國立交通大學

交通運輸研究所

博士論文

應用格位傳送模式建構預測型高速公路動態起迄矩陣推估演算法

Predictive Estimation of Dynamic Freeway Origin-Destination Matrices with Cell Transmission Modeling

研 究 生： 曾群明

指導教授： 藍武王 博士

邱裕鈞 博士

中 華 民 國 一〇一 年 六 月

應用格位傳送模式建構預測型高速公路動態起迄矩陣推估演算法

Predictive Estimation of Dynamic Freeway Origin-Destination Matrices with Cell Transmission Modeling

研　究　生：曾群明	Student: Chun-Ming Tseng
指導教授：藍武王　博士	Advisors: Dr. Lawrence W. Lan
邱裕鈞　博士	Dr. Yu-Chiun Chiou

國立交通大學

交通運輸研究所

博士論文

A Dissertation

Submitted to Institute of Traffic and Transportation

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in

Management

June 2012

Taipei, Taiwan, Republic of China

中　華　民　國　一〇一　年　六　月

應用格位傳送模式建構預測型高速公路
動態起迄矩陣推估演算法

研究生：曾群明　　　　　　　　　　　指導教授：　藍武王　博士
邱裕鈞　博士

國立交通大學交通運輸研究所

摘要

正確和有效率的動態起迄矩陣推估對運輸規劃、交通管理及交通策略研擬是很重要的資訊，近年來隨著智慧型運輸系統之快速發展，動態起迄矩陣之推估及即時性交通控制之執行如路徑導引及號誌控制等績效均有顯著提昇。過去許多文獻利用所偵測到的高速公路主線與上、下匝道流量，進行高速公路動態起訖矩陣之推估，或以增加額外的假設或是外生的資訊，如路徑選擇行為或已知的依時性交通流量資料等提昇績效，然而，在此課題中，所面臨的挑戰是欲推估之參數數量遠大於資料所提供之訊息及如何處理旅行時間變異影響之問題。

基此，本研究提出一個預測型高速公路動態起迄矩陣推估方法，包括一個二階段交通量預測模式及一個整合型動態起迄矩陣推估演算法，交通量預測模式係以滾動式自我建構交通模型來預測中、長期交通特性，其中，包含以成長型階層式自我組織放射圖(GHSOM)將所有交通量樣本分成若干群，並針對每一群以基因規劃法(GP)建構相對應之非線性交通量預測模式去預測交通特性。而整合型動態起迄矩陣推估演算法係為了能在不同的交通情況下，有效率且準確的獲取車輛到達型態，藉由結合格位傳送模式(Cell Transmission Model，CTM)及卡門濾波(Extended Kalman Filtering，EKF)，來建構遞迴式動態起迄矩陣推估演算法，由CTM 模擬車輛運行行為，預測各依時起迄對之到達率，再以 EKF 推估動態起迄矩陣。

為驗證此演算法之績效及實用性，本文以所設計之 6 個起迄對之小型路網推估 90 分鐘的起迄矩陣作為驗證範例。同時為比較本模式於旅行時間預測之績效，以 Greenshields 巨觀模式假設在進入路網之車輛會於兩時階範圍內到達迄點之條件下，預測車輛旅行時間。結果顯示本模式推估結果之 RMSE 為 0.069 遠較 Greenshields 之 0.145 為優。在實例應用上，本文先以國道 1 號泰山收費站至楊梅交流道計 6 個交流道 36 公里長之 3 車道高速公路中型路網進行實驗驗證，結果顯示格位傳送模式在自由流至擁擠流等不同情境交通狀況下，有效模擬車輛到

達率，而本模式並以相當低的 RMSE 績效，精確推估動態起迄矩陣。

最後，本研究再以國道 1 號頭份交流道至北斗交流道計 15 個交流道長 110 公里之 3 車道高速公路大型路網作為實例驗證，結果也證明本文所提二階段交通量預測模式的績效及整合型動態起迄矩陣推估演算法的實用性。另外，本文亦比較所提二階段交通量預測模式之績效優於傳統 ARIMA 模式。並進一步針對交通量樣本長度進行敏感度分析，發現 5 分鐘為一時階之交通量樣本長度在 120 個時階（10 小時），即能達到相當之預測績效。

關鍵字：交通量預測、基因規劃法、成長型階層式自我組織放射圖、滾動式自我建構交通模型、動態起迄矩陣推估、進階卡門濾波、格位傳送模式

Predictive Estimation of Dynamic Freeway Origin-Destination Matrices with Cell Transmission Modeling

Student: Chun-Ming Tseng Advisors: Dr. Lawrence W. Lan
Dr. Yu-Chiun Chiou

Institute of Traffic and Transportation
National Chiao Tung University

ABSTRACT

Accurate and effective dynamic origin-destination (O-D) matrices estimation is important for transport planning, traffic management and strategic planning. Recently, the rapid development of intelligent transport systems has enhanced accurate dynamic O-D information and the implementation of real-time traffic control, such as real-time route guidance and signal control. Numerous studies have devoted to developing estimation algorithms based on observable mainline and ramp flow rates, with constraints dependent on time series traffic flow and extra system equations and using recursive or non-recursive system solution techniques to estimate O-D matrices. However, this dynamic O-D matrices estimation issue remains challenging in that the number of parameters to be estimated is always far greater than the available information, and the impact of travel time variability on the time-varying O-D matrices.

In light of this, the study proposes a novel approach to estimate medium-to long-term freeway dynamic O-D matrices. The proposed approach includes a two-stage prediction model with an integrated algorithm. The traffic prediction model predicts medium- to long-term traffic features based on rolling self-structured traffic patterns. The rationales include using the growing hierarchical self-organizing map model (GHSOM) to partition unlabeled traffic patterns into clusters and then developing an associated genetic programming (GP) model to predict the traffic features in each cluster. And then, the integration algorithm, which combined cell transmission model (CTM) with extended Kalman filtering (EKF) to respectively and iteratively estimate the arrival distributions and O-D proportions.

To demonstrate the performance and applicability of the proposed approach, a seminal example with six O-D pairs of 90 minutes estimation is designed. The

performance of this mode in terms of travel time prediction is compared to the Greenshields macroscopic model prediction. The results showed that the propose approach is better than the Greenshields model. In the field study, a medium-scale networks and a large-scale network of on-ramp traffic patterns on a freeway are examined. The medium-scale network covers a section of Taiwan No.1 Freeway (Taishan toll station to Yangmei toll station), a 36 km three-lane freeway section with 6 interchanges, and the results showed that the CTM can accurately capture the degree of traffic dispersion under traffic scenarios ranging from free-flow to congested-flow conditions and that the proposed EKF algorithm can accurately estimate the O-D proportions with rather low RMSE.

For the large-scale network, 15 interchanges from Toufen interchange to Beidou interchange, a 110-kilometer stretch of Taiwan No.1 Freeway, were tested and the results indicated the practical applicability of the proposed algorithm. In addition, the proposed method has performed much better than the conventional ARIMA model. The sensitive analysis has also revealed the necessity of acquiring five-minute traffic patterns longer than 120 time intervals (10 hours) in order to achieve sufficient high prediction accuracy.

KeyWords: *Traffic prediction, Genetic programming, Growing hierarchical self-organizing map, Rolling self-structured traffic patterns, Dynamic origin-destination, Cell transmission model, Extended Kalman filtering*

誌謝

人家讀書像坐高鐵，咻一下就到達目的地，我的求學生涯像坐慢車，站站停且停很久，速度慢挫折也多，然也因此有停車下站的機會，豁然發現站站均有不一樣的景緻，欣賞美景及思考未來是人生另一番收穫，也認識很多我生命中的貴人，得到支持、鼓勵及指導，讓我終也能達到目標且行囊滿滿。回想進修期間，修課、資格考、提計畫書、投稿及撰寫論文，可謂關關難過關關過，除了自己的埋首苦讀外，是你們的適時支援，才能成就今日的我，我想大聲的說，謝謝你們。

首先，要感謝交通部各級長官及辦公室的同仁與好友，同意我在職進修攻讀博士學位，並給予個人在公務上的協助；求學期間，最要感謝恩師藍武王教授及邱裕鈞教授悉心指導，除提升我專業能力外，更強化個人思考邏輯及研究精神。藍老師嚴謹的治學精神及邱老師的諄諄教誨，讓我在治學態度及待人處世上獲益良多，未來在職場上必將受用無窮。博士班授課期間，蒙黃承傳教授、黃台生教授、馮正民教授、汪進財教授及陳穆臻所長在專業知識的指導；論文研討、審查及口試期間，承蒙臺灣大學工商管理學系許鉅秉教授、成功大學交通管理科學系胡大瀛教授及淡江大學運輸管理學系董啟崇教授悉心斧正並提供寶貴意見，使本論文更臻完整，在此致上誠摯的謝意。

感謝銘德、益三、孟佑、易詩、沛儒、國洲、世昌、昭榮、素如、昭宏、魏瑜等學長姐、啟椿、士軒、聖章、嘉陽、牧民、裕文及季森的支持及鼓勵；同學承憲、永祥、文斌、立欽陪我一起努力；同門的志誠、日新、彥蘅、姿慧、彥斐、傳強及志偉，適時給予火力支援；好朋友茂陽、青嶽、美娟、炳榮、文成、天岑的鼎力相助，在此一併致謝。

另外，要感謝母校警察大學王文麟院長及交通系全體老師、開南大學陳武正教務長、公務人員保障暨培訓委員會邱華君委員、警察專科學校陳連禎校長、臺北市政府交通局林志盈局長、警官二隊陳建聰隊長等長官一貫的愛護及指導，特表謝意。

最後要感謝支持我的家人，特別是愛妻靜宜、寶貝暄晴及揚哲、岳父母，謝謝你們的支持、安慰及包容，是我堅持向前行的最大動力。

謹將這本論文獻給在天上的父母，並與所有愛我及我所愛的親朋好友分享這份喜悅。

曾群明

2012 年 6 月交通大學台北校區

CONTENTS

LIST OF TABLE

LIST OF FIGURE

CHAPTER 1 INTRODUCTION

1.1 Research Background

Accurate dynamic origin-destination (O-D) information is required for the implementation of real-time traffic control measures, such as real-time route guidance and signal control. Numerous studies have devoted to developing estimation algorithms for the dynamic O-D matrix based mainly on observable mainline and ramp flow rates. The dynamic O-D matrices estimation algorithms can be divided into two categories (Ho, 2008): assignment-based (e.g. Ashok and Ben-Akiva, 2000, 2002) and non-assignment-based (e.g. Chang and Wu, 1994; Chang and Tao, 1996, 1999; Lin and Chang, 2005, 2007). The assignment-based method primarily relies on a dynamic traffic assignment algorithm to generate link flows; while the non-assignment-based method directly estimates O-D matrices. However, this issue remains challenging in that the number of parameters to be estimated is always far greater than the available information, thus additional assumption or exogenous information, such as route choice behaviors, priori O-D matrix information, sequence of observational periods of traffic counts data (e.g. Bell, 1983, 1991; Yang et al., 1992, 1995; Vardi, 1996; Lo, et al., 1996; Hazelton, 2001), should be further considered.

One of the most challenging issues remained to be tackled in the context of dynamic O-D matrices estimation is the impact of travel time variability on the time-varying O-D matrices. Chang and Wu (1994) assumed that the vehicles entering the freeway in a time interval are distributed in a small range (within two time intervals). However, if O-D pair traffic traverses a sufficiently long distance or experiences moderate to heavy congestions, then the travel time variability may be rather large, which can result in a serious traffic dispersion phenomenon. As a result, the O-D pair traffic entering the freeway in a specific narrow time interval will reach their destinations over a wide time interval, which will greatly increase the difficulty in accurately estimating the dynamic O-D matrices. In other words, an accurate prediction model for the arrival distribution of entering O-D pair traffic under various traffic conditions is undoubtedly imperial for dynamic O-D matrices estimation.

Based on this, Chang and Tao (1995) assumed a macroscopic traffic model to efficiently predict the travel time according to concurrent traffic conditions and used the predicted travel time to estimate traffic arrival distributions and then to estimate the O-D matrices. Lin and Chang (2005, 2007) further assumed the travel time of drivers following a certain distribution and then used such a distribution to estimate their arrival patterns. However, the studies of Chang and Wu (1994), Chang and Tao

(1995) and Lin and Chang (2005, 2007) all made strong assumptions regarding the prediction of traffic dispersion, which might not be valid for various conditions from free-flow to gridlock. In addition, the state equations in the abovementioned studies may involve relatively too many parameters, largely increasing the model complexity to be implemented.

The dynamic O-D matrices estimation algorithms can be divided into two categories: assignment-based and non-assignment-based. Differ to the assignment-based method primarily relies on a dynamic traffic assignment algorithm to generate link flows; the non-assignment-based method directly estimates O-D matrices. Taking into account the difficulty of non-assignment method to obtain a priori dynamic O-D information, some studies have developed different methods of O-D estimation, using available time series traffic flow to reduce the dependence on the priori time-dependent O-D and dynamic traffic assignment models. However, dynamic O-D estimation model is still limited to small-scale networks. The main reason is the relationship of time-dependent O-D and the flow of sections are difficult to establish, and travel time is unrealistically assumed as constant in the model.

To remedy these gaps in dynamic O-D matrices estimation, it is deemed necessary and important to develop an integrated simulation-based dynamic O-D estimation model by using an efficient traffic prediction and simulation models to replicate traffic behaviors as well as a dynamic O-D estimation model to determine the O-D pair shares of entering traffic (i.e. on-ramp traffic).

To do so, an efficient and accurate traffic simulation model should be incorporated into the commonly adopted O-D matrices estimation method, extended Kalman filtering (EFK) algorithm. The cell transmission model (CTM), proposed by Daganzo(1994, 1995) was developed as a discrete approximation to the hydrodynamic theory of traffic flow. It is capable of automatically tracking shocks and acceleration waves and thus capturing traffic behavior in the process of the formation, propagation, and dissipation of queues. In that, the LWR continuum model is discretized into cells. The road is represented by a number of small sections (cells). The simulation model keeps tracking the number of vehicles in each cell, and in each time-step it calculates the number of vehicles that cross the boundaries between adjacent cells. Therefore, how to incorporate CTM models into the dynamic O-D estimation model is imperial.

Additionally, to successfully estimate the dynamic O-D matrices, a sufficiently long period of entering traffic has to be accurately predicted so as to be the input of the traffic simulation model to simulate the arrival pattern of the entering traffic after necessary transverse time between the origin and destination interchanges. Due to the rapid fluctuations of traffic flow and obvious peak and off-peak traffic, a efficient and effective medium-to-long term traffic prediction model is also required.

1.2 Research Objectives

In view of the importance of arrival distribution prediction in estimating the O-D matrices and to efficiently and accurately capture the traffic behaviors along with their arrival distributions under various traffic conditions, this study uses CTM to simulate the traffic movement behaviors, to predict the arrival distributions of all O-D pair traffic in various time intervals, and then to estimate the dynamic O-D matrices. Moreover, the conceptual representation of spatial (cell) and temporal (discrete time interval) conditions of traffic makes CTM especially suitable for dynamic O-D matrices estimation. Our proposed model intends not only to result in a substantial increase of system observability with significantly less parameters than those in literature, but also to contribute enhancing the quality of dynamic O-D matrices estimation.

However, it would be unrealistic to predict the arrival distribution using CTM by assuming that on-ramp traffic flow along a freeway remains unchanged over time. To rectify this unrealistic assumption, a medium-to-long term (e.g. next two to four hours) prediction model of on-ramp traffic along a freeway is required.

Most of the existing traffic prediction models use statistical methods or artificial intelligent methods to conduct a short-term prediction (e.g. next 5 minutes). Such short-term prediction models may experience low performance for medium-to-long term traffic prediction since traffic patterns can change dramatically (e.g., from peak hours to off-peak and vice versa). According to field observation, daily traffic patterns do repeat spatially and temporally over and over again. To enhance the prediction performance should the historical traffic data be clustered into appropriate different traffic patterns, it would become possible to accurately predict the traffic features in a rolling manner for a medium-to-long term traffic.

This study proposes a novel approach, based on rolling self-structured traffic patterns, to make a medium-to-long prediction for the traffic features along a freeway corridor, in the prediction process, traffic of historical traffic data are collected and used Growing hierarchical self-organizing map (GHSOM) to identify the similar cluster of traffic patterns into a cluster without the need to pre-determine the number of clusters. The input values of learned Genetic programming (GP) model belonging to that cluster are used to perform the traffic prediction in each cluster without the need of prior knowledge regarding data distribution or model specification. With the predicted medium-to-long on-ramp traffic, CTM is used to simulate the arrival patterns, and then, EKF is employed to estimate the O-D proportions. To validate the accuracy and applicability of the proposed approach, an empirical study on a freeway is examined.

Based on abovementioned motivations, and think about the unique properties of freeway network, for example: traffic conservation features on and off ramp, estimated time dependent flow, and no complicated path selection problem, that enable to provide information of value to estimate the dynamic the O-D, the study aims to propose a novel approach to estimate dynamic O-D matrices, based on Chang and Wu's (1994) model, taking into consideration the mainline traffic information and travel time delays to further amend travel time estimation methods. The approach includes a two-stage prediction model with the GHSOM algorithm to extract clusters of traffic patterns and GP to predict the traffic for each cluster separately, and then proposes an integrated algorithm which combines the CTM with the EKF to respectively and iteratively estimate the arrival distributions and O-D proportions. The objectives of this study can be stated below:

1. Propose a two-stage prediction model with GHSOM algorithm and GP to predict a medium-to long-term (e.g. next two to four hours) traffic flow of on-ramp traffic along a freeway, supporting the CTM to estimate arrival distributions.

2. Based on the predicted entering traffic over a sufficient long period, the CTM model is used to estimate the arrival patterns of entering traffic at every time click.

3. Based on the predicted entering traffic, arrival patterns, the EKF algorithm is used to estimate dynamic O-D ratios.

4. Demonstrate the performance and applicability of the proposed integrated dynamic O-D matrices estimation model, a seminal example designed, a medium-scale networks and a large-scale network are studied, respectively.

1.3 Chapter Organization

The research framework is depicted in Figure 1. This dissertation is organized as follows. Chapter 1 introduces the background and research objectives of the research. Chapter 2 briefly summarizes related studies. Chapter 3 proposes the research framework of this study, including the introductions to the study freeway corridor, and variables/parameters used in the proposed models. Chapter 4 introduces the proposed models, including a two-stage traffic prediction model, integrated arrival distribution model and EKF-based dynamic O-D matrices estimation. Chapter 5 presents case studies to demonstrate the performance and applicability of the proposed model. Chapter 6 gives the conclusions of this study and suggestions for future studies.

As shown in Figure 1, this study contains the following parts:

1. Objective definition

Clearly identify the research topics and objectives of this study.

2. Literature review

To review the following research topics: (1) Static and dynamic O-D matrices estimation models. (2) Traffic arrival distribution predict model. (3) Traffic pattern clustering model. (4) Traffic pattern prediction model.

3. Problem statement

Definitions of the typical freeway corridor, variables and parameters, and introducing the process of estimation of O-D estimation matrices.

4. Model formulation

The proposed model in this study can be divided into two parts. One is a two-stage prediction model based on GHSOM algorithm and GP, the other is combined CTM with EKF to respectively and iteratively estimate the arrival distributions and O-D proportions.

5. Seminal example

A seminal example is designed to demonstrate the performance and applicability of the proposed estimation algorithm.

6. Case study

Case studies on to two different sized freeway networks, one is a medium, another is a large-scale, are conducted to demonstrate the performance and applicability of the proposed estimation algorithm.

7. Concluding remarks

The major findings and contributions of this study are narrated. The suggestions for future studies are also addressed.

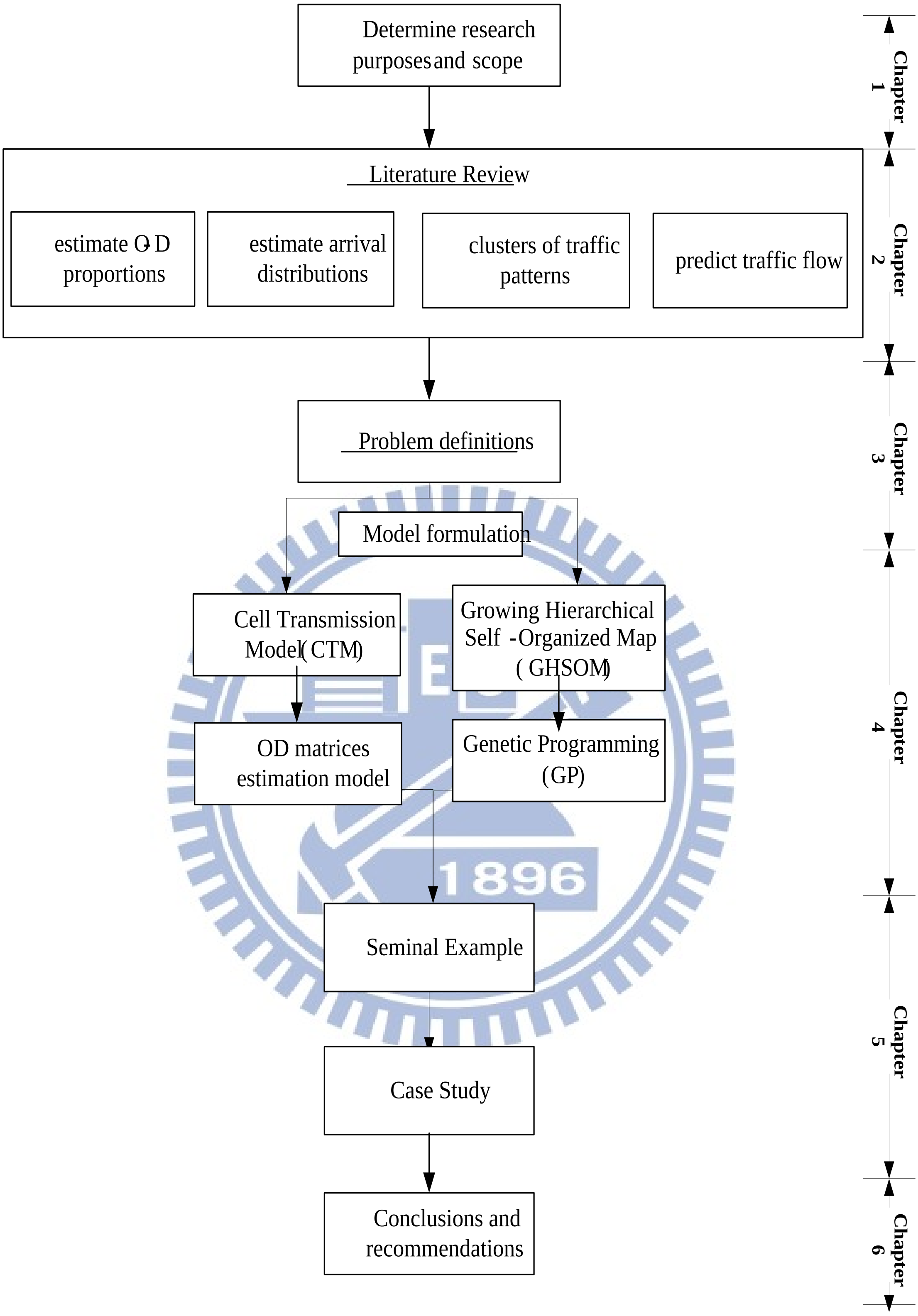


Figure 1 Research framework

CHAPTER 2 LITERATURE REVIEW

The chapter consists of four sections. Section 2.1 addresses the O-D matrices estimation. Section 2.2 discusses the traffic pattern clustering theory. Section 2.3 addresses the traffic prediction and section 2.4 discusses the traffic simulation models.

2.1 O-D Matrices Estimation

This section reviews various approaches for estimating the static and dynamic O-D matrix.

2.1.1 Static O-D Matrices Estimation

There are several methods for formulating estimators for an unique matrix.

2.1.1.1 Maximum entropy model

In transportation and regional planning, the most popular approach for estimating the static O-D matrix is the maximum entropy model. Van Zuylen and Willumsen (1980), the assumption of the approach is that all of the combinations of individual travel decisions, so called states, are equally likely to occur. The set of O-D flows with the highest likelihood of occurring is therefore the set with the maximum number of states.

2.1.1.2 Maximum likelihood and generalized least-squares

Another method is classical statistical inference techniques. The two main estimators are the maximum likelihood and the generalized least-squares. The maximum likelihood estimator maximizes the likelihood of observing the experimental data condition on the true trip matrix. For this method, distributional assumptions need to be made for the sample and traffic counts, Zhang and Maher （1998） addresses the problem of estimating an O-D matrix with platoon dispersion from fully disaggregate data that the passage times of vehicles at the entries and exits or the origins and destinations of a network, and shown that the maximum likelihood estimation can be formulated more generally as a transportation problem.

On the other hand, no distributional assumptions need to be made for the generalized least squares approach. Wang （2005）developed a GLS method to estimate the O-D matrix that extracted from the partial information of the sub O-D matrix and the means of population information of traffic counts of roads implemented by the sensor of traffic flow and both the location and movement of vehicle recorded by ATIS, and the implementation is demonstrated by a numerical example.

2.1.1.3 Others static O-D matrices estimation

Others method for estimate O-D matrix is including bayesian inference, gradient based solution techniques, simultaneous and column generation algorithm.

Bayesian inference uses a priori probabilities on the trip demands, by combining these probabilities with the conditional probability on the traffic counts; one can obtain the posterior probability of the demand conditioned on the traffic count. The arguments of this probability can then be maximized by different methods. Codina (2004) presents an algorithmic based on a method for nondifferentiable optimization due to Wolfe that can be interpreted as a conjugate directions method with better convergence properties as shown with a set of computational tests, alternative to the O-D matrix adjustment problem from observed link volumes when it is formulated as a mathematical programming problem with a bi-level structure, that the upper level function is approximating gradients.

Lo and Chan（2003）proposed a procedure for the simultaneous estimation of an O-D matrix and link choice proportions in a network change with traffic conditions from traffic counts for congested network, this procedure performs statistical estimation and traffic assignment alternately until convergence to obtain the best estimators for both the O-D matrix and link choice proportions, which are consistent with the traffic counts.

Ricardo（2008） presented a column generation algorithm for the demand adjustment problem iteratively solves a deterministic user equilibrium model for a given O–D matrix and a restricted DAP is formulated via a single level optimization problem. The convergence on local minimum of the proposed algorithm requires only the continuity of the link travel cost functions and the gauges used in the definition of the DAP. To analyze the convergence and performance of the proposed algorithm, various numerical tests were carried out on small scale problems.

2.1.2 Dynamic OD Matrices Estimation

Recently, there has been increased research in the area of dynamic O-D matrix estimation. Cremer and Keller (1987) highlighted the causal relationships that exist between the time variable sequences of entrance flow volumes and the sequences of short-time exit flow counts. They claim that enough information can be obtained from the counts at the entrances and the exits to obtain unique and bias-free estimates for the unknown O-D flows without further a priori information.

2.1.2.1 Bi-level

Lundgren and Peterson (2008) use a general nonlinear bi-level minimization formulation of the problem, where the lower level problem is to assign a given O-D matrix onto the network according to the user equilibrium principle. After reformulating the problem to a single level problem, the objective function includes implicitly given link flow variables, corresponding to the given O-D matrix. They propose a descent heuristic to solve the problem, which is an adaptation of the well-known projected gradient method. Numerical experiments are presented which indicate that the solution approach can be applied in practice to medium to large size networks.

