

第三章 資料處理模式與旅行時間預測模式發展

3.1 資料來源

本研究利用高速公路上的交通資訊蒐集設備包括車輛偵測器 (Vehicle detector, VD)、探針車(Probe vehicle)、電子收費系統(ETC) 與自動車輛辨識系統(AVI)，以下分別說明各種資料來源之特性與資料前處理：

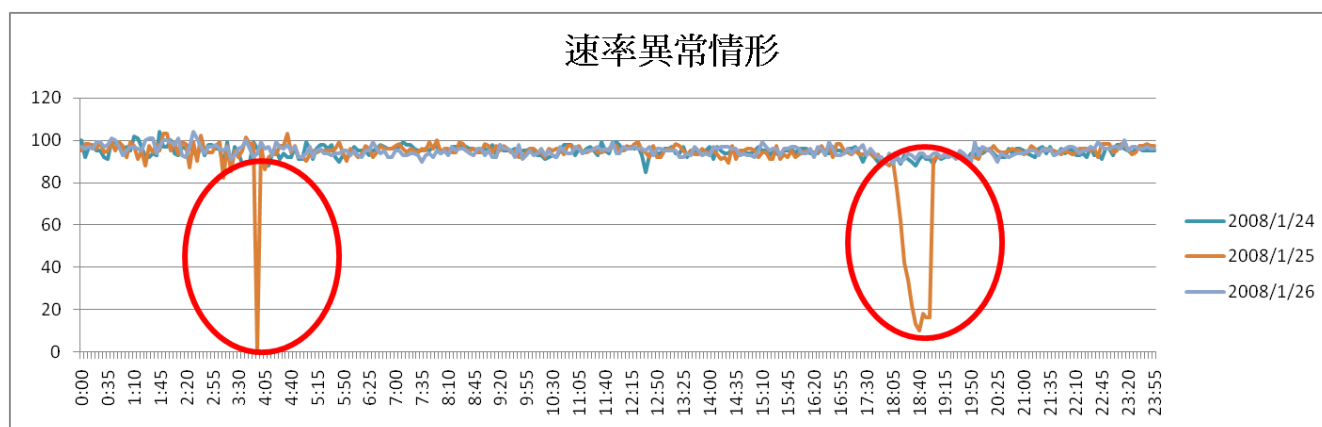
(1) 車輛偵測器 (VD)

高速公路的車輛偵測器具有以下特性：

- A. 佈設位置與密度視路段交通狀況而定，偵測器間約間隔 1~2 公里。
- B. 車輛偵測器回傳頻率為 5 分鐘/次，回傳的資訊有偵測器編號、5 分鐘內各車種的平均速度、五分鐘內各車種的流量、回傳時間等
- C. 偵測器傳回的為點位資料，非路段資料。
- D. 偵測器有可能因為通訊或是設備問題，導致遺失資料，相關差補方法於 3.4 節中詳述。

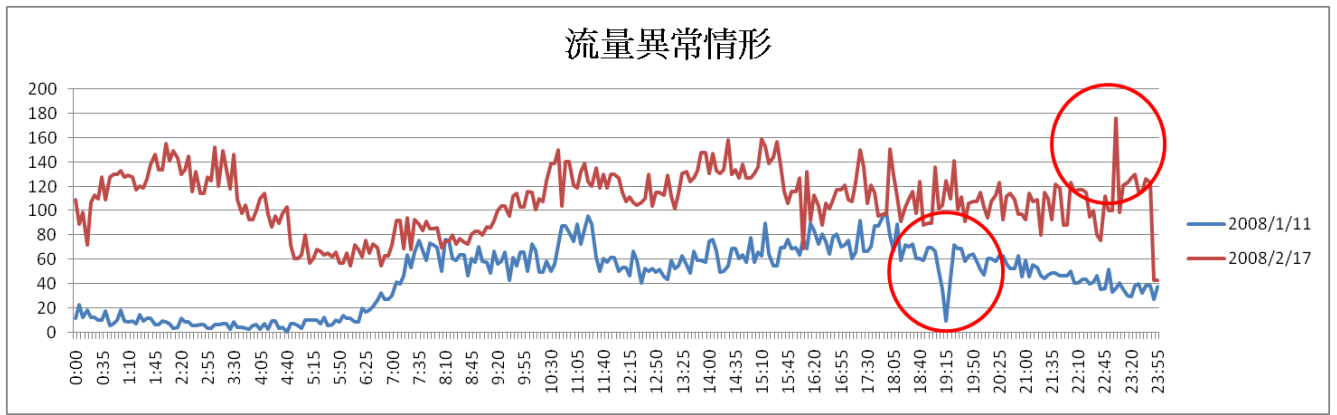
車輛偵測器在蒐集資料的過程中可能發生下列問題：

- A. 速率、流量突然增加或降低(如圖 3.1-1、圖 3.1-2)
- B. 長時間速率偏低或偏高(如圖 3.1-3)
- C. 長時間流量偏低或偏高(如圖 3.1-4)
- D. 資料缺漏(如圖 3.1-5)



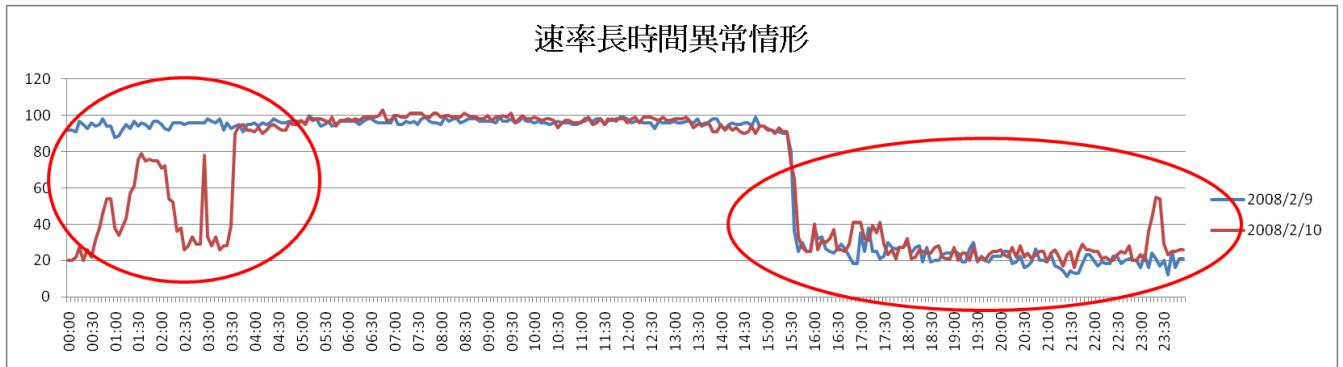
[資料來源：本研究整理]

