psi(8)$. The time-dependent link flows are illustrated in Table 6.2. Link flows on b and c are treated as the observation vector, y_t, in this test. Both time-invariant and time-varying coefficient dynamic model proposed in chapter 3 and 4 are conducted in this test.

Table 6.2 Time-dependent link flows of MRT network

Time Interval	Link flow a	Link flow b	Link flow c	Link flow d	Link flow e	Link flow f	Link Flow g	Link Flow h
1	1	12	31	27	27	17	13	7
2	3	14	116	108	103	50	30	23
3	0	12	85	84	77	39	14	9
4	6	15	82	77	72	33	21	11
5	5	25	74	69	63	42	16	10
6	2	10	71	69	65	34	17	11
7	3	33	115	108	101	56	28	12
8	3	26	62	58	56	29	12	7
9	6	29	62	59	58	34	23	15
10	1	28	84	81	71	45	21	10
11	3	19	80	72	65	28	15	7
12	3	29	113	107	103	63	31	20
13	7	35	129	119	110	51	25	17
14	2	18	50	45	44	18	7	7
15	6	39	128	121	113	58	20	12
16	6	35	97	91	83	43	15	11
17	1	12	107	100	93	58	35	25
18	3	16	108	101	97	50	25	15
19	0	22	76	76	71	32	19	8
20	4	29	69	66	63	26	13	7
21	1	17	101	100	96	63	18	12
22	2	9	41	40	37	29	10	5
23	2	6	59	57	55	35	14	8

6.1.1 Time-invariant coefficient model

In the time-invariant coefficient model, transition matrix is assumed to be fixed. The estimated path flows compare to real path flows are demonstrated in Figure 6.2. The black line indicates the real O-D data derived from the Taipei Mass Rapid Transit company, and the grey line indicates the estimation result of the proposed algorithm.

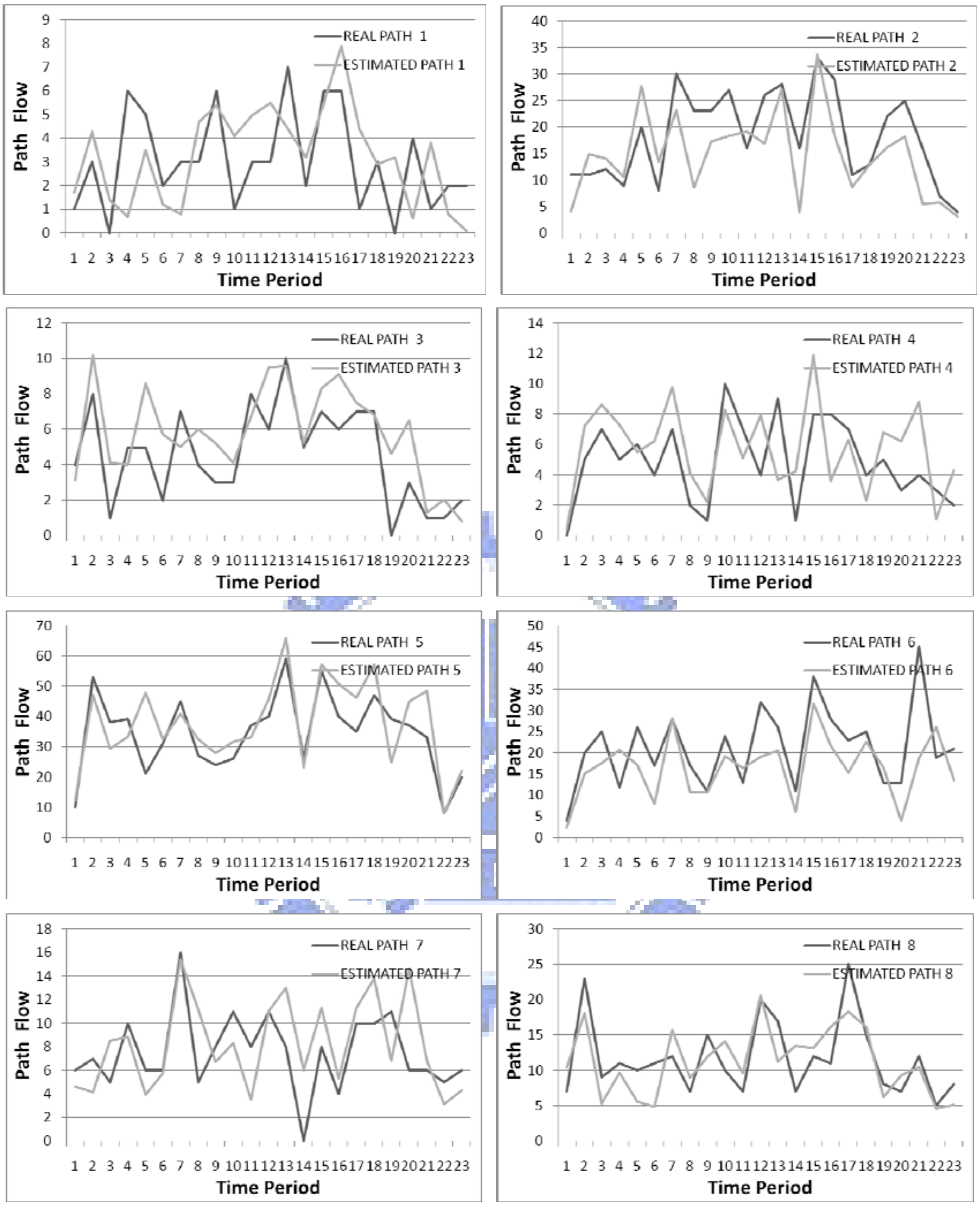
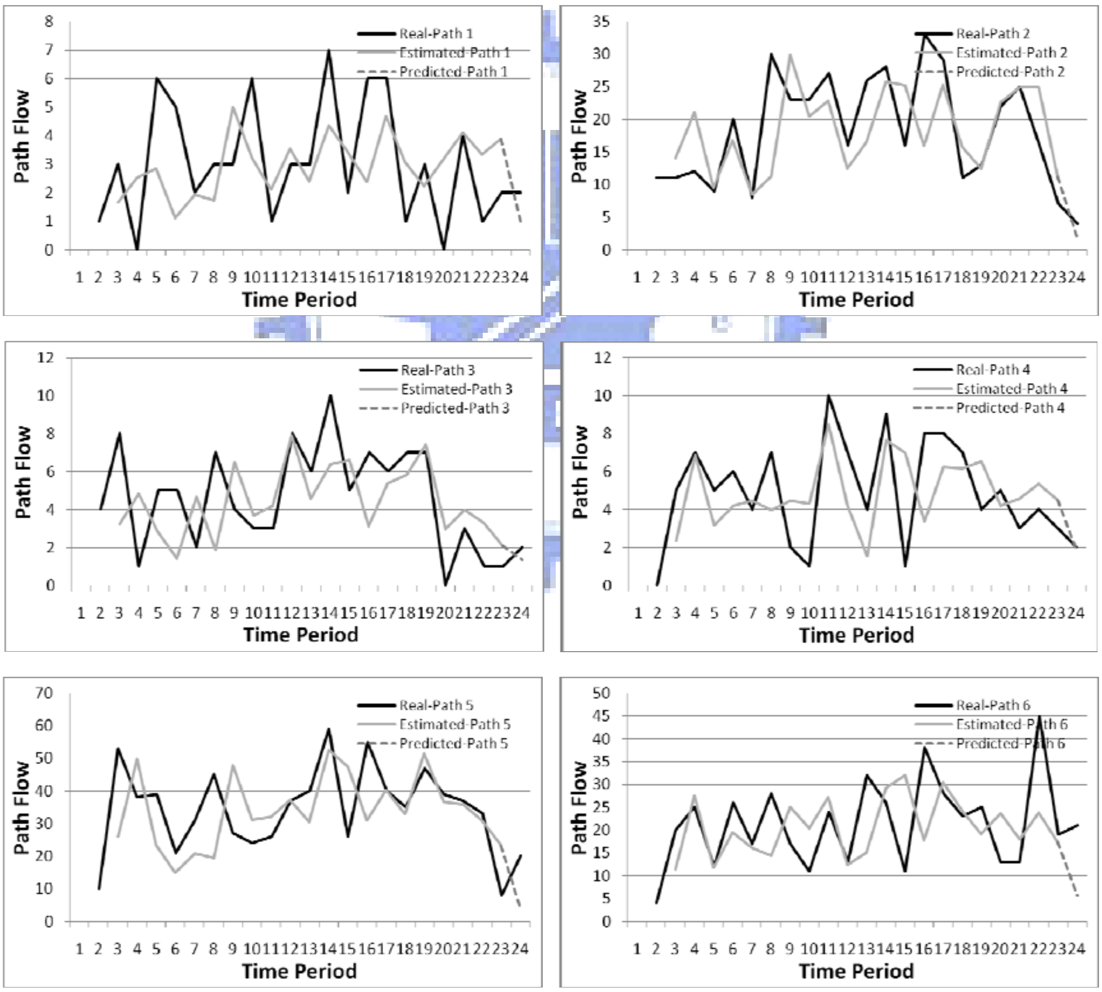


Figure 6.2 The comparison of real and estimated data on MRT network by time-invariant coefficient model.

The correlation between real and estimated data is 0.816, which indicates a medium to strong association between real and estimated data. The mean absolute error (MAE) is 4.78, and the mean absolute percentage error (MAPE) is 49%.

To evaluate the prediction result of this model, observations at the last time-period is deducted from the observation vector. After applying the proposed model to the rest of observation vector, estimated transition matrix, $\hat{F}$, and state vector at time-period 22, $\hat{\psi}_{22}$, are obtained. The predicted path flows at time-period 23 can be calculated by $\hat{\psi}_{23} = \hat{F}\hat{\psi}_{22}$. The mean absolute error of the prediction is 5.43; and the mean absolute percentage error is 51.3%. The comparisons of real and predicted path flows are demonstrated in figure 6.3. The predicted path flows are indicated as dash lines.



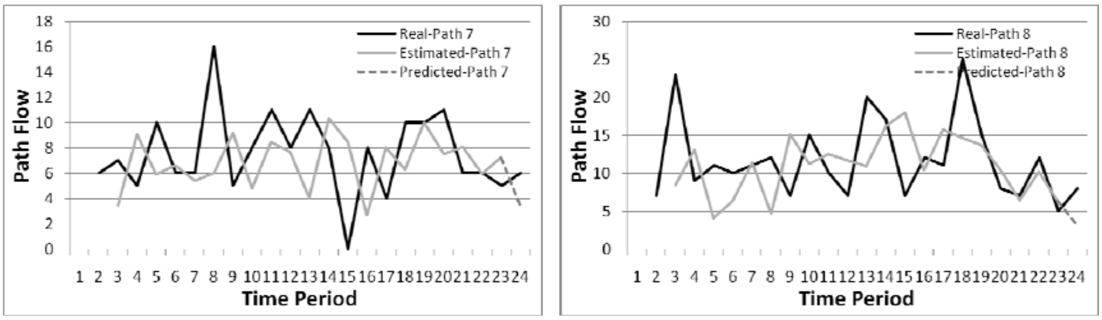


Figure 6.3 The comparison of real and predicted data on MRT network by time-invariant coefficient model.