2.1.2.2 Entropy

Wu（1997）develops an improved O-D algorithm based on the existing multiplicative algebraic reconstruction technique with the entropy- maximizing approach. The improvement of the algorithm in numerical stability and convergence speed s obtained by incorporating a normalization technique and a diagonal searching strategy, and demonstrated that the proposed new algorithm hold much promise for efficient application in advanced traffic management systems.

2.1.2.3 Others approach of dynamic O-D matrix estimation

Others approach of dynamic O-D matrix estimation including recursive estimation, column generation approach and Statistical inference. Li and Moor（1999） proposed a fast constrained recursive identification (CRI) algorithm to estimate intersection O-D matrices dynamically, the basic idea of the CRI algorithm is to estimate intersection O-D matrices based on equality-constrained optimization. Numerical results show that the accuracy of estimates by the CRI algorithm is fairly good the solutions obtained by the CRI are optimal in majority of the cases, compared with the ordinary recursive least squares method, the CRI algorithm with its reasonable balance between accuracy and computational simplicity is very suitable for practical use.

Sherali and Park（2001） develop a column generation approach that uses a sequence of dynamic shortest path sub-problems in order to estimate time-dependent path flows, or O-D trip tables, using available data on link traffic volumes for a general road network. Hazelton（2008）consider the problem of estimating a sequence of O-D matrices from link count data collected on a daily basis, they recommend a parsimonious parameterization for the time varying matrices so as to permit application of standard statistical estimation theory. A number of examples of suitably parameterized matrices are provided.

2.1.3 Related EKF Approaches

The Kalman filtering (KF) algorithms are known as a "prediction-correction" technique, which is based on the criterion of least square unbiased estimation of the state and measurement vectors.

In the pioneering study by Cremer and Keller, they applied the idea of dynamic OD estimation, where time dependant traffic counts were used in a recursive KF based OD estimator. And then, Chang and Wu (1994) applied the KF estimate for time-varying freeway O-D matrices, the proposed model employs information from mainline traffic counts, ramp flow measurements, and macroscopic traffic characteristics to construct a set of dynamic equations, which realistically consider the interrelations between O-D distributions and observed flows under congested conditions. To improve the operational efficiency necessary for real-time applications, a revised model with some approximation have also been developed, Wu and Chang (1996) suggested the improved KF based O-D estimator with time series of link and screenline flows, with properly selected screenlines and efficient computing algorithms, the proposed model also offers the potential for real-time applications.

Chang and Miaoul (1999) present an integrated method based on KF for estimating time-varying O-Ds in urban networks. The proposed method starts with their's previously developed two-stage, non-assignment-based model that can yield a time-varying O-Ds without a reliable prior O-D set and a dynamic traffic assignment model (DTA). To further improve the estimation accuracy and also account for the impact of urban signals, they have developed an intersection O-D estimation model that can produce an additional set of system observation constraints based on either existing or estimated intersection turning fractions, and the results of simulation experiments have clearly indicated that proposed method for integrated estimation of time-varying network O-D distributions is quite promising.

Lin and Chang (2006) present two robust algorithms, one for estimation of an initial O-D set and the other for tackling the input measurement errors with an extended estimation algorithm. The core concept of the initial O-D estimation algorithm is to decompose the target network in a number of sub-networks based on proposed rules, and then execute the estimation of the initial O-D set iteratively with the observable information at the first time interval. To contend with the inevitable detector measurement error, this study proposes an interval-based estimation algorithm that converts each model input data as an interval with its boundaries being set based on some prior knowledge. The performance of both proposed algorithms has been tested with a simulated system, and the results are quite promising.

Lin and Chang (2007) presents a KF based approach for estimating the dynamic freeway O-D distribution, based on measurable time series of mainline and ramp flows, and estimated travel time distributions. The proposed model captures the speed discrepancy among drivers, due either to their desired speeds or responses to congestion, with an embedded travel time distribution function and the identified interrelations between time-varying ramp and mainline flows. With the employed mainline information and travel time function, the proposed system equation has increased its observability with less parameter. Extensive numerical analyses with respect to the sensitivity of both input measurement errors and the selection of initial parameters on the estimation results have revealed that the proposed model is sufficiently robust for real-world applications.

Zhou and Mahmassani (2007) present a structural state space model to systematically incorporate regular demand pattern information, structural deviations and random fluctuations. By considering demand deviations from the a priori estimate of the regular pattern as a time-varying process with smooth trend, a polynomial trend filter is developed to capture possible structural deviations in real-time demand. Based on a KF framework, an optimal adaptive procedure is further proposed to capture day-to-day demand evolution, and update the a priori regular demand pattern estimate using new real-time estimates and observations obtained every day. These models can be naturally integrated into a real-time dynamic traffic assignment system and provide an effective and efficient approach to utilize the real-time traffic data continuously in operational settings.

2.2 Traffic Pattern Clustering

Numerous heuristic algorithms for clustering have been developed, which can generally be divided into three categories: statistics clustering, metaheuristics and neural network.

2.2.1 Statistical Methods

2.2.1.1 K-means algorithm

K-means algorithm is one of the most popular clustering algorithms for discovering clusters in data. Kantabutra et al (2000) offer a parallel solution to the K-means problem by taking advantage of a cluster of inexpensive workstations and a relatively low price of hard disk. And experiments show that this parallel algorithm achieves a much faster speed than the existing algorithms.

Frahling (2005) develop an efficient implementation for a k-means clustering algorithm, the novel feature of their algorithm is that it uses coresets to speed up the algorithm. A coreset is a small weighted set of points that approximates the original point set with respect to the considered problem. The main strength of the algorithm is that it can quickly determine clustering of the same point set for many values of k. This is necessary in many applications, since, typically, one does not know a good value for k in advance. They study have clustering for many different values of k and can determine a good choice of k using a quality measure of clustering that is independent of k, for example the average silhouette coefficient. The average silhouette coefficient can be approximated using coresets.

2.2.1.2 Fuzzy c-means algorithm

Fuzzy c-means (FCM) algorithm is another popular clustering techniques and subject of active research in several real world applications, because it is efficient, straightforward, and easy to implement.

Pal et al (2005) propose a new model called possibilistic-fuzzy c-means (PFCM) model. PFCM produces memberships and possibilities simultaneously, along with the usual point prototypes or cluster centers for each cluster. PFCM is a hybridization of possibilistic c-means (PCM) and FCM that often avoids various problems of PCM, FCM and FPCM. PFCM solves the noise sensitivity defect of FCM, overcomes the coincident clusters problem of PCM and eliminates the row sum constraints of FPCM. And show that PFCM compares favorably to both of the previous models.

Yu et al (2007) implement the horizontal collaborative clustering with the partial supervision clustering approach where the clustering is carried by the guidance of some labeled patterns. In this approach, they selected the patterns and interested in to provide FCM with collaborative information and control the degree of the influence of the selected patterns on the clustering. This new method is called partially horizontal collaborative fuzzy c-means (PHC-FCM). After presenting two approaches to realizing the selection of the labeled patterns, named cutset based approach and entropy based approach, and experiments are carried and show the performance of the new method.

2.2.2 Metaheuristic Methods

Chiou and Lan (2001) employs genetic algorithms to solve clustering problems three models, SICM, STCM, CSPM, are developed according to different coding/decoding techniques. The effectiveness and efficiency of these models under varying problem sizes are analyzed in comparison to a conventional statistics clustering method. The results for small scale problems indicate that CSPM is the most effective but least efficient method, medium to large scale indicate that CSPM is still the most effective method. They have applied CSPM to solve an exemplified p-Median problem, and demonstrate that CSPM is usefully applicable.

Yang et al (2005) developed a Genetic k-means-algorithm-based classification of direct load-control curves, the aim of reducing the number of variables for classification and enhancing the classification effectiveness, autoregression moving-averaging (ARMA) modelling techniques are employed to extract the features of the DLC curves. Based on the features extracted, the genetic k-means algorithm is then adopted for classification owing to its ability to partition given global data optimally into a specified number of clusters. Through the proposed approaches, categories are derived of the DLC curves complying and noncomplying with the control pattern. The results obtained from the comparisons with the artificial-neural-network approach show that the clusters divided using the proposed approach exhibit very high classification rates for the practical data on Taiwan Power Company DLC programmes.

An ant-based self-organizing map (ABSOM) was proposed on 2008. The ABSOM embeds the exploitation and exploration rules of state transition into the conventional SOM algorithm to avoid falling into local minima. To examine the usefulness of the proposed method, the ABSOM is combined with K-means into a two stage clustering method, i.e. ABSOM+K-means. Applied four public data sets, the ABSOM has been proved that it performs better than Kohonen’s SOM and it also works very well in the two-stage cluster analysis when it is taken as a preprocessing technique.

2.2.3 Artificial Neural Network Methods

2.2.3.1 Self-organizing map

The Self-Organizing Map (SOM) s a very popular unsupervised neural network model for the analysis of high-dimensional input data as in data mining applications, on the one hand, it is very simple to write down and to simulate, its practical properties are clear and easy to observe.

Yang et al (2004) presents an efficient approach to clustering the DLC curves through a structure of SOM. Aiming at selecting significant features of DLC curves, methods of nonlinear principal component analysis (NLPCA) and periodic analysis are proposed for feature extraction. The dual multilayer neural networks (DMNN) model is employed in the proposed NLPCA method. In the periodic analysis method, the periodic characteristics of the DLC curves are investigated. Results obtained from the comparison of six different approaches show that the clusters obtained from the proposed approach exhibit lowest degrees of misclassification for the practical data on Taiwan Power Company (TPC) DLC programs.

The SOM has shown to be a stable neural network model for high-dimensional data analysis. However, at least two limitations have to be noted, which are caused, on the one hand, by the static architecture of this model, as well as, on the other hand, by the limited capabilities for the representation of hierarchical relations of the data.

2.2.3.2 GHSOM

The GHSOM, an unsupervised learning neural network, is a powerful data mining technique for the clustering and visualization of large and complex data sets, which is an improvement over the basic SOM, can adapt its architecture during its learning process and expose the hierarchical structure that exists in the original data.

Pampalk et al (2003) present a novel approach to reveal the inherent hierarchical structure of data using multiple SOMs together with heuristics which optimize the stability, they evaluate the approach using data from real-world data mining projects in the music domain.

Tangsripairoj and Samadzadeh (2006) demonstrate the potential of the GHSOM for the organization and visualization of a collection of reusable components stored in a software repository, and compared the results with the ones obtained by using the traditional SOM.

Yang et al (2010) developed the GHSOM to overcome the limitations of SOM, to obtain higher detection rate and improve the stability of intrusion detection, some improvements on GHSOM algorithm are made: (1) they introduce a new metric that includes both numerical and symbolic data as input patterns. (2) using Tension and Mapping Ratio (TMR) instead of parameter $\tau 1$, the growth of a map is automatically controlled. This improved GHSOM is implemented and applied to intrusion detection. Their experimental results show that the detection rate has been increased by employing the improved GHSOM compared to the original SOM and GHSOM.

2.2.3.3 Support vector machine

Asa et al (2001) present a novel clustering method using the approach of support vector machines, and present a simple algorithm for identifying these clusters, and then demonstrate the performance of our algorithm on several datasets.

Chen et al (2001) investigates the connection between fuzzy classifiers and kernel machines, establishes a link between fuzzy rules and kernels, and proposes a learning algorithm for fuzzy classifiers. The corresponding fuzzy classifier is named positive definite fuzzy classifier (PDFC). A PDFC can be built from the given training samples based on a support vector learning approach with the IF-part fuzzy rules given by the support vectors.

Yella et al (2007) presents a comparison of several pattern recognition techniques combined with various stationary feature extraction techniques for classification of impact acoustic emissions. Results from support vector machines in combination with linear predictive cepstral coefficients delivered good classification rates.

2.3 Traffic Prediction

This section reviews various approaches to predict traffic flow, including Statistical Methods, Metaheuristic Method and Artificial Neural Network Methods.

2.3.1 Statistical Methods

2.3.1.1 Autoregressive integrated moving average

Autoregressive integrated moving average (ARIMA) is one of the popular linear models in time series forecasting during the past three decades.

Dervoort et al (1996) introduced A hybrid method of short-term traffic forecasting, the KARIMA method. The technique uses a Kohonen self-organizing map as an initial classifier, each class has an individually tuned ARIMA model associated with it. The explicit separation of the tasks of classification and functional approximation greatly improves forecasting performance compared to either a single ARIMA model or a backpropagation neural network, and demonstrated by producing forecasts of traffic flow, at horizons of half an hour and an hour, for a French motorway.

Smith et al (2002) effort seeks to examine the theoretical foundation of nonparametric regression and to answer the question of whether nonparametric regression based on heuristically improved forecast generation methods approach the single interval traffic flow prediction performance of seasonal ARIMA models.

Zhang (2003) proposed a hybrid methodology that combines both ARIMA and ANN models to take advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. Experimental results with real data sets indicate that the combined model can be an effective way to improve forecasting accuracy achieved by either of the models used separately.

Tseng and Tzeng (2005) proposes a fuzzy seasonal ARIMA (FSARIMA) forecasting model, which combines the advantages of the seasonal time series ARIMA (SARIMA) model and the fuzzy regression model. It is used to forecast two seasonal time series data of the total production value of the Taiwan machinery industry and the soft drink time series.

Stathopoulos and Karlaftis (2003) present a flexible and explicitly multivariate time-series state space models using core urban area loop detector data. The results clearly suggest that different model specifications are appropriate for different time periods of the day. Further, it also appears that the use of multivariate state space models improves on the prediction accuracy over univariate time series ones.

2.3.1.2 Grey forecasting model

Chang et al (2005) constructed a rolling grey forecasting model (RGM) to predict Taiwan's annual semiconductor production. The univariate grey forecasting model (GM) makes forecast of a time series of data without considering possible correlation with any leading indicators. It was expected that the annual semiconductor production in Taiwan should be closely tied with U.S. demand.

Vlahogianni et al (2006) offers a set of tools and methods to assess on underlying statistical properties of short-term traffic volume data, and results indicate that the statistical characteristics of traffic volume can be identified from prevailing traffic conditions; for example, volume data exhibit frequent shifts from deterministic to stochastic structures as well as transitions between cyclic and strongly nonlinear behaviors. These findings could be valuable in the implementation of a variable prediction strategy according to the statistical characteristics of the prevailing traffic volume states.

2.3.2 Metaheuristic Method

Chen (2007) applies a novel neural network technique, support vector regression (SVR), to forecast tourism demand forecasting, the approach, and known as genetic algorithm (GA)-SVR, which searches for SVR's optimal parameters using real value GAs, and then adopts the optimal parameters to construct the SVR models.

Dimitriou et al (2008) presents an adaptive hybrid fuzzy rule-based system (FRBS) approach for the modeling and short-term forecasting of traffic flow in urban arterial networks, and employs univariate and multivariate data structures and uses a genetic algorithm for the offline and online tuning of the FRBS membership functions according to the prevailing traffic conditions. The results obtained from the online application of the proposed model are found to overperform those of the offline application and conventional statistical techniques.

Brezocnik et al (2004) propose GP to predict surface roughness in end-milling. Accuracy of the best model was proved with the testing data. It was established that the surface roughness is most influenced by the feed rate, whereas the vibrations increase the prediction accuracy.

Yao and Lin (2007) developed the GP algorithm, that is utilised to search for the optimal Volterra filter structure, and several experiments are made to justify the effectiveness and efficiency of the proposed modified by the GP algorithm.

Gaur and Deo (2008) presents an application of a relatively new soft computing tool called GP to forecasting the ocean waves on real-time or online basis while carrying out any operational activity in the ocean. In order to obtain forecasts that are station-specific a time-series based approach like stochastic modeling or artificial neural network was attempted by some investigators in the past.

Afzal and Torkar (2011) performed a systematic review of literature that compared genetic programming models with comparative techniques based on different independent project variables. The objective is to investigate the evidence for symbolic regression using genetic programming being an effective method for prediction and estimation in software engineering, when compared with regression/machine learning models and other comparison groups (including comparisons with different improvements over the standard GP algorithm).

2.3.3 Artificial Neural Network Methods

The trend in the literature is to apply artificial intelligence based soft computing techniques for complex prediction problems. Artificial neural networks which are a member of soft computing techniques were applied to strength prediction of several types of domain in the literature with considerable success.

2.3.3.1 Neural networks

Ledoux (1997) propose cooperation based neural networks traffic flow model, which aims at being integrated into a real time adaptive urban traffic control system. Concludes on the potentials of neural networks applied to traffic flow modeling, one minute ahead predictions of the queue length and the output flows have been obtained with fairly good accuracy.

Dougherty and Cobbett (1997) developed a technique of stepwise reduction of network size by elasticity testing the large neural network, the back-propagation neural network to forecast traffic flow, speed have to select the vast number of possible input parameters, and show a way of out-perform naive predictors.

Kirby et al (1997) has demonstrated that a straightforward application of neural networks can be use to forecast traffic flow along a motorway link. Dia (2001) discusses an object-oriented neural network model that was developed for predicting short-term traffic conditions, the feasibility of the approach is demonstrated through a time-lag recurrent network which was developed for predicting speed data up to 15 minuites into the future, and results obtained indicate that the model is capable of predicting speed up and with high degree of accuracy.

2.3.3.2 Support vector machines

Pai and Lin (2005) consider the ARIMA model cannot easily capture the nonlinear patterns, applied support vector machines (SVMs), a novel neural network technique, successfully to solve nonlinear regression estimation problems, the proposed hybrid methodology that exploits the unique strength of the ARIMA model and the SVMs model in forecasting stock prices problems.

Huang and Sadek (2009) develops a novel forecasting approach inspired by human memory, called the spinning network (SPN), the approach is then used for short-term traffic volume forecasting, utilizing a data set compiled from real-world traffic volume data. The results of the performance testing conducted demonstrates the superior predictive accuracy and drastically lower computational requirements of the SPN compared to either the neural network or the nearest neighbor approach.

Ishak (2003) presents an approach to optimize the short-term traffic prediction performance on freeways using multiple artificial neural network topologies under different network and traffic condition settings; the study was conducted to encourage multi-model techniques that are capable of improving the performance over single-model approaches. Using a mix of traditional and modern neural network topologies, the short-term speed prediction performance was extensively evaluated under different input settings and various prediction horizons (from 5 to 20 minutes).

2.4 Traffic Simulation Models

The Section consists of three parts. 2.4.1 Addresses the macroscopic approach of traffic flow theory. 2.4.2 discusses the microscopic models and 2.4.3 addresses the vehicular traffic of CTM models.

2.4.1 Macroscopic Models

The conventional traffic flow models have two different conceptual frameworks, macroscopic and microscopic traffic flow models. The macroscopic traffic flow models is view as a compressible fluid and mostly devoted to elucidating the relations between speed, density and flow in various traffic conditions and roadway environments. The fluid-dynamical description models analogize vehicular flows to fluids by assuming the aggregate homogeneous behavior of drivers.

Lighthill and Whitham (1955) and Richard (1956) treated the flow as function of only the local density and developed the most well-known one-order fluid-dynamical (LWR) models. The shock wave theory used by LWR model was to explain many traffic phenomena such as congested traffic upstream of a freeway bottleneck. Other researchers such as Payne (1971), Liu et al (1998) and Zhang (1998) derived high-order similar models. Wong and Wong (2002) formulated a multi-class traffic flow model as an extension of LWR model with heterogeneous drivers. The main concepts are described as follow: When viewing from an aircraft at a freeway, one can imagine the vehicular traffic flow is regarded as a stream. Due to this analogy, traffic is explained in relationship of flow, density, and speed (Kühne, 1998).

For traffic prevailed on freeway, since longitudinal movement is the main concern, the continuity equation is further simplified into a two-dimensional equation with two independent variables—the location at an instant of time. Upon this, Lighthill and Whitham (1955) first conjectured that density is the function of the two above-mentioned independent variables. After that, they adopted the traditional formula for estimating fluid flow rate.

Obviously LWR model is an over-simplification of traffic phenomena, since it assumes a homogeneous and deterministic traffic flow and it implies smooth and concave functionso for both speed and density. Therefore in the past three decades, many efforts were devoted in improving the LWR model. Most notable studies in this regard include the work by Bick and Newell (1960) for two-lane bidirectional road, and those by Liu et al (1998) and Zhang (2005) of high-order similar models. In the same spirit, some other proposed systems of finite different equations (FPE) to model freeway traffic, such as Payne (1971) and Daganzo (1995). Wong and Wong (2002) further formulated a multi-class traffic flow model as an extension of LWR model with heterogeneous drivers.

2.4.2 Microscopic Models

The microscopic traffic flow models, in contrast to the macroscopic traffic flow models, describe the interrelationship of individual vehicle movements with other vehicles.

These models of vehicular traffic attention are focused on individual vehicle each of which is represented by a “particle”. Car-following models are the most pertinent ones to explicate the one-dimensional movements in a longitudinal lane such that the following vehicle adjusts its speed to maintain desirable or safe distance headways with the lead vehicle. Stimulus-response model is perhaps the most prominent type developed in the 1950s and 1960s by the General Motors (GM) research group, which is still being applied or extended.

The concept of car following on a motorway indicates a driver reacts to the altering headway with his predecessor the behavior of a driver. The main idea is that a driver will through control of the vehicle deceleration and acceleration to maintain a suitable distance and time gap between it and the vehicle that precedes it in the same lane. Many studies have devoted to find mathematical formulas to descript the following behavior of the individual driver. The starting point of define the equation of motion is usually the analogue of the Newton’s equation for each individual vehicle. This approach assumes within a range of distance, a stimulus-response relationship exist.

The stimulus function can be composed of many factors: the speed of the vehicle, distance headway, relative speed, etc. Each driver can respond to the surrounding traffic conditions only by accelerating or decelerating the vehicle. Different forms of the equations of motion of the vehicles in the different versions of the car-following models arise from the differences in their postulates regarding the nature of the stimulus.

Chandler et al (1958) proposed the first follow-the-leader model, the different in the velocities of the following nth and the leading n+1th vehicle was supposed to be the stimulus for the nth vehicle. It was assumed that every driver be likely to keep with the synchronized velocity as that of the front vehicle.

Pipes (1953) car-following theory is a linear model that depicts vehicular traffic behavior. That with the purpose of avoid crash with the leading vehicle, each driver must keep a safe distance from the leading vehicle. If the velocity of the nth vehicle is higher, the spacing to its leading vehicle needs to keep larger.

A series of models have been developed in the 1950s and 1960s by Herman and his colleagues at the General Motors Research Laboratories to address microscopic approaches that focused on describing the driver car-following behaviors. Five generations of the GM car-following models are recognized and they provided an essential contribution to realize the traffic flow. Nowadays they are still applied in various aspects, including traffic stability and safety studies, level of service and capacity analysis, driver’s reaction times, etc.

Recently, various cellular automata (CA) models, comprehensive based on the aforementioned conventional traffic flow theory, have been developed to describe the phenomena of real traffic flows characterized with complex dynamic behaviors.

Cellular automaton (CA) is a powerful tool to describe the phenomena of real traffic flows characterized with complex dynamic behaviors. Nagel and Schreckenberg (1992) first proposed the renowned NaSch model to reproduce the basic features of real traffic. In their model, the road is divided into squared-cells of length 7.5 meters. Each cell can either be empty or occupied by at most one car. The space, speed, acceleration and even the time are treated as discrete variables. The state of the road at any time-step is derived from one time-step ahead by applying acceleration, braking, randomization and driving rules for all cars at the same time. Obviously, their coarse description of cells is an extreme simplification of the real world conditions. Therefore, a considerable number of modified NaSch CA rules have been found in the past decade. Other related works that improved NaSch coarse cells with finer cells have also been found. In addition, Wolf (1999) employed a modified NaSch model to address the metastable states at the jamming transition in detail. Wang et al (2000) introduced NaSch model and Fukui-Ishibashi model to investigate the asymptotic self-organization phenomena of one-dimensional traffic flow. Pottmeier et al (2002) studied the impact of localized defect in a CA model for traffic flow exhibiting metastable states and phase separation.

NaSch model was proposed by K. Nagel & M. Schreckenber (1992). Capable of reproduce important entities of real traffic flow, e.g. density-flow relation. Model description, one-lane traffic, divided into cells of length 7.5 m. Each cell can either be empty or occupied by at most one car with discrete velocity.

The NaSch model is a minimal model in the sense that all the four steps are necessary to reproduce the basic features of real traffic; however, additional rules need to be formulated to capture more complex situations.

Barlovic et al (1998) based on the original NaSch model, and further proposed the velocity dependent randomized (VDR) model which is basically analogous to the BJH model and endeavors to establish proper update rule for randomization (Pn) in accordance with the velocity variation of vehicles. The VDR model did successfully exhibit the metastable states and consequently, the hysteresis effect that was never visible in the previous simulations. Therefore the VDR model has been widely adopted worldwide since then.

Knospe et al (2000) based on original VDR model and the following driving strategy. At large distances the cars move with their desired velocity vmax. At intermediate distances drivers react to velocity changes of the next vehicle downstream, i.e. to ‘brake lights’. At small distances the drivers adjust their velocity such that safe driving is possible. The acceleration is delayed for standing vehicles and directly after braking events.

Kerner (2004), a German traffic physician, introduced a three-phase traffic theory which consists of free flow, synchronized flow and wide moving jam phases. The later two phases exist in congested states where downstream front of the synchronized flow phase is often fixed at a bottleneck but the wide moving jam will propagate through the spatial locations of the bottleneck.

To explore the emergence of such traffic patterns, Kerner and partners have shown complex spatiotemporal behaviors based on empirical freeway traffic analysis. The effect of slow cars in two-lane systems was further studied by Knospe et al (1999) who found that even few slow cars could initiate the formation of platoons at low densities.

Moreover, Kerner (2005) also compared the congested pattern control approach with the free flow control approach at an on-ramp bottleneck with ramp metering. It was found that the congested pattern control approach has higher throughputs on the main road downstream of the bottleneck and considerably lower vehicle waiting times at the light signal on the on-ramp. The upstream propagation of congestion does not occur even if large amplitude perturbations appear in traffic flow.

2.4.3 Related Cell Transmission Model

The CTM was developed as a discrete approximation to the hydrodynamic theory of traffic flow. It is capable of automatically tracking shocks and acceleration waves and thus capturing traffic behavior in the process of the formation, propagation, and dissipation of queues. One famous approach in this regard is the CTM model proposed by Daganzo (1994, 1995). In that, the LWR continuum model is discretized into cells. The road is represented by a number of small sections (cells). The simulation model keeps tracking the number of vehicles in each cell, and in each time-step it calculates the number of vehicles that cross the boundaries between adjacent cells. The flow from one cell to the other depends on how many vehicles can be sent by the upstream cell and how many can be received by the downstream cell. The amount of vehicles that can be sent is a function of the density in the upstream cell whereas the number can be received depends on the density in the receiving cell. The lagged CTM (Daganzo, 1999) is a refinement of this scheme, where the amount of vehicles a cell can receive (from the adjacent upstream cell) is also affected by the density some time earlier in the cell.

Lo et al (2001) developed a dynamic traffic-control formulation and transformed the CTM to a set of mixed-integer constraints and subsequently cast the dynamic signal-control problem to a mixed-integer linear program. This study produced results to show the benefit of dynamic timing plans and demonstrated that some of the existing practice on signal coordination could be further improved.

Lo and Szeto (2002) developed a cell-based DTA formulation and through defining an appropriate gap function, and transformed a formulation based on the nonlinear complementarity problem to an equivalent mathematical program, this formulation encapsulates a network version of the CTM, the formulation is able to capture dynamic traffic phenomena, such as shock-waves, queue formation, and dissipation. Moreover, it is capable of capturing dynamic traffic interactions across multiple links.

Lo and Szeto (2002) developed a cell-based dynamic traffic assignment formulation that through a variational inequality approach. This formulation satisfies the first-in-first-out (FIFO) conditions through the CTM, and employed an alternating direction method developed for co-coercive variational inequality problems.