圖 3.1-1 速率異常分佈圖



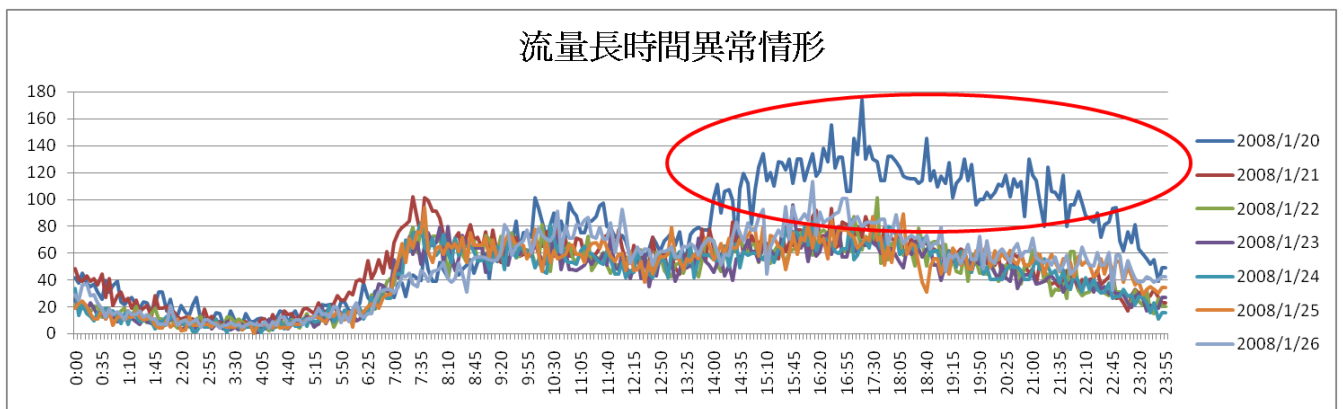
[資料來源：本研究整理]

圖 3.1-2 流量異常分佈圖



[資料來源：本研究整理]

圖 3.1-3 速率異常分佈圖(長時間偏低)



[資料來源：本研究整理]

圖 3.1-4 流量異常分佈圖(長時間偏高)



[資料來源：本研究整理]

圖 3.1- 5 流量資料遺失圖

(2) 探針車(Probe vehicle)

探針車之資料具有以下特性：

- A. 探針車資訊回傳頻率為 30 秒/次。回傳的資訊有車機編號、座標、速度、方位、回傳時間等。
- B. 探針車回傳之資料為點位資料，非路段資料。
- C. 探針車回傳之位置或速度資料可能會產生漂移的現象。
- D. 探針車主要以客運車輛為主，由於客運車輛行駛國道時，有車道限制，不能行駛最內側車道，且速限較一般小客車低。
- E. 同時段出發之探針車之旅行時間會有差異
- F. 起迄點可能與目標路徑不同
- G. 探針車可能會於目標路徑中下匝道載客後再上匝道

探針車在蒐集資料的過程中可能發生 GPS 漂移的現象，如圖 3.1-6。



[資料來源：本研究整理]

圖 3.1- 6 GPS 座標漂移

因應上述特性，探針車前處理之項目如下：

- A. 刪除座標點位在路徑 100 公尺以外的資料。
- B. 依照出發時段分類資料。
- C. 若是探針車於目標路徑中有下匝道載客，則需利用 GPS 資料扣除下匝道與上匝道之間的一般道路旅行時間。
- D. 計算同一時段中旅行時間之平均數、變異數、最大值與最小值。

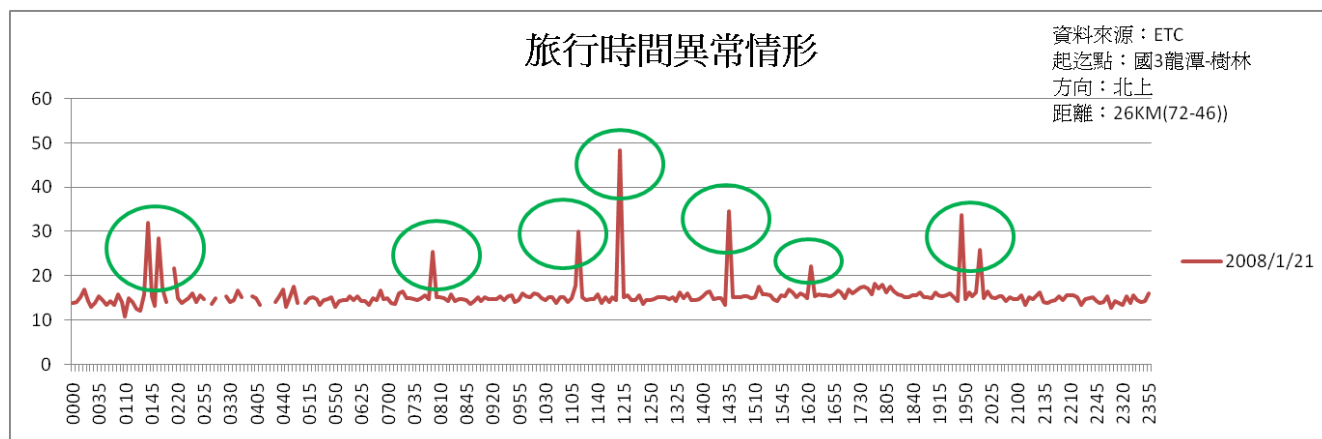
(3) 電子收費系統(ETC)