6.1.2Time-varying coefficient model

This subsection demonstrates the result of proposed time-varying coefficient model. The path flow estimation is based on a rolling horizon concept; in this example, there are nine states in a horizon period, is illustrated on table 6.3.

To evaluate the estimation and prediction result of this model, same procedures are conducted as that of time-invariant coefficient model. The correlation between real and estimated data is 0.969, which indicates a strong association between real and estimated data. The mean absolute error is 1.57, and the mean absolute percentage error is 15.1%. Six out of eight path flows, without path 5 and 8, pass the chi-square test of 95% confidence interval. From the solution framework, we know the first time-period of estimation tends to have larger error, because the ψ_0 is generated randomly. If the estimation of path flows in first time-period are deducted from the dataset, all of the path flows pass the chi-square test of 95% confidence interval.

The predicted path flows at time-period 23 can be calculated by $\hat{\psi}_{23} = \hat{F}_{22}\hat{\psi}_{22}$. The mean absolute error of the prediction is 3.54; and the mean absolute percentage error is 72.2%. The comparison of real, estimated and predicted path flows are demonstrated in figure 6.4.

Table 6.3 An example of rolling horizon estimation (9 elements in a horizon period).

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Real Path	11	11	12	9	20	8	30	23	23	27	16	26	28	16	33	29	11	13	22	25	16	7	4
Estimated Path		10.5	15.0	14.9	21.1	9.2	26.3	27.1	23.6	27.8	17.0	24.0	29.9	20.9	32.2	27.6	11.3	12.5	21.5	23.6	14.8	5.6	0
Period 1		10.5	13.4	9.2	20.5	8.9	31.1	24.8	23.7	28.2													
Period 2			16.6	20.0	23.6	13.0	22.0	29.9	20.8	34.6	16.2												
Period 3				15.5	19.2	9.9	23.3	26.2	18.0	31.7	13.2	28.0											
Period 4					21.2	4.9	24.8	19.1	22.8	23.2	15.6	22.9	28.2										
Period 5						9.2	26.7	34.5	19.0	28.1	18.5	17.3	31.3	29.5									
Period 6							29.7	31.2	18.5	28.6	17.7	21.2	28.6	22.5	38.1								
Period 7								23.9	32.9	24.4	16.6	23.7	29.8	23.0	28.5	25.9							
Period 8									32.8	24.2	16.4	23.3	29.6	23.0	28.4	25.4	12.1						
Period 9										27.3	20.2	26.7	37.2	26.3	34.1	27.1	14.7	12.5					
Period 10											18.5	28.1	30.3	18.1	35.2	30.5	14.3	15.0	25.2				
Period 11												24.9	26.9	14.7	31.9	27.6	10.5	11.8	21.2	23.9			
Period 12													26.9	14.7	31.7	27.3	10.2	11.5	21.0	23.8	14.4		
Period 13														16.8	33.4	29.8	11.0	13.6	22.2	25.4	16.6	7.5	
Period 14															28.9	27.4	6.4	10.3	17.9	21.1	13.4	3.8	0

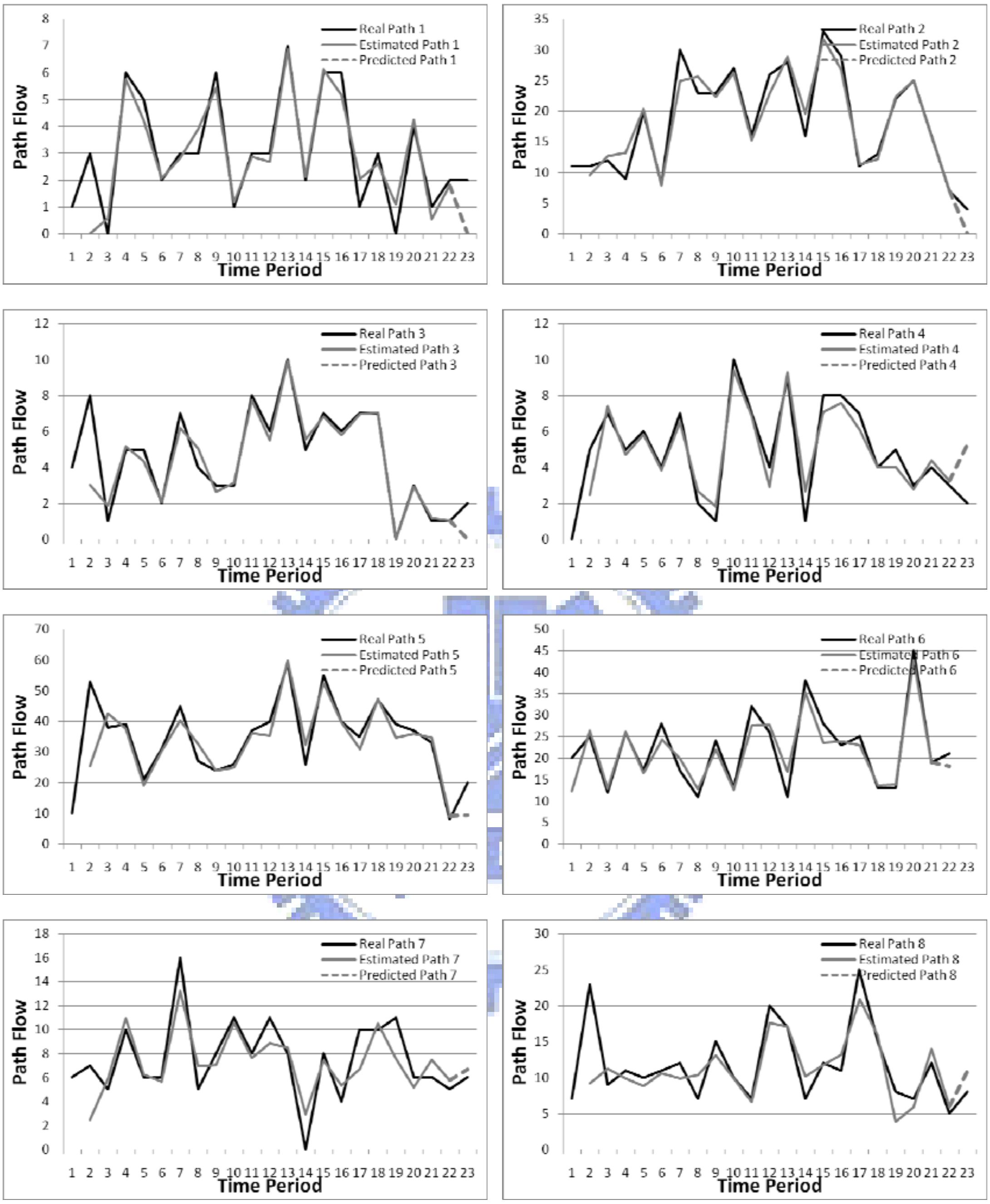


Figure 6.4 The comparison of real, estimated, and predicted path flows on MRT network by time-varying coefficient model with 9 states in a horizon period.

6.1.3 Discussion of MRT network example

From the numerical example of MRT test network, we could discover a significant improvement comparing the time-varying coefficient model to the time-invariant coefficient mode. As for the estimation, the time-invariant model has a MAE of 4.78 and MAPE of 49%; while the time-varying model has a MAE 1.57 of and MAPE of 15.1%. As for the prediction, the time-varying coefficient model also has smaller mean absolute error. However, the mean absolute percentage error of time-varying coefficient model is larger.

6.2 Examples 2 (Taiwan Freeway Network)

Models considering travel time effect are demonstrated in this section. The test network is a real freeway network with Electronic Toll Collection (ETC) information. ETC can provide where and when the particular vehicle appeared, this information can help us tracing a particular vehicle. The test network consists of northern part of Taiwan freeway network, including part of freeway No.1, No. 2, and No. 3. The network is illustrated as figure 6.5.

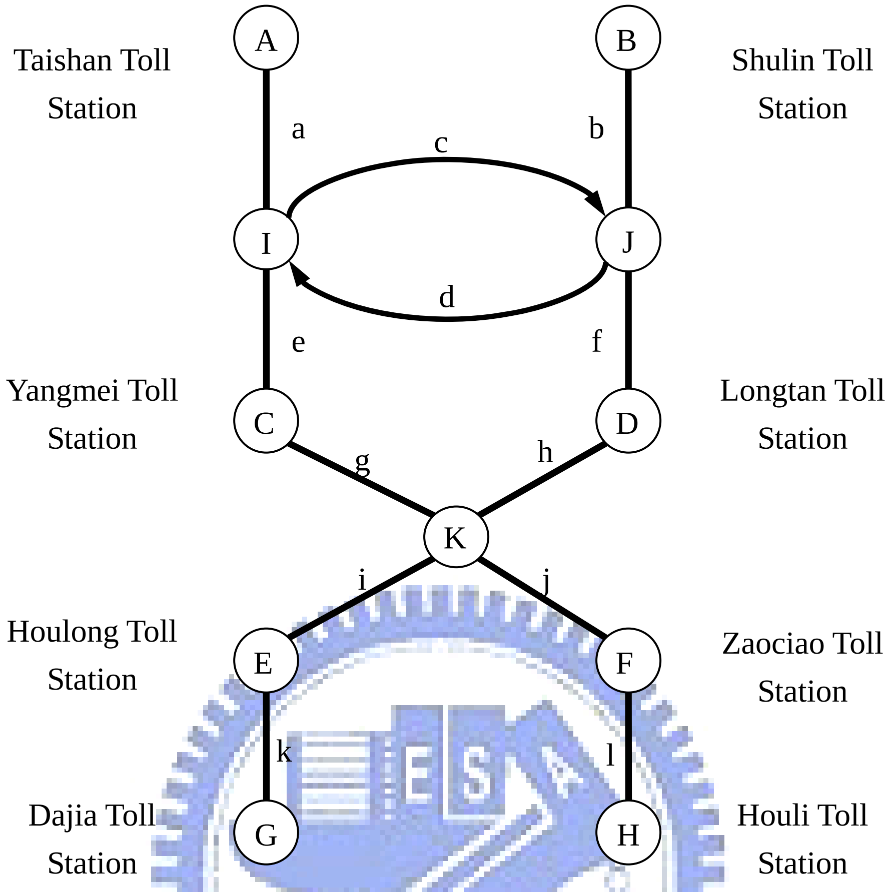


Figure 6.5 The freeway test network.