Lin and Lo (2003) show that these moving jams are not particularly peculiar but can be explained with the hydrodynamic theory of traffic flow, or the Lighthill-Whitham-Richards model, and the merge and diverge models in the cell transmission model. In fact, and demonstrate that this stationary jam phenomenon can be replicated with a simple two-wave velocity (or triangular) flow-density relationship in conjunction with the hydrodynamic theory. This finding provides some evidence to support that a triangular flow-density relationship is a good approximation of field observations.

Szeto and Lo (2004) develops a cell-based formulation for the simultaneous route and departure time choice problem with elastic demands through a variational inequality problem (VIP), and prove that the O-D first-in-first-out (FIFO) property is only maintained under certain conditions of the travel time and schedule delay costs. the theoretical analyses together with the empirical results indicate that O-D FIFO should hold in reality.

Boel and Mihaylova (2006) present a stochastic model that extends the CTM, of freeway traffic at a time scale and of a level of detail suitable for on-line estimation, routing and ramp metering control. The freeway is considered as a network of interconnected components, corresponding to one-way road links consisting of consecutively connected short sections (cells). The model is validated over synthetic data with abrupt changes in the number of lanes and over real traffic data sets collected from a Belgian freeway.

Gomes and Horowitz (2006) using a cell transmission-like model called the asymmetric cell transmission model (ACTM) to solved onramp metering control problem, formulation captures both free flow and congested conditions, and includes upper bounds on the metering rates and on the onramp queue lengths.

2.5 Summary

Base on the above literature review, this study proposes a novel approach, based on rolling self-structured traffic patterns, to make a long prediction for the traffic features along a freeway corridor. In the prediction process, a sequence of historical traffic data are collected and used to identify the similar cluster of traffic patterns. The input values of learned GP model belonging to that cluster are used to predict the subsequent many hours. With the predicted medium-to-long on-ramp traffic, CTM is used to simulate the arrival patterns. And then, EKF is employed to estimate the O-D proportions.

CHAPTER 3 PROBLEM DEFINITIONS

This chapter introduces the problem definitions in this research. Section 3.1 addresses the problem statement including the definitions of typical freeway corridor, variables, parameters and the rolling concept; the proposed model framework is introduced in Section 3.2.

3.1 Problem Statement

3.1.1 Definitions

3.1.1.1 A typical freeway corridor

The dynamic O-D estimation model of this research is based on Chang *et al*. (1994, 2006, 2007), consider a typical linear freeway corridor with N segments, coding 0 to N-1, as shown in Figure 2. Assume that detectors are installed at all on-ramps, off-ramps, and mainline links. The information that is readily available for estimation of dynamic O-D distribution is the time series of entering flow, $q_i(k)$, exiting flow, $y_j(k)$, and mainline flow, $U_l(k)$.

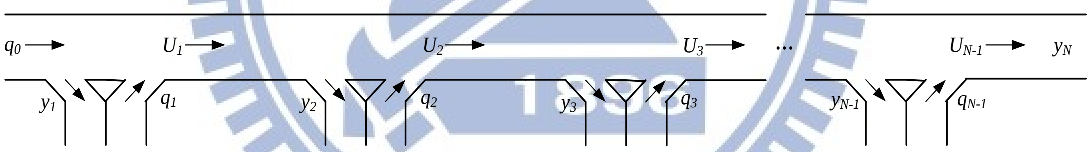


Figure 2 A typical linear freeway corridor

The relationship between the dynamic O-D pattern, resulting link flow, and arrival pattern can be expressed by the following equations:

$$y_j(k) = \sum_{m=0}^{M}\sum_{i=0}^{j-1} q_i(k-m)\cdot\rho_{ij}^{m}(k)\cdot b_{ij}(k-m) \tag{1}$$

$$U_l(k) - q_l(k) = \sum_{m=0}^{M}\sum_{i=0}^{j-1}\sum_{j=l+1}^{N}\left[q_i(k-m)\rho_{il}^{m}(k)\right]b_{ij}(k-m) \tag{2}$$

Eqs. (1) and (2) express the relationship between entering traffic ($q_i(k-m)$), mainline traffic volume ($U_l(k)$-$q_l(k)$), O-D matrices proportion ($b_{ij}(k-m)$), and arrival pattern ($\rho_{ij}^{m}(k)$).

3.1.1.2 Definition of variables and parameters

Obviously, the system formulation has a large number of state parameters, *i.e.*, $b_{ij}(k)$ and $\rho_{ij}^{m}(k)$, causing low efficiency in estimation. As such, more information is required to ensure the proposed model to be computationally efficient and tractable.

The variables and parameters of O-D estimation, GHSOM and GP, used in this study are defined in Table 1.

Table 1 Definition of variables and parameters

Variables/ parameters	Definition
$q_0(k)$	The number of vehicles entering the upstream boundary of the freeway section during time interval *k*.
$q_i(k)$	The number of vehicles entering freeway from on-ramp *i* during time interval *k*, $i = 1,2,...,N-1$.
$y_j(k)$	The number of vehicles leaving freeway from off-ramp *j* during time interval *k*, $j = 1,2,...,N-1$.
$y_N(k)$	The mainline volume at the downstream end of the freeway section during time interval *k*.
$U_l(k)$	The number of vehicles crossing the upstream boundary of segment *l* during time interval *k*, $l = 1,2,...,N-1$.
$T_{ij}(k)$	The number of vehicles entering the freeway from on-ramp *i* during time interval *k* that are destined to off-ramp *j*, where $0 \le i < j \le N$.
$b_{ij}(k)$	The proportion of $q_i(k)$ heading toward destination node *j* during time interval *k*.
$\rho_{ij}^{m}(k)$	The fraction of $T_{ij}(k-m)$ vehicles departing from entry node *i* during time interval *k* that takes *m* time intervals to exiting node *j*.
$\rho_{il}^{m}(k)$	The fraction of $T_{ij}(k-m)$ trips from entry node *i* during time interval *k* that takes *m* time intervals to mainline node.
$\rho_{ij}^{-}(k)$	The fraction of $T_{ij}(k-m)$ trips from entry node *i* during time interval *k* that takes *m* ($m = t_{ij}^{-}(k)$, $t_{ij}^{-}(k) = \mathrm{int}[\mu_{ij}(k)/t_0]$) time intervals to mainline node.
$\rho_{ij}^{+}(k)$	The fraction of $T_{ij}(k-m)$ trips from entry node *i* during time interval *k* that takes *m* ($m = t_{ij}^{+}(k)$, $t_{ij}^{+}(k) = t_{ij}^{-}(k)+1$) time intervals to mainline node.
$X_k(t+1)$	A sequence of *r*-period historical traffic features observed at location *k* starting from time $t+1$.
$\hat{x}_k(t+r+h)$	Predicted the flow features *h* periods ahead.
τ_1	Breadth of GHSOM, a threshold to specify the desired level of detail that is to be shown in a particular SOM.
τ_2	Depth of GHSOM, a threshold to specify the desired quality of input data representation at the end of the learning process.
$\alpha(l)$	The learning rate function which controls the amount of weight

Variables/ parameters	Definition
	vector adjustment and decreases with the iterations.
$\sigma(l)$	Defines the width of the neighborhood function and also decreases monotonically.
$w_{ij}(k)$	The white noise of the state variable at time k, and $E(w_{ij}(k)) = 0$, $Var(w_{ij}(k)) = D(k)$ follow Gaussian distribution.
e(k)	The observation error term of the measurement variable, following Gaussian distribution, and $E(e(k)) = 0$, $Var(e(k)) = R(k)$.

3.1.1.3 The rolling concept

Let $X_k(t+1)=[x_k(t+1), x_k(t+2), \ldots, x_k(t+r)]$ denote a sequence of *r*-period historical traffic features (*e.g.*, five-minute flow rates in the study) observed at location *k* starting from time *t*+1. The proposed method aims to predict the flow features *h* periods ahead, denoted as $\hat{x}_k(t+r+1)$, $\hat{x}_k(t+r+2)$,..., $\hat{x}_k(t+r+h)$. For a short-term prediction (*i.e.*, *h*=1 or 2), the predicted results have provided useful information for some ITS applications like traffic responsive control. However, if the purpose is to develop advanced traveler information systems, we need to predict the above flow sequences at different locations (*i.e., k*>>1) with longer periods ahead (*i.e.*, *h*>>1).

This study aims to predict $\hat{x}_k(t+r+1)$, $\hat{x}_k(t+r+2)$,..., $\hat{x}_k(t+r+h)$ with relatively long period ahead (*h*>>1). Undoubtedly, the prediction may be rather inaccurate if *h*>>*r*. To overcome this difficulty, a rolling-horizon concept is incorporated into the proposed prediction method. First, we employ the GHSOM model to classify the given traffic patterns into appropriate clusters. Then, used the GP model to prediction traffic sequence in each cluster such that a portion of the most updated (*r*-*s*) periods of traffic sequence, say $x_k(t+s)$, $x_k(t+s+1)$, …, $x_k(t+r)$, are used to predict the traffic at the very next time period, $\hat{x}_k(t+r+1)$. This predicted traffic flow together with the previous traffic flows are further used to predict the consecutive periods in a rolling manner. Specifically, $\hat{x}_k(t+r+1)$ is predicted based on $x_k(t+s)$, $x_k(t+s+1)$,…, $x_k(t+r)$; $\hat{x}_k(t+r+2)$ is predicted based on $x_k(t+s+1)$, $x_k(t+s+2)$,…, $x_k(t+r)$, $\hat{x}_k(t+r+1)$; $\hat{x}_k(t+r+3)$ is predicted based on $x_k(t+s+2)$, $x_k(t+s+3)$,…, $\hat{x}_k(t+r+1)$, $\hat{x}_k(t+r+2)$; and so on. The details of traffic pattern clustering and traffic prediction are further elaborated next chapter.

3.1.2 Estimation of an O-D Matrix

The core logic of the proposed model is how to estimate dynamic OD matrices $T_{ij}(k)$ or $b_{ij}(k)$ by observing the number of vehicles on ramp $q_i(k)$, off ramp $y_j(k)$ and main line $U_l(k)$

According to the definition, the relation between the dynamic O-D pattern and resulting link flow can be expressed by equations (Lin and Chang, 2007):

$$q_i(k) = \sum_{j=i+1}^{N} T_{ij}(k), \qquad i = 0,1,\ldots,N-1 \tag{3}$$

$$T_{ij}(k) = q_i(k) \cdot b_{ij}(k), \qquad 0 \le i < j \le N \tag{4}$$

The above two equations are subjected to the following natural constraints:

$$0 \le b_{ij}(k) \le 1, \qquad 0 \le i < j \le N \tag{5}$$

$$\sum_{j=i+1}^{N} b_{ij}(k) = 1 \qquad i = 0,1,2,\ldots N-1 \tag{6}$$

Consider the speed variation among drivers, it is reasonable to assume that the travel time of vehicles from node i to node j during time interval k are distributed among time intervals k-M,…, k-1, and k where M is the maximum number of intervals required for vehicles to traverse the entire freeway section. The traffic volume leaving freeway from off-ramp j, $y_j(k)$, can thus be expressed as:

$$y_j(k) = \sum_{m=0}^{M} \sum_{i=0}^{j-1} T_{ij}(k-m)\, \rho_{ij}^{m}(k) \tag{7}$$

$$U_l(k) = \sum_{m=0}^{M} \sum_{i=0}^{l-1} \rho_{il}^{m}(k) \left[\sum_{j=l+1}^{N} T_{ij}(k-m) \right] + q_1(k) \tag{8}$$

where $\rho_{ij}^{m}(k)$ shall satisfy the following relations:

$$0 \le \rho_{ij}^{m}(k) \le 1, \qquad 0 \le i \le j \le N, \quad m = 0,1,\ldots,M \tag{9}$$

$$\sum_{m=0}^{M} \rho_{ij}^{m}(k+m) = 1, \;\; 0 \le i < j \le N \tag{10}$$

Thus, taking into consideration the difference of travel time caused by different driver speed factors, Chang and Wu (1994) add a variable $\theta_{ij}^{m}(k)$ to reflect the effects of estimated results on different travel times, and using the parameters $\rho_{ij}^{m}(k)$ to predict the vehicle arrival distribution, which is defined as follows:

$\rho_{ij}^{m}(k)$ ：The fraction of $T_{ij}(k\text{-}m)$ vehicles departing from entry node i during time interval k that takes m time intervals to exiting node j.

Therefore, considering this argument, the number of vehicles on the off ramp at time k can be rewritten as Eq. (11). It represents the dynamic relationship between O-D distribution pattern and road traffic.

$$
\begin{aligned}
y_j(k) &= \sum_{m=0}^{M}\sum_{i=0}^{j-1} T_{ij}(k-m)\cdot\rho_{ij}^{m}(k) \\
&= \sum_{m=0}^{M}\sum_{i=0}^{j-1} q_i(k-m)\cdot b_{ij}(k-m)\cdot\rho_{ij}^{m}(k)\ ,\quad j=1,2,\ldots,N
\end{aligned}
\tag{11}
$$

The relationship of the number of vehicles on the main line, must also comply with the constraints of flow conservation, and significantly increases the performance of the estimation model.

$$
\begin{aligned}
U_l(k) &= \sum_{m=0}^{M}\sum_{i=0}^{l-1}\rho_{il}^{m}(k)\left[\sum_{j=l+1}^{N} T_{ij}(k-m)\right] + q_1(k) \\
U_l(k) - q_1(k) &= \sum_{m=0}^{m}\sum_{i=0}^{l-1}\sum_{j=l+1}^{N} q_i(k-m)b_{ij}(k-m)\rho_{il}^{m}(k),\qquad l=1,2,\ldots,N-1
\end{aligned}
\tag{12}
$$

Eq. (12) is a traffic equation to express the relationship between entering traffic ($q_i(k-m)$), arrival pattern ($\rho_{ij}^{m}(k)$), O-D matrices proportion ($b_{ij}(k-m)$), and mainline traffic volume ($U_l(k)$-$q_l(k)$). Obviously, the system formulation has a large number of state parameters, *i.e.*, $b_{ij}(k)$ and $\rho_{ij}^{m}(k)$. The number of these unknown parameters increases with the necessary M value. As such, some more information is required to ensure this proposed model to be computationally efficient and tractable.

To deal with the large number of unknown parameters, Chang and Wu (1994) simplified the formulations by assuming that the speeds of vehicles entering the freeway at the same time interval are distributed in a small range. Therefore, eqs. (11) and (12) can be rewritten as:

$$
y_j(k) = \sum_{i=0}^{j-1}\left[q_i(k-t_{ij}^{+}(k))\right]\rho_{ij}^{+}(k)b_{ij}(k-t_{ij}^{+}(k))) + \sum_{i=0}^{j-1}\left[q_i(k-t_{ij}^{-}(k))\right]\rho_{ij}^{-}(k)b_{ij}(k-t_{ij}^{-}(k)))
\tag{13}
$$

$$
U_l - q_l(k) = \sum_{i=0}^{l-1}\sum_{j=l+1}^{N}\left[q_i(k-t_{ij}^{+}(k))\right]\rho_{ij}^{+}(k)b_{ij}(k-t_{ij}^{+}(k))) + \sum_{i=0}^{j-1}\left[q_i(k-t_{ij}^{-}(k))\right]\rho_{ij}^{-}(k)b_{ij}(k-t_{ij}^{-}(k)))
\tag{14}
$$

As such, the number of unknown parameters reduces from $(M + 1)N(N + 1)/2$ to $3N(N+1)/2$. However, if the target freeway corridor is sufficiently long and experiences moderate congestion, the speeds of vehicles for the same O-D may vary in a wide range. Then, eqs. (13) and (14) are not adequate to capture all complex interrelations between traffic flows and O-D patterns. To overcome these limitations, Lin and Chang (2005) proposed a new set of generalized formulations by employing a distribution to represent the potential variation of travel times among drivers due to the impact of congestion and the difference in their desired speeds. They assumed that the travel times of drivers departing from node *i* during time interval *k* to node *j* follow a specific distribution. Since the travel times for the same O-D pair drivers departing during the same time interval follow a distribution, Lin and Chang (2007) replaced $\rho_{ij}^{m}(k)$ with a cumulative density function for one time interval as follows:

$$\rho_{ij}^{m}(k) = \int_{m\cdot t_0}^{(m+1)t_0} f_{ij}^{m}(x)dx \tag{15}$$

By applying the above travel time distribution concept, the relationships between ramp volumes and O-D proportions can be rewritten as:

$$\begin{aligned} y_j(k) &= \sum_{m=0}^{M}\sum_{i=0}^{j-1} q_i(k-m)\cdot\rho_{ij}^{m}(k)\cdot b_{ij}(k-m) \\ &= \sum_{m=0}^{M}\sum_{i=0}^{j-1} q_i(k-m)\cdot\left[\int_{m\cdot t_0}^{(m+1)t_0} f_{ij}^{m}(x)dx\right]\cdot b_{ij}(k-m) \end{aligned} \tag{16}$$

$$\begin{aligned} U_l(k)-q_l(k) &= \sum_{m=0}^{M}\sum_{i=0}^{j-1}\sum_{j=l+1}^{N}\left[q_i(k-m)\rho_{il}^{m}(k)\right]b_{ij}(k-m) \\ &= \sum_{m=0}^{M}\sum_{i=0}^{j-1}\sum_{j=l+1}^{N} q_i(k-m)\left[\int_{m\cdot t_0}^{(m+1)t_0} f_{ij}^{m}(x)dx\right]b_{ij}(k-m) \end{aligned} \tag{17}$$

Compared to Chang and Wu (1994), the number of unknown parameters for Eqs. (16) and (17) has reduced from $3N(N+1)/2$ to $2N(N+1)/2$. On the other hand, Lin and Chang (2005) represented the different speeds of vehicles for the same O-D pair by using the distribution of travel time.

Although the relevant studies (e.g. Chang and Wu, 1994; Chang and Tao, 1995; Lin and Chang, 2005, 2007) have shed light on the dynamic O-D matrices estimation, most of them made subjectively assumptions regarding arrival distributions, which may not be valid for various conditions from free-flow to gridlock. In addition, most of these models are too complex, causing low efficiency in estimation. In view of the importance of the arrival distribution prediction and the estimation efficiency required for real-time implementation, this study aims to develop a model that can accurate capture the traffic hydrodynamics under various traffic conditions in an efficient manner.

The macroscopic pattern only calculates the average travel time dependent on average link flows on larger scale road networks, while traffic flows increase leading to road network congestion. This method of travel time estimation is too innaccurate and cannot reflect actual vehicle trips. In terms of microscopic traffic flow model estimates, detailed observation of driver behavior of all vehicles on the road, and a complicated calculation and update will result in a lack of efficiency and timeliness. Therefore, in view of the mediumscopic traffic model travel time estimation methods, not only effectively observe vehicle fleet operator behavior, but also reduce computer calculation time and storage space, improving the efficiency and objectivity of dynamic O-D matrices estimation. The CTM is a mediumscop traffic mode for predicting travel time, developed by Daganzo (1994), based on the basic concepts of fluid dynamics to derive a single direction and single entrance road density process, not making any additional assumptions for effective and accurate forecasting of vehicle operating behavior on networks to reflect the true dissipation under different traffic conditions.

In the past, many scholars validated the effectiveness of the CTM model, for instance, the estimation of the traffic density and congestion patterns in difficult to measure road sections based on the use of CTM mode conversion state - space model (e.g. Sun et al, 2003). According to the hybrid Monte Carlo development of the Mixture KF algorithm to solve the discrete approximation of non-observation of the transition state of the state-space model. The estimated results show that real-time filtering algorithm is feasible and efficient, and the advantages of this model is average estimation error under 10%, the estimate fitting in with daily variation of traffic information.

Compared to other assumptions, this study using CTM simulates the behavior of real traffic, fitting in with real traffic operating conditions, in order to simulate the vehicle's arrival distribution in the road network, and then replacing the assumptions of travel time to no more than one time interval.

Therefore, the research proposes CTM to predict traffic arrive patterns, more easily than simulating travel time with different traffic conditions than the macroscopic model, but also more quickly than the microscopic solution.

However, the arrival distribution estimations using CTM are based on an unrealistic assumption that the on-ramp traffic along a freeway remains unchanged over time. To rectify this unrealistic assumption, a medium-to-long term (e.g. next two to four hours) prediction model of on-ramp traffic along a freeway is required.

According to field observation, daily traffic patterns do repeat spatially and temporally over and over again, this research proposes a two-stage prediction model, based on rolling self-structured traffic patterns, to make a long prediction for the traffic features along a freeway corridor, first employs the GHSOM model to partition unlabeled traffic patterns into appropriate number of clusters and then develops a GP model associated with each cluster to predict the traffic features based on rolling self-structured traffic patterns. With the predicted medium-to-long on-ramp traffic, CTM is used to simulate the arrival patterns. And then, EKF is employed to estimate the O-D proportions.

3.2 The Proposed Model Framework

Figure 3 presents the detailed process of the proposed approach. Throughout the prediction process, the step 1, inputting the rolling on ramp traffic patterns and use GHSOM algorithm to cluster traffic patterns, and step 2, taking 240 time interval traffic pattern to matching on the similar cluster and use GP to predict 48 time interval traffic information. Moreover, CTM is used to simulate the arrival patterns and EKF is used to estimate the O-D proportions, the program of the propose method was coded in Visual Basic 6.0, and refer to the GHSOM package, developed jointly by the University of Aberdeen and Vienna University of Technology.

To replicate traffic behaviors by CTM, traffic demand of each O-D pair must be given by GHSOM and GP to predict traffic on ramp in advance. That is, a set of $b_{ij}(k)$ is to be determined and used to assign the detected on-ramp traffic to different downstream interchanges. Once the arrival distributions of all entering traffic have been successfully simulated, $\rho_{ij}^{m}(k)$ can be computed and used to calibrate the O-D proportions of entering traffic $q_i(k)$ by EKF, namely $b_{ij}'(k)$. Then, the new O-D proportions $b_{ij}'(k)$ will be used to replicate a revised arrival distribution $\rho_{ij}^{m}{}'(k)$ in an iterative manner.

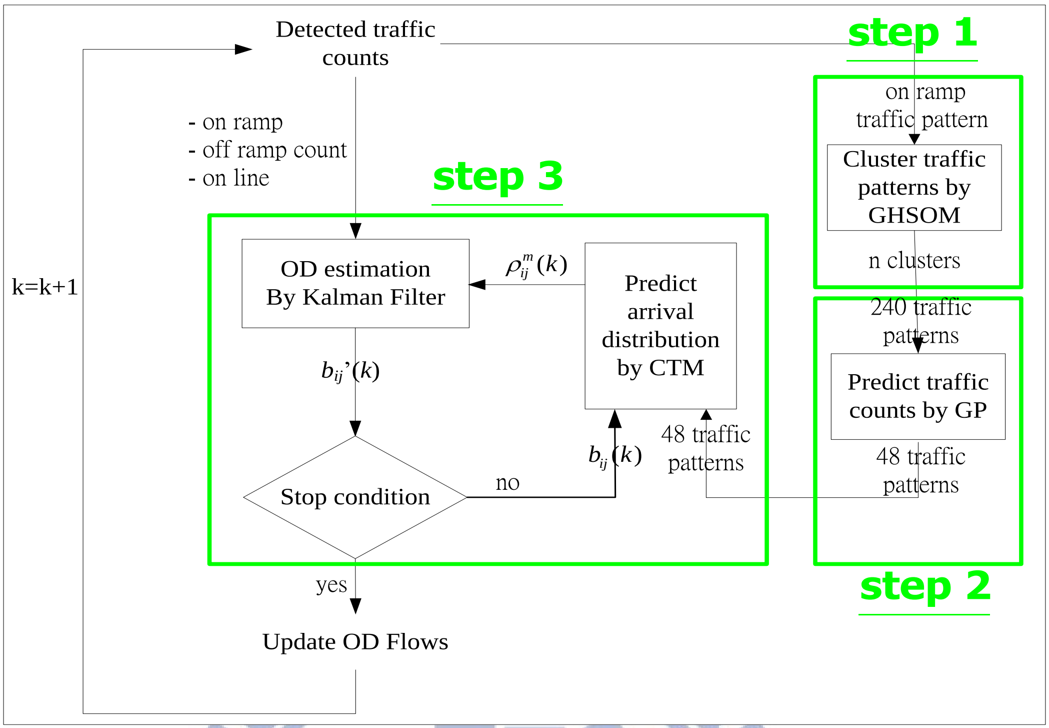


Figure 3 Process of the proposed approach

The core logic of the proposed model is build recursive dynamic OD matrix estimation model, using GHSOM and GP to predict traffic on ramp in advance, and then supply the CTM to predict arrival patterns by simulated traffic flow on freeway time and space, and finally using EKF to estimate OD matrices.

Among them, the stop condition: $\sum_i \sum_j [b_{ij}^{n-1}(k) - b_{ij}^{n}(k)]^2 < 0.05$ where n is the operation result of n-th recursive, or 500 times the number of recursion. Under the recursive conditions program will result in a new $b_{ij}'(k)$, otherwise $b_{ij}'(k)$ to continue to do the operation until the convergence condition is reached.

CHAPTER 4 MODEL FORMULATION

This chapter introduces the dynamic freeway O-D matrices estimation algorithm in this research. Section 4.1 addresses the self-structured medium-to-long term traffic prediction model, the arrival distribution modeling was introduced in Section 4.2, and the dynamic OD matrices estimation algorithm was introduced in Section 4.3.

4.1 Self-Structured Medium-to-long Term Traffic Prediction Model

4.1.1 Traffic Pattern Clustering: GHSOM

According to our observation, daily traffic patterns do repeat spatially and temporally over and over again, this historical traffic data are collected and used GHSOM to identify the similar cluster of traffic patterns into appropriate different traffic patterns.

Pattern clustering is also known as cluster analysis, set partitioning, Q-analysis, typology, grouping, clumping, classification, numerical taxonomy, or unsupervised pattern recognition. In traffic literature, traffic pattern at a specific location can represent a sequence of traffic features such as flow, speed, occupancy, etc. Hence, traffic pattern clustering is a classification process wherein a group of unlabeled traffic patterns are partitioned into a number of sets—similar patterns in the same cluster and dissimilar patterns in different clusters.

Brucker (1978) and Welch (1983) proved that, for specific objective functions, clustering becomes an NP-hard problem when the number of clusters exceeds three, if one aims to find the optimal clusters. Numerous heuristic algorithms for clustering have been developed, which can generally be divided into five categories: statistics clustering, mathematical programming, network programming, neural network and metaheuristics (Chiou and Lan, 2001; Chiou and Chou, 2010). Rauber *et al.* (2002) proposed the GHSOM model and proved that it possesses excellent performance in pattern clustering; thus, this study will employ GHSOM to conduct traffic pattern clustering.

In fact, GHSOM is an extension of self-organized map (SOM), an artificial neural network that performs clustering by means of unsupervised competitive learning algorithm, initiated by Kohonen (1982). During the learning process, the network performs clustering and the model vectors change to reflect the similarity of neighboring clusters. The goal of SOM is to represent the points in the source space by corresponding points in a lower dimensional target space, often in a two-dimensional lattice. However, SOM can neither capture the inherent hierarchical structure of data, nor determine the size of the preset map ignoring the characteristics of data distribution. To overcome these shortcomings, the GHSOM (Rauber *et al.*, 2002) has a hierarchical structure of multiple layers, where each layer consists of several independent growing SOMs.