ETC 之資料具有以下特性：

- A. 由於通過收費站的速度較快、時間較短，因此會低估整體車流的旅行時間。
- B. ETC 蒐集到之旅行時間與目標路徑可能不同。

C. ETC 蒐集到之旅行時間可能會包含下匝道之後再上匝道或是有進入休息站之車輛之旅行時間。

ETC 在蒐集資料的過程中可能發生旅行時間過高的問題(如圖 3.1-7)。



[資料來源：本研究整理]

圖 3.1-7 旅行時間異常分佈圖

因應上述特性，ETC 之前處理項目如下：

- A. 將資料依照出發時段分類。
- B. 計算每時段旅行時間之平均值、變異數、最大值與最小值。
- C. 刪除上一筆與下一筆旅行時間差異超過 15% 之資料。
- D. 將 VD 計算出的旅行時間或探針車之旅行時間與 ETC 之旅行時間延遲加總使其起迄點與目標路徑一致。

(4) 自動車輛辨識系統(AVI)

AVI 之資料具有以下特性：

- (1) 取得路徑旅行時間之頻率不固定。
- (2) 同一時段比對到之旅行時間會有差異。
- (3) 若 AVI 架設於匝道與平面道路連接處，則旅行時間包含行駛匝道的時間，資料仍需進行修改。
- (4) AVI 蒐集到之旅行時間可能會包含下匝道之後再上匝道或是有進入休息站之車輛之旅行時間。

因應上述特性，AVI 之前處理項目如下：

- (1) 依照出發時段分類資料
- (2) 計算同一時段中旅行時間之平均數、變異數、最大值與最小值
- (3) 利用統計方法找出差異過大之門檻值，並刪除上一筆與下一筆旅行時間差異超過 15% 之資料

圖 3.1-8 說明本研究如何取得實測路段--新竹系統交流道至汐止系統交流道--之旅行時間：

- a. ETC 蒐集到汐止收費站至龍潭收費站之旅行時間。
- b. 扣除利用 VD 資料計算而得之汐止系統交流道至汐止收費站之旅行時間。
- c. 加上利用探針車蒐集到之新竹系統交流道至龍潭收費站之旅行時間。
- d. 得到新竹系統交流道至汐止系統交流道之旅行時間。

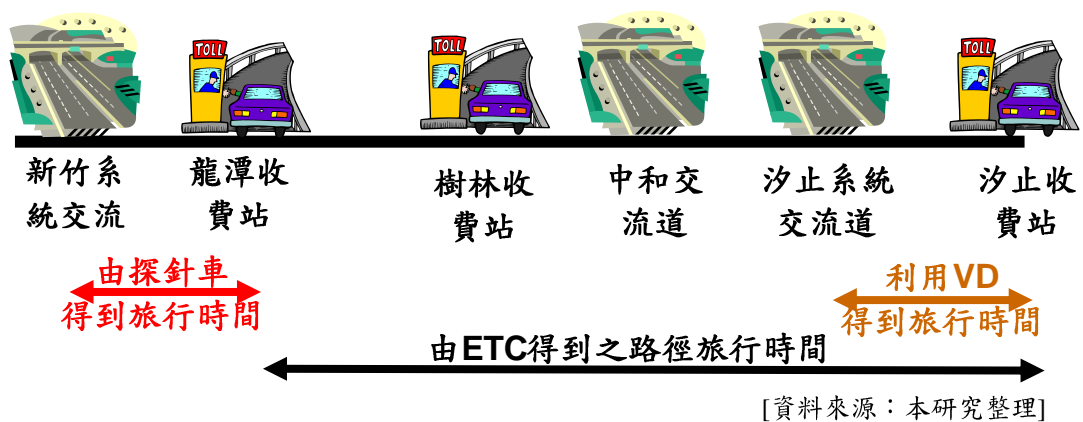
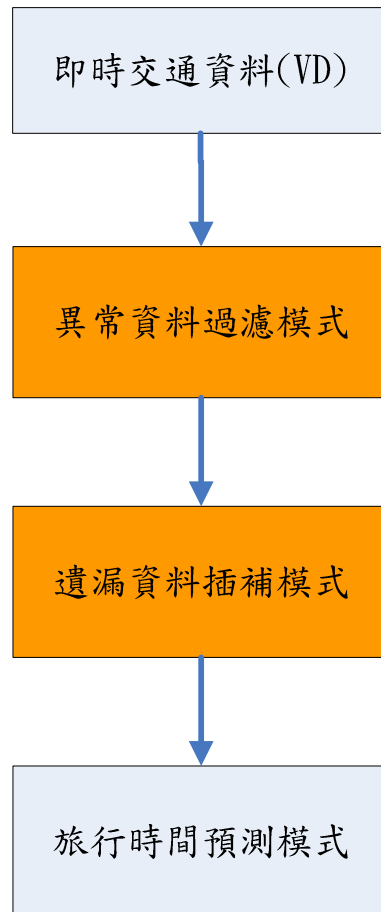


圖 3.1- 8 ETC 與探針車整合方式示意圖

前述已說明探針車與 ETC 之資料前處理過程，由於 VD 之資料前處理過程較為複雜，因此特別分節詳述之(於 3.2 至 3.3 節中詳述)。

3.2 偵測器異常資料過濾模式

若是採用異常資料則會造成預測之旅行時間不準確，因此於資料回傳之後必須先過濾異常之速度、流量、佔有率或是旅行時間，才能作為旅行時間預測之用。資料處理模式如圖 3.2-1 所示，包括：異常資料過濾模式與遺漏資料插補模式。分別說明如下：



[資料來源：本研究整理]