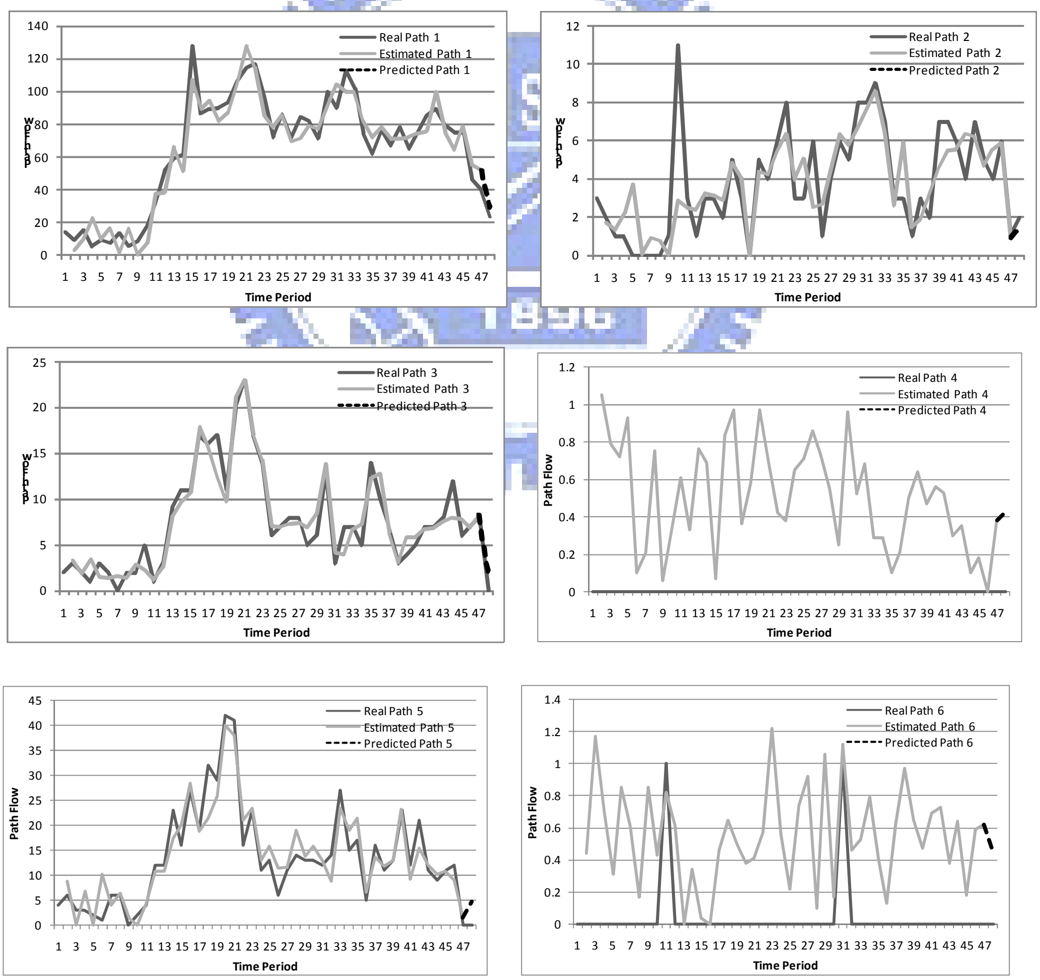
The test network consists of 22 O-D pairs and 30 paths, demonstrated in Table 6.4. Path flow data are the number of vehicles depart from their origin toll station to their destination toll station, only vehicles with ETC equipment and travel longer than two toll stations will be count. Since ETC information provides when and where a particular vehicle appears, travel time between every two stations can be calculated. Therefore, link flow on certain time can be obtained by the summation of corresponding path flows departs from their origin on a particular time. For example, flow on link e at time t can be calculated by the summation of path 2, 4, 6, 8, 10 at time $t-2$ and path 12, 14, 16, 18, 20 at time $t-1$.

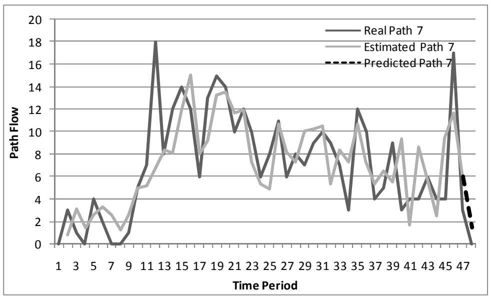
Table 6.4 Path set of example freeway network.

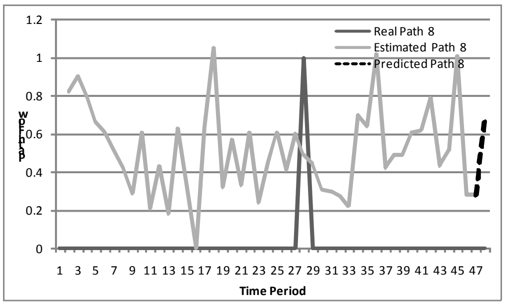
Path	Origin-Destination	Link Set
$\psi(1)$	B → D	b
$\psi(2)$	B → C	b, d ,e
$\psi(3)$	B → E	b, f, h, i
$\psi(4)$	B → E	b, d, e, g, i
$\psi(5)$	B → G	b, f, h, i, k
$\psi(6)$	B → G	b, d, e, g, i, k
$\psi(7)$	B → F	b, f, h, j
$\psi(8)$	B→ F	b, d, e, g, j
$\psi(9)$	B → H	b, f, h, j, l
$\psi(10)$	B → H	b, d, e, g, j, l
$\psi(11)$	A → D	a, c, f
$\psi(12)$	A → C	a, e
$\psi(13)$	A → E	a, c, f, h, i
$\psi(14)$	A → E	a, e, g, i
$\psi(15)$	A → G	a, c, f, h, i, k
$\psi(16)$	A → G	a, e, g, i, k
$\psi(17)$	A → F	a, c, f, h, j
$\psi(18)$	A → F	a, e, g, j
$\psi(19)$	A → H	a, c, f, h, j, l
$\psi(20)$	A → H	a, e, g, j, l
$\psi(21)$	D → E	h, i
$\psi(22)$	D → G	h, i, k
$\psi(23)$	D → F	h, j
$\psi(24)$	D → H	h, j, l
$\psi(25)$	C → E	g, i
$\psi(26)$	C → G	g, i, k
$\psi(27)$	C → F	g, j
$\psi(28)$	C→ H	g, j, l
$\psi(29)$	E → G	k
$\psi(30)$	F → H	l

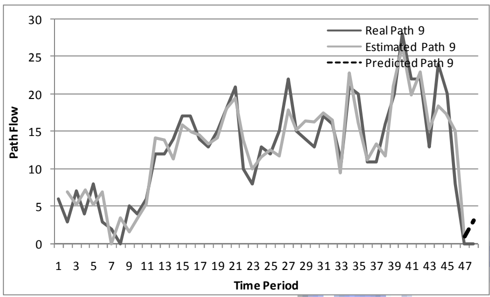
6.2.1 Time-invariant coefficient model considering travel time effect

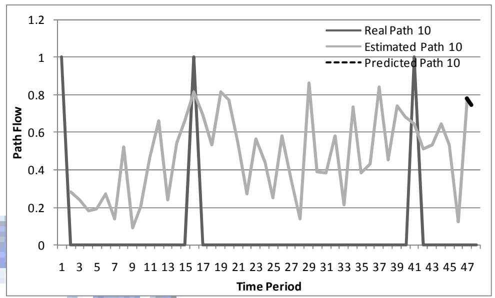
Time-invariant coefficient model considering travel time effect is discussed in this subsection. In this subsection, transition matrix is assumed to be fixed, while path-link incidence matrix, H_t, is generated by real data. The observation vector is time-dependent link flows on the network, aggregated as 30-minute interval. The estimated path flows compare to real path flows are demonstrated in Figure 6.6 as follows.