The GHSOM architecture starts from a top-level map, which grows in size to represent a collection of data at different specific levels. For instance, Layer 1 contains 2×2 units and provides a rather rough organization of the main clusters in the input data. The four independent maps in Layer 2 give a more detailed data information. The three identified units in Layer 2, which have diversified input data mapped onto them, are further expanded to form a new independent SOM at the subsequent layer (Layer 3), and so on, depicted in Fig. 4.

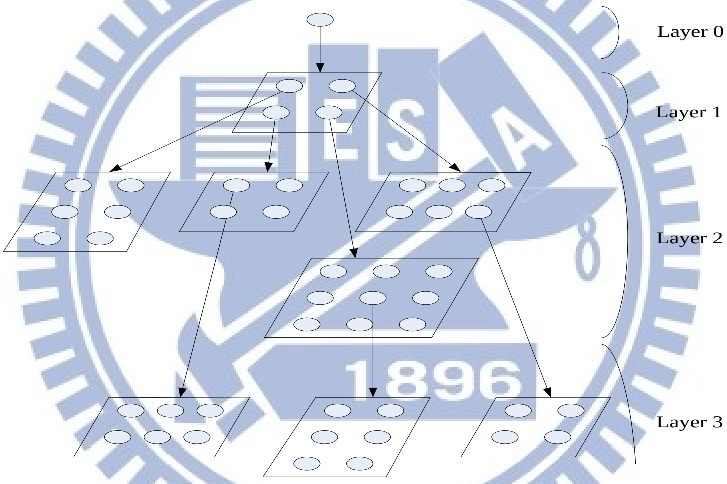


Figure 4 An illustration of the GHSOM architecture

To elucidate the training algorithm of GHSOM, the training algorithm of conventional SOM is given below. A typical SOM network consists of an input layer and an output or competitive layer. The input layer is composed of a set of r-dimensional input vectors $\boldsymbol{x_k}=[x_k(t+1),\ x_k(t+2),\ \ldots,\ x_k(t+r)]$, where r indicates the number of features (i.e. the flow series at consecutive time intervals in this study) contained in each input vector. The output layer is an m-dimensional (oftentimes m=2) grid, which consists of a set of neurons, each associated with a r-dimensional weight vector $\boldsymbol{w_i} = [w_{i1},\ w_{i2},\ldots,\ w_{ir}]$ with the same dimension as the input vector. The arrangement of the neurons can be rectangular or hexagonal. Conceptually, SOM takes a set of inputs mapping them onto the neurons of two-dimensional grid. Randomly initializing the weight vectors, the SOM network then performs learning as the following steps.

Step 1: Randomly initialize the weight vector of each neuron.

Step 2: Determine the winning neuron. The SOM network determines the winning neuron for a given input vector, selected randomly from the set of all input vectors. For every neuron on the grid, its weight vector is compared with the input vector by using some similarity measures, e.g., Euclidean distance. The neuron whose weight vector is closest to the input vector is selected as the winning neuron. Eq. (18) shows how to determine the winning neuron b.

$$b: \|x_{kt} - w_b(l)\| = \min_i \{\|x_{kt} - w_i(l)\|\} \tag{18}$$

where l denotes the number of current learning iteration.

Step 3: Update the weights. After a winning neuron is determined, the weight vectors of winning neuron along with its neighboring neurons are updated so as to "move" toward the input vectors according to the following equation:

$$w_i(l+1) = w_i(l) + h_{bi}(l)(x_{kt} - w_i(l)) \tag{19}$$

where $h_{bi}(l)$ is the neighborhood function. A widely used neighborhood function is based on the Gaussian function:

$$h_{bi}(l) = \alpha(l)\exp(-\frac{\|r_i - r_b\|^2}{2\sigma(l)^2}) \tag{20}$$

where $\alpha(l)$ is the learning rate function which controls the amount of weight vector adjustment and decreases with the iterations; r_i and r_b are the locations of the neuron i and winning neuron b in the lattice; $\sigma(l)$ defines the width of the neighborhood function and also decreases monotonically.

Step 4: Test the stop condition. Steps 2 and 3 are repeated until all the patterns in the training set have been processed. In addition, to achieve a better convergence towards the desired mapping, it is usually required to repeat the previous loop until some convergence criteria are met.

Based on the concept of above SOM learning process, the training algorithm of GHSOM grows in two dimensions: horizontally (by increasing the size of each SOM) and hierarchically (by increasing the number of layers). In the horizontal growth, each SOM modifies itself in a systematic way similar to the growing grid so that each neuron does not represent too large an input space. In the hierarchical growth, on the other hand, the principle is to periodically check whether the lowest layer of SOMs have achieved sufficient coverage for the underlying input data. The basic steps of the horizontal growth and the hierarchical growth of the GHSOM are delineated below (Tangsripairoj and Samadzadeh, 2006):

Horizontal growth:

Step 1: Randomly initialize the weight vector of each neuron.

Step 2: Perform the conventional SOM learning algorithm for a preset number of iterations.

Step 3: Find the error unit *e* and its most dissimilar neighbor unit *d*. The error unit *e* is the neuron with the largest deviation between its weight vector and the input vectors it represents.

Step 4: Insert a new row or a new column between *e* and *d*. The weight vectors of these new neurons are initialized as the average of their neighbors.

Step 5: Repeat steps 2-4 until the mean quantization error of the map (MQE_m) is less than ($\tau_1 \cdot mqe_u$). τ_1 is a threshold to specify the desired level of detail that is to be shown in a particular SOM. mqe_u is the mean quantization error of the neuron *u* in the preceding layer of the hierarchy. Eq. (21) calculates mqe_u, which is the average distance between the weight vector of neuron *u* and the input vector mapped onto this neuron:

$$mqe_u = \frac{1}{n_{C_u}} \sum_{x_j \in C_u} \left\| x_j - w_u \right\|,\ n_{C_u} = \left| C_u \right| \tag{21}$$

where C_u denotes the set of input vectors that are mapped onto unit *u*; w_i denotes the weight vector of unit *i*; $\left\| x_j - w_u \right\|$ denotes the distance between input vector x_j and weight vector w_u; $\left| C_u \right|$ denotes the cardinality of the set C_u. Furthermore, MQE_m, the mean of all neurons' quantization errors in the map, is calculated as follows:

$$MQE_m = \frac{1}{n_U} \sum_{i \in U} mqe_i,\quad n_U = \left| U \right| \tag{22}$$

where *U* denotes the subset of map units.

Hierarchical growth:

Step 1: Check each neuron to find out if its mqe_u is greater than ($\tau_2 \cdot mqe_0$). τ_2 is a threshold to specify the desired quality of input data representation at the end of the learning process; mqe_0 is the mean quantization error of the single neuron of Layer 0, then assign a new SOM at a subsequent layer of the hierarchy. mqe_0 is computed as follows:

$$mqe_0 = \frac{1}{n_I} \sum_{x_j \in I} \left\| x_j - m_0 \right\|, \; n_I = |I| \tag{23}$$

where I is the set of all input vectors. mqe_0 is regarded as a measurement of the overall dissimilarity of input data.

Step 2: Train the SOM with input vectors mapped to this neuron.

4.1.2 Traffic Pattern Matching

Due to the scope of the research exclude nonrecurring congested flow, such as random irregular events as accidents, disabled vehicles, and other special situations. Base on the concept of rolling and only consider the varying of traffic counts but the location of interchange and detect of timing, the study assume all input on ramp traffic patterns can identify to an exclusive cluster.

The pattern matching stage, first employs the average of individual time periods (e.g. t=1,t=2,…) of every pattern produced by *GHSOM* as cluster seed, then calculates the squared Euclidean distance of the input sequence and cluster seed pattern of every clusters. It then assigns the input objects to specific cluster according to the nearest cluster using a squared Euclidean distance measure. For each input pattern F_i, compute its membership $m(C_j|F_i)$ in each cluster C_j. The membership function $m(C_j|F_i)$ defines the proportion of pattern F_i that belongs to the j^{th} cluster C_j. The GHSOM algorithm uses the membership $m(C_j|F_i) \in \{0,1\}$, if the pattern F_i closest to the cluster C_j (minimum squared Euclidean distance), then $m(C_j|F_i)$ =1; otherwise $m(C_j|F_i)$ =0.

The process can be summarized in the following steps, and show as Fig 5:

Step 0: Input a traffic pattern F of r-periods point set.

Step 1: The cluster pattern average of every group produced by GHSOM used as cluster seed.

Step 2: Compute the squared Euclidean distance of each input objects such that

$$Min(F_t, F_{ct}) = \sum_t (f_t - f_{ct})^2 \tag{24}$$

Step 3: Assign the input objects to specific cluster according to minimum squared Euclidean distance.

Step 4: Use the prediction patterns of the specific cluster to predict traffic flow.

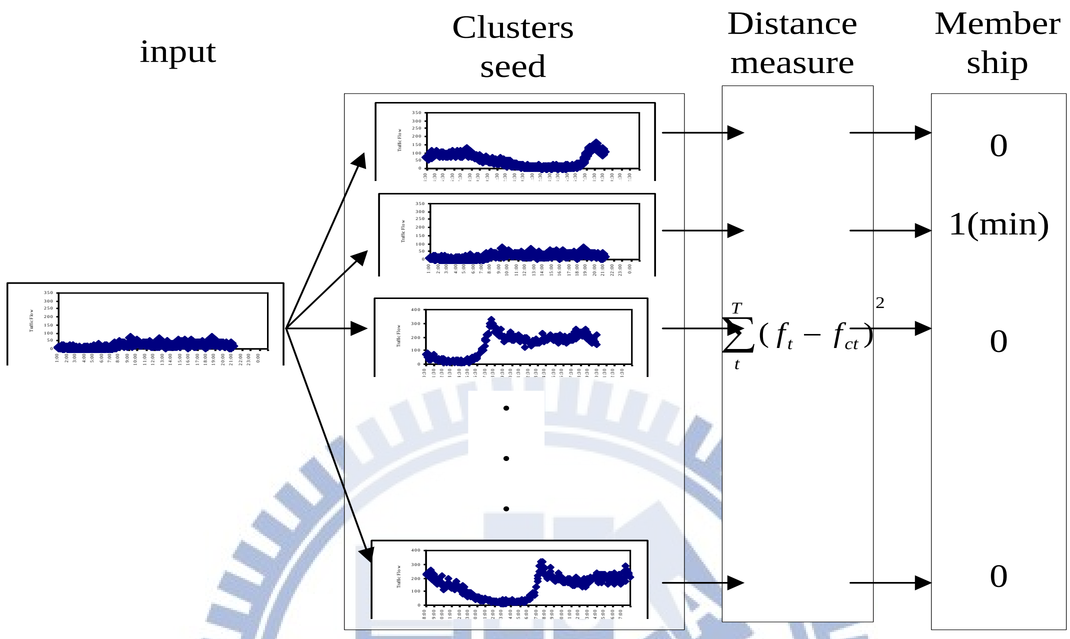


Figure 5 The process of traffic pattern matching

4.1.3 Traffic Prediction: GP

After dividing the traffic patterns into several clusters, a GP traffic prediction model is then developed for each cluster, which predicts h periods ahead based on historical *r* periods.

The GP model is a global optimization algorithm based on the mechanism of natural selection and offspring generation (Koza, 1992). It starts with a population of randomly generated individual trees, each corresponding to a linear combination of traffic flows in the previous periods. Every generated tree is evaluated for its fitness value, which is further utilized for the selection of generated offspring trees.

For ease of explanation, assume there are a total of *I* traffic patterns to be assigned to cluster *l*, each traffic pattern denoted as $X_{li}(1)=[x_{li}(1), x_{li}(2), \ldots, x_{li}(r)]$, i=1, 2,…, *I*. The learning process of GP is described below:

Step 0: Define function set and terminal set. The function set consists of the arithmetic functions of addition, subtraction, multiplication, division, as well as a conditional branching operator. The terminal set is set as the latest *r* periods of traffic flow data.

Step 1: Initialize random population size.

Step 2: Evaluate fitness values of the trees. Randomly select trees from the population, evaluate them with training traffic patterns belonging to this cluster, and then rank them according to their fitness values. A fitness measure is defined as follows:

$$E_{lq} = \sqrt{\frac{\sum_{i=1}^{I}\sum_{t=1}^{h}\left(x_{li}(t+r+1) - f_q(X_{li}(t))\right)^2}{h}} \tag{25}$$

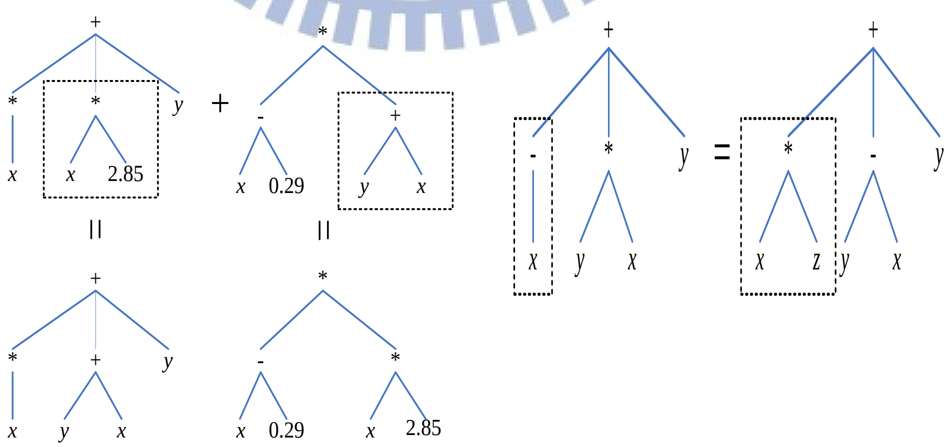
where $f_q(\cdot)$ denotes the mathematical expression of tree q predicting the traffic flow at next time period based on the input historical data at previous time periods, i.e., $f_q(X_{li}(1)) = \hat{x}_{liq}(t+r+1)$.

Step 3: Create new individual by applying genetic operations. The genetic operations further include reproduction, crossover and mutation as follows.

Step 3-1: Reproduction. Replace the least-fit two traffic patterns by the best-fit two.

Step 3-2: Crossover. Create new offspring by randomly combining the chosen parts of two selected trees in each parent tree and swapping the sub-tree rooted at crossover points, illustrated in Fig. 6(a).

Step 3-3: Mutation. Randomly select a mutation point in a tree and substitute the sub-tree rooted there with a randomly generated sub-tree, illustrated in Fig. 6(b).



(a) Crossover of GP (b) Mutation of GP

Figure 6 Crossover and mutation operations of GP

Step 4: If the fitness tends to zero, then stop the procedure. Otherwise, proceed to the next step.

Step 5: Generate new population by using genetic operations, and return to Step 2.

Once a new traffic pattern with r periods is collected, it automatically assigns to the closest cluster, into which all traffic patterns have been classified by GHSOM. The traffic pattern is then fed into the tuned GP model in this cluster to predict the next h periods in a rolling manner.

4.2 Arrival Distribution Modeling: CTM

With the predicted entering traffic over a sufficient long period, the CTM model is used to estimate the arrival patterns of entering traffic at every time click.

The CTM is developed by Daganzo (1994) that conceptual representation of spatial and temporal conditions of traffic fleet formation and relief, assuming homogeneous sections, only one enterance, one exit and no other ramps on the road for vehicle access. CTM is especially suitable for dynamic O-D matrices estimation. The present paper employs CTM to predict the arrival distribution of an O-D pair traffic, which will then be used to compute $\rho_{ij}^{m}(k)$.

4.2.1 Cell-Based Arrival Distribution Modeling

As shown in Figure 7, a freeway is equally discretized into homogeneous sections (cells), numbered consecutively from i = 1 to I starting with the upstream end of the road, where the length of each cell is the distance traveled by a vehicle in one time interval under free-flow traffic. For instance, if set time interval as 6 seconds and free-flow speed as 100 km/hr. The cell length is then determined as 1/6 km.

Origin	cell	cell	cell	…	…	cell	cell	cell	Destination
	i=1	*, 2*	*, …,*		*j*	*, …,*		*I*	

Figure 7 Cell representation of a freeway corridor

In light traffic, all vehicles in a cell can be assumed to advance to the next cell with each interval. It is unnecessary to know where within the cell they are located. Therefore, the system's evolution obeys:

$$n_{i+1}(t+1) = n_i(t) \qquad \text{for } t = 0, 1, 2, \ldots, T \tag{26}$$

Where $n_i(t)$ is the number of vehicles in cell i at time t ; $n_{i+1}(t+1)$ is the number of vehicles in cell *i+1* at time *t+1*.

It is assumed that this equation holds true for all traffic flows unless queuing occurs. When speed reduced due to vehicles entering a bottleneck in the road and queuing occurring, The following two variables are introduced to incorporate queuing in the model, simulating traffic flow variation caused by congestion: (1) Q_i, the maximum flow from cell i – 1 to i during time interval t (when the clock advances from t to t + 1), which also known as "capacity," and is assumed to be a constant under the whole simulation period. (2) $N_i(t)$, the maximum number of vehicles that can be present in cell i in time t. Thus, $N_i(t) – n_i(t)$ is the amount of empty space in cell i at time t.

With these, we define $c_i(t)$ as the number of vehicles that can flow into i for time interval t as:

$$c_i(t) = \min\{n_{i-1}(t)\ ,\ Q_i\ ,\ \frac{w}{v}\,[N_i(t) - n_i(t)]\} \tag{27}$$

The recursive relationship of the CTM model is generated based on the above formula with continuous time-variance, therefore, we can understand the number of vehicles of every cell at every time interval through the recursive mode.

The characteristics of CTM are the number of vehicles on cell at time t is the function of the last cell and time, similar to the LWR fluid model of relationship density and flow rate. Assumes a simplified version of the fundamental diagram, usually based on a Isosceles trapezium form, as shown in Figure 9, and provides simple solutions for realistic networks. It is assumed that a free-flow speed *v* at low densities and a backward shockwave speed *-v* for high densities are constant, the relationship is:

$$\mathrm{q} = \min\{\mathrm{vk}, \mathrm{q}_{\max}, \mathrm{v}(\mathrm{k}_\mathrm{j} - \mathrm{k})\} \qquad \text{for } 0 \le \mathrm{k} \le \mathrm{k}_\mathrm{j} \tag{28}$$

Where v represents free flow speed, k_j: saturation density, and $q_{\max} \le k_j v/2$, maximum flow rate. Four kinds of traffic state and its position in Figure 8, respectively, as follows:

(1) Free flow: speed is v, density is between [0~k_j/3], flow is between [0 ~ q_{max}];

(2) Light Synchronized Flow: speed is between [v/2, v], density is between [k_j/3, 2k_j/3], flow is q_{max};

(3) Heavy Synchronized Flow: speed is between [v/5, v/2], density is between [2k_j

/3, 5 k_j /6], flow is between [q_{max}/2, q_{max}];

(4) Synchronized Flow: speed is between [0, v/5], density is between [5k_j /6, k_j], flow is between [0, q_{max}/2]

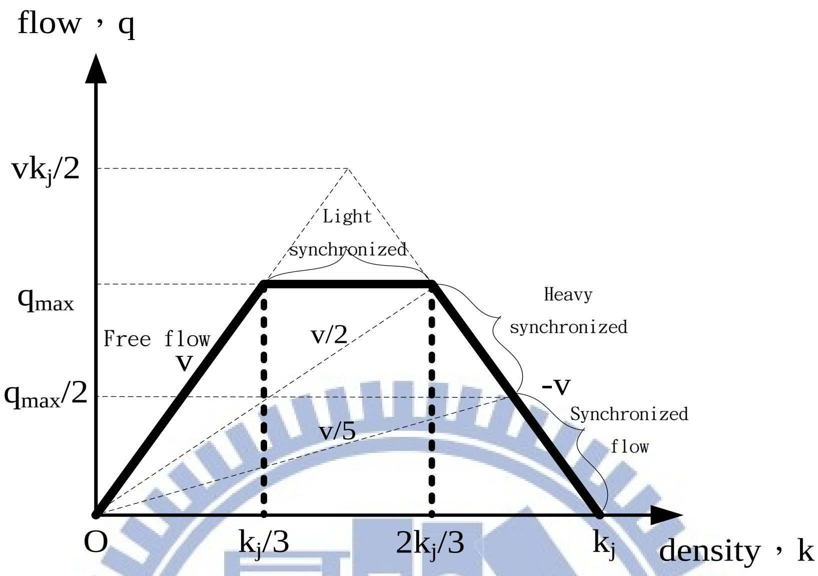


Figure 8 Fundamental diagram of CTM

Based on the above relationship, can finding when queuing occurs; the CTM is based on a recursion where the cell occupancy at time t+1 equals its occupancy at time t, plus its inflow and minus the outflow:

$$n_i(t+1) = n_i(t) + c_i(t) - c_{i+1}(t) \tag{29}$$

If the remaining storage capacity and flow capacity of next cell is sufficient, all vehicles will move forward to the next cell; otherwise, only part of them can move to next cell proportionally, the logic is presented as follows:

$$\begin{aligned} &if \quad c_i(t+1)+q_i(t+1) \le \min[Q_i, N-n_i(t+1)] \\ &then \quad c_{i+1}(t+1) = c_i(t+1)+r_i(t+1) \end{aligned} \tag{30}$$

$$\begin{aligned} &if \quad c_i(t+1)+r_i(t+1) > \min[Q_i, N-n_i(t+1)] \\ &then \quad c_{i+1}(t+1) = 1-\left[\min[Q_i, N-n_i(t+1)] \Big/ c_i(t+1)+r_i(t+1)\right] \end{aligned} \tag{31}$$

4.2.2 Application of vehicle arrival pattern prediction

In this study, the length of time interval is determined by the length of cell and freeway geometric conditions. If time interval is greatly lengthened, it will lead to the cell to be too long to analyze traffic operate behavior between 2 interchanges. However, if cell is too short, the capacity of the cell is insufficient, and will lead to large error due to estimated traffic flow. Therefore, this study in simulation accuracy and time, six seconds is selected as clock tick length, and sensitivity analysis for different time interval was not tested.

The operation concept of the CTM model is to predict vehicle arrival distribution is shown as Figure 9 of the relationship of cell storage, equally discretized into homogeneous sections (cells), numbered consecutively from i to j starting with the upstream end of the road, because each driver and travel time inconsistency resulting in off ramp time variables, probably from m, m+1, ..., M time interval, but the number of vehicles starting with the upstream end of the road not decided by cell i to j but decided by complete road traffic conditions. Therefore, CTM model to predict the arrival distribution ρ_{ij}^{m} is to calculate and record the proportion of the number of vehicles off ramp at every time interval and total number of vehicles.

The variation relationship of actual computing process and the number of vehicles on cell is based on the basic concept of CTM, and then further research into expansion demand. This research takes into account aditional factors for on ramp, off ramp, and main road, to respond to the real behavior of traffic operations and different driver driving behavior leading to different vehicle speed that may cause queuing on road network as the basic concept of CTM only consider linear sections. Therefore, the concept of CTM of the vehicle on the main line, on ramp, and off ramp is shown as Figure 9, with every cell divided into several small cells based on vehicle sources. Each small cell records the number of vehicles that come from different sources q_i at different time intervals. These vehicles operate according to the congestion level of roads and the capacity limitations of cells. For instance, under free flow of traffic, the travel time of vehicles is only possible during one or two periods. However, when queuing occurs on road networks such as synchronized flow, the travel time will vary according to the different reactions of the driver, so the arrival time may be distributed in the k +1, ..., k + m, time periods.

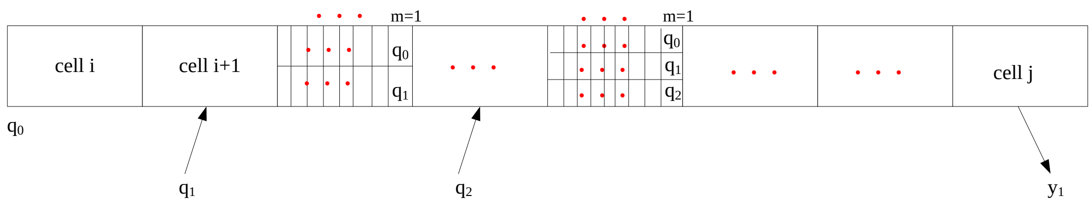


Figure 9 The concept of cell storage of CTM

Eq.(32) shows the cell transmission relationship of vehicles extended by considering the impact of on ramp and off ramp. The research is based on a recursion where the cell occupancy at time t+1 equals its occupancy at time t, plus its inflow and minus the outflow, the following relationship:

$$n_c(k+1) = (n_c(k) - P_c(k)) + Y_c(k) - y_{c+1}(k) \tag{32}$$

$$Y_c(k) = y_c(k) + r_c(k) \tag{33}$$

$$y_c(k) = \min\left\{n_{c-1}(k), Q_c(k), \frac{1}{2}[N_c(k) - n_c(k)]\right\} \tag{34}$$

$$r_c(k) = \min\left\{q_i(k), Q_c(k), \frac{1}{2}[N_c(k) - n_c(k)]\right\} \tag{35}$$

In the equation, $n_c(k+1)$ is the number of vehicles in cell c at time $k+1$; $P_c(k)$ is the number of vehicles off ramp at time k; $Y_c(k)$ is the capacity in cell c at time k; $y_c(k)$ is the number of vehicles of mainline entering cell c at time k; $r_c(k)$ is the number of vehicles on the on ramp entering cell c at time k.

The above relationship is the base concept of CTM and the following is a step by step calculation for CTM:

Step 1: Initialization: setting the length of road, free flow, saturation density, maximum flow, maximum capacity and other relevant parameters, as well as the cell of the on and off ramp, and the flow of the on ramp.

Step 2: Identify the O-D ratio as $b_{ij}(k)$ estimated by EKF.

Step 3: According to eq. (32)-(35), calculate the number of vehicles for each cell at each time interval.

Step 4: Calculate arrival ratio $\rho_{ij}^m(k)$ at the both same starting and different origins according to respective time interval.

4.3 Dynamic OD Matrices Estimation Algorithm: EKF

Based on the predicted entering traffic, arrival patterns, and then proposes an integrated algorithm which combines the CTM with the EKF to respectively and iteratively estimate the arrival distributions and O-D proportions.

Extended Kalman filtering algorithm (EKF) proposed by Kalman (1960) has been widely used in various fields, most commonly in the field of transport to estimate O-D matrices, traffic density and ramp metering. EKF algorithm is an optimal recursive data processing algorithm, which is used to estimate current values of the variables according to system noise statistics, the uncertainty of measurement error, and any available initial conditions of variables.

The main concept of EKF is to estimate the stated variables based on recursive anaysis at given time k periods, the use of the state vector estimate equation to predict the state estimation of the next time $x^-(k)$, and forecast the observation value $\hat{z}(k)$ at time k, later updating the state variable $x^-(k)$ to $x^+(k)$ base on error term of the predictd observing value and actual observing values, and then use $x^+(k)$ to predict the state $x^-(k+1)$ of next period (k+1), and continue doing the recursive algorithm and estimation according to the calculation procedure shown in Figure 10.