圖 3.2- 1 資料處理模式

蒐集速率、流量、佔有率或旅行時間資料時，可能會發生下列幾種異常情形：

- 速率、流量或旅行時間突然增加或降低
- 長時間速率偏低或偏高
- 長時間流量偏低或偏高

造成異常資料的原因有很多，必須找出造成異常之原因以利後續處理，其可能原因通常有下列幾種：

(1) 設備因素

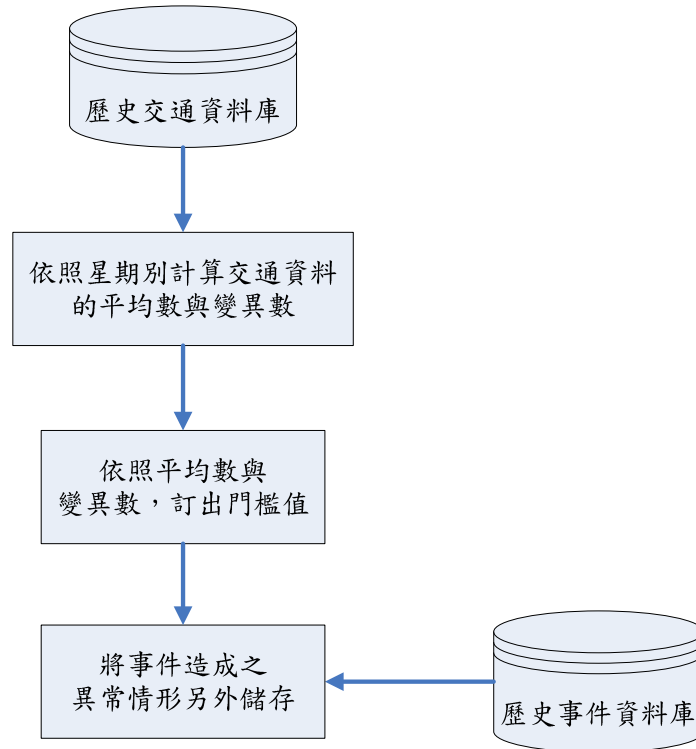
- 偵測器故障
- 通訊因素

(2) 其他因素

- 天候因素，如：某路段的豪雨、起霧、路段積水等
- 駕駛行為異常之車輛

- 異常之車輛組成，例如：短時間內有多輛大車進入主線
- 發生事件

本研究將依照下列流程建立異常資料過濾模式，流程圖如下圖所示：



[資料來源：本研究整理]

圖 3.2-2 異常資料過濾模式建立流程圖

步驟一：

將流量、速度與佔有率分別依照星期別分別計算不同時段與不同路段之平均數與變異數

步驟二：制定門檻值

計算上述不同時間不同路段的平均數與標準差後，刪除在三倍標準差以外的資料，以三倍標準差當作門檻值。

本研究所蒐集資料屬於大樣本性質，依照統計學原理當樣本數大於 30 時，參數估計式抽樣分配趨近常態分配。而常態分配中大部分之觀測值，會落在距離平均數正負三倍標準差的範圍內。因此落在三倍標準差外的值可判定為離群值 (Outliers)。由於前後時段之數值不可能相差過大，因此可將這些離群值視為錯誤資料，並予以刪除。

步驟三：蒐集事件資料，並將由於事件造成之異常分開儲存

上述幾種可能的異常資料發生原因中，除了事件造成的異常外，其餘原因造成之異常才是本模式過濾之目標，因此蒐集事件資料有助於界定由於事件所造成之異常有哪些。當即時資料進來時，本研究將先比對全國路況資料庫，看看當時是否有發生事件，若有事件發生，則直接將資料存入事件資料庫，若無，即可依照此門檻值過濾資料。事件資料庫將於 3.4 節中詳述。

本計畫採用探針車為推估旅行時間的資料來源，所回傳的資訊有車機編號、座標、速度、方位、時間資料，但常因天氣因素或傳輸問題而產生座標飄移、起迄點與預測範圍不同等現象。因此為確保資料的正確性，本研究發展探針車的過濾模式，採用運研所提供「99 年版交通路網數值地圖」的點位資料，所要過濾範圍為國道三號新竹系統至汐止系統間，因此刪除國道客運所回傳 GPS 點位不在範圍內的資料。此外在設定刪除與高速公路垂直距離大於 100 公尺的座標點的過濾條件，讓所有 GPS 資料都落在主線道的範圍內。

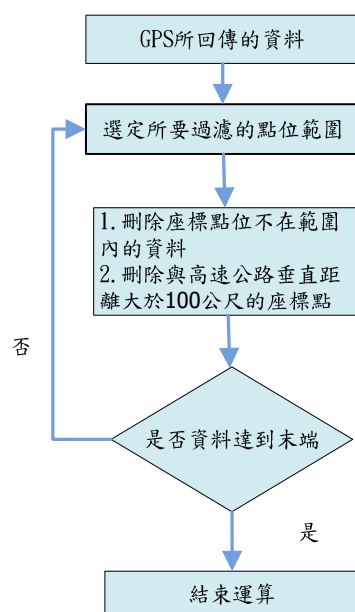


圖 3.2- 3 探針車資料過濾模式

此外取得 ETC 資料之方式，並非以即時抓取的方式來取得，而是取得其歷史資料後匯入本研究之資料庫當中，取得方式除了透過光碟燒錄的方式、也透過遠通電收內部的 FTP 擷取資料。本研究以非即時的方式，將這樣的 ETC 資料抓入資料庫當中進行旅行時間的推估

與校正。但 ETC 即時資料與歷史資料具有時間延遲，故需要發展資料過濾模式，其說明如下（圖 3.2- 4）：

- A. 與前一段時間（如五分鐘）的平均旅行時間比較，旅行時間不可能有劇烈變化
- B. 刪除正負 15% 以外的資料(遠通電收建議)

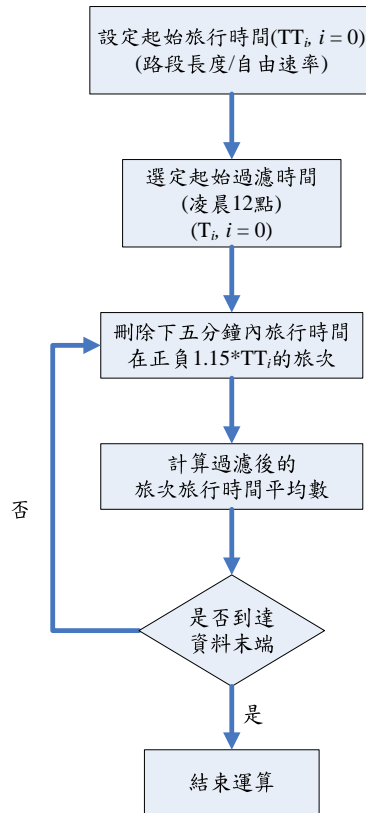


圖 3.2- 4 ETC 資料過濾模式

ETC 大小車資料過濾與比較

利用收費站之間大客車與小客車之速度資料，進行旅行時間推估模式之校正。原始資料來源為非即時抓取的方式，並透過遠通電收內部的 FTP 擷取資料，而歷史之速度資料與實際速度有所差異，因此需要發展速度過濾模式將異常值刪除，其說明如下：