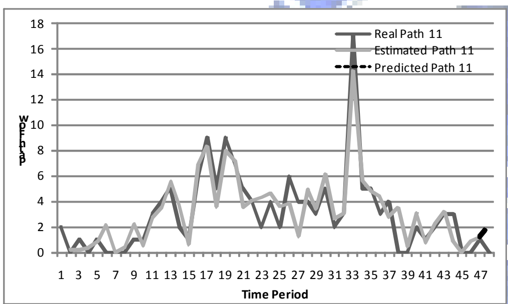


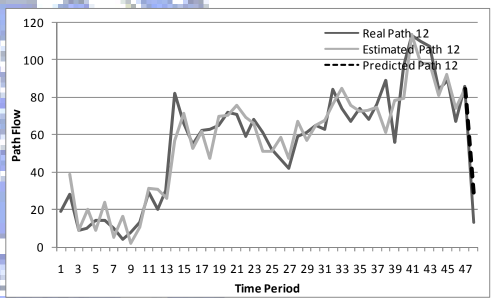

Real Path 7
Estimated Path 7
Predicted Path 7
Path Flow
Time Period

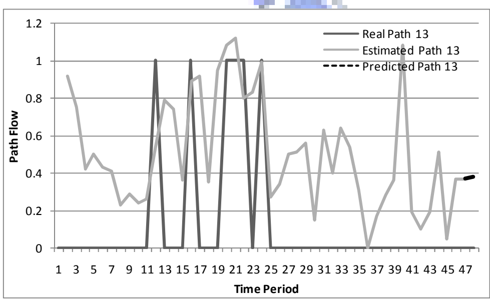

Real Path 8
Estimated Path 8
Predicted Path 8
Time Period

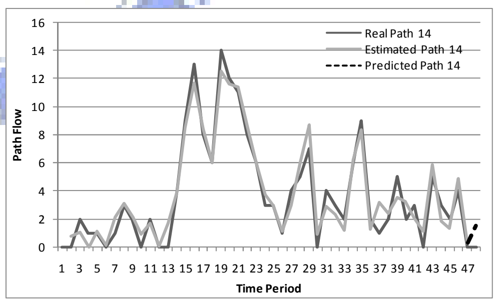

Real Path 9
Estimated Path 9
Predicted Path 9
Path Flow
Time Period

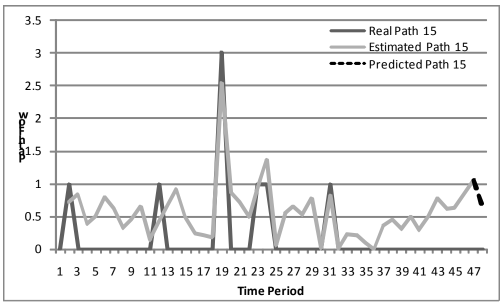

Real Path 10
Estimated Path 10
Predicted Path 10
Path Flow
Time Period

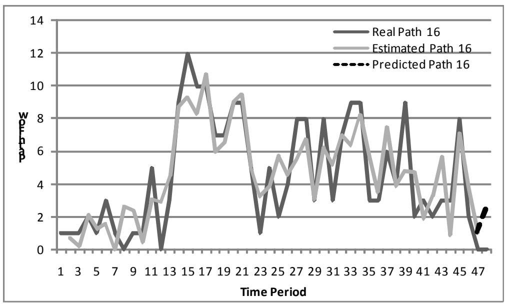

Real Path 11
Estimated Path 11
Predicted Path 11
Time Period

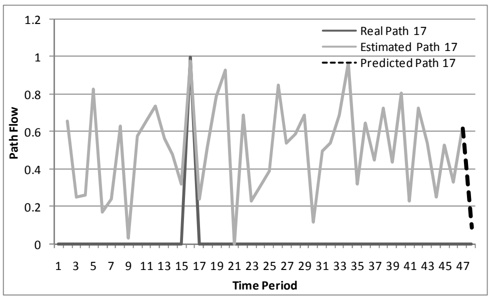
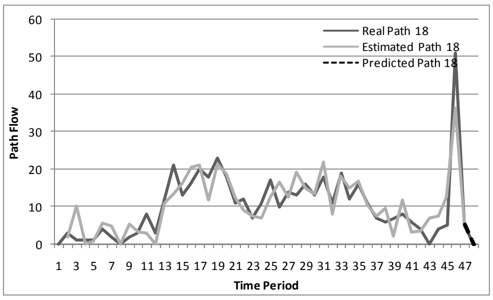
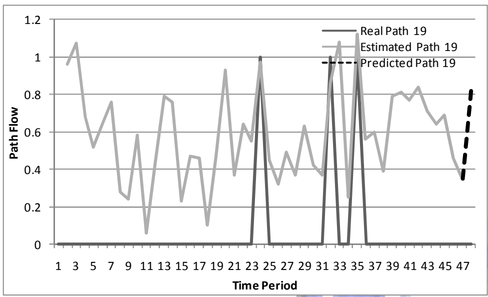
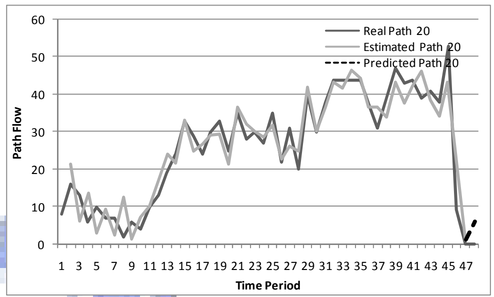
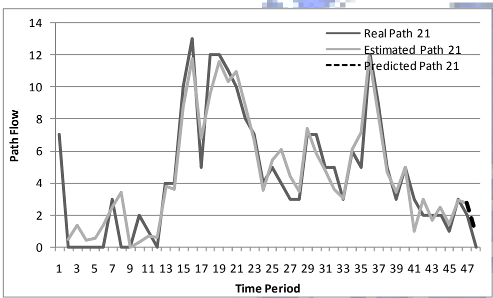
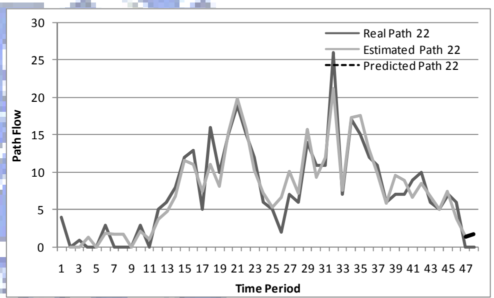
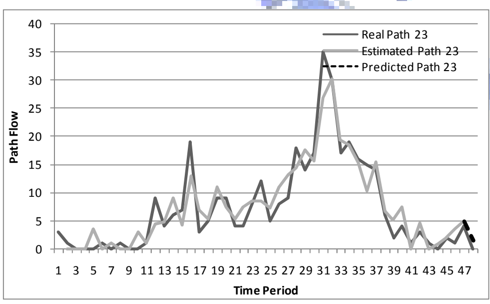
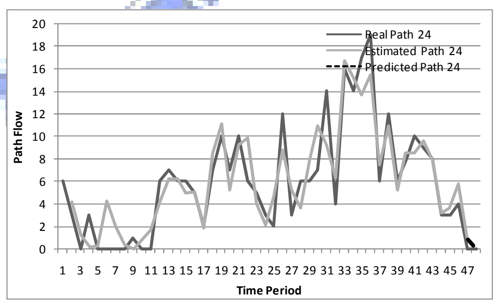
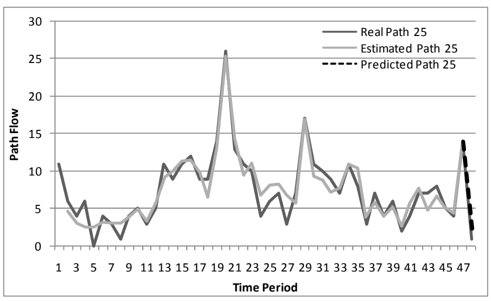
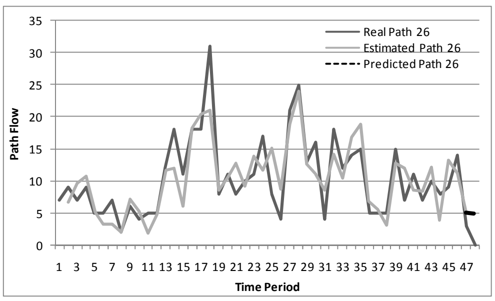