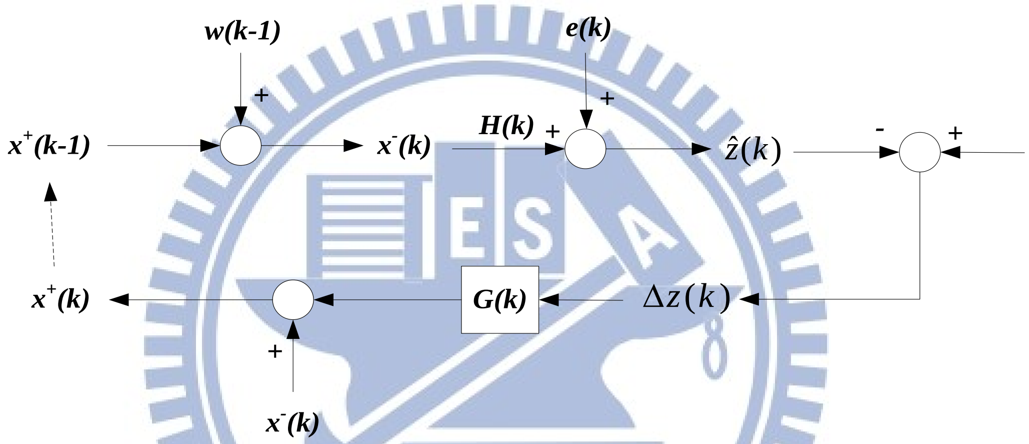


Figure 10 State variables estimate process of EKF algorithm

EKF algorithm is composed by the state equation and measurement equation. State equation is the relationship composed by estimation variables of next time interval, and measured equation is the relationship of the traffic flow of actual section and state variables. In this mode, state variables are $b_{ij}(k)$ in Eq. (36), while the observed value is the relation formed by number of vehicles on ramp $q_i(k)$, the number of vehicles off ramp $y_j(k)$ and the number of vehicles of main road $U_l(k)$. Therefore, the state variables $b_{ij}(k)$ is usually assumed to be random and an independent random walk process, while state equation can be expressed in the following:

$$b_{ij}(k+1) = b_{ij}(k) + w_{ij}(k) \tag{36}$$

Where, $w_{ij}(k)$ is the white noise of the state variable at time k, and $E(w_{ij}(k)) = 0$, $Var(w_{ij}(k)) = D(k)$ follow Gaussian distribution. $\boldsymbol{D(k)} = diag[D_b ,..., D_b]$ is a N (N +1)

/ 2 dimensions covariance matrix of $w_{ij}(k)$.

Measurement equation is the relationship formed by the traffic flow of actual road and main road $y_j(k)$, $q_i(k)$, $U_l(k)$, and expressed as the following formula:

$$z'(k) = H(k)b(k) + e(k) \tag{37}$$

In this equation, z '(k) is the measurement variable, e(k) is a 2N-1 dimensional observation error term, also following Gaussian distribution, and E(e(k)) = 0, Var(e (k)) = R(k). R(k) = diag [r_1 ,..., r_{2N-1}], is a 2N-1 dimensions covariance matrix of e(k), H(k) is a (2N-1)*(N(N+1)/2) dimensions transformation matrix, and H(k) = $H^k = \left[H^k_{rs}\right]_{(2N-1)*N(N+1)/2}$, in which, $\left[H^k_{rs}\right]$ matrix elements are 0, except the following matrix elements.

$$[H^k_{rs}] = \begin{matrix} H^k_{j,Ni+j-i(i+1)/2} = \sum_{m=0}^{M} q_i(k-m)\rho^m_{ij}(k) \quad for\ 0 \le i < j \le N \\ H^k_{N+l,Ni+j-i(i+1)/2} = \sum_{m=0}^{M} q_i(k-m)\rho^m_{il}(k) \quad for\ 0 \le i < l < j \le N \end{matrix} \tag{38}$$

In the above model formulation, the information of each O-D pair can be estimated using the data provided by the surveillance system or historical information, and the unknown set of parameters are O-D proportions, $b_{ij}(k)$.

As used in most existing approaches, the dynamic O-D parameters, $b_{ij}(k)$, are assumed to follow the random walk process between successive time intervals:

$$b_{ij}(k+1) = b_{ij}(k) + w_{ij}(k), \quad 0 \le i < j \le N \tag{39}$$

$$B(k+1) = B(k) + W(k) \tag{40}$$

$$Z(k) = H(k) \cdot B(k) + W(k) \tag{41}$$

$$Z(k) = \left[y_1(k), y_2(k), ..., y_N(k); U_1(k) - q_1(k), ..., U_{N-1}(k) - q_{N-1}(k)\right]^T \tag{42}$$

Where, $w_{ij}(k)$, a random term, is an independent Gaussian white noise sequence with zero mean and its covariance, $Z(k)$, is a column vector, $H(k)$ is a matrix with its entries given by the corresponding coefficients in eqs. (37) and (41), and $e(k)$ is an observation noise vector, which can be defined as a Gaussian white noise with zero mean and its covariance matrix, and R = Var[$e(k)$] = diag[$r_1, \ldots, r_{2N-1}$] is a diagonal positive definite matrix. $B(k)$ is a matrix of the O-D proportions of entering flows $b_{ij}(k)$. $W(k)$ is a matrix of white noise $w_{ij}(k)$.

The proposed estimation algorithm, based on the EKF concept, is presented as follows.

Step 0: Initialization.

Parameters settings include cell length L_i, i = 0,1,. . . , N-1, and time interval, t_0. $\mathrm{var}[e(k)] = diag[r_1, r_2 \ldots]$. $X(0) = E[b(0)]$. $P(0) = Var[b(0)]$. Besides, on-ramp, link and off-ramp flows are given.

Step 1: Determine $\rho_{ij}^m(k)$ by CTM.

Step 2: Compute the linearized transformation matrix based on the determinant $\rho_{ij}^m(k)$.

$$H^{K-1} = \left[H_{rs}^{k-1}\right] \tag{43}$$

$$H_{j,Ni+j-i(i+1)}^{k} = \sum_{m=0}^{M} q_i(k-m) \cdot \rho_{ij}^m(k) \qquad for \quad 0 \le i < j \le N \tag{44}$$

$$H_{N+l,Ni+j-i(i+1)}^{k} = \sum_{m=0}^{M} q_i(k-m) \cdot \rho_{ij}^m(k) \qquad for \quad 0 \le i < j \le N \tag{45}$$

$$\left[H^{K-1}\right] = \left[h_1, h_2, \ldots, h_{2N-1}\right]^T \tag{46}$$

$$Z'(k) = \left[y_1(k), y_2(k), \ldots, y_N(k); U_1(k) - q_1(k), \ldots, U_{N-1}(k) - q_{N-1}(k)\right]^T \tag{47}$$

Step 3: Initialization of the sequential Kalman filtering method.

Set $b_0 = b(k+1)$

$p_0 = p_{k+1} + D$ where $D = \left[d_b, \ldots, d_b\right]$ is a covariance matrix of $W(k)$

Step 4: Sequential Kalman filtering iterations.

For i = 1, 2, …, 2N-1

$$g^i = p^{i-1} h_i^T \left[h_i p^{i-1} h_i^T + r_i\right]^{-1} \tag{48}$$

$$p^i = p^{i-1} - g^i h_i p^{i-1} \tag{49}$$

$$\delta^i = y_i(k) - h_i b(k-1) \tag{50}$$

Truncation:

$$\alpha' = \underset{0 \le \alpha \le 1}{MAX} \left[\alpha \middle| 0 \le \left[b^{i-1} \right] + \alpha \delta^i g^i \le 1 \right] \tag{51}$$

Set $\left[b^i \right] = \left[b^{i-1} \right] + \alpha \delta^i g^i$

Normalization:

For m=1, 2, …, N-2

$$\beta_m = \sum_{j=m+1}^{N} b^i_{mj} \tag{52}$$

$$b^i_{mj} = b^i_{mj} / \beta_m \qquad j=m+1,\ldots,N. \tag{53}$$

Step 5: Stop condition test.

Check the convergence of estimated O-D proportions. If preset stop conditions (convergence level or number of iterations) has not been met, then go to Step 1. Otherwise, go to Step 6.

Step 6: Prediction of the states.

Set $p_k = p^{2N-1}$ and $\left[b(k) \right] = \left[b^{2N-1} \right]$, $k = k + 1$, go to Step 1.

CHAPTER 5 CASE STUDY

To demonstrate the performance and applicability of the proposed approach, a seminal example with six O-D pairs estimation is designed on section 5.1. A medium-scale networks and a large-scale network of on-ramp traffic patterns on a freeway are examined on section 5.2 and 5.3.

5.1 Case Study 1: Seminal Example

5.1.1 Network design and Parameter settings

To demonstrate the performance and applicability of the proposed estimation algorithm, a small closed network is studied. First, a description of the network design and second is data simulation, followed by a verified model by small road network; and then analysis the performance of estimation results, and sensitive analysis of proposed model with different traffic scenarios, to check the accuracy of estimates of different traffic flows and interchanges.

5.1.1.1 Network Design

The model validated with design network in this study, as shown in Figure 11. The total length of section road network is 10,000 m, the main line is 3 lanes and a total of eight nodes, three starting points and three ending points, six O-D pairs, respectively, $b_{1,7}$、$b_{1,8}$、$b_{1,6}$、$b_{7,8}$、$b_{7,6}$、$b_{8,6}$, with the $b_{i,j}$ representing O-D ratio of flow from origin i to destination j and the relationship of road length and cell location shown in Figure 11.

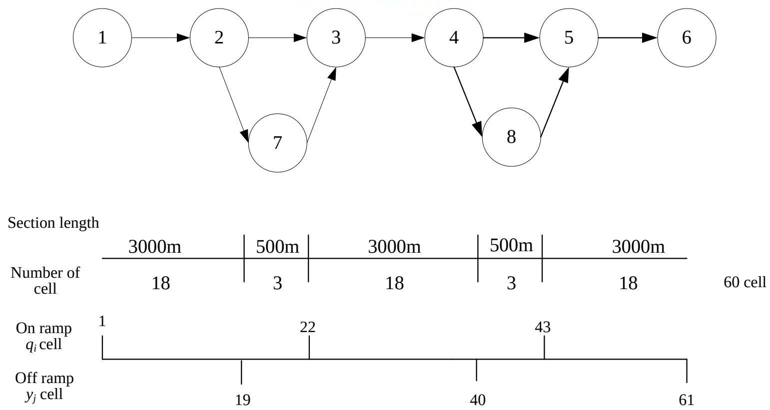


Figure 11 Small design network

5.1.1.2 Parameter settings

The parameters of CTM are set as follow: time interval=6 seconds, free flow speed=100 km/hr, jam density=400 vehicles per kilometre, capacity N =6,000 vehicles per hour, cell storage capability=67 vehicles, maximum flow rate Q = 10 (veh / time click) (equivalent to traffic 6,000 pcu / hr or per lane 2,000 pcu/hr), cell length=1/6 km and assumed the interval on and off ramp of the same interchange is 500 m (3 cell).

5.1.2 Data Simulation

This study used traffic simulation software to generate time-dependent links and flow data, based on this simulation of the initial input data that is required for the estimation model constructed in this study, such as time-dependent traffic flow of main road, on ramp and off ramp, and simulated by traffic simulation software DynaTaiwan, that is a local traffic estimation and prediction system, developed by domestic scholar.

The traffic flow of main roads, on and off ramps is estimated as $b_{ij}(k)$, however, the dynamic O-D matrix of actual road network is difficult to obtain. To generate highly accurate dynamic O-D matrix estimation, transport planning simulation software adapted in this study, generates the traffic flow that conforms to real network conditions.

To generate real time-dependent traffic flows, the time series of 90 minutes O-D traffic under different traffic conditions are given based on the assumed O-D pair flows, DynaTaiwan, traffic simulation software modified from DynaSmart to account for the traffic behaviors in Taiwan, to generating real-time on-ramp, link, and off-ramp traffic flows at every 6-second time interval. The three simulated real-time detected traffic flows are then inputted into the proposed estimation algorithm.

First, the initial data set uses a assumptions of initial O-D information as a basis to co-operate with the attributes of data nodes and line road network on Figure 11, simulating 90 minute peak and off-peak features of traffic demand (Figure 12), to produce time-dependent sections of traffic data, then using the initial data to enforce schema validation.

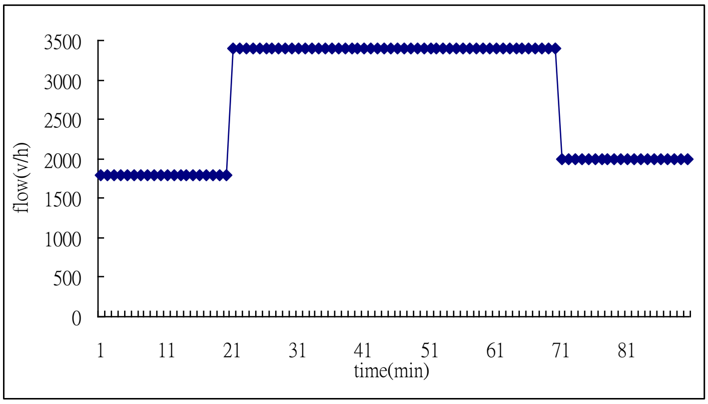


Figure 12 Peak and off-peak features of traffic demand

Both DynaTAIWAN and CTM are mediumscopic traffic models, so the two models have 6-second time periods, however, the EKF model is coordinates with CTM simulation to estimate 5 minutes of dynamic O-D matrices, to facilitate the traffic management personnel to implement appropriate management strategy.

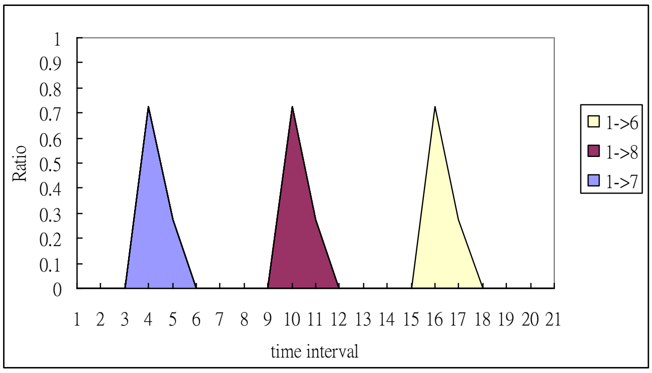
5.1.3 Performance and Estimation Results

5.1.3.1 Traffic Dispersion Phenomenon

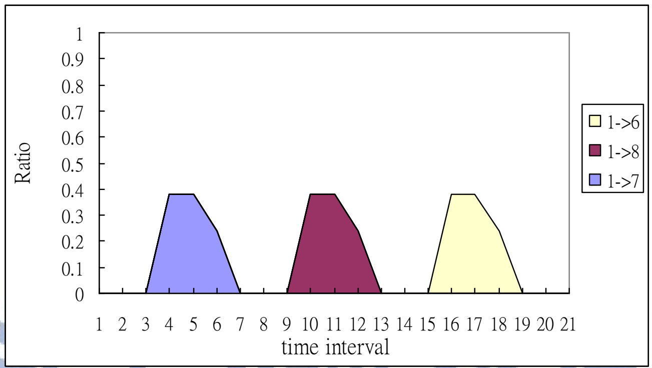
In order to verify the performance of CTM model on the fleet dissipate types in different traffic conditions. Four scenarios with various traffic conditions are simulated, including free-flow, light synchronized flow, heavy synchronized flow and congested flow.

The synchronized flow concept proposed by Kerner and Herrmann (1998), while traffic flow increases, the tiny interference of fleet will result in more than the critical impact of traffic, meaning that, when the traffic flow was gradually increased, the vehicles on the road will be limited to the impact of nearby vehicles, driving not in accordance with the desired speed of travel, and thus result in a synchronization driving behavior with other vehicles, the distribution of vary traffic condition shown in Figure 13.

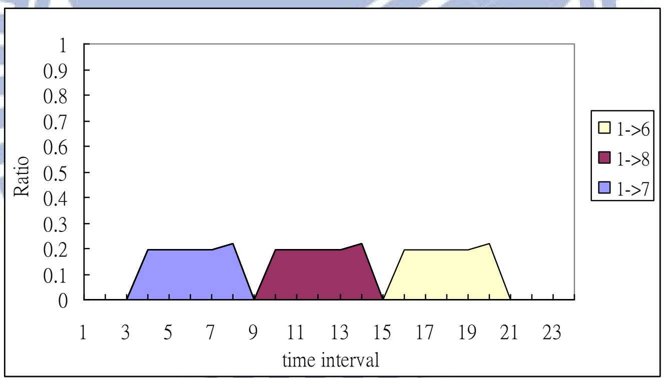
All entering traffic is increased step-by-step by a same ratio ranging from 10% to 100%. The traffic condition of the roadway is determined by its most critical segment (*i.e.* the most congested segment which is usually located at the middle of the roadway). Until the traffic flow at the most critical segment reaches its capacity, it is under free-flow condition. After oversaturation, traffic conditions are equally divided into three traffic phases: light synchronized, heavy synchronized and congested flows. For instance, for the traffic leaving from the same origin 1 and heading for various destinations in time interval t=1, their arrival distributions under various traffic conditions are graphically depicted in Figure 13. As shown in Figure 13 (a), almost all the ranges of arrival times cover only one or two intervals under free-flow condition. Once the traffic flow increases, the degree of traffic dispersion will remarkly appear. As shown in Figures 13(b)-(d), in light synchronized flow, the remaining storage capacity and flow capacity of next cell is sufficient, all vehicles in a cell can be assumed to advance to the next cell with each interval, the same entering traffic will arrive at destination among a wider range of time intervals ranging from two to three time intervals, when the entering traffic increasing to the remaining storage capacity and flow capacity of next cell is not sufficient, only part of them can move proportionally, four to five time intervals under heavy synchronized flow, and six to eight time intervals under congested flow, suggesting the capability of the CTM model in replicating traffic dispersion phenomenon.



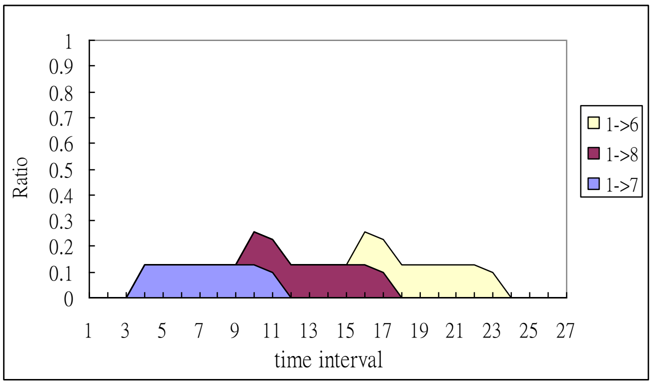
(a) Free-flow



(b) Light synchronized flow



(c) Heavy synchronized flow



(d) Congested flow

Figure 13 Arrival distributions of entering traffic from origin 1 to various destinations (7,8,6)

5.1.3.2 O-D Estimation

1. Performance

The deviation of the estimated O-D proportions of each time interval and each O-D pair from given O-D proportions is used as a measure of model performance, this research is based on Chang and Wu's (1994) model, following and to compare the accuracy of dynamic OD estimation, the root-mean-square error (RMSE) is used to evaluate the performance of the proposed algorithm, which is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sum_{k=1}^{T}(b_{ij}(k)-\hat{b}_{ij}(k))^2}{N(N-1)T}} \tag{54}$$

Where $\hat{b}_{ij}(k)$ is the estimated O-D proportions of traffic entering interchange i and heading to interchange j.

In addition, to analyze the significant statistical performance of these estimated results, this study used the Chi-Square Test as an alternative measure in the following manner.

$$\chi^2 = \sum_{k=1}^{T}\frac{(b_{ij}-b^*_{ij})^2}{b_{ij}} \tag{55}$$

2. Result analysis

To investigate the effects of initial value settings of O-D proportions on the performance of the proposed algorithm, two initial value setting approaches are adopted and compared: randomly generated (RG) approach and equal share (ES) approach. Take origin No.1 interchange as an example, the associated O-D proportions are denoted as $b_{17}(k)$, $b_{18}(k)$, and $b_{16}(k)$. For the RG approach, three random numbers 0.123, 0.341, and 0.782 are generated and then normalized such that the sum of three proportions equals 1. Thus, $b_{17}(k)$=0.099, $b_{18}(k)$=0.274, and $b_{16}(k)$=0.628. In contrast, for the ES approach, three proportions for the same example is simply set as $b_{17}(k)$=0.333, $b_{18}(k)$=0.333, and $b_{16}(k)$=0.334.

The distributions of real b_{18} proportions (from No.1 interchange to No.8 interchange) along with estimated O-D proportions by RG and ES approaches are given in Figure 14. Note that the proposed algorithm can predict real O-D proportions accurately regardless which initial value setting approaches being adopted. However, the predicted result by RG approach is slightly superior to that by ES approach. Thus, the RG approach will be adopted in predicting other O-D proportions.

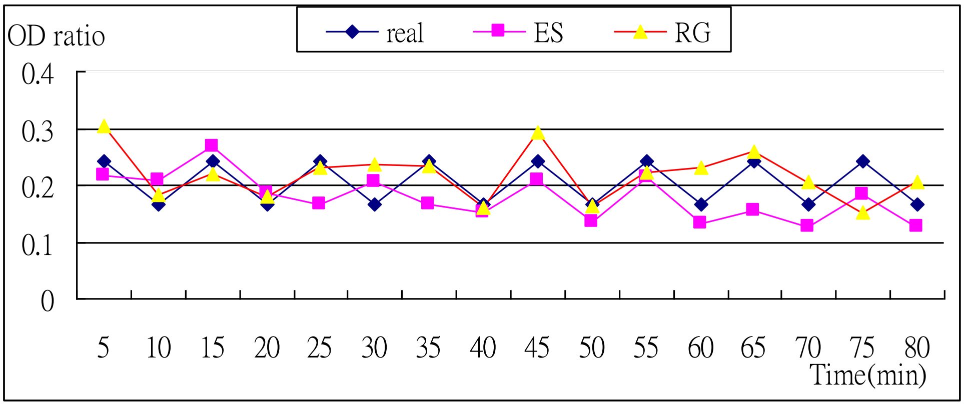


Figure 14 Distributions of real and two predicted b_{18} proportions by EG and ES approaches

Further with $b_{1,8}$, for example as Figure 15, when the 256 time interval, the convergence process of the iteration, can present the convergence results about 64 times recursion. Table 2 further lists the estimation results of mean and standard deviation of the various origins and destinations in each time interval. Know from the table, there is insignificant different between on the mean and standard deviation of the estimates of the two setting ways initial value. The estimation performance of randomly generated initial values is better than others, and average performance indicators RMSE 0.086 and 0.069, respectively. However, according to chi-square test results show that the estimation has not significant differences between two different initial setting for all the origin and destination with the actual value. This result also shows that this study provides models less sensitivity for the initial setting value.

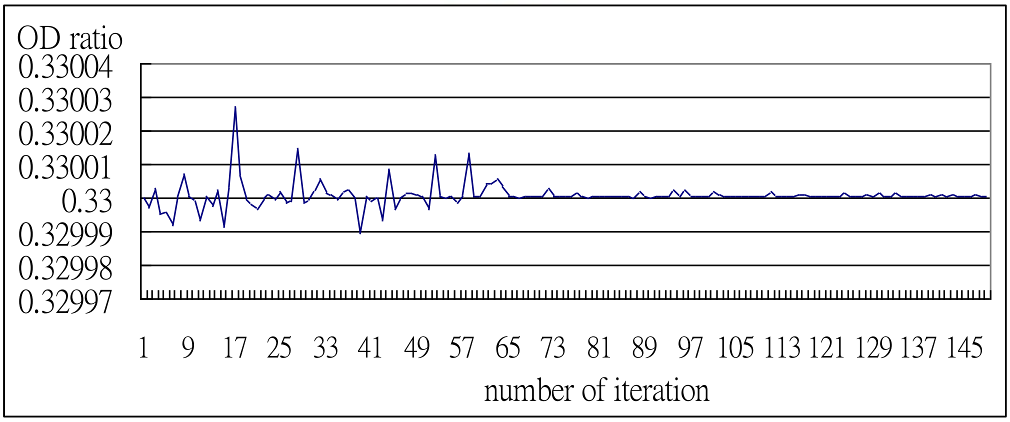


Fig 15 The convergence process of the iteration when the 256 time interval

Table 2 The estimation results of the various origins and destinations

O-D pair	real		ES		RG	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
$b_{1,7}$	0.225	0.049	0.192	0.037	0.216	0.043
$b_{1,8}$	0.324	0.013	0.337	0.084	0.288	0.055
$b_{1,6}$	0.451	0.056	0.471	0.064	0.496	0.062
$b_{7,8}$	0.426	0.027	0.453	0.103	0.45	0.097
$b_{7,6}$	0.574	0.027	0.547	0.103	0.55	0.097
$b_{8,6}$	1	0	1	0	1	0
RMSE	–		0.086		0.069	
χ^2	–		1.89		2.042	

註：$^*\chi^2 > \chi^2_{0.05,14} = 23.685$ 。

5.1.4 Travel Time Analysis

All of driver's driving behavior and speed are not alike, resulting in the travel time in the road network is dissimilar with others, in the past literature of non-assignment mode dynamic OD estimation, not only continued to improve model construction methods , but also to execute the number of surveys and studies for the travel time estimation, such as scholars Chang et al from 1994 until 2007 made a series of studies, according to the pattern established in 1994 to improve the travel time estimation method, gradually increase the accuracy of dynamic OD estimation.

To consider the impact of travel time on the estimated results, this study based on Chang and Wu’s (1994) model, too. However, the model to estimate the travel time is using the average traffic flow of two end points to stand for the road sections traffic flow, and then obtained average travel time, and assuming the distribution during two time period. This way of estimating travel time is too simple to be a true reflection of the vehicle actual operation on road network; base on this, this study uses the mediumscopic traffic model CTM to predict the vehicle arrival patterns.

To compare the accuracy of prediction by CTM model and the results exactitude of dynamic OD estimation, in addition, this study using Greenshields's speed-density model to estimate travel time base on the assumption that the same as Chang and Wu’s (1994) arrive distribution within two time interval, where, the speed-density models as:

$$u = u_f (1 - k / k_f) \quad (56)$$

Based on the the road network and verify situation of above simple example, to explore several different methods of estimating the travel time on many time interval, first, can be seen from Figure 16, the distribution of different sections of the traffic flow, and the flow to inputting the Greenshields's speed-density model to estimate the speed and then deduce the road travel time, with estimation results shown in Figure 17, using Greenshields model to calculate travel time for the O-D pairs, and distributed during two time periods, where the horizontal axis expresses time interval, and the vertical axis represents road traffic.

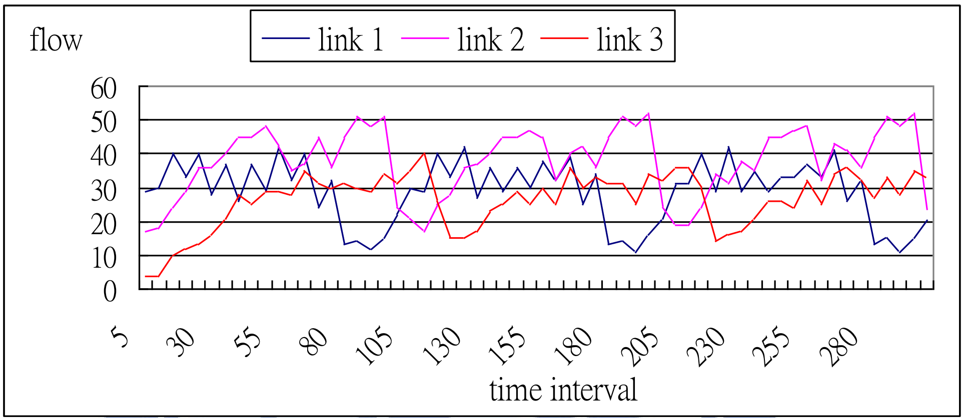


Figure 16 Distribution of different sections of the traffic flow

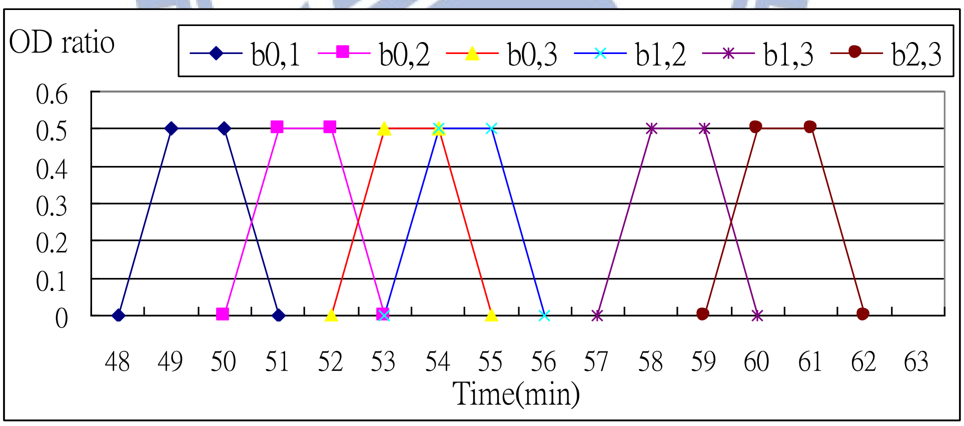


Figure 17 Distribution of the O-D pairs calculated by Greenshields model

To predict vehicles arrival distribution through O-D matrices, $b_{1,8}$, represented by the illustration, in this case, traffic flow is less at the first 20 minutes, so the results are more accurately estimated. While the traffic flow gradually increases to show congested conditions, it appears more unreasonable that the travel time calculated by Greenshields model and the assumption of arrival distribution within two time intervals. Therefore, shown as Figure 18, the estimate results will be distorted when the back time interval, leading to an overall average of RMSE up to 0.145, is much higher than the estimated of performance (RMSE = 0.069) provided by this study.