- A. 與前一段時間（如五分鐘）的平均速度比較，平均速度不可能有劇烈變化
- B. 刪除超過正常速限之資料（低於 10kpm、高於 120kpm）

本研究分析龍潭-樹林收費站 2011 年一月份 ETC 之資料，以分析大小車之速度差值

表 3.2- 1 一月份 ETC 之資料

速度差值	北上	南下
平日尖峰	16.81	12.74
平日離峰	23.86	10.46
假日尖峰	15.83	9.94
假日離峰	16.82	10.58

(單位：km/hr)

因國道客運有車道限制，不能行駛最內側車道，且速限較一般小客車低。而本研究所發佈對象為小客車，因此利用簡單迴歸方式修正旅行時間推估模式，透過其迴歸參數將大客車平均速度轉換成小客車平均速度。以利後續旅行時間推估模組使用，其模式如下：

$$(小客車平均速度)=\alpha+\beta(大客車平均速度)$$

表 3.2- 2 大小車迴歸參數

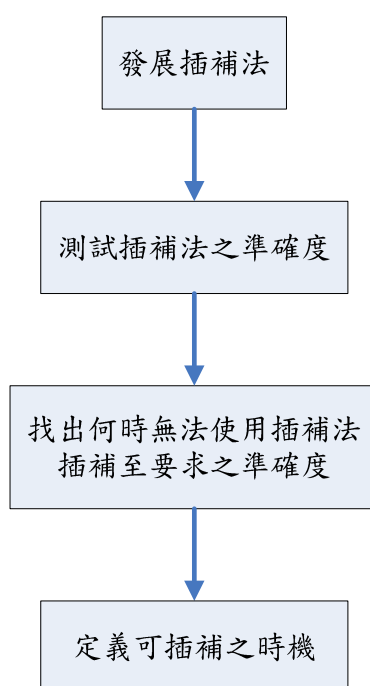
北上	α	β	南下	α	β
平日尖峰	88.0489	0.135855	平日尖峰	90.40584	0.097041
平日離峰	104.9501	-0.01207	平日離峰	101.115	-0.00255
假日尖峰	89.86883	0.088273	假日尖峰	108.5274	-0.1181
假日離峰	100.3914	0.003374	假日離峰	97.77981	0.025441

3.3 遺漏資料插補模式

需要遺漏資料插補模式的原因有二：

1. 當過濾完異常資料後，異常資料即被刪除，因此需要資料插補模式插補被刪除之資料。
2. 偵測器之資料可能會因為偵測器本身或是通訊問題而發生資料遺漏之情形，因此需要一模式來插補遺漏之資料。

本研究將依照下列流程建立遺漏資料插補模式，流程圖如下圖所示：



[資料來源：本研究整理]

圖 3.3- 1 遺漏資料處理流程圖

步驟一：發展插補法

資料插補法有很多，大致可分為兩類，本研究係利用多重插補法：

(1) 單一插補法

利用資料集之分配與平均數來估計遺漏的資料。使用單一插補法之假設為完整資料與遺漏資料間為一對一關係，一對一關係指一個遺漏的資料只有一個可能的正確值。

其缺點為完整資料與遺漏資料為多對一關係，亦即遺漏的這筆資料有有多個可能的數值，並非一個單一常數，視遺漏值為單一常數會低估變異數與共變數。例如：上游偵測器的速率是 80km/hr，下游偵測到 100km/hr，而中游偵測器資料遺漏，一對一的關係是假設中游偵測器的速率只有一個可能的值，例如：90km/hr。但多對一個關係則假設中游偵測器有多個可能的速率值，例如有可能是 90km/hr 也有可能是 85km/hr。

(2) 多重插補法

多重插補法可改善單一插補法之缺點，其係利用一些模擬的方法，如馬可夫鍊，對於同一遺漏資料點進行多次估計，在分析這些資料的平均數與變異數後產生最終的估計結果。

步驟二：測試插補法之準確度

本研究將依照下列步驟測試插補法之準確度：

步驟 2.1：將達到資料精確度與資料可用性要求之 VD 與歷史旅行時間資料依照不同交通需求與不同路段加以分類，分成平日離峰、平日尖峰、假日離峰、假日尖峰。

步驟 2.2：為了得知在不同之資料遺漏比例下插補法之準確度，因此必須於此步驟中將達到品質要求之資料分別移除不同比例之資料，並且在不同之遺漏比例下分別測試插補法之準確度，例如：可以分別移除 20%、40%、60%、80%、100% 之資料來測試。

步驟 2.3：若是使用單一插補法，需測試不同資料遺漏比例下，插補之資料與原始資料的絕對誤差。若是使用多重插補法則需測試在不同插補運算次數與不同資料遺漏比例下，插補之資料與原始資料的絕對誤差。

步驟 2.4：分別利用插補過後之資料集與原始資料集推估路徑旅行時間，並計算兩者之間之最大絕對誤差與平均絕對誤差。

步驟三：找出不同資料遺失比例下，插補法插補之準確度

依據上述步驟可以得知資料遺漏比率與插補資料絕對誤差之關係，可以知道不同路段與不同交通型態下，資料遺漏比例與插補後資料精確度之關係。

發展旅行時間預測模式時，可分析不同資料精確度與旅行時間預測精確度之間之關係，如此便能得知遺失資料程度大於多少時，無法滿足旅行時間預測精確度之要求，此時便停止發佈預測之旅行時間。例如：資料遺漏比例與插補之精準度可能如下表：