Real Path 12
Estimated Path 12
Predicted Path 12
Path Flow
Time Period


Real Path 13
Estimated Path 13
Predicted Path 13
Path Flow
Time Period


Real Path 14
Estimated Path 14
Predicted Path 14
Path Flow
Time Period


Real Path 15
Estimated Path 15
Predicted Path 15
Time Period


Real Path 16
Estimated Path 16
Predicted Path 16
Time Period


Real Path 17
Estimated Path 17
Predicted Path 17
Path Flow
Time Period

Real Path 18
Estimated Path 18
Predicted Path 18
Path Flow
Time Period

Real Path 19
Estimated Path 19
Predicted Path 19
Path Flow
Time Period

Real Path 20
Estimated Path 20
Predicted Path 20
Path Flow
Time Period

Real Path 21
Estimated Path 21
Predicted Path 21
Path Flow
Time Period

Real Path 22
Estimated Path 22
Predicted Path 22
Path Flow
Time Period

Real Path 23
Estimated Path 23
Predicted Path 23
Path Flow
Time Period

Real Path 24
Estimated Path 24
Predicted Path 24
Path Flow
Time Period

Real Path 25
Estimated Path 25
Predicted Path 25
Path Flow
Time Period

Real Path 26
Estimated Path 26
Predicted Path 26
Path Flow
Time Period

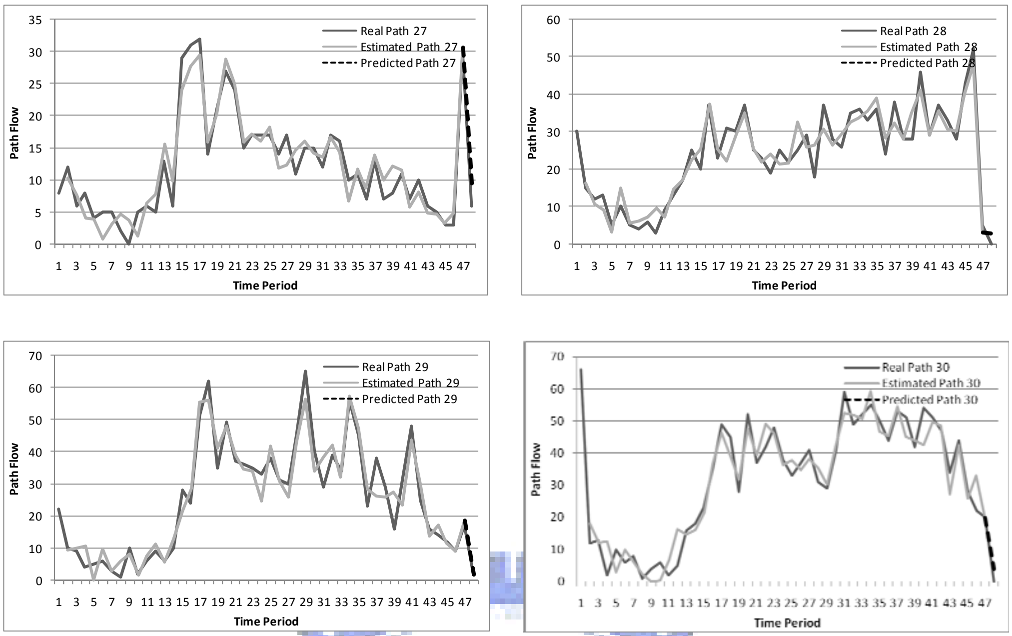


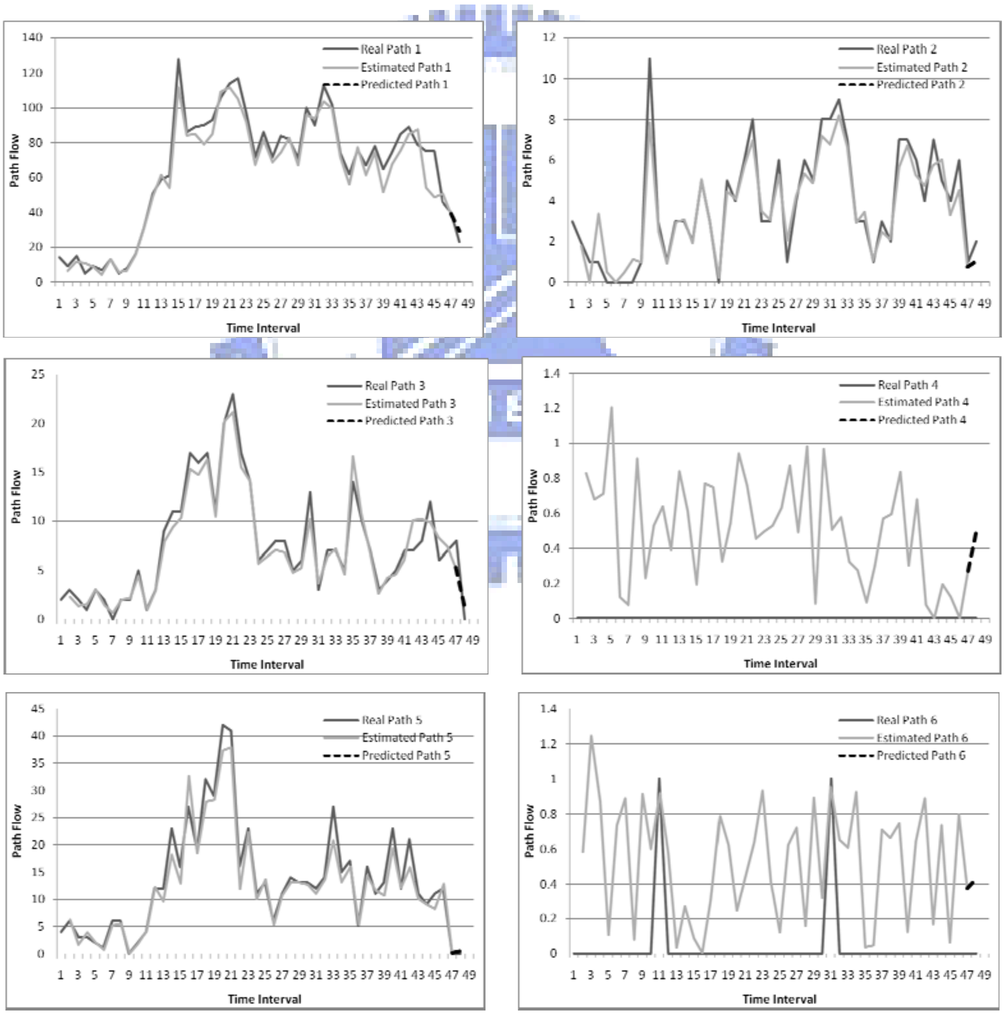
Figure 6.6 The comparison of real and estimated data on freeway network by time-invariant coefficient model considering travel time effect.

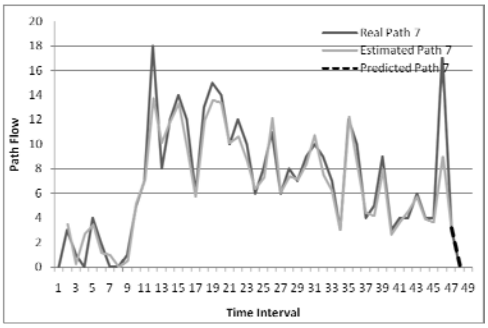
It is observed from the data, there exists nearly zero flows on path 4, 6, 8, 10, 13, 15, 17, and 19. These paths are excluded from the MAE and MAPE calculation, therefore, only 22 paths are considered in the calculation.

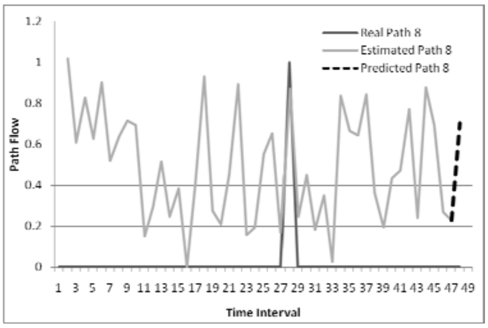
The correlation between real and estimated data is 0.982, which indicates a strong association between real and estimated data. The mean absolute error is 2.48, and the mean absolute percentage error is 31.6%. Eight out of 22 paths pass the paired-sample T test of 90% confidence interval; we further test the rest 14 paths, 10 out of them pass the chi-square test of 90% interval.

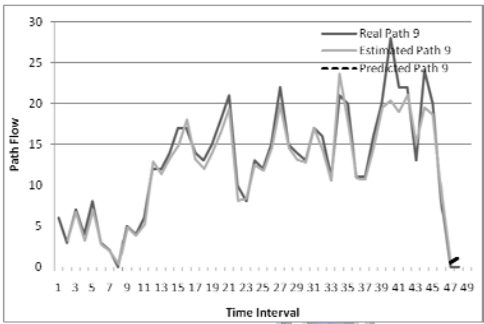
6.2.2 Time-varying coefficient model considering travel time effect

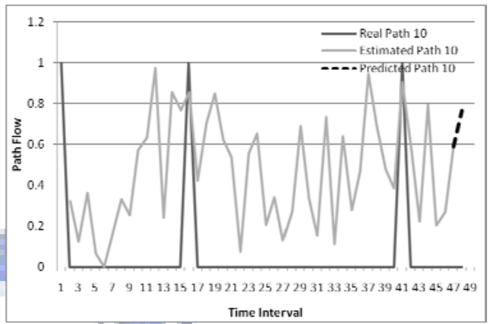
This subsection demonstrated the result of proposed time-varying coefficient model considering travel time effect on the test freeway network. The estimation of path flow is also based on a rolling horizon concept as the MRT network, except of 33 states in a horizon period. The estimated path flows compare to real path flows are illustrated in figure 6.7.