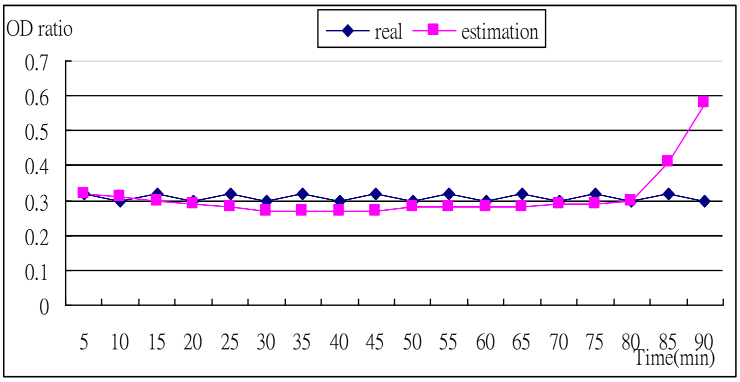


Figure 18 Distort of the distribution on O-D pairs calculated by Greenshields model

From the above results based on Chang and Wu (1994) assumption that the vehicles travel time on the road network not over 2 time intervals, using Greenshields to estimate the average travel time in off-peak traffic demand to keep a certain degree of accuracy. However, during peak traffic demand, the estimated travel time and arrival patterns, allow results to vary greatly with actual traffic behavior, resulting in the estimation error of the O-D ratio becoming larger. It can be seen that this study predicts vehicle arrival distribution by CTM model, more conforming to vehicle operating behavior on real road networks and more accurately estimating dynamic O-D matrices.

5.1.5 Sensitivity Analysis

In order to understand the performance and limitations of this study, a small network was used to execute the model's sensitivity analysis. First, review whether the model is suitable for congestion on actual networks, and base demand varying on networks to adjust the network's traffic patterns, to enforce model sensitivity analysis. In addition, test the impact that different ratio restrictions on ramp for vehicles arrival distribution and check the changes in estimation on CTM models.

In this section, Figure 11, the small network is used to understand the impact of model estimation for different traffic demand on the network. Assuming the main line road network has three lanes, the speed is 100KPH, and the on and off ramp speed is 40KPH, with a simulation time interval of 6-seconds. To estimate the O-D proportion during 90-minutes, others parameters are set as above, and then execute sensitivity analysis based on different traffic flow, free flow, light synchronized flow, heavy synchronization flow and congested flow, with the results shown in Figure 19 (represented by b_{76})

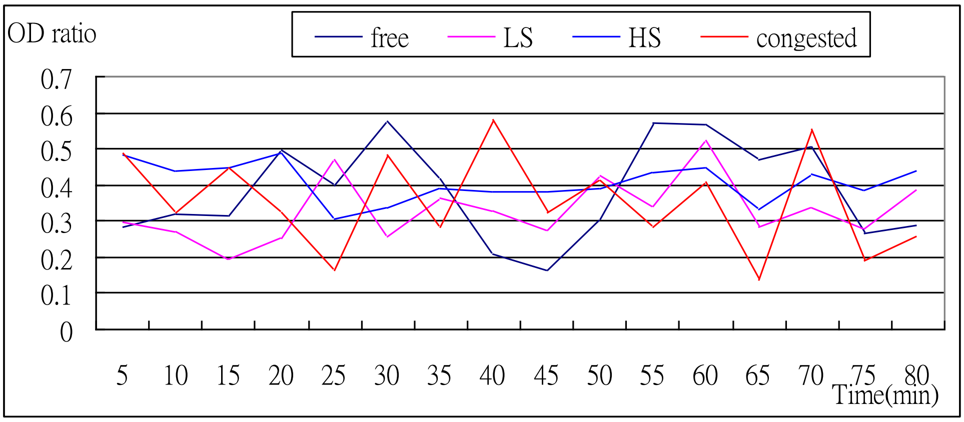


Figure 19 O-D estimation under different traffic flow

Table 3 is the comparison table of the actual value with estimate results under different traffic flows. On the table, we can observe the standard deviation and average are not largely different under two different initial value sets. Performance indicators of RMSE are about 0.08, following the traffic flow increase, its estimated accuracy will be slightly reduced. In Chi-square test, assuming 95% confidence interval, all the O-D pairs with actual values are not significantly different statistically.

Table 3 Estimation comparison table of actual values with different traffic flow

O-D pairs	actual values		Free flow		light synchronized flow		heavy synchronization flow		congested flow	
	Average	standard deviation	Average	standard deviation	Average	standard deviation	Average	standard deviation	Average	standard deviation
b_{17}	0.225	0.049	0.224	0.061	0.19	0.056	0.249	0.031	0.229	0.029
b_{18}	0.324	0.013	0.353	0.079	0.326	0.085	0.276	0.072	0.283	0.028
b_{16}	0.451	0.056	0.423	0.06	0.483	0.065	0.475	0.072	0.488	0.051
b_{78}	0.426	0.027	0.389	0.11	0.431	0.121	0.426	0.114	0.328	0.055
b_{76}	0.574	0.027	0.611	0.11	0.569	0.121	0.574	0.114	0.672	0.055
b_{86}	1	0	1	0	1	0	1	0	1	0
RMSE			0.08		0.082		0.083		0.069	
χ^2			2.811		2.623		2.341		1.709	

ps： $^*\chi^2 > \chi^2_{0.05,16} = 26.296$ 。

5.2 Case Study 2: A Medium-scale Network

5.2.1 The Study Corridor and Parameter settings

5.2.1.1 The Study Corridor

This study used a section of Taiwan No.1 Freeway from Taishan toll station to Yangmei toll station, to demonstrate the performance and applicability of the proposed estimation algorithm. This is a 36 km three-lane freeway section with 6 interchanges, which in order include, Linkou, Taoyuan, Neili, Jhongli, Youth and Yangmei Interchange, with a total of 28 O-D pairs, as shown in Figure 20.

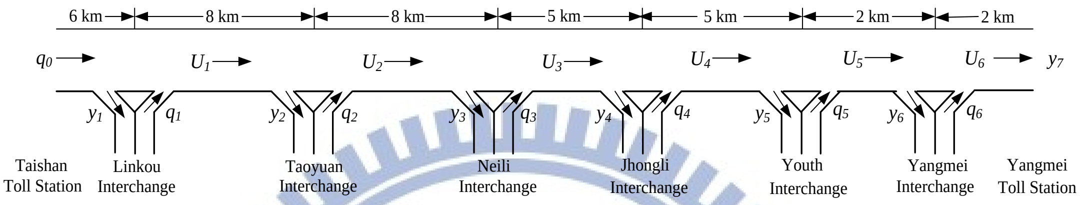


Figure 20 The case study of Taiwan No.1 freeway northern section

5.2.1.2 Parameter settings

To show the capability of CTM in replicating the traffic hydrodynamics and to investigate the degree of traffic dispersion under various traffic conditions, a simulation on the above-mentioned three-lane freeway section with 6 interchanges is conducted. Parameters are set as follows: free flow speed=100 km/hr, jam density=400 vehicles per kilometre, capacity=6,000 vehicles per hour, cell storage capability=67 vehicles, time interval=6 seconds, and cell length=1/6 km.

5.2.2 Data Collection and Traffic Simulation

5.2.2.1 Data Collection

Peak and off-peak ratio of full day trip demand on actual Taiwan freeway is represented, and as much as possible meets the actual road traffic flow. The full day trip flow data of freeway is based on the full day trip demand survey and estimate by Dui-Ji Chen (2010).

The survey results of full day traffic information, includes the traffic flow of Taoyuan Interchange going south is the largest traffic flow among the six interchanges, and the traffic flow of Neili Interchange going south is the lowest. In order to understand the proportion of traffic flow for every interchange per hour accounting for traffic flow of full day traffic counts. To generate the ratio of traffic counts of each-hour accounting for the traffic flow of full-day based on the sub-periods flow conversion by detector data collection shown in Figure 21. The ratio of morning peak 7-9 accounts for full-time traffic about 0.168; the ratio of flow largest at 15-17 is about 0.153.

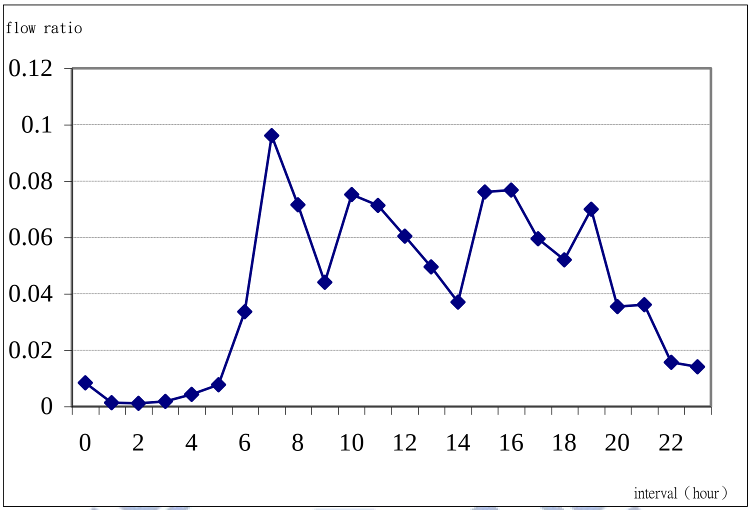


Figure 21 The proportion of traffic per hour as calculated by Dui-Ji Chen (2010).

5.2.2.2 Traffic Simulation

To generate highly accurate dynamic O-D matrix estimation, transport planning simulation software DynaTaiwan, adapted in this study, and generates the traffic flow real-time on-ramp, link, and off-ramp traffic flows at every 6-second time interval that conforms to real network conditions, the time series of four hours O-D traffic under different traffic conditions are given based on the assumed O-D pair flows. The three simulated real-time detected traffic flows are then inputted into the proposed estimation algorithm.

As above section, the initial data set uses an assumptions of initial O-D information as a basis to co-operate with the attributes of data nodes and line road network on Figure 22, simulating 4 hours peak and off-peak features of traffic demand (Figure 23), to produce time-dependent sections of traffic data, then using the initial data to enforce schema validation.

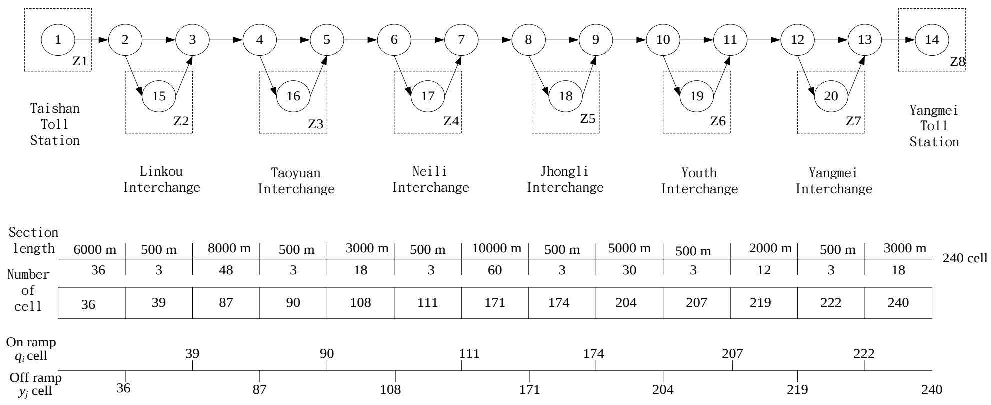


Figure 22 Taishan to Yangmei toll network cell map

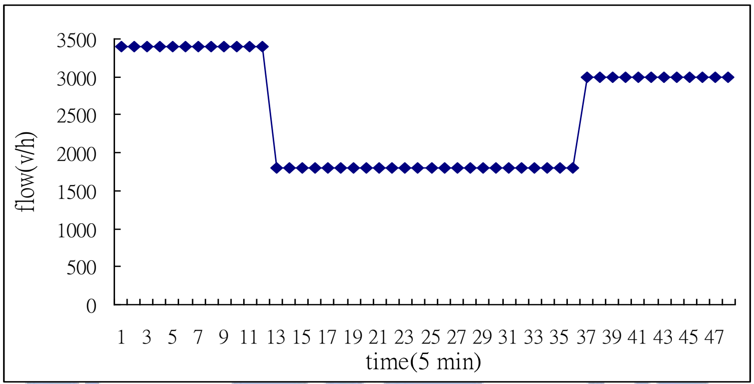


Figure 23 peak and off-peak features of traffic demand

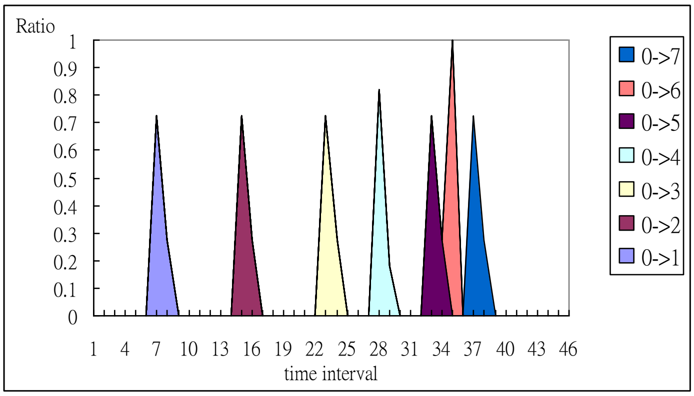
As the section 5.1, the two models have 6-second time periods, however, the EKF model is coordinates with CTM simulation to estimate 5 minutes of dynamic O-D matrices, to facilitate the traffic management personnel to implement appropriate management strategy.

5.2.3 Performance and Estimation Results

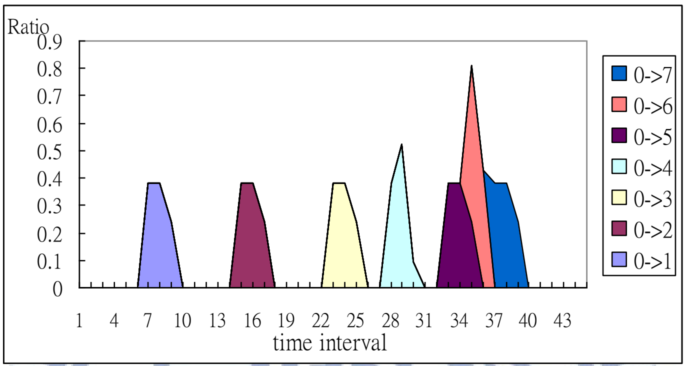
5.2.3.1Traffic Dispersion Phenomenon

As the section 5.1, all entering traffic is increased step-by-step by a same ratio ranging from 10% to 100%. The traffic condition of the roadway is determined by its most critical segment. Until the traffic flow at the most critical segment reaches its capacity, it is under free-flow condition. After oversaturation, divided into three traffic phases: light synchronized, heavy synchronized, and congested flows. For example, for the traffic leaving from the same origin 0 (Taishan Toll station) and heading for various destinations (Linkou interchange to Yangmei Toll station) in time interval t=1, their arrival distributions under various traffic conditions are graphically depicted in

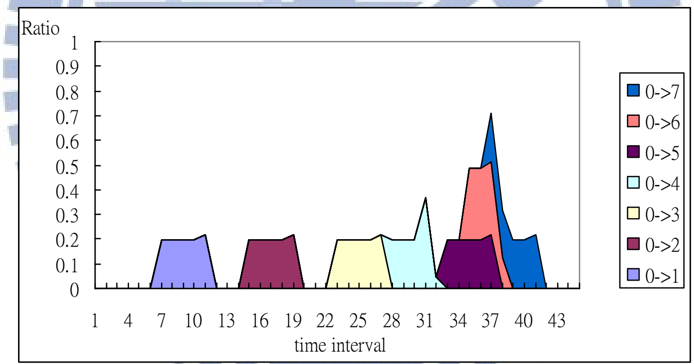
Figure 24. Figure 24 (a), the ranges of arrival times cover only one or two intervals under free-flow condition. As shown in Figures 24(b), in light synchronized flow, the same entering traffic will arrive at destination among a wider range of time intervals ranging from two to three time intervals. Figures 24(c), four to five time intervals under heavy synchronized flow, and Figures 24(d), six to eight time intervals under congested flow.



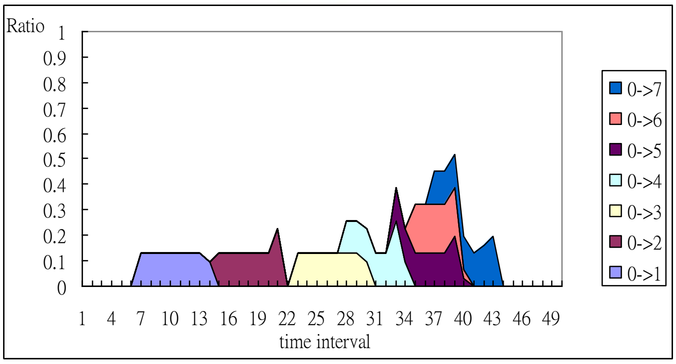
(a) Free-flow



(b) Light synchronized flow



(c) Heavy synchronized flow



(d) Congested flow

Figure 24 Arrival distributions of entering traffic from origin 0 to various destinations (1~7)

5.2.3.2 O-D Estimation

1. Performance

To measure the performance of the model, this study used the deviation of the estimated O-D proportions of each time interval and each O-D pair from given O-D proportions. The root-mean-square error (RMSE) is used to evaluate the performance of the proposed algorithm, which is defined as Eq. 54.

2. Result analysis

To compare the effects of initial value settings of O-D proportions on the performance of the proposed algorithm, as the section 5.1, two initial value setting approaches are adopted and compared: randomly generated (RG) approach and equal share (ES) approach. Take origin No.4 interchange as an example, the associated O-D proportions are denoted as $b_{45}(k)$, $b_{46}(k)$, and $b_{47}(k)$. For the RG approach, three random numbers 0.123, 0.341, and 0.782 are generated and then normalized such that the sum of three proportions equals 1. Thus, $b_{45}(k)$=0.099, $b_{46}(k)$=0.274, and $b_{47}(k)$=0.628. In contrast, for the ES approach, three proportions for the same example is simply set as $b_{45}(k)$=0.333, $b_{46}(k)$=0.333, and $b_{47}(k)$=0.334.

The distributions of real b_{15} proportions (from Linkou interchange to Youth interchange) along with estimated O-D proportions by RG and ES approaches are given in Figure 25, the RG approach will be adopted in predicting other O-D proportions.

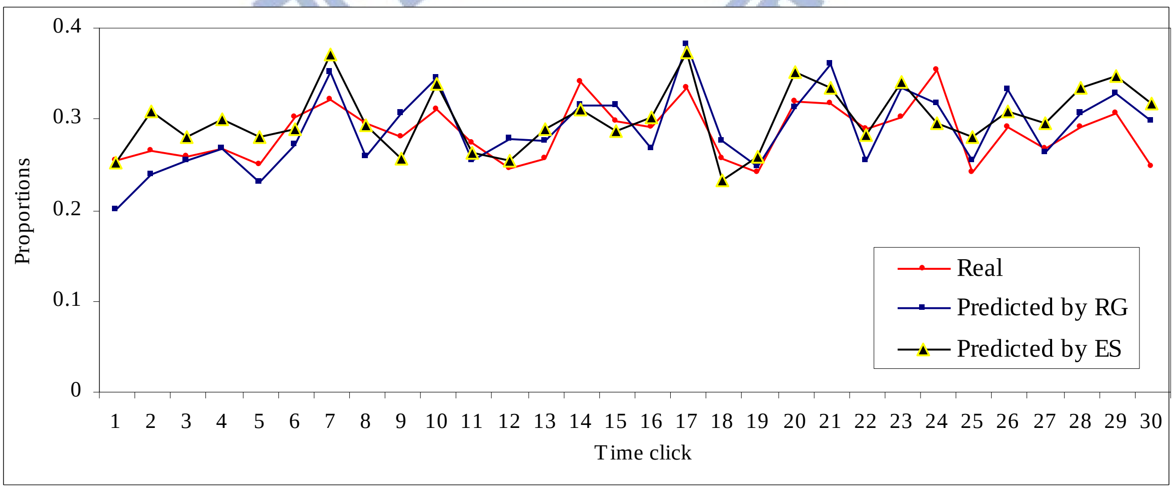


Figure 25 Distributions of real and two predicted b_{15} proportions by EG and ES approaches

Table 4 presents the *RMSE* values for the 28 O-D proportions. The results show that the overall average *RMSE* is 0.0443, indicating a good fitness and practical applicability of the proposed algorithm.

Table 4 RMSE values for 28 O-D proportions of the proposed integrated algorithm.

To From	Linkou interchange	Taoyuan interchange	Neili interchange	Jhongli interchange	Youth interchange	Yangmei interchange	Yangmei toll station
Taishan toll station	0.032724	0.010306	0.056023	0.038756	0.046837	0.055235	0.006009
Linkou interchange	-	0.064495	0.010955	0.036353	0.065464	0.013809	0.039796
Taoyuan interchange	-	-	0.055185	0.028696	0.067517	0.031295	0.054170
Neili interchange	-	-	-	0.063319	0.033135	0.038757	0.066709
Jhongli interchange	-	-	-	-	0.048676	0.046625	0.079683
Youth interchange	-	-	-	-	-	0.075176	0.075176
Yangmei interchange	-	-	-	-	-	-	0

5.3 Case Study 3: A Large-scale Network

5.3.1 The Study Corridor and Parameter settings

5.3.1.1 The Study Corridor

In order to demonstrate the performance and applicability of the proposed estimation algorithm, a section of Taiwan No.1 Freeway from Toufen Interchange to Beidou Interchange is studied. This is a 110 km three-lane freeway section with 15 interchanges, which in order include, Toufen, Miaoli, Sanyi, Houli, Taichung systematic, Fengyuan, Daya, Taichung, Nantun, Wangtain, Chunghua systematic, Chunghua, Puyian systematic,Yanlin and Beidou Interchange with a total of 136 O-D pairs, as shown in Figure 26.

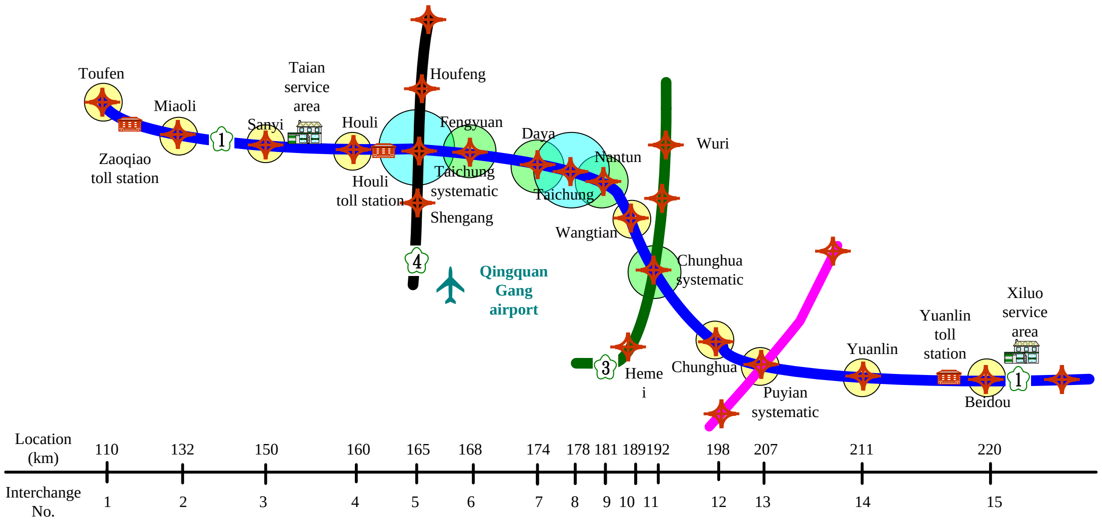


Figure 26 The case study of Taiwan No.1 freeway northern section

5.3.1.2 Parameter settings

1. GHSOM

The two important parameters of GHSOM are set as: τ_1 = 0.85 and τ_2 = 0.0035. In addition, both of the learning rate function $\alpha(l)$ and the neighborhood function $\sigma(l)$ are set as linear, monotonically decreasing over iterations. The parameters setting of GHSOM are refering to Rauber et al. (2002), and sensitivity analysis for different numbers was not tested.

2. GP

Table 5 Parameter settings for GP

Parameter	Setting
Fitness	Mean square error
Terminal set	$x(t)$, $x(t$-1),…, $x(t$-240) and random number b
Function set	+, -, ×
Population size	50
Reproduction rate	0.08
Crossover rate	0.60
Mutation rate	0.01
Initial minimum depth	2
Number of generations	300
Initialization method	Direct method

The parameters of GP model are detailed in Table 5. To avoid producing too complicated traffic prediction function, only three operators (+, - and ×) are considered in this study. Terminal set contains the traffic flow data in the previous 240 time intervals with a randomly generated number b. The parameters setting of GP are refer to Yao and Lin (2009), and sensitivity analysis for different numbers was not tested.

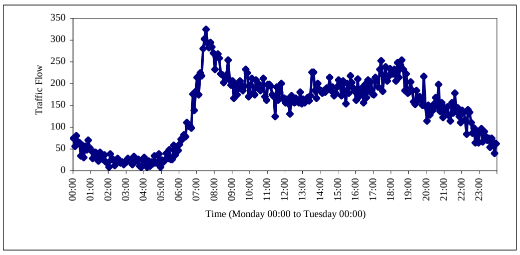
3. CTM

Used a simulation on the above-mentioned three-lane freeway section with 15 interchanges, show the capability of CTM in replicating the traffic hydrodynamics and to investigate the degree of traffic dispersion under various traffic conditions. Parameters are set as follows: free flow speed=100 km/hr, jam density=400 vehicles per kilometre, capacity=6,000 vehicles per hour, cell storage capability=67 vehicles, time interval=6 seconds, and cell length=1/6 km.