表 3.3- 1 遺漏比例與插補之精準度

遺漏比例 (百分比)	精準度 (百分比)
20	90
40	85
60	60
80	20
100	0

假設模式對於速率的精準度要求為 90%，從上表中可知當遺漏比例為 20%時，插補之後資料精準度可達要求的 90%，但當遺漏比例再往上升時，插補之後的資料即無法達到精準度之要求，因此當遺漏比例超過 20%時，則停止預測旅行時間。其說明如(下圖 3.3- 2)為插補 VD 資料流程

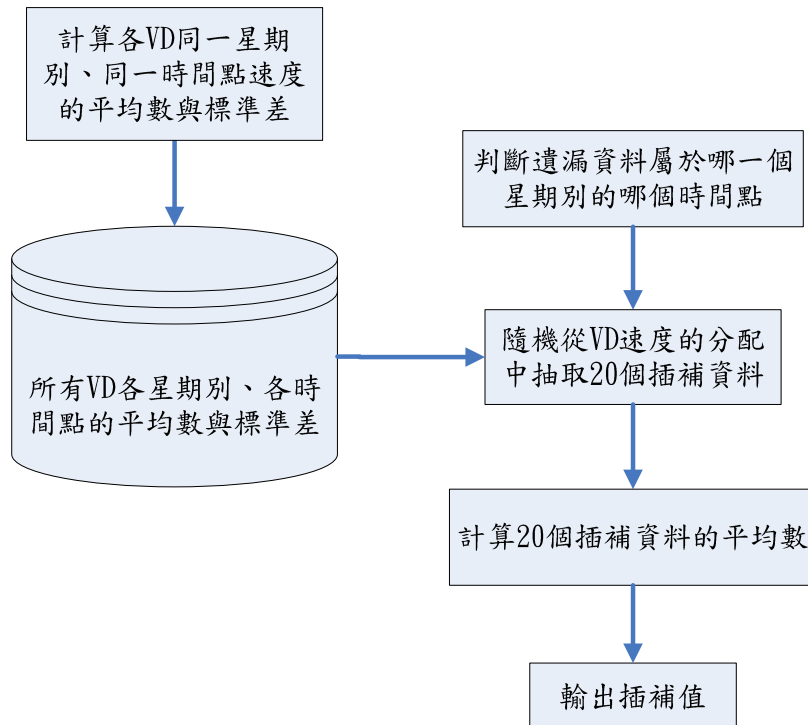


圖 3.3- 3 插補 VD 資料流程

3.4 資料結構分析設計與相關軟硬體設施

3.4.1 資料儲存格式

在本研究中，蒐集之交通資料來源有 VD、探針車（以客運車輛為資料來源）與 ETC 等，其中 VD 與探針車可分為即時資料與歷史資料兩部份，其中歷史資料又可依照是否有事件區分為一般歷史資料與事件歷史資料。而 ETC 目前只有歷史資料可供使用。

各類型資料蒐集頻率如下：

- ✓ VD：5 分鐘/次
- ✓ 探針車：30 秒/次
- ✓ ETC：5 分鐘/次
- ✓ AVI：5 分鐘/次

各項資料的資料庫格式如下：

表 3.4- 1 VD 資料表

編號	欄位	型態	長度	欄位說明
1	Fdatetime	nvarchar	50	收集日期
2	Ftime	nvarchar	50	收集時間
3	Vdid	nvarchar	50	VD 編號
4	Ftspd	nvarchar	50	平均速度
5	Favol	nvarchar	50	平均流量
6	UseTime	datetime	8	時間
7	source	nvarchar	50	資料來源

表 3.4- 2 客運車輛 GPS 訊號資料表單

編號	欄位	型態	長度	欄位說明
1	liGPSID	char	10	GPS 編號
2	iVehicleID	char	10	車輛編號
3	iDriverID	int	10	駕駛編號
4	iOriginalDriverID	int	10	原始駕駛編號

5	dtTime	datetime	8	時間
6	fLatitude	real	50	緯度
7	fLongitude	real	50	經度
8	iHeading	int	10	方向(度)
9	ucVelocity	int	10	速度(KPH)

表 3.4- 3 ETC 資料表

編號	欄位	型態	長度	欄位說明
1	DetectionDatetime	datetime	8	日期時間
2	Start_TollNo	int	10	收費站起點代碼
3	End_TollNo	int	10	收費站迄點代碼
4	Dir	int	10	方向
5	Cartype	int	10	車種
6	Weekday	int	10	星期別

表 3.4- 4 AVI 資料表

編號	欄位型態	型態	長度	欄位說明
1	id	Integer	4	編號
2	Startdatetime	datetime	8	經過起點的日期
3	Direct	nvarchar	50	方向
4	CarNum	nvarchar	50	車牌序號
5	FThroughPt	nvarchar	50	經過點 A
6	TThroughPt	nvarchar	50	經過點 B
7	TroughATime	datetime	8	經過 A 點的時間
8	TroughBTime	datetime	8	經過 B 點的時間
9	DiffTime	datetime	8	時間差

表 3.4- 5 事件資料表

編號	欄位型態	型態	長度	欄位說明
1	Start_VD_ID	nchar	10	起點偵測器編號
2	End_VD_ID	nchar	10	迄點偵測器編號
3	TT_DateTime	datetime	8	日期
4	Total_Travel_Time	real	50	旅行時間
5	Direction	nchar	10	方向
6	Weekday	int	4	星期別
7	VD_Up_AveSpeed	real	50	上游 VD 平均速度
8	VD_Down_AveSpeed	real	50	下游 VD 平均速度
9	Probe_AveSpeed	real	50	探針車平均速度
10	Probe_Speed_1	real	50	小客車平均速度
11	Probe_Speed_2	real	50	大客車平均速度
12	Path	nvarchar	50	路段
13	Incident	nchar	10	是否事件發生
14	Incident type	nvarchar	50	事件發生類型