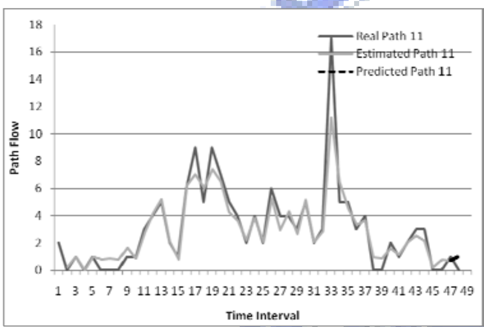


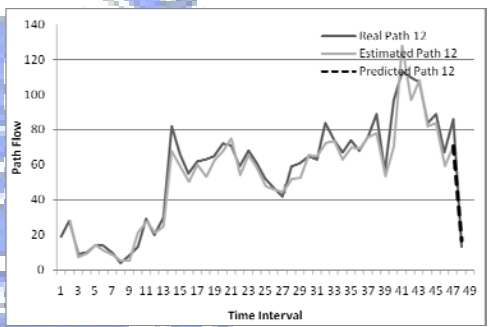

Real Path 7
Estimated Path 7
Predicted Path 7
Path Flow
Time Interval

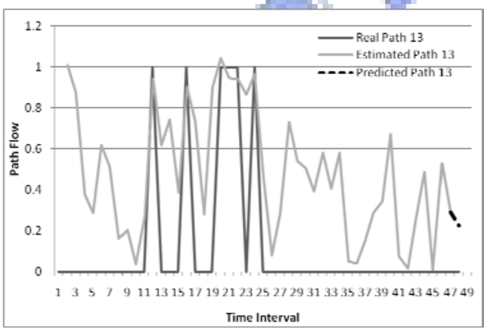

Real Path 8
Estimated Path 8
Predicted Path 8
Path Flow
Time Interval

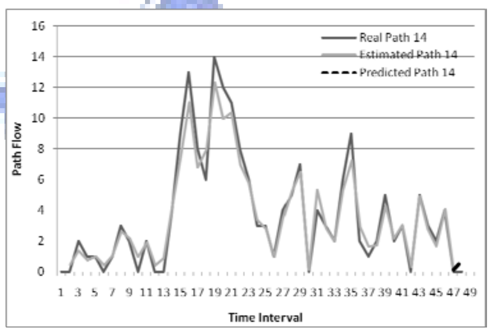

Real Path 9
Estimated Path 9
Predicted Path 9
Path Flow
Time Interval

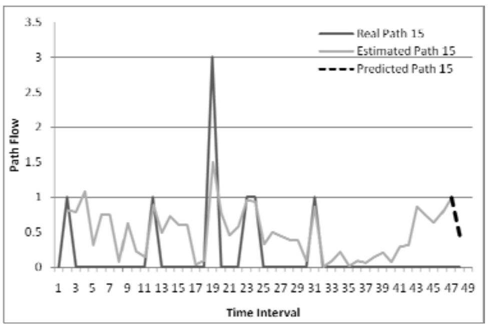

Real Path 10
Estimated Path 10
Predicted Path 10
Path Flow
Time Interval

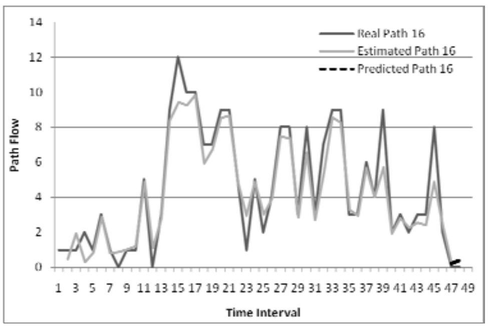

Real Path 11
Estimated Path 11
Predicted Path 11
Path Flow
Time Interval

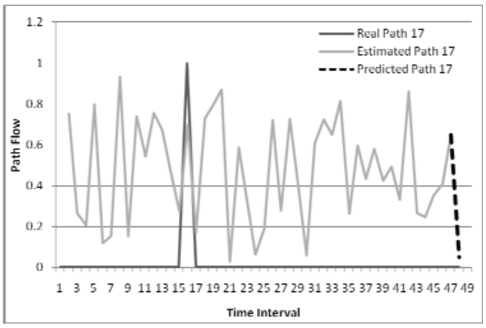

Real Path 12
Estimated Path 12
Predicted Path 12
Path Flow
Time Interval

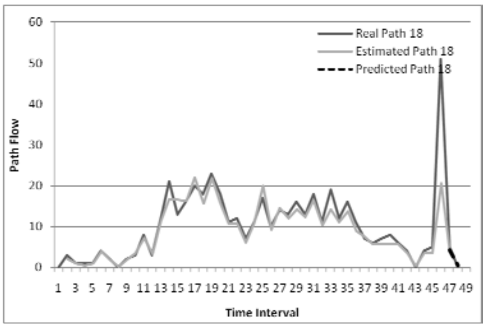

Real Path 13
Estimated Path 13
Predicted Path 13
Path Flow
Time Interval

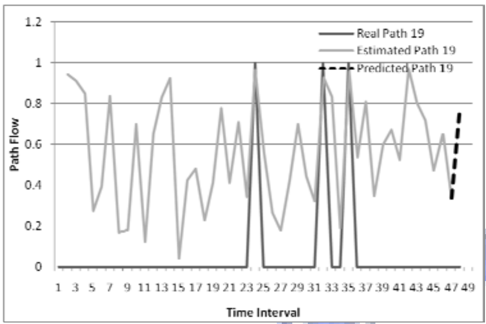

Real Path 14
Estimated Path 14
Predicted Path 14
Path Flow
Time Interval

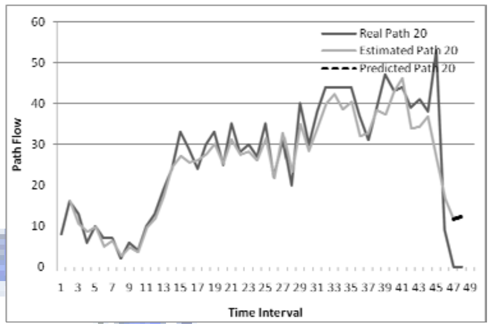

Real Path 15
Estimated Path 15
Predicted Path 15
Path Flow
Time Interval

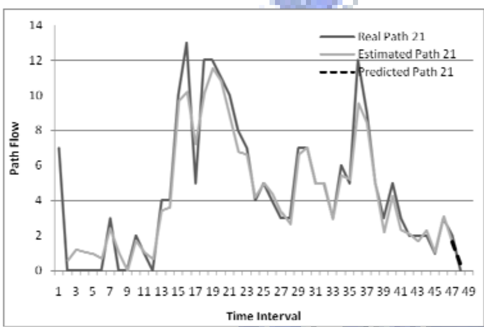

Real Path 16
Estimated Path 16
Predicted Path 16
Path Flow
Time Interval

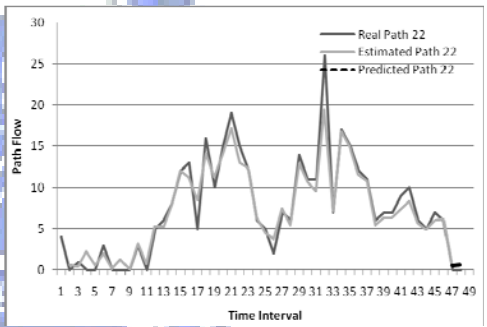

Real Path 17
Estimated Path 17
Predicted Path 17
Path Flow
Time Interval

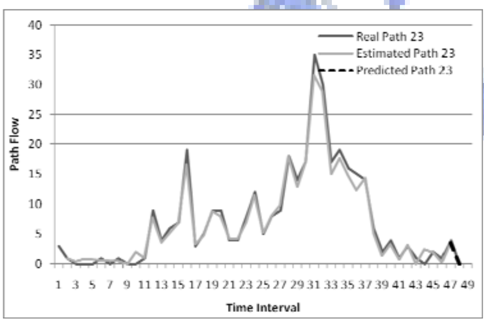

Real Path 18
Estimated Path 18
Predicted Path 18
Path Flow
Time Interval

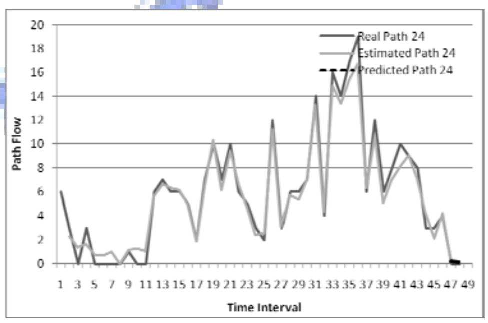

Real Path 19
Estimated Path 19
Predicted Path 19
Path Flow
Time Interval

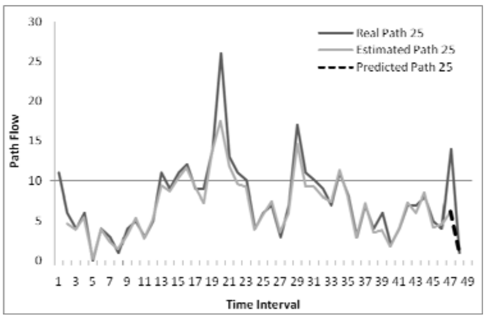

Real Path 20
Estimated Path 20
Predicted Path 20
Path Flow
Time Interval

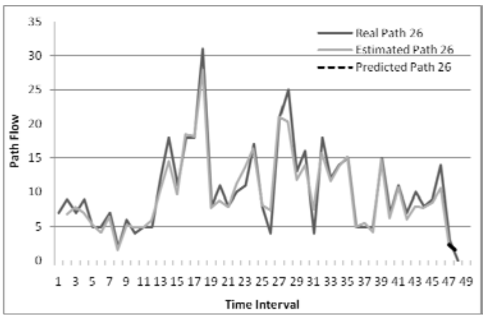

Real Path 21
Estimated Path 21
Predicted Path 21
Path Flow
Time Interval


Real Path 22
Estimated Path 22
Predicted Path 22
Path Flow
Time Interval


Real Path 23
Estimated Path 23
Predicted Path 23
Path Flow
Time Interval


Real Path 24
Estimated Path 24
Predicted Path 24
Path Flow
Time Interval


Real Path 25
Estimated Path 25
Predicted Path 25
Path Flow
Time Interval


Real Path 26
Estimated Path 26
Predicted Path 26
Path Flow
Time Interval

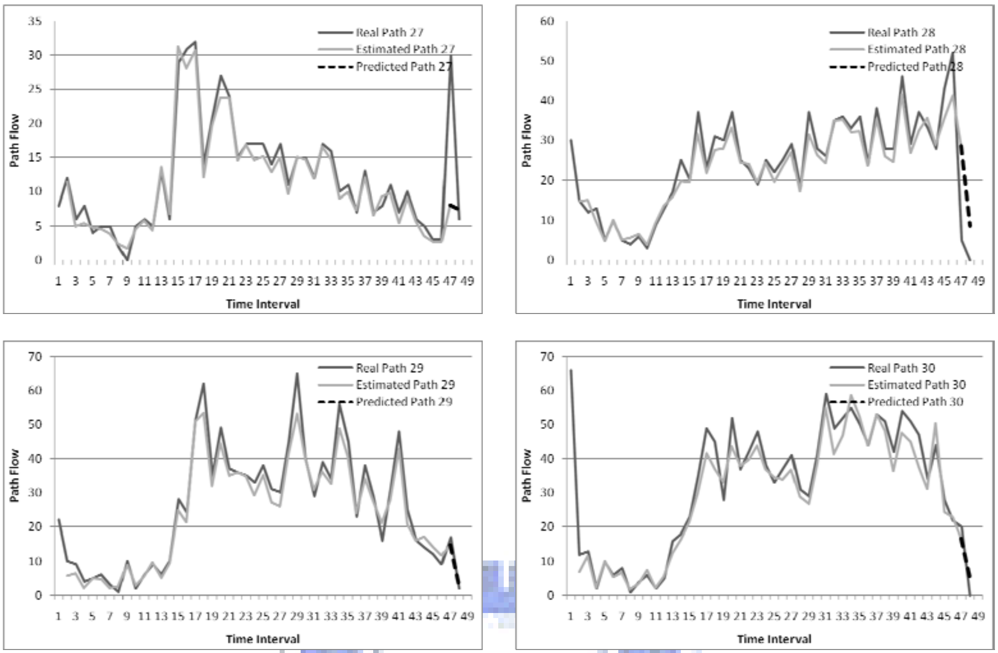


Figure 6.7 The comparison of real and estimated data on freeway network by time-varying coefficient model considering travel time effect.

Same as section 6.2.1, path 4, 6, 8, 10, 13, 15, 17, and 19 are excluded from the MAE and MAPE calculation with their nearly zero flows. The correlation between real and estimated data is 0.989, which indicates a strong association between real and estimated data. The mean absolute error is 1.66, and the mean absolute percentage error is 13.29%. Nine out of 22 paths pass the paired-sample T test of 90% confidence interval; we further test the rest 13 paths, 10 out of them pass the chi-square test of 90% confidence interval.

In the numerical example of freeway test network, the time-varying model also has a better performance compared to time-invariant model in estimation. However, the difference between these models is not as significant as that of MRT example. Some of the paths in this numerical example have zero or very little flows during the whole estimation period, i.e. path 4, 6, 8, 10 and so on. Neither time-invariant nor time-varying model are able to describe those paths.

6.2.3 Comparison with Predetermined Transition Matrix Method

Existing researches concerning time-dependent O-D estimation usually assume the prior information of the O-D matrix (or transition matrix) is know (or at least partially known). This section compares the algorithm proposed in this study with predetermined transition matrix method. The example is based on the freeway network as previous section. However, the proposed algorithm no longer generates the transition matrix; it is now calculated by historical data. The transition matrix in this example is a time-dependent diagonal matrix that

$$F_t = \begin{bmatrix} \frac{\bar{x}(1)_{t+1}}{\bar{x}(1)_t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\bar{x}(p)_{t+1}}{\bar{x}(p)_t} \end{bmatrix} \qquad (6.1)$$

where $\bar{x}(1)_t$ denotes the path flow of 1st path on time t.

The comparison of predetermined transition matrix and algorithm proposed in this research is demonstrated in Figure 6.8; only path 1 to 4 are illustrated for simplicity. In figure 6.8, the real path flow, flow estimated by predetermined transition matrix, and flow estimated by the proposed model are indicated in black, dash, and grey lines respectively.

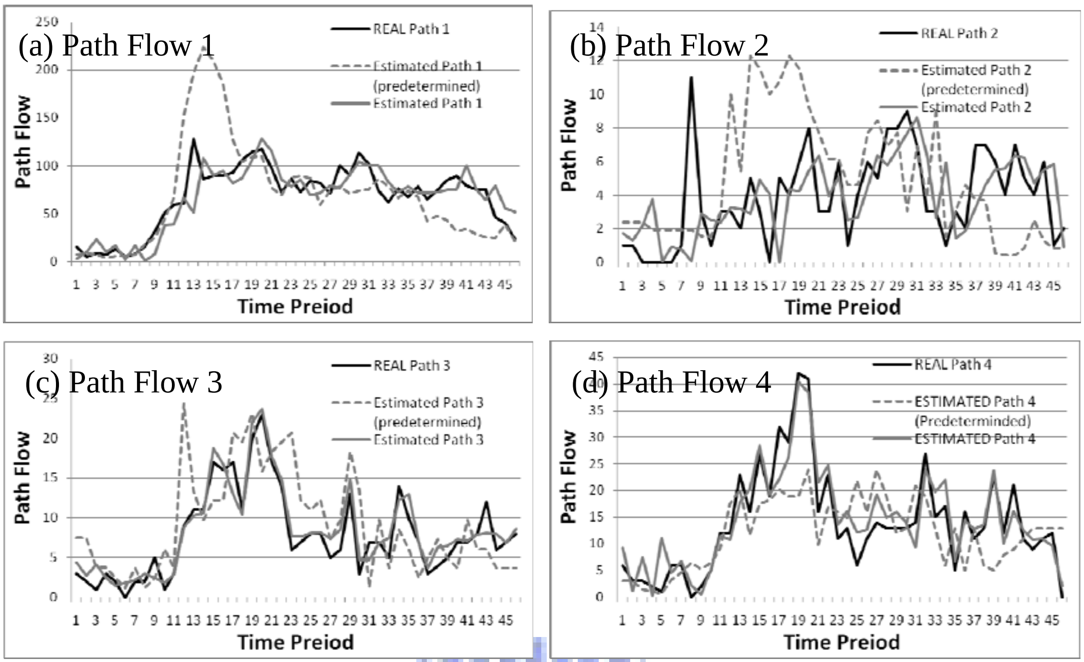


Figure 6.8 The result comparison with predetermined transition matrix.