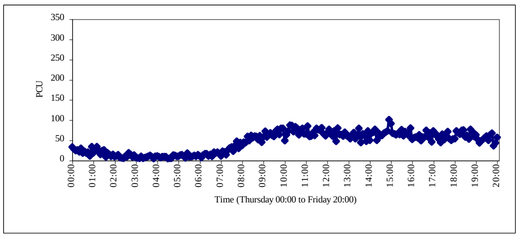
5.3.2 Data Collection

The five-minute on-ramp traffic flow data at 15 interchanges from Toufen interchange to Beidou interchange, a 110-kilometer stretch of Taiwan No.1 Freeway (Fig. 26), over a week from May 25th to May 31st (Monday to Sunday) 2009 were used for the case study. At each interchange, southbound traffic flows were first aggregated from different ramps. The traffic pattern is composed of 288 consecutive five-minute traffic flow data (24 hours). There were 2,016 time intervals in a week, thus can form 1,729 (=2,016-287) traffic patterns at each interchange. A total of 25,935 (=1,729×15) traffic patterns have been generated for the entire study corridor.

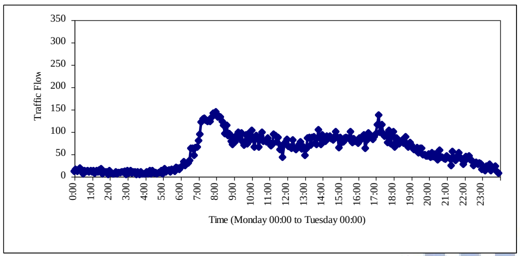
With five-minute time interval, the proposed method aims to predict the next 48 time intervals (4 hours) based on previous 240 time intervals (20 hours). In other words, the previous 240 traffic flow data are used to determine the closest cluster and then to feed into the corresponding tuned GP model to predict the traffic flow at the next 48 time intervals in a rolling manner. Looking into an example of traffic patterns during the same periods (Monday 00:00am to Tuesday 00:00am) at different interchanges, Fig 27 shows that the traffic patterns remarkably differ from each other. For instance, Taichung and Taichung system interchanges do exhibit significant peak and off-peak traffic patterns, but Wangtian and Fengyuan interchanges do not.



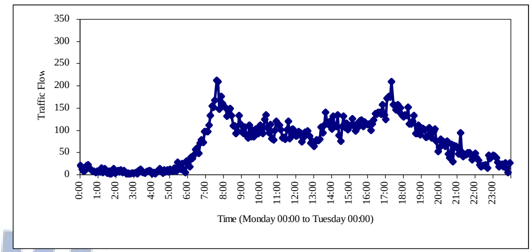
(a) Taichung interchange



(b) Wangtian interchange



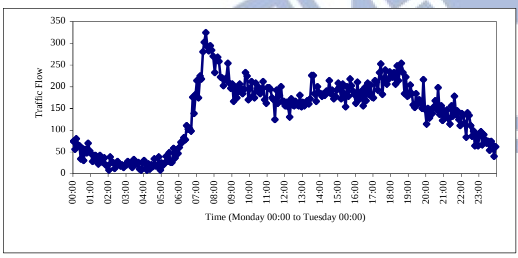
(c) Fengyuan Interchange



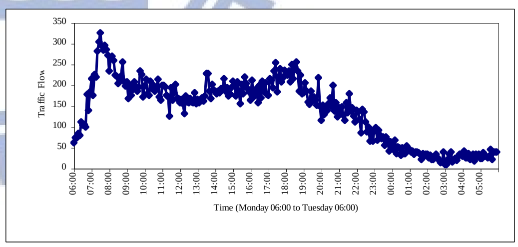
(d) Taichung system Interchange

Figure 27 Traffic patterns at different interchanges (Wednesday 00:00 to Thursday 00:00)

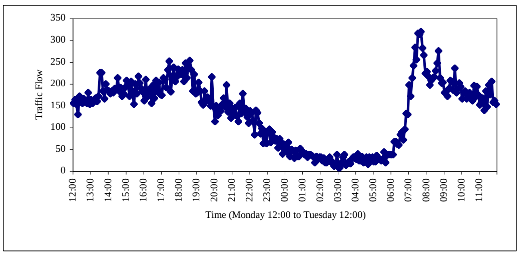
A more detailed traffic patterns at Taichung interchange during different time periods are further illustrated in Fig. 28, which reveals that the traffic patterns at a specific location are also remarkly different, but similar patterns may repeat themselves over different time periods.



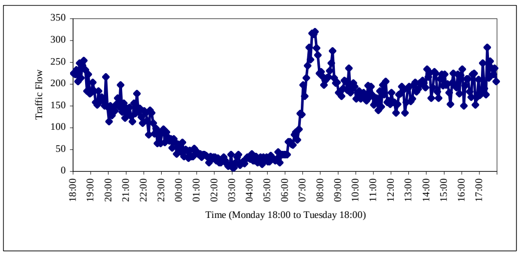
(a) Monday 00:00 to Tuesday 00:00



(b) Monday 06:00 to Tuesday 06:00



(c) Monday12:00 to Tuesday 12:00



(d) Monday 18:00 to Tuesday 18:00

Figure 28 Traffic patterns at different time periods (Taichung interchange)

To train and validate the proposed model, the traffic patterns are randomly divided into two sets: training set (18,155 traffic patterns) and validation set (7,780 traffic patterns) at a ratio of 7:3. The parameter settings and the results of clustering and prediction are presented below.

5.3.3 Performance and Estimation Results

5.3.3.1Traffic patterns clusters

The illustration of the GHSOM architecture, as Fig 29, a total of 3 layer, layer 1 contain 6 clusters, layer 2 contain 20 clusters, layer 3 contain 22 clusters, 36 different clusters have been identified by GHSOM.

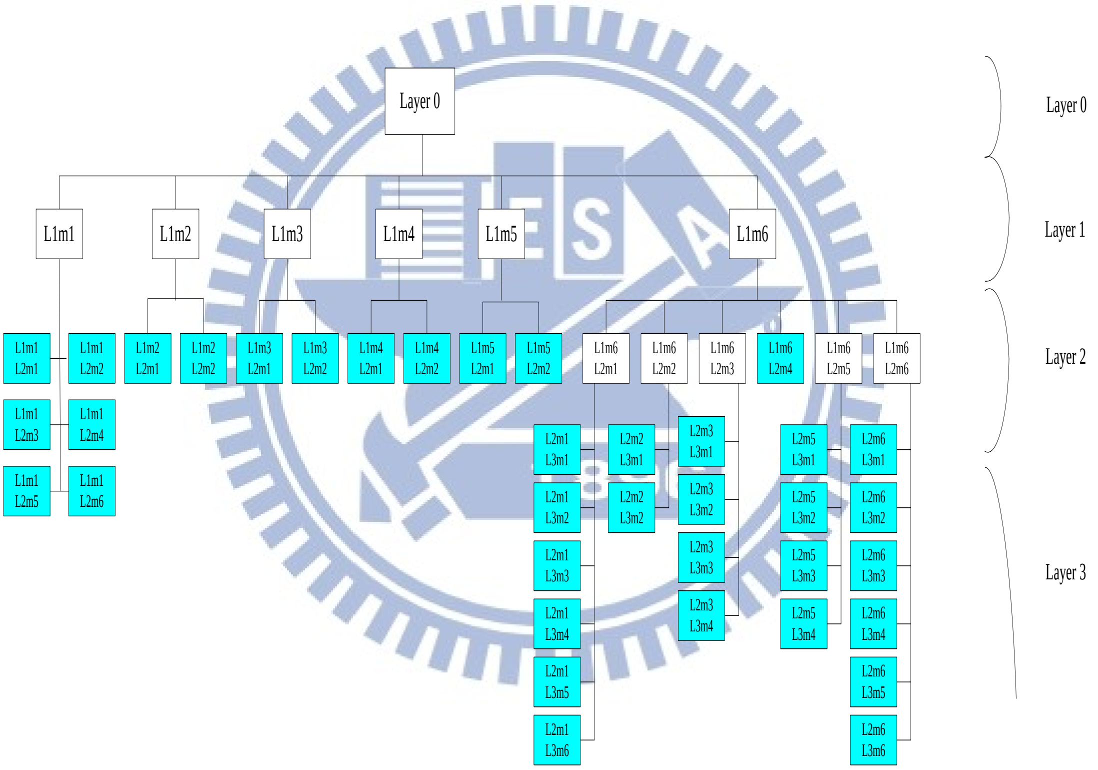


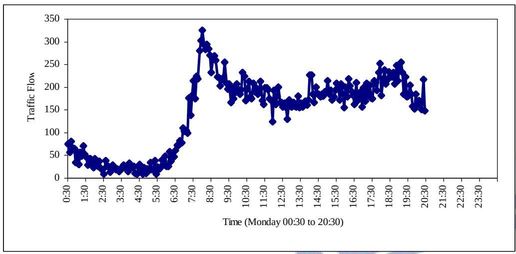
Figure 29 The illustration of the GHSOM architecture

Note that the number of traffic patterns in each of the 36 clusters ranges from 143 (Cluster 1) to 2,488 (Cluster 6), detailed in Table 6. Such self-structured traffic patterns will be used for prediction in the GP model.

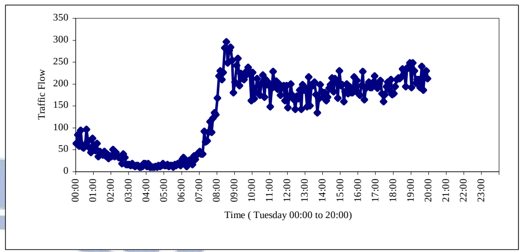
Table 6 Clustering results of traffic patterns

Cluster	Number of patterns	Cluster	Number of patterns
1	143	19	155
2	381	20	214
3	429	21	429
4	1238	22	607
5	440	23	1381
6	2488	24	429
7	155	25	155
8	155	26	238
9	179	27	429
10	405	28	464
11	1298	29	381
12	1238	30	345
13	155	31	155
14	167	32	167
15	464	33	214
16	417	34	179
17	1333	35	321
18	321	36	488

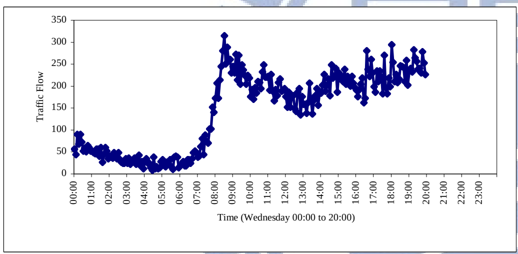
To display the similarity of traffic patterns in the same cluster, the traffic patterns in Cluster 1 (urban area), Cluster 15 (suburban area) and Cluster 30 (rural area) are demonstrated in Fig. 30 through Fig. 32, respectively. To avoid lengthy discussion, we only present four randomly selected traffic patterns from each of these three clusters. In Fig. 30, Cluster 1 contains traffic patterns starting from 00:00 to 20:00 on weekdays in the urban area (e.g. Taichung interchange and Taichung system interchange) where maximum five-minute flow rates can exceed 300 pcu.



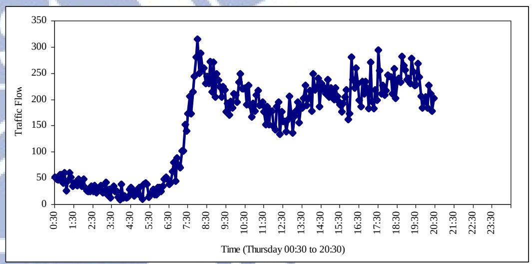
(a) Taichung interchange (Monday 00:30 to Monday 20:30)



(b) Taichung interchange (Tuesday 00:00 to Tuesday 20:00)



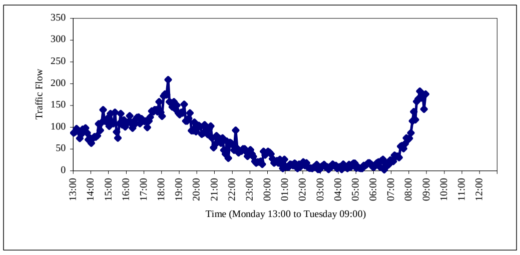
(c) Taichung interchange (Wednesday 00:00 to 20:00)



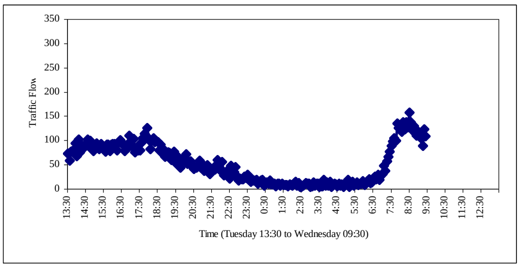
(d) Taichung interchange (Thursday 00:30 to 20:30)

Figure 30 Four randomly selected traffic patterns from Cluster 1 (urban area)

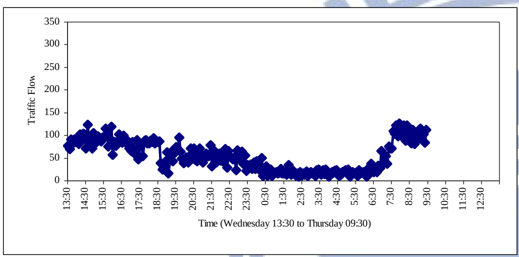
In Fig. 31, Cluster 15 contains traffic patterns starting from 13:00 to 09:00 on weekdays in the suburban area (e.g. Chunghua, Fengyuan, Daya, and Nantun) where most of five-minute flow rates are below 150 pcu. It is obvious that the peak and off-peak phenomena of the traffic patterns in Cluster 15 are not as sharp as those in Cluster 1.



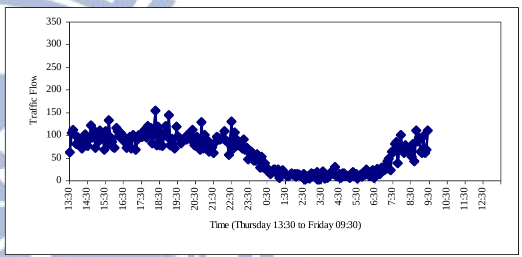
(a) Chunghua interchange (Monday 13:00 to Tuesday 9:00)



(b) Fengyuan interchange (Tuesday 13:30 to Wednesday 9:30)



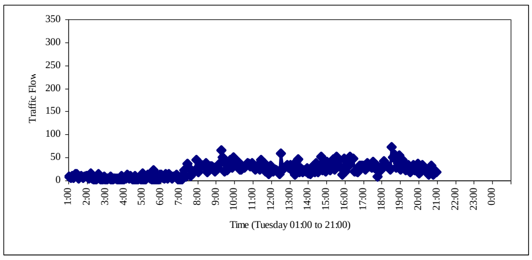
(c) Daya interchange (Wednesday 13:30 to Thursday 9:30)



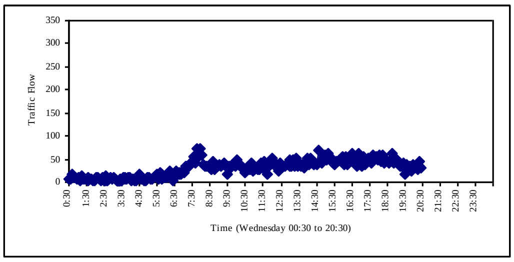
(d) Nantun interchange (Thursday 13:00 to Friday 9:00)

Figure 31 Four randomly selected traffic patterns from Cluster 15 (suburban area)

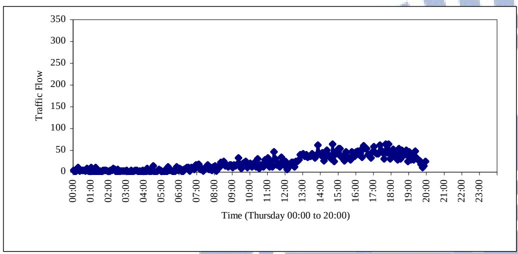
In Fig. 32, Cluster 30 contains traffic patterns starting from 00:00 to 20:00 on weekdays or weekends in the rural area (e.g. Toufen, Miaoli, Sanyi, and Wangtian Interchanges) where most of five-minute flow rates are lower than 50 pcu. No significant peak and off-peak phenomena can be identified.



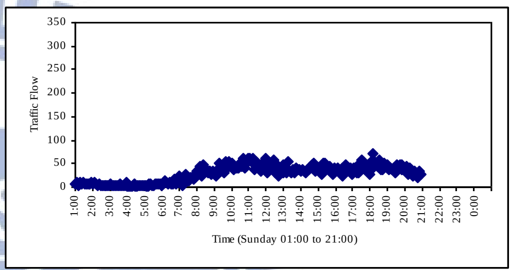
(a) Toufen interchange (Tuesday 01:00 to 21:00)



(b) Miaoli interchange (Wednesday 00:30 to 20:30)



(c) Sanyi interchange (Tuesday 00:00 to 20:00)



(d) Wangtian interchange (Sunday 01:00 to 21:00)

Figure 32 Four randomly selected traffic patterns from Cluster 30 (rural area)

Based on the clustering results, the traffic patterns in the same cluster are similar and those in different clusters are remarkly dissimilar, suggesting the correctness of our clustering model.

5.3.3.2 Traffic Flow Prediction

1. GP traffic prediction models

On the self-structured traffic patterns associated in the 36 clusters, a total of 36 GP traffic prediction models are further developed—one prediction model for each cluster. For brevity, we only demonstrate three clusters (1, 15, and 30). In Cluster 1 where 143 traffic patterns are contained. We randomly divide these 143 patterns into two sets: training set (100 patterns) and validation set (43 patterns). Based on the training traffic patterns, the GP model for Cluster 1 is tuned as follows:

$$x(t+1)=1.01x(t)+5.41\times10^{-6}x(t-1)x(t-2)x(t-7)-1.75\times10^{-6}x(t-1)x(t-4)^{2}-2.02\times10^{-8}x(t)^{3}x(t-6)+4.05\times10^{-8}x(t-1)x(t-5)^{3} \quad (57)$$

According to Eq.(57), only the traffic flow data at previous seven time intervals are required to predict the traffic flow of next time interval. For instance, to predict the traffic flow rate in Fig. 33(a) at time interval (t+1), say, 20:05, we need to input seven detected flow rates at $x(t\text{-}7)$=19:25, $x(t\text{-}6)$=19:30, $x(t\text{-}5)$=19:35, $x(t\text{-}4)$=19:40, $x(t\text{-}2)$=19:50, $x(t\text{-}1)$=19:55, and $x(t)$=20:00. The traffic flow rate at 20:05 can therefore be calculated as $x(t+1)$ according to Eq.(57). To predict in a rolling manner for the next time interval (t+2), the six detected flow data from 19:30 to 20:00 together with the above predicted traffic flow at 20:05 are inputted into Eq.(57) to calculate the traffic flow $x(t+2)$ at 20:10. This process continues until all of the traffic flow data for the next four hours (48 time intervals) have been obtained.

Following the same vein, the GP models for Cluster 15 and Cluster 30 are tuned as below, respectively:

$$\begin{aligned} x(t+1)=\,&1.0483x(t)-0.0942x(t-2)+0.0031x(t-1)x(t-2)-4.897\times10^{-5}x(t-5)x(t-7)^2-3.733\times10^{-7}x(t)^2x(t-1)x(t-2)\\ &+3.165\times10^{-7}x(t-1)x(t-2)x(t-5)x(t-7)+1.255\times10^{-7}x(t-4)x(t-7)^3+4.97\times10^{-8}x(t)x(t-2)x(t-7)^2\\ &+1.038\times10^{-7}x(t-1)x(t-4)x(t-5)x(t-6) \end{aligned} \tag{58}$$

$$\begin{aligned} x(t+1)=\,&0.8301x(t)+0.166x(t-2)+0.0195x(t)x(t-1)-0.0181x(t)x(t-2)-1.24\times10^{-6}x(t-6)x(t-7)^2\\ &-4.53\times10^{-6}x(t)x(t-1)x(t-2)^2-3.81\times10^{-6}x(t-1)^2x(t-5)x(t-6)-3.91\times10^{-6}x(t-1)x(t-5)x(t-6)^2\\ &-3.55\times10^{-6}x(t)^2x(t-3)x(t-6)+7.51\times10^{-6}x(t-2)^2x(t-5)x(t-6)+4.16\times10^{-6}x(t)^2x(t-1)x(t-3)\\ &+3.74\times10^{-6}x(t-1)^2x(t-6)^2 \end{aligned} \tag{59}$$

According to Eqs.(58) and (59), Clusters 15 and 30 require the inputs of traffic flow data at previous six and seven time intervals, respectively. Our results show that the 36 GP traffic prediction models require the inputs of traffic flow data at most the previous twelve time intervals.

2. Performance

The mean absolute percentage error (MAPE) is used to evaluate the performance of the proposed method:

$$MAPE=\frac{1}{T\times J}\sum_{j=1}^{J}\sum_{t=1}^{T}\left|\frac{x_j(t)-\hat{x}_j(t)}{x_j(t)}\right| \tag{60}$$

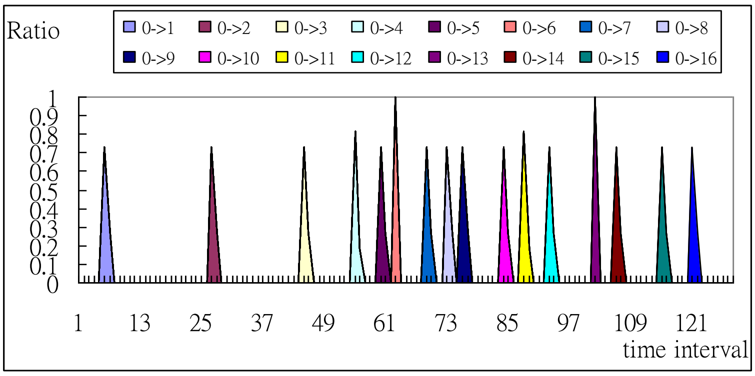
Where $x_j(t)$ and $\hat{x}_j(t)$ are the real and predicted traffic flow at time interval t at interchange j; T is the total prediction time intervals; J is the total number of interchanges in the study corridor (T=48 and J=15).

Our results show that the *MAPE* values of Clusters 1, 15 and 30 are 5.10%, 4.85% and 5.03% for training and 10.15%, 7.18% and 8.17% for validation, respectively. Of the total 36 clusters, the average training and validation *MAPE* values are 4.58% and 10.07%, respectively. It suggests a satisfactory prediction accuracy of the proposed method.

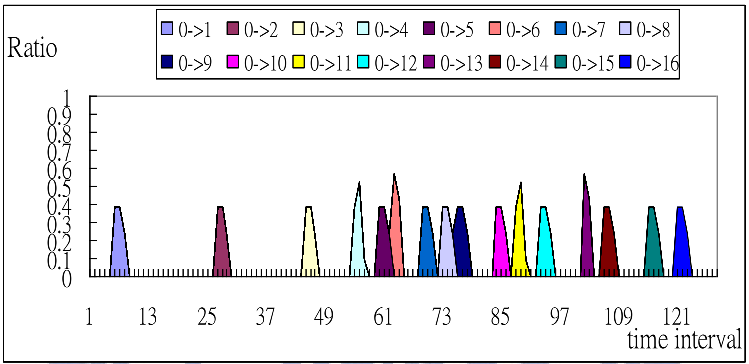
5.3.3.3Traffic Dispersion Phenomenon

As section 5.1, different scenarios with four traffic conditions, including free-flow, lightly synchronized flow, heavily synchronized flow, and congested flow, are simulated. And entering traffic flows are increased step-by-step by the same ratio from 10% to 100%. Under free-flow condition, almost all ranges of arrival times cover only one or two intervals. Under lightly synchronized flow condition, the remaining storage capacity and flow capacity of the next cell are sufficient; hence all vehicles in a cell can advance into the next cell within each interval. As the traffic keeps increasing, the degree of traffic dispersion will become significantly. Under heavily synchronized flow condition, the entering traffic has increased over the remained storage capacity, flow capacity of the next cell is not sufficient; thus, only a portion of them can move forward proportionally. Under congested flow, the arrival times of traffic dispersion can be as long as six to eight time intervals.

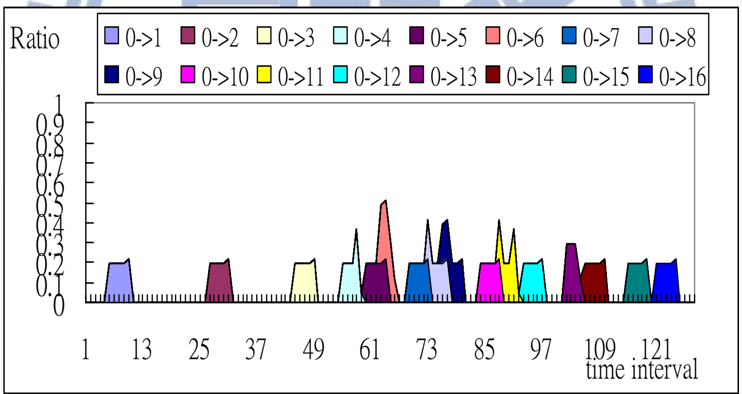
For instance, for the traffic leaving from the same origin 0 (Toufen) and heading for various destinations (Toufen interchange to Beidou interchange) in time interval t=1, their arrival distributions under various traffic conditions are graphically depicted in Figure 33. As shown in Figure 33 (a), the ranges of arrival times cover only one or two intervals under free-flow condition. Once the traffic flow increases, as shown in Figures 33(b), in light synchronized flow, the time intervals ranging from two to three time intervals. As Figures 33(c), four to five time intervals under heavy synchronized flow, and as Figures 33(d) six to eight time intervals under congested flow.



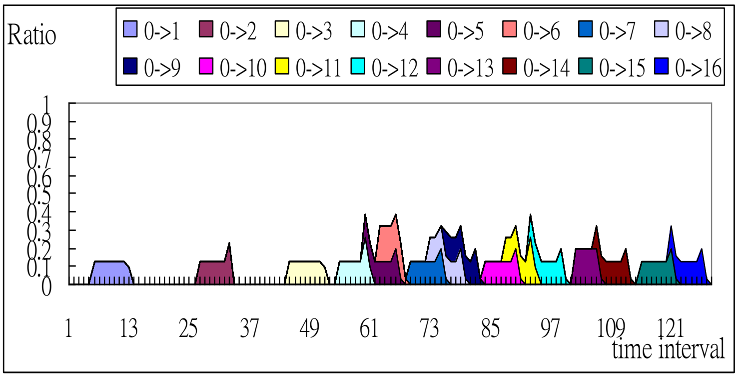
(a) Free-flow



(b) Light synchronized flow



(c) Heavy synchronized flow



(d) Congested flow

Figure 33 Arrival distributions of entering traffic from origin 0 to various destinations (1~16)

5.3.3.4 O-D Estimation

1. Performance

To measure the performance of the model, estimated O-D proportions of each time interval and each O-D pair from given O-D proportions. The root-mean-square error (RMSE) is used to evaluate the performance of the proposed algorithm, which is defined as Eq.54.

2. Result analysis

The results show that the overall *RMSE* values for the 136 O-D proportions is 0.1043, indicating a rather good fitness of the proposed approach. Figure 34 displays the process of convergence for the time interval k=986, $b_{1,15}$, Toufen Interchange to Beidou Interchange.