3.4.2 相關軟硬體設施

本研究之硬體設施如下：

1. 伺服器乙台，規格如下：

中央處理器：Intel E5300 以上

主機板：MSI X58M 以上

記憶體：DDRII 2G 以上

硬碟：320G 以上

電源供應器：400W 以上

螢幕：15 吋以上

2. 本研究之軟體清單如下：

SQL Server 乙套

Microsoft Visual Studio 2005 乙套

OS: Windows XP

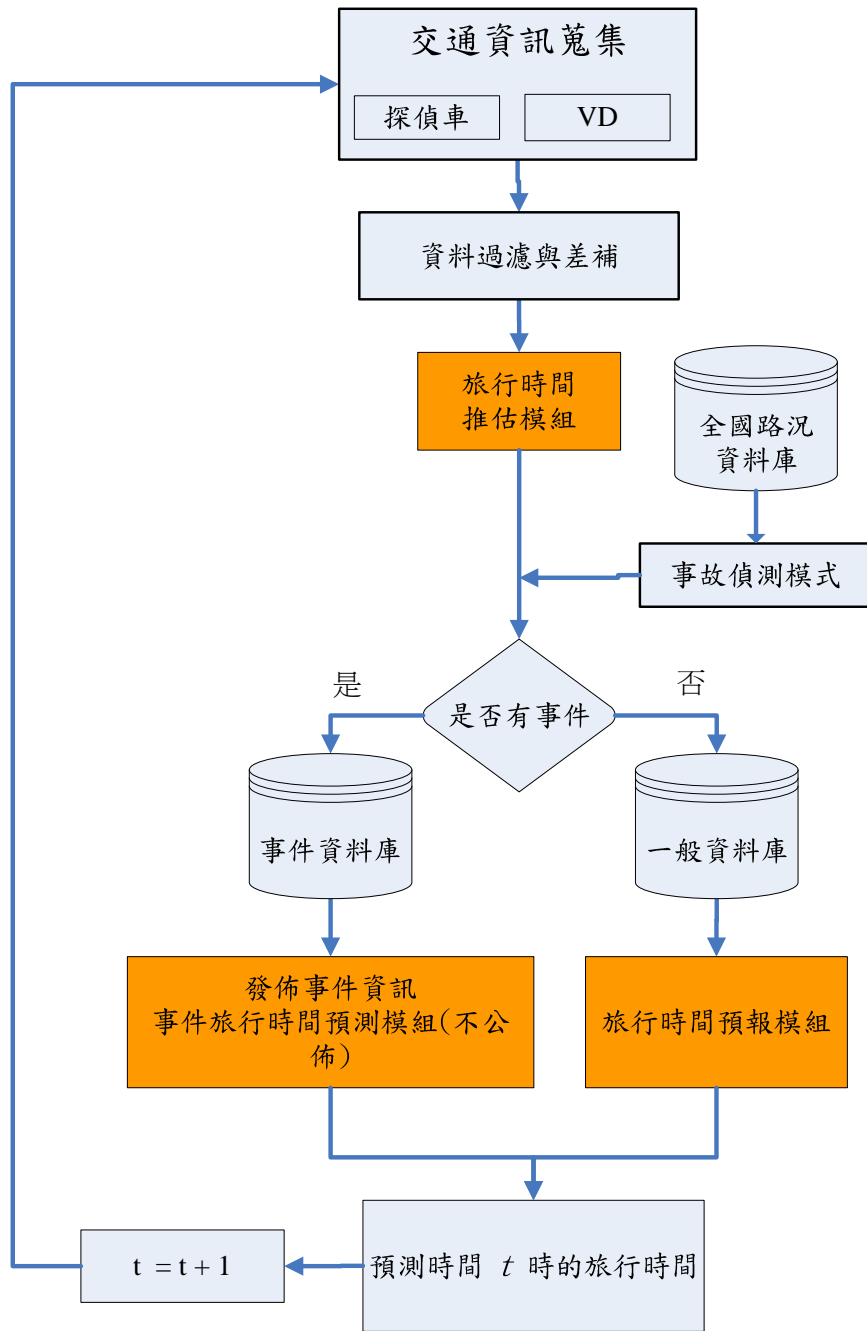
3.5 旅行時間預測模式

為了整合多種資料來源並能預測非重現性擁塞（Nonrecurrent Incidents）時之旅行時間，本研究之模式主要分為兩大模組：

1. 旅行時間推估模組
2. 旅行時間預測模組(一般與事件)

就事件定義主要討範圍為非重現性擁塞部份，一般又可分為預期性事件與突發性事件。所謂預期性事件是指事先計畫已知之管制措施，如道路封閉、施工管制、路段清潔管制等，至於突發性事件則是指一般交通意外事件，主要包括事故（Accident）、車輛故障、散落物...等。此外針對預報和預測定義如下，預測為對未來某時段內某一地區或部分路段可能出現的交通狀況所作的推斷。它是根據對全國路況資料庫分析，結合有關民眾通報、國道警察和警廣通報等綜合研究後作出的。所以，不是絕對準確僅供參考，以不發佈形式提供給內部同仁作為交管策略之用或是替代道路的指引。而預報是經由內部運算後可提供給使用者了解目前該路段的旅行時間為何，可讓使用者自行選擇，如同氣象預報的觀念。

首先先過濾與處理偵測器與探針車取得的交通資料，接著利用處理後之資料推估路徑旅行時間。推估出路徑旅行時間後，先查詢全國路況中心資料庫，看此時是否有事件產生。並透過事故偵測模式查詢，若是全國路況資料庫查詢到路徑上有事件發生時，則將推估出之路徑旅行時間存放至事件旅行時間資料庫，並發佈事件資訊，此時仍會透過事件旅行時間預測模組預測旅行時間，但此旅行時間由於誤差較大，因此僅供內部參考，不會公佈；若是事件偵測模組沒有偵測到事件發生，將推估出之路徑旅行時間存放至一般旅行時間資料庫，並採用旅行時間預測模組來預測旅行時間。整體流程圖如圖 3.5-1：



[資料來源：本研究整理]