The mean absolute error is 4.79, and the mean absolute percentage error is 61.87%. Predetermined transition matrix tends to preserve the O-D pattern of historical data; however, the O-D pattern may vary from day to day. Although predetermined transition matrix might have the advantage of preserving O-D pattern, but it might sometimes misleading the results.

Chapter 7

Conclusions and Perspectives

In this chapter, some conclusions and perspective of this study is made. Network flow information is a fundamental input of the Advanced Traffic Management System (ATMS). To meet the need of ATMS, it is necessary to develop a suitable model that estimate network status based on partial information provided by surveillance system installed on networks. Existing works that address this problem often involves user equilibrium behavior or system optimal status. User equilibrium is a way to describe flow on network; however, the existence of such status is questionable. Therefore, this study suggests a statistic approach, rather than a behavior approach, to estimate path flows on network. In this closing chapter, the conclusions are mentioned in section 7.1, and the future researches of this study are suggested in section 7.2.

7.1 Conclusions

The main objective of this study is to develop a path-base assignment model by linear dynamic system with Bayesian Approach. Results of this study are summarized as follows.

(1) Path-base assignment model with linear dynamic system.

This study formulates the path-base assignment problem by both time-invariant and time-varying coefficient state space model. The model is capable to estimate time-dependent path flows with partially observed link flows without prior information. Most existing works on path-flow (or O-D flow) estimation focused on surveillance data usually assume the prior information of O-D matrix (or transition matrix) is known (or at least partially known). In this study, we relax such assumption by combining Gibbs sampler and Kalman filter in the solution algorithm.

(2) Convergence assessment of Gibbs sampler

Gibbs sampler, a particular type of Markov Chain Monte Carlo method, is a powerful tool that being widely applied in many discipline. However, if the method had been used naively without convergence control, it might misleading the answer. Most single chain convergence control method focused on monitoring whether the cumulative sum of sequence is stable or not. Although the concept of single chain methods are straight forward, but they cannot bring information on regions it does not visit. Parallel chain methods try to overcome such defect by generating multiple chains, aiming at eliminating the dependence on initial conditions. The convergence control of parallel chain is then based on the comparison of the estimations of different quantities for the parallel chains.

In this study, we suggest a method to monitor the convergence by estimate the factor by which the scale of the current distribution to the target distribution might be reduced if the simulations were continued to the infinity.

(3) Parallel implementation

Gibbs sampler requires tremendous iterations to reach convergence; normally, there would be tens of thousands iterations in the proposed algorithm. To enhance the performance, a parallel computing technique is introduced in this research.

The parallel technique is combined with the parallel chain convergence control; each computing nodes accounts for a unique chain, while the server node monitors their convergence. Since every computing node generates its own chain, only initial data and computed outputs are transferred between served node and computing nodes. Therefore, communication between computing nodes are minimum; that means computing power can be easily increased without communication bandwidth limitation.

7.2 Perspectives

This study suggests dynamic systems to estimate the time-dependent path flows; both time-invariant and time-varying coefficient dynamic systems are discussed. The original motivation of this research is to develop a path-base assignment model that is capable to provide real-time network flow information. Therefore, it can provide real-time information for both ATIS and ATMS to enhance the network performance. The long-term objective of this research is to develop a dynamic traffic management system that can manage the traffic in real-time. However, this study is only the beginning; the roadmap of future researching topics is illustrated as figure 7.1. Future research areas can be categorized as modeling development and implementation issues. The further development of the model is mentioned as follows.

(1) Path-set generation

The path-observation incidence matrix mentioned in this study, H, is artificially given. To make the assignment model more applicable for general network, a path generation model must be proposed.

(2) Considering the continuous dynamic systems.

The path-base assignment combined model in this study is based on the discrete dynamic system. However, the path flows and transition matrix might be considered changing smoothly in time. Therefore, consider a dynamic system that continuous in time but partially observed might be a suitable extension of the original model.

(3) Multiple user class

Travelers in this research are assumed similar in their characteristics; however, to obtain more realistic models, travelers are often divided into classes with different characteristics. An example of the classification is dividing the travelers into buses, passenger cars, and trucks. In real world, some vehicle detectors could provide vehicle type information while other could not. A model that can handle both information sources should be developed.

Secondly, implementation issues of real-time deployment is also in important work. According to the computation experience of this study, the computing time will increase rapidly when the possible path set becomes large. The traffic flow theories that describe the link dynamics to address the travel time is also an important issue. Thus, the perspectives of implementation issues are discussed as follows.

(1) Different parallel schemes

Different parallel schemes may be proposed to enhance the efficiency. Advanced partitioning method could also lead to a better solution. For example, partitioning on networks rather than generating parallel chains could reduce the dimension of state vector.

(2) Considering the intersection delays

In the urban traffic network, intersection delays contribute most to the travel times. A model that describes intersection delays could be considered while implementing the assignment model in urban networks.

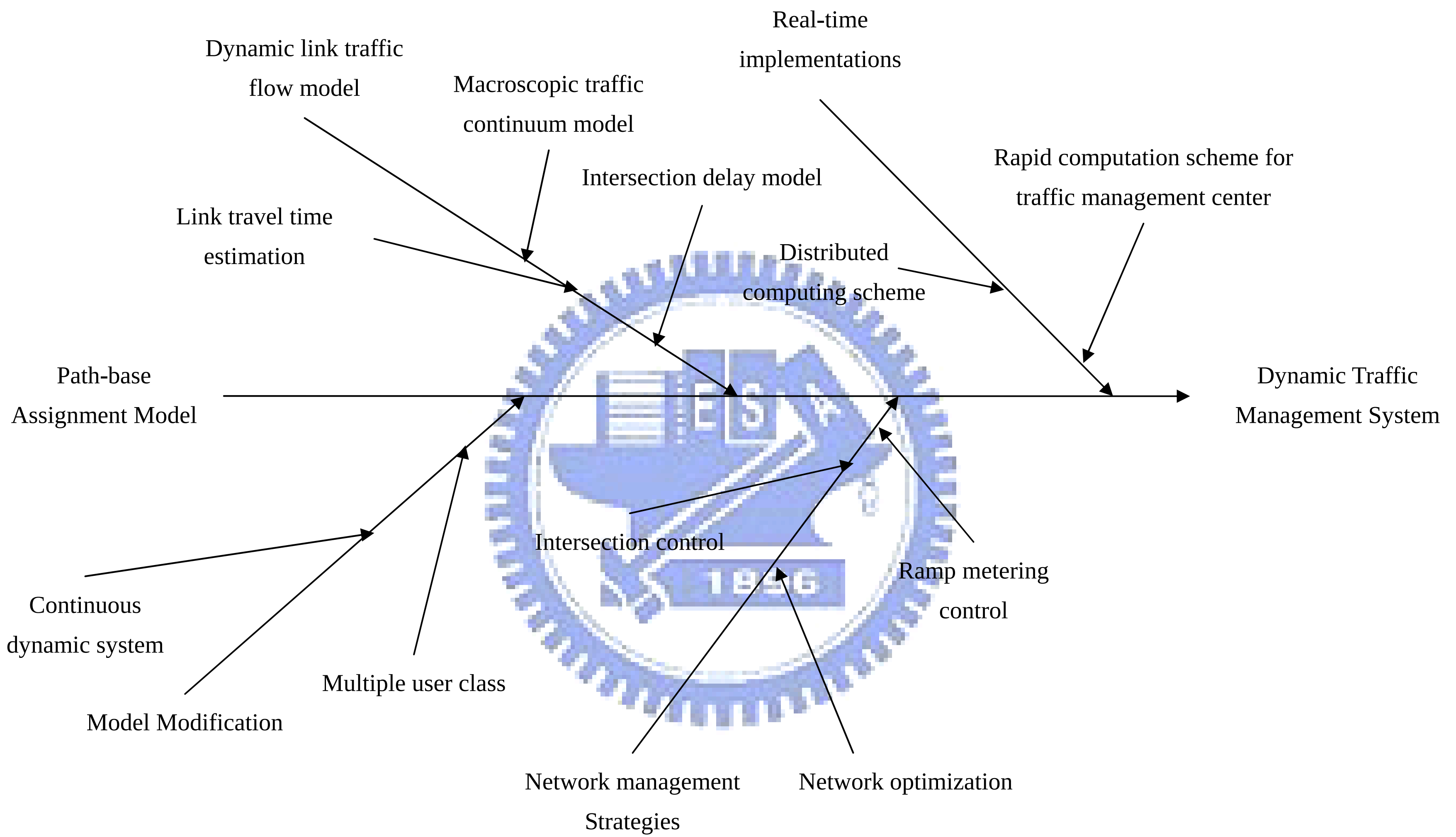


Figure 7.1 Roadmap of future researching topics.