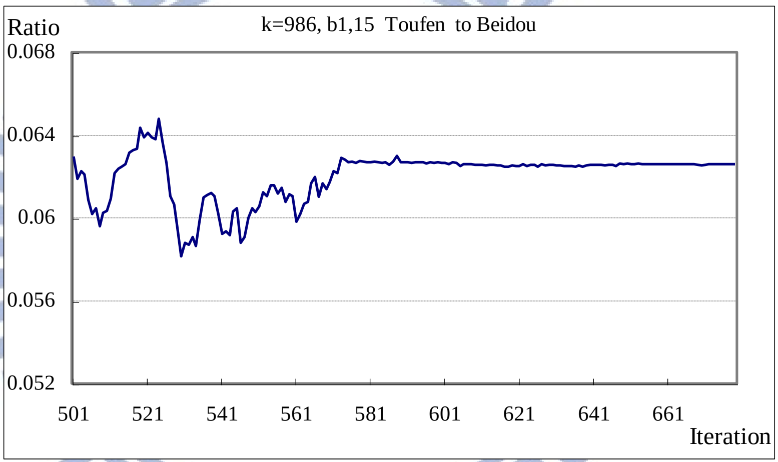


Figure 34 The process of convergence

5.3.4 Sensitivity Analysis

Our proposed method heavily depends on the real-time fed-in traffic flow data, which are used for determining which clusters it belongs to, and furthermore, for predicting the traffic flows at the next 48 time intervals. In the previous settings, the traffic patterns have been defined as traffic flow data at consecutive 240 time intervals, or 20 hours. In the following, a sensitivity analysis with different traffic pattern lengths: 48 (4 hours), 72 (6 hours), 96 (8 hours), 120 (10 hours), 144 (12 hours), 192 (18 hours) and 240 (20 hours) time intervals is further conducted. The MAPE values for training and validation are presented in Table 7. We note that shorter lengths (*e.g.*, *L*=48 and 72) have relatively lower prediction accuracy than longer ones, suggesting the necessity to input a sufficient long traffic pattern for both pattern recognition and prediction. It is also interesting to note that there are no significant changes in prediction accuracy once the length of traffic patterns is longer than 120 time intervals.

Table 7 The MAPE values with different traffic pattern lengths

Lengths	Training	Validation
48	7.29%	19.72%
72	7.78%	14.74%
96	5.32%	12.01%
120	5.72%	10.34%
144	5.91%	10.86%
192	5.63%	10.38%
240	4.58%	10.07%

5.3.5 Comparison

To show the superior performance of the proposed method, a commonly-used traffic prediction model—the autoregressive integrated moving average (ARIMA) model is further developed for comparison. Following the same data basis as the proposed method, the ARIMA model is also developed on the previous 240 time intervals and predicts the following 48 time intervals. Taking Cluster 1 as an example, its corresponding ARIMA model can be calibrated as follows:

$$(1-B)Z_t = 1.18 + \frac{(1+0.606B)}{(1+0.710B)}a_t + \varepsilon_t \tag{61}$$

Where the three important parameters of ARIMA are set as: $p=1$, $d=1$ and $q=1$. Z_t and ε_t are the actual value and random error at time period t, respectively, B is backward difference operator. The *MAPE* values of training and validation datasets are 21.77% and 28.65%, respectively, which are much higher than those of our proposed method based on Eq.(60), which are 4.58% and 10.07%, respectively.

With the self-structured traffic patterns, a simplified prediction model was to be naively developed by averaging the traffic flow data at the last 48 time intervals. For example, if one traffic pattern is of interest in Cluster 1 where 143 traffic patterns have been identified, then the traffic flows at the next 48 time intervals are predicted by taking the average traffic flow of 143 traffic patterns. The *MAPE* values of training and validation datasets for all clusters with this simplified model are 25.21% and 32.73%, respectively, which are much higher than those of our proposed method

(4.58% and 10.07%). Again, this comparison further confirms the superiority of the proposed method and it suggests the necessity of GP model.

CHAPTER 6 CONCLUDING REMARKS

This research proposes a two-stage prediction model with an integrated algorithm to estimate dynamic O-D matrices. The contributions and findings were concluded in Section 6.1. Recommendations for further research were addressed in Section 6.2.

6.1 Conclusions

The conclusions in this study are summarized in the following points:

1. This study has developed an integrated estimation algorithm by combining cell transmission model (CTM) and extended Kalman filtering (EKF) to respectively and iteratively estimate the arrival distributions and the O-D proportions. Our proposed model intends not only to result in a substantial increase of system observability with significantly less parameters than those in literature, but also to contribute enhancing the quality of dynamic O-D matrices estimation.

2. According to field observation, daily traffic patterns do repeat spatially and temporally over and over again. This research proposes a two-stage prediction model, employs the GHSOM model to partition unlabeled traffic patterns into appropriate number of clusters and then develops a GP model associated with each cluster to predict the traffic features based on rolling self-structured traffic patterns, to enhance the prediction performance to accurately predict the traffic features in a rolling manner for a medium-to-long term traffic.

3. A case study is undertaken on a 110-kilometer freeway stretch with 15 interchanges. The historical 240 five-minute southbound traffic flows (20 hours) at each interchange are used to determine the closest cluster and then fed into the corresponding GP model to predict the five-minute traffic flows for the next 48 time intervals (4 hours). The results show that the proposed method have achieved relatively high prediction accuracy in urban area, suburban area and rural area interchange (average *MAPE*= 4.58% in training and 10.07% in validation, respectively). In addition, the proposed method has performed much better than the conventional ARIMA model.

4. The results from a case study on Taiwan’s freeway have also shown that the CTM can satisfactorily capture the traffic dispersion under various traffic conditions, free flow, light synchronized flow, heavy synchronization flow and congested flow and the proposed algorithm can accurately estimate the O-D proportions with a

rather low *RMSE*, indicating the practical applicability of the proposed algorithm.

5. To compare the performance of this mode of travel time prediction to the Greenshields macroscopic model to predict vehicle travel time and assume that vehicles will enter the road network within the scope of the two time step until the point is reached. The results showed that the predicted results in this mode RMSE of 6.9% than Greenshields of 14.5%.

6.2 Recommendations

Several directions for future research can be identified.

1. About the field study:

The proposed algorithm is only valid for the case of linear freeway corridor, the applicability and efficiency of the proposed algorithm can be extended to a large scale network, and also can be incorporated route choice behaviors to elaborate the application to the complicated networks.

And, to demonstrate the efficiency of the proposed algorithm, four different traffic flow, free flow, light synchronized flow, heavy synchronization flow and congested flow was used to execute the model's sensitivity analysis. However, the proposed algorithm is only sensitivity analysis based on recurring congestion, the future research can be extended to nonrecurring congested flow, for instance random irregular events as accidents, disabled vehicles, and other special situations.

2. About the data source:

Due to data availability in the case study, the O-D matrices are arbitrarily given and then used to generate "real time" detected traffic flows by traffic simulation software. With advanced traffic surveillance technologies, however, it is feasible to collect real-time traffic information to further examine the applicability of the proposed algorithm.

3. About traffic predict:

Firstly, the proposed method can be readily applied to predict the traffic flows in a larger freeway network with more time intervals ahead. Secondly, it is interesting to compare the prediction accuracy made by the proposed method with other methods such as genetic clustering model (GCM), artificial neural network (ANN), and support vector machine (SVM). Finally yet importantly, incorporating our proposed method with dynamic O-D matrices information to predict other traffic features (e.g., travel times for various O-D pairs) over a relatively long freeway corridor can be very useful for developing advanced traveler information systems, which calls for further exploration.

4. Additional information:

Future study can incorporate the innovative technique to automatically record and match the license plate numbers of passing vehicles so as to determine the partial trails of the vehicles such as License plate recognition (LPR), other logical assumptions such as route choice behaviors or user equilibrium, into more complex techniques such as generalized least squares or bi-level programming, so as to further improve the accuracy of O-D matrix estimation.

5. Comparison with other algorithms:

A comparison of the proposed approach can be made with other existent O-D estimation algorithms to demonstrate the superiority of different algorithms, and further to explore the characteristic features of varying on results and parameter.

6. The disadvantage of performance indicator:

To measure the performance of O-D estimation, the root-mean-square error (RMSE) is used to evaluate the performance of the proposed algorithm of the research. However, the O-D proportions is a minimal value while the number of the interchange increasing, will result in the square of the deviation of the estimated O-D proportions of each time interval and each O-D pair from given O-D proportions to tend to zero, others indicator will be consider to evaluate the performance of the O-D estimation on future, for instance mean absolute percentage error (MAPE).

REFERENCES

曾國雄、卓訓榮、周幼珍、江勁毅 (1997),「動態流量推估動態 O-D 方法之研究」, *運輸計劃季刊*，第二十六卷第四期，頁 615-638。

陳敦基 (2010)，高速公路按里程電子收費通行費率之研究，交通部臺灣區國道高速公路局委託研究計畫。

邱裕鈞、藍武王、許珮珊、曾群明(2011)，「應用格位傳送模式建構高速公路動態起迄矩陣推估演算法」，運輸學刊，第二十三卷第一期， 頁 97-128。

Afzal, W., Torkar, R. (2011) On the application of genetic programming for software engineering predictivemodeling: A systematic review. Expert Systems with Applications 38, 11984 – 11997.

Asa, B.H., David, H., Hava, T. I., Vapnik, V. (2001) Support vector clustering. Journal of Machine Learning Research 21, 25-137.

Ashok, K., Ben-Akiva, M. (2000) Alternative approaches for real-time estimation and prediction of time-dependent original-destination flows. Transportation Science 34, 21-36.

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

Barlović, R., Santen, L., Schadschneider, A., Schreckenberg, M. (1998) Metastable states in cellular automata for traffic flow. European Physical Journal 5B, 793–800.

Baykasog ̌lu, A., Llu, H.G., Anakc,H.C., zbakr, L.O. (2008) Prediction of compressive and tensile strength of limestone via genetic programming. Expert Systems with Applications 35, 111-123.

Bell, M.G.H. (1983) The estimation of an origin-destination matrix from traffic counts. Transportation Science 17, 198-217.

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

Bick, J. H., Newll, G.F. (1960) A continuum model for two-directional traffic flow. Applied Mathematics 18, 191-204.

Boel, R., Mihaylova, L. (2006) A compositional stochastic model for real time freeway traffic simulation. Transportation Research 40B, 319-334.

Brezocnik, M., Kovacic, M., Ficko, M. (2004) Prediction of surface roughness with genetic programming. Journal of Materials Processing Technology 157, 28–36.

Brucker, P., (1978) On the complexity of clustering problems. In Optimization and Operation Research, 45-54.

Chandler, R.E., Herman, R., Montroll, E. (1958) Traffic dynamics: Studies in car following. Operations Research 6, 165–184.

Chang, G.L., Wu, J. (1994) Recursive estimation of time-varying origin-destination distributions with dynamic screenline flows. Transportation Research 34B, 277-290.

Chang, G.L., Tao, X. (1996) Estimation of dynamic network O-D distribution. Proceedings of 13th Symposium on Transportation and Traffic Flow Theory, Elsevier Science.

Chang, G.L., Miaou, S.P. (1999) Real-time prediction of traffic flows using dynamic generalized linear models. Transportation Research Record 1678, 168-178.

Chang, G.L., Tao, X. (1999) An integrated model for estimating time-varying network O-D distribution. Transportation Research 33A, 381-399.

Chang, S.C., Lai , H.C., Yu, H.C. (2005) A variable P value rolling grey forecasting model for Taiwan semiconductor industry production. Technological Forecasting & Social Change 72, 623–640.

Chen, H., Grant-Muller, S., (2001) Use of sequential learning for short-term traffic flow forecasting. Transportation Research 9C, 319-336.

Chen, K.Y. (2007) Forecasting systems reliability based on support vector regression with genetic algorithms. Reliability Engineering and System Safety 92, 423-432.

Chiou, Y.C., Lan, L.W. (2001) Genetic clustering algorithms. European Journal of Operational Research 135(2), 413-427.

Chiou, Y.C., Ai, C.M. (2007) Rolling grey forecasting models for short-term traffics. Journal of the Eastern Asia Society for Transportation Studies 7, 2486-2501.

Chiou, Y.C.,Chou, S.T. (2010) Ant clustering algorithms. International Journal of Applied Evolutionary Computation 1, 1-15.

Chiou, Y.C., Lan, L.W., Tseng, C.M. (2010) Estimation of dynamic freeway origin-destination matrices with cell-based arrival distribution modeling. Journal of Eastern Asia Society for Transportation Studies 7, 1-16.

Codina, E. (2004) Adjustment of O–D trip matrices from observed volumes: An algorithmic approach based on conjugate directions. European Journal of Operational Research 155, 535–557.

Cremer, M., Keller, H. (1987) A new class of dynamic methods for the identification of origin-destination flows. Transportation Research 21B, 117-132.

Daganzo, C.F. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation research 28B, 269-287.

Daganzo, C. F. (1995) The cell transmission model: network traffic. Transportation Research 29B, 79-93.

Daganzo, C. F. (1999) The lagged cell-transmission model. Inst. of Transportation Studies, University of California, Berkeley, California, USA.

Dervoort, M., Dougherty, M., Watson, S. (1996) Combining kohonen maps with ARIMA time series models to forecast traffic flow. Transportation Research 4C, 307-318.

Dia, H. (2001) An object-oriented neural network approach to short-term traffic forecasting. European Journal of Operational Research 131, 253-261.

Dimitriou, L., Tsekeris, T., Stathopoulos, A. (2008) Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transportation Research 16C , 554-573.

Dougherty, M., Cobbett, M. (1997) Short-term inter-urban traffic forecasts using neural networks. International Journal of Forecasting 13, 21-31.

Frahling, G. (2005) A fast k-means implementation using corsets. Heinz Nixdorf Institute and Department of Computer ScienceUniversity of Paderborn.

Gaur, S., Deo, M.C. (2008) Real-time wave forecasting using genetic programming. Ocean Engineering 35, 1166-1172.

Gomes, G., Horowitz, R. (2006) Optimal freeway ramp metering using the asymmetric cell transmission model. Transportation Research 14C, 244-262.

Hazelton, M.L. (2001) Inference for origin-destination matrices: Estimation, prediction and reconstruction. Transportation Research 35B, 667-676.

Hazelton, M.L. (2008) Statistical inference for time varying origin-destination matrices. Transportation Research 42B, 542-552.

Ho, W.M. (2008) Comparisons of assignment-based and non-assignment-based time dependent O-D estimation models for general networks. Unpublished thesis, National Cheng Kung University (in Chinese).

Huang, S., Sadek, A.W. (2009) A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting. Transportation Research 17C, 510-525.

Ishak, S., Baton, B. (2003) Optimization of dynamic neural networks performance for short-term traffic prediction. TRB 2003 Annual Meeting.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. Transactions of the ASME Journal of Basic Engineering 82, 35-45.

Kantabutra, S., Naramittakapong, C., Kornpitak, P. (2000) Pipelined k-means algorithm on COWs. The Theory of Computation Group.

Kerner, B.S. (2004) The physics of traffic, Springer-Verlag Berlin Heidelberg.

Kerner, B.S. (2005) Control of spatiotemporal congested traffic patterns at highway bottlenecks. Physica 355A, 565-601.

Kirby, H., Dougherty, M., Watson, S. (1997) Should we use neural networks or statistical models for short term motorway traffic forecasting. International Journal of Forecasting 13, 43-50.

Knospe, W., Santen, L., Schadschneider, A., Schreckenberg, M. (1999) Disorder effects in cellular automata for two-lane traffic. Physica 265A, 614-633.

Knospe, W., Santen, L., Schadschneider, A., Schhreckenberg, M. (2000) Towards a realistic microscopic description of highway traffic. Journal of Physics 33A, 477-485.

Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics 43, 50-69.

Koza, J.R. (1992) Genetic programming: on the programming of computers by means of natural selection. The MIT Press.

Kühne, R. in: Gartner, N., Messner, C.J., Rathi, A.J. (1998) Traffic Flow Theory, 2nd ed. Transportation Research Board, Special Report 165.

Lam, W.H.K., Chan, K.S., Shi, J.W. (2002) A traffic flow simulator for short-term travel time forecasting. Journal of Advanced Transportation 36, 265-291.

Lam, W.H.K., Chan, K.S., Tam, M.L., Shi, J.W. (2005) Short-term travel time forecasts for transport information system in Hong Kong. Journal of Advanced Transportation, 39, 289-306.

Lam, W.H.K., Tang, Y.F., Tam, M.L. (2006) Comparison of two non-parametric models for daily traffic forecasting in Hong Kong. Journal of Forecasting 25, 173-192.

Lan, L.W., Sheu, J.B., Huang, Y.S. (2007) Prediction of short-interval traffic dynamics in multidimensional spaces. Journal of the Eastern Asia Society for Transportation Studies 7, 2353-2367.

Lan, L.W., Lin, F.Y., Kuo, A.Y. (2010) Three novel methods to predict traffic time series in reconstructed state spaces. International Journal of Applied Evolutionary Computation 1, 16-35.

Ledoux, C. (1997) An urban traffic flow model integrating neural network. Transportation Research Part 5C, 287-300.

Li, B., Moor, B.D. (1999) Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. Transportation Research 33B, 203-214.

Lighthill, M. H., Whitham, G. B. (1955) On kinematic waves: a theory of traffic flow on long crowed roads. Proceedings of the Royal Society, Series 229A, 317-345.

Lin, P.W., Chang, G.L. (2006) Modeling measurement errors and missing initial values in freeway dynamic origin-destination estimation systems. Transportation Research 14C, 384-402.

Lin, P.W., Chang, G.L. (2005) A robust model for estimating freeway dynamic origin-destination matrix. Transportation Research Record 1923, 110-118.

Lin, P.W., Chang, G.L. (2007) A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix. Transportation Research 41B, 554-572.

Lin, W.H., Lo, H.K. (2003) A theoretical probe of a German experiment on stationary moving traffic jams. Transportation Research 37B, 251-261.

Liu, G., Lyrintzis, A. S., Michalopoulus, P. G. (1998) Improved high-order model for freeway traffic flow. Transportation Research Record 1644, 37-46.

Lo, H.K. (2001) A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transportation Science 35.

Lo, H.K., Szeto, W. Y. (2002) A cell-based dynamic traffic assignment model: formulation and properties. Mathematical and Computer Modelling 35, 849-865.

Lo, H.K., Szeto, W.Y. (2002) A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transportation Research 36B, 421-443.

Lo, H.K., Zhang, N., Lam, W. (1996) Estimation of an origin-destination matrix with random link choice proportions: A statistical approach. Transportation Research 30B, 309-324.

Lo, H.P., Zhang, N. (1996) Estimation of an origin-destimation matrix with random with choice proportions : a statistical approach. Transportation Research 30B, 309-324.

Lo, H.P., Chan, C.P. (2003) Simultaneous estimation of an origin–destination matrix and link choice proportions using traffic counts. Transportation Research 37A, 771–788.

Lundgren, J.T., Peterson, A. (2008) A heuristic for the bi-level origin-destination matrix estimation problem. Transportation Research 42B, 339-354.

Nagel, K., Schreckenberg, M. (1992) A cellular automaton model for freeway traffic. Physics 2I, 2221-2229.

Pai, P.F., Lin, C.S. (2005) A hybrid ARIMA and support vector machines model in stock price forecasting, Omega 33, 497 – 505.

Pal, N.R. Pal, K., Keller, J.M., Bezdek, J.C. (2005) A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems 13, 517-532.

Pampalk, E., Widmer, G., Chan, A. (2003) A new approach to hierarchical clustering and structuring of data with self-organizing maps. Technical Report OeFAI-TR-2003-09.

Payne, H. J. (1971) Model of freeway traffic and control. Mathematics of Public Systems 1, 51-61.

Pipes, L.A. (1953) An operational analysis of traffic dynamics. Journal of Applied Physics 24 (3), 74–281.

Pottmeier, A., Barlovic, R., Knospe, W., Schadschneider, A. (2002) Localized defects in a cellular automaton model for traffic flow with phase separation. Physica 308A, 471-482.

Rauber, A., Merkl, D., Dittenbach, M. (2002) The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. Neural Networks, IEEE Transactions 13(6), 1331-1341.

Ricardo, G.R. (2008) A column generation algorithm for the estimation of origin–destination matrices in congested traffic networks. European Journal of Operational Research 184, 860–878.

Richards, P. I. (1956) Shock waves on highway. Operations Research 4, 42-51.

Sherali, H.D. Park. T. (2001) Estimation of dynamic origin-destination trip tables for a general network. Transportation Research 35B, 217-235.

Smith, B.L., Williams, B.M., Oswald, R.K. (2002) Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research 10C, 303-321.

Stathopoulos, A., Karlaftis, M.G. (2003) A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research 11C, 121-135.

Sun, X., Munoz, L., Horowitz, R. (2003) Highway traffic state estimation using improved mixture kalman filters for effective ramp metering control. Proceedings 42nd IEEE Conference on Decision and Control, 6333-6338.

Szeto, W.Y., Lo, H.K. (2004) A cell-based simultaneous route and departure time choice model with elastic demand. Transportation Research 38B, 593-612.

Tangsripairoj, S., Samadzadeh, M.H. (2006) Organizing and visualizing software repositories using the growing hierarchical self-organizing map. Journal of Information Science and Engineering 22, 283- 295.

Tseng, F.M., Tzeng, G.H. (2005) A fuzzy seasonal ARIMA model for forecasting. Omega 33, 497-505.

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

Vardi, Y. (1996) Network tomography: estimating source-destination traffic intensities from link data. Journal of the American Statistical Association 91, 365-377.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C. (2006) Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume. Transportation Research 14C, 351-367.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C. (2006) Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume, Transportation Research 14C, 351 – 367.

Wang, B.H., Wang, L., Hui, P.M., Hu, B. (2000) The asymptotic steady states of deterministic one-dimensioned traffic flow models. Physica 27B, 237-239.

Wang, S.C., Tong, C.O. (1998) Estimation of time-dependent origin-destination matrices for transit network. Transportation Research 32B, 35-48.

Wang, X. (2005) GLS estimation of O-D matrix with traffic counts and information from ATIS. Proceedings of the Eastern Asia Society for Transportation Studies 5, 1188-1196.

Welch, W.J. (1983) A mean squared error criterion for the design of experiment. Biometrika 70, 205-213.

Wolf, D.E. (1999) Cellular automata for traffic simulations. Physica 263A, 438-451.

Wong, G. C. K., Wong, S. C. (2002). A multi-class traffic flow model–an extension of LWR model with heterogeneous drivers. Transportation Research 36A, 827-841.

Wu, J. (1997) A real-time origin-destination matrix updating algorithm for on-line approach. Transportation Research 31B, 381-396.

Wu, J., Chang, G.L. (1996) Estimation of time-varying origin-destination distributions with dynamic screenline flows. Transportation Research 30B, 277-290.

Yang, H., Sasaki, T., Iida, Y., Akiyama, T. (1992) Estimation of origin-destination matrices from link traffic counts on congested networks. Transportation Research 26B, 417-434.

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

Yang, H.T., Chen, S.C., Tsai, W.N. (2004) Classification of direct load control curves for performance evaluation. IEEE Transactions on Power Systems 19, 811-829.

Yang, H.T., Chen, S.C., Peng, P.C. (2005) Genetic k-means-algorithm-based classification of direct load-control curves. IEE Proc.-Gener. Transm. Distrib. 152, 32-49.

Yang, Y., Jiang, D., Xia, M. (2010) Using improved GHSOM for intrusion detection. Journal of Information Assurance and Security 5232-239.

Yao, L., Lin, C.C. (2009) Identification of nonlinear systems by the genetic programming-based Volterra filter. IET signal process 3, 93-105.

Yella, S., Gupta, N.K., Dougherty, M.S. (2007) Comparison of pattern recognition techniques for the classification of impact acoustic emissions. Transportation Research 15C, 345–360.

Yu, N., Zhang,H.M., Recker, W.W. (2005) Inferring origin–destination trip matrices with a decoupled GLS path flow estimator. Transportation Research 39B, 497-518.

Yu, F., Tang, J., Cai, R. (2007) Partially horizontal collaborative fuzzy c-means, International Journal of Fuzzy Systems 9, 13-28.

Zhang, G.P. (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159-175.

Zhang, H. M., Kim, T. (2005) A car-following theory for multiphase vehicular traffic flow. Transportation Research 39B, 385-399.

Zhang, H. M. (1998) A theory of nonequilibrium traffic flow. Transportation Research 32B, 485-498.

Zhang, X., Maher, M.J. (1998) The evaluation and application of a fully disaggregate method for trip matrix estimation with platoon dispersion, Transportation Research 32B, 261-276.

Zhou, X., Mahmassani, H.S. (2007) A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. Transportation Research 41B, 823-840.

APPENDIX VITA

曾群明(Chun-Ming Tseng)

教育

私立南光高職機工科建教合作班	(1983)
中央警察大學交通系 學士	(1993)
國立交通大學交通運輸研究所 碩士	(1998)
國立交通大學交通運輸研究所 博士	(2012)

考試與證照

特種考試警察人員乙等考試 交通警察人員交通組	(1993)
專門職業及技術人員高等考試 交通工程技師	(1999)
公務人員高等考試三級 交通工程科	(2001)
公務人員普通考試 交通工程科(錄取)	(2001)

獲獎

東亞運輸協會(EASTS)年會 最佳論文獎	(2011)

經歷

經濟部臺糖公司新營副產加工廠、永康糖廠 技術員	(1983-1988)
臺北縣政府警察局 分隊長	(1993)
警政署保六總隊第二警官隊 組員	(1994-2001)
臺北市政府交通局 技士	(2001-2004)
臺灣警察專科學校 兼任講師	(1993、1995)
交通部 科員、專員、編審、秘書	(2004-)

Publications related to dissertation research

A. Journal Papers

1. Chiou, Y.C., Lan, L.W. and Tseng, C.M. (2010) "Estimation of Dynamic Freeway Origin-Destination Matrices with Cell-Based Arrival Distribution Modeling," Journal of Eastern Asia Society for Transportation Studies, 7, 1-16.
2. 邱裕鈞、藍武王、許珮珊、曾群明(2011)，「應用格位傳送模式建構高速公路動態起迄矩陣推估演算法」，運輸學刊，第二十三卷第一期， 頁 97-128。
3. Chiou, Y.C., Lan, L.W. and Tseng, C.M. (2011) "Optimal Locations of License Plate Recognition to Enhance the Origin-Destination Matrix Estimation," Journal of Asian Transport Studies (Accepted).
4. Chiou, Y.C., Lan, L.W. and Tseng, C.M. (2011) "A Novel Method to Predict Traffic Features based on Rolling Self-Structured Traffic Patterns," Submitted to Journal of ITS.

B. Conference Papers

1. Chiou, Y.C., Lan, L.W. and Tseng, C.M. "Estimation of Dynamic Freeway Origin-Destination Matrices with Cell-Based Arrival Distribution Modeling," Eastern Asia Society for Transportation 2009.
2. 邱裕鈞、藍武王、許珮珊、曾群明，「應用格位傳送模式建構高速公路動態起迄矩陣推估演算法」，中華民國運輸學會第24屆論文研討會，民國98年12月。
3. Chiou, Y.C., Lan, L.W. and Tseng, C.M. "Estimation of Freeway Dynamic Origin-Destination Matrices: A Novel Approach," The 15th Hong Kong Society of Transportation Studies.
4. Chiou, Y.C., Lan, L.W. and Tseng, C.M. "Optimal Location of License Plate Recognition to Enhance the Origin-Destination Matrices Estimation," Eastern Asia Society for Transportation 2011.

C. Research Reports

1. 邱裕鈞、曾群明(2010)，應用車輛辨識系統提昇起訖屢次矩陣推估之研究(III)：國科會研究報告，計畫編號：NSC 96-2628-E-009-171-MY3。
2. 邱裕鈞、曾群明(2009)，應用車輛辨識系統提昇起訖屢次矩陣推估之研究(II)：國科會研究報告，計畫編號：NSC 96-2628-E-009-171-MY3。
3. 邱裕鈞、曾群明(2008)，應用車輛辨識系統提昇起訖屢次矩陣推估之研究(I)：國科會研究報告，計畫編號：NSC 96-2628-E-009-171-MY3。