References

Ashok K, Ben-Akiva ME (1993) Dynamic O-D matrix estimation and prediction for real-time traffic management systems. In: C. F. Daganzo (ed) Transportation and Traffic Theory, Elsevier, New York, pp 465-484

Ashok K, Ben-Akiva ME (2002) Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. Transportation Science 36:184-198

Bar-Gera H, Boyce DE (2003) Origin-based algorithms for combined travel forecasting models. Transportation Research 34(B):405-422

Bell MGH (1991) The estimation of origin-destination matrices by constrained generalized least squares. Transportation Research 25(B):13-22

Ben-Akive ME, Koutsopoulos HN, Mishalani R, Yang Q (1997) Simulation laboratory for evaluating dynamic traffic management systems. ASCE Journal of Transportation Engineering 123:283-289

Benameur N, Roberts JW (2004) Traffic matrix inference in IP Networks. Networks and Spatial Economics 4:103-114

Bernstein D, Friesz TL, Stough R (1996) Variational inequalities, dynamical systems and control theoretic models for predicting time-varying urban network flows. Transportation Science 30:14-31

Bernstein D, Friesz TL, Suo Z, Tobin RL (2001) Dynamic network user equilibrium with state-dependent time lags. Networks and Spatial Economics 1:319-347

Birge JR, Ho JK (1993) Optimal flows in stochastic dynamic networks with congestion. Operations Research 41:203-216

Boyce DE, Chon KS, Lee YJ, Lin KT, LeBlanc LJ (1983) Implementation and computational issues for combined models of location, destination, mode and route choice. Environment and Planning 15(A):1219-1230

Boyce DE, Zhang YF, Lupa MR (1994) Introducing "Feedback" into four-step travel forecasting procedure versus equilibrium solution of combined model. Transportation Research Record 1443:65-74

Boyce DE, Ran B, LeBlanc LJ (1995) Solving an instantaneous dynamic user-optimal route choice model. Transportation Science 29:128-142

Boyce DE (2002) Is the sequential travel forecasting paradigm counterproductive? Journal of Urban Planning & Development 128:169-183

Boyce DE, O'Neill CR, Scherr W (2007) New results on solving the sequential travel forecasting procedure with feedback. In: 17th PTV Vision User Group Meeting, Berlin, Germany

Carey M (1987) Optimal time varying flows on congested networks. Operations Research 35:58-69

Carey M (1992) Nonconvexity of the dynamic assignment problem. Transportation Research 26(B):127-133

Carter CK, Kohn R (1994) On Gibbs sampling for state space models. Biometrika 81:541-553

Cascetta E (1984) Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. Transportation Research 18(B):289-299

Cascetta E, Inaudi D, Marquis G (1993) Dynamic estimators of origin-destination matrices using traffic counts. Transportation Science 27:363-373

Cascetta E, Nguyes S (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. Transportation Research 22(B):437-455

Chang GL, Tao X (1995) An advanced computing architecture for large-scale network O-D estimation. In: Proceedings of 6th International Conference on Vehicle Navigation and Information Systems, pp 317-327

Chang GL, Wu J (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. Transportation Research 28(B):141-160

Chen HK, Hsueh CF (1998) A model and an algorithm for the dynamic user-optimal route choice problem. Transportation Research 32(B):219-234

Chen HK, Chang MS, Wang CY (2001) Dynamic capacitated user-optimal departure time/route choice problem with time-window. European Journal of Operational Research 132:603-618

Cho HJ, Lan CL, Jou YJ, Hwang MC, Lee TT (2006) A framework of real-time traffic information system. WSEAS Transactions on Mathematics 5:117-122

Cremer M, Keller H (1981) Dynamic identification of O-D flows from traffic counts at complex intersections. In: Proceedings of 8th International Symposium on Transportation and Traffic Theory, University of Toronto press

Dafermos S (1980) Traffic equilibrium and variational inequalities. Transportation Science 14:42-54

Daganzo C (1994) Cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Science 17:430-453

Dixon MP, Rilett LR (2005) Population origin-destination estimation using automatic vehicle identification and volume data. Journal of Transportation Engineering-ASCE 131:75-82

Doblas J, Benitez FG (2005) An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix. Transportation Research 39(B):565-591

Erlander S, Nguyen S, Stewart NF (1979) On the calibration of the combined distribution-assignment model. Transportation Research 13(B):259-267

Erlander S (1990) Efficient behavior and the simultaneous choices of origin, destinations and routes. Transportation Research 24(B):363-373

Evans SP (1973) A relationship between the gravity model for trip distribution and the transportation problem in linear programming. Transportation Research 7:39-61

Evans SP (1976) Derivation and analysis of some models for combining trip distribution and assignment. Transportation Research 10:37-57

Fisk CS, Boyce DE (1983) A note of trip matrix estimation from link traffic count data. Transportation Research 17(B):245-250

Florian M, Nguyen S, Ferland J (1975) On the combined distribution-assignment of traffic. Transportation Science 9:43-53

Friesz TL, Luque J, Tobin RL, Wie BW (1989) Dynamic network traffic assignment considered as a continuous time optimal control problem. Operations Research 37:893-901

Friesz TL, Tobin RL, Cho HJ, Mehta NJ (1990) Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. Mathematical Programming 48:265-284

Friesz TL, Cho HJ, Mehta NJ, Tobin RL, Anandalingam G (1992) A simulated annealing approach to the network design problem with variational inequality constraints. Transportation Science 26:18-26

Friesz TL, Bernstein D, Smith TE, Tobin RL, Wie BW (1993) A variational inequality formulation of the dynamic network user equilibrium problem. Operations Research 41:179-191

Friesz TL, Bernstein D, Mehta NJ, Tobin RL, Ganjalizadeh S (1994) Day-to-day dynamic network disequilibria and idealized traveler information systems. Operation Research 42:1120-1136

Friesz TL, Shah S (2001) An overview of nontraditional formulations of static and dynamic equilibrium network design. Transportation Research 35(B):5-21

Frühwirth-Schnatter S (1992) Data augmentation and dynamic linear models. Journal of Timer Series Analysis 15:183-202

Gamerman D (1998) Markov chain monte carlo for dynamic generalized linear models. Biometrika 85:215-227

Geman S, Geman D (1984) Stochastic relaxation, gibbs distribution, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721-741

Ghali MO, Smith MJ (1995) A model for the dynamic system optimum traffic assignment problem. Transportation Research 29(B):155-170

Grewal MS, Andrews AP (1993) Kalman filtering : theory and practice. Prentice-Hall.

Helbing D (1995) Improved fluid-dynamic model for vehicular traffic. Physical Review 51(B):3164-3169

Hwang MC, Cho HJ (2006) Numerical simulation of link travel time with hyperbolic traffic behavior. In: Proceedings of the IEEE ITSC 2006 Conference, pp.400-405

Janson BN (1991) Dynamic traffic assignment for urban road networks. Transportation Research 25(B):143-161

Jou YJ, Cho HJ, Lin PW, Wang CY (2006) Incomplete information analysis for the origin-destination survey table. Journal of Transportation Engineering-ASCE 132:193-200

Jou YJ, Hwang MC (2002) Rapid transit system O/D pattern calculation with statistical Gibbs sampling and Kalman techniques. In: Taipei Rapid Transit Corp (ed) The proceedings of 2002 World Metro Symposium & Exhibition.

Jungbacker B, Koopman SJ (2007) Monte Carlo estimation for nonlinear non-Gaussian state space models. Biometrika 94:827-839

Lam WHK, Huang HJ (1992) A combined trip distribution and assignment model for multiple user classes. Transportation Research 17(B):449-462

Lee JYS, Lam WHK, Wong SC (2001) Pedestrian simulation model for Hong Kong underground stations. In: Proceedings of the IEEE International Conference on Intelligent Transportation Systems, pp 554-558

Leo CJ, Pretty RL (1990) Numerical simulation of macroscopic continuum traffic models. Transportation Research 26(B):207-220

Li Y, Ziliaskopoulos T, Boyce DE (2002) Combined model for time-dependent trip distribution and traffic assignment. Transportation Research Record 1783:98-109

Lighthill MH, Witham GB (1955) On kinematic waves II: a theory of traffic flow on long crowded roads. In: Proceeding of the Royal Society, Vol: A229, pp.317-345

Lo HP, Chan CP (2003) Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. Transportation Research 37(A):771-788

Lundgren JT (1989) Optimization Approaches to Travel Demand Modelling. Ph. D Dissertation, Linköping University

Lundgren JT (1989) A comparison of two algorithms for the combined trip distribution and assignment problem. In: Research Report Lith-MAT-R-88-O9.

Lundgren JT, Patriksson M (1996) An algorithm for the combined distribution and assignment model. In: Bell, M.G.H. (ed.), Transportation Networks: Recent Methodological Advances. Pergamon Press, Oxford.

Lundgren JT, Patriksson M (1998) The combined distribution and stochastic assignment problem. Annals of Operations Research 92:309-329

Maher MJ (1983) Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. Transportation Research 17(B):435-447

Mahmassani HS, Peeta S (1993) Network performance under system optimal and user equilibrium dynamic assignments: Implications for ATIS. Transportation Research Record 1408:83-93

Merchant DK, Nemhauser GL (1978a) A model and an algorithm for the dynamic traffic assignment problems. Transportation Science 12:183-199

Merchang DK, Nemhauser GL (1978b) Optimality conditions for a dynamic traffic assignment model. Transportation Science 12:200-207

Nie YM, Zhang HM (2008) A relaxation approach for estimating origin-destination trip tables. Network and Spatial Economics Doi:10.1007/s11067-007-9059-y

Okutani I (1987) The kalman filter approach in some transportation and traffic problems. In: N. H. Gartner and N. H. M. Wilson (ed) Transportation and Traffic Theory, Elsevier, New York, pp 397-416

Oppenheim N (1995) Urban Travel Demand Modeling. Wiley, Johm & Sons.

Payne HJ (1979) A macroscopic simulation model of freeway traffic. Transportation Research Record 772:68-75

Pitt MK, Shephard N (1999) Analytic convergence rates and parameterization issues for the gibbs sampler applied to state space models. Journal of Time Series Analysis 20:63-85

Peeta S, Yang TH (2003) Stability issues for dynamic traffic assignment. Automatica 39:21-34

Ran B, Boyce DE (1996) A link-based variational inequality formulation of ideal dynamic user-optimal route choice problem. Transportation Research 4(C):1-12

Ran B, Boyce DE, LeBlanc LJ (1993) A new class of instantaneous dynamic user-optimal traffic assignment models. Operations Research 41:192-202

Richards PJ (1956) Shock waves on the highway. Operations Research 4:42-51

Robert CP, Cellier D (1998) Convergence control of MCMC algorithms. In: Robert CP (Ed.) Discretization and MCMC Convergence Assessment. Lecture Notes in Statistics 135, Springer.

Tomlin JA (1971) A mathematical programming model for the combined distribution-assignment of traffic. Transportation Science 5:122-130

Wie BW, Tobin RL, Friesz TL, Bernstein D (1995) A discrete time, nested cost operator approach to the dynamic network user equilibrium problem. Transportation Science 29:19-92

Xu M, Chen A, Gao Z (2008) An improved origin-based algorithm for solving the combined distribution and assignment problem. European Journal of Operational Research 188:354-369

Yang H, Meng Q (1996) A mixed, stochastic trip distribution and assignment model with variable origin and destination costs. In: The first conference of Hong Kong Society for transportation studies, pp. 125-140

Yang Q, Koutsopoulos HN (1996) A microscopic traffic simulator for evaluation of dynamic traffic management systems. Transportation Research 4(C):113-129

Ziliaskopoulos AK (2000) A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transportation Science 34:37-49

Zuylen HJ, Willumsen LG (1980) The most likely trip matrix estimated from traffic counts. Transportation Research 14(B):281-293