圖 3.5-1 模式流程圖

以下針對各模組分別說明之。

3.5.1 旅行時間推估模組

旅行時間推估模組如圖 3.5-2 所示，另外本研究也利用探針車與 ETC 之資料來驗證旅行時間推估模組。

重要流程說明如下：

(1) 蒐集偵測器、探針車與 ETC 之資料

即時蒐集下列資料，並將其之資料寫入資料庫中：

- A. 偵測器之速度資料
- B. 探針車速度、點位等資料
- C. ETC 旅行時間

(2) 對於高速公路路段先判斷路段中是否有探針車之資料，如果沒有則用推估模式 1 推估偵測器間的旅行時間，如果有則用推估模式 2 推估偵測器間的旅行時間。

(3) 推估模式 1

對於沒有探針車的路段，本研究利用偵測器所取得之所有車種平均速度資料推估兩偵測器間的旅行時間：

$$T_i = \frac{x_{i+1} - x_i}{(V_{i+1} + V_i) / 2} \quad (3.5-1)$$

其中，

T_i ：偵測器 i 到偵測器 i+1 間的旅行時間

x_i ：偵測器 i 的位置

V_i ：偵測器 i 偵測到的速度

(4) 推估模式 2

對於有探針車的路段，則利用即時探針車速度資料加上路段兩端的 VD 速度資料來推估各偵測器間的速度曲

線：

$$V_{i,j}(t) = \alpha_{i,j} * [v_{i,j}(t_{i,j})] + \beta_{i,j} \quad (3.5-2)$$

其中，

$V_{i,j}(t)$ ：在時間 t 時，偵測器 i 到偵測器 $i+1$ 間第 j 段的速度函數

$t_{i,j}$ ：探針車通過偵測器 i 到偵測器 $i+1$ 間第 j 段下游點的時間

$v_{i,j}(t_{i,j})$ ：探針車通過偵測器 i 到偵測器 $i+1$ 間第 j 段下游點的速度，或是偵測器 $i+1$ 偵測到的速度

$\alpha_{i,j,t}$ 、 $\beta_{i,j,t}$ ：在時間 t 時，偵測器 i 到偵測器 $i+1$ 間第 j 段的參數

利用探針車的速度函數即可求得於偵測器間的平均旅行時間 Δt_i ：

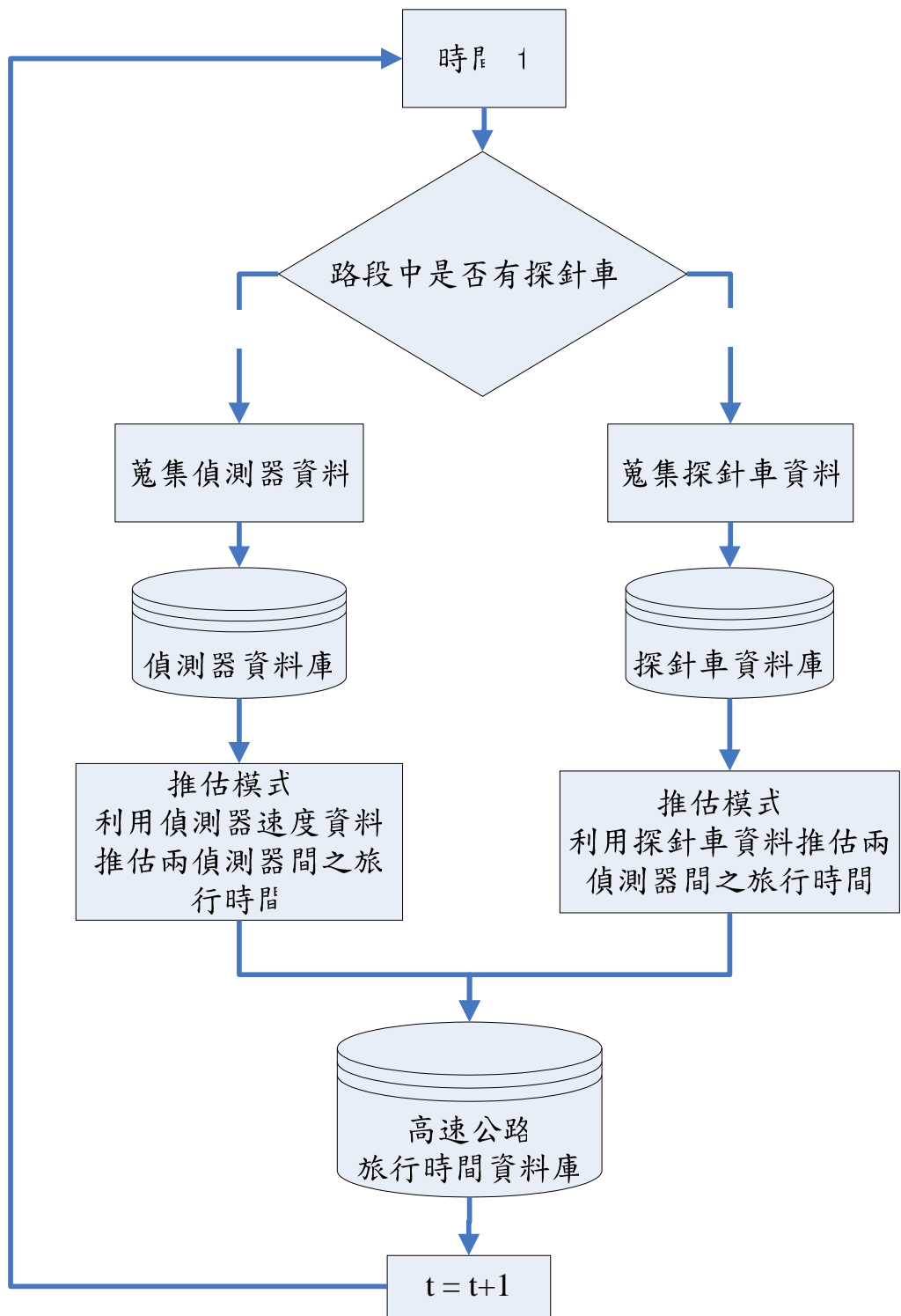
$$x_{i+1} - x_i = \sum_j \int_t^{t+\Delta t_i} V_{i,j}(t) dt \quad (3.5-3)$$

其中

x_i ：偵測器 i 的位置

Δt_i ：偵測器 i 到偵測器 $i+1$ 間的平均旅行時間

(5) 將推估出的旅行時間輸入旅行時間資料庫中。



[資料來源：本研究整理]

圖 3.5-2 旅行時間推估模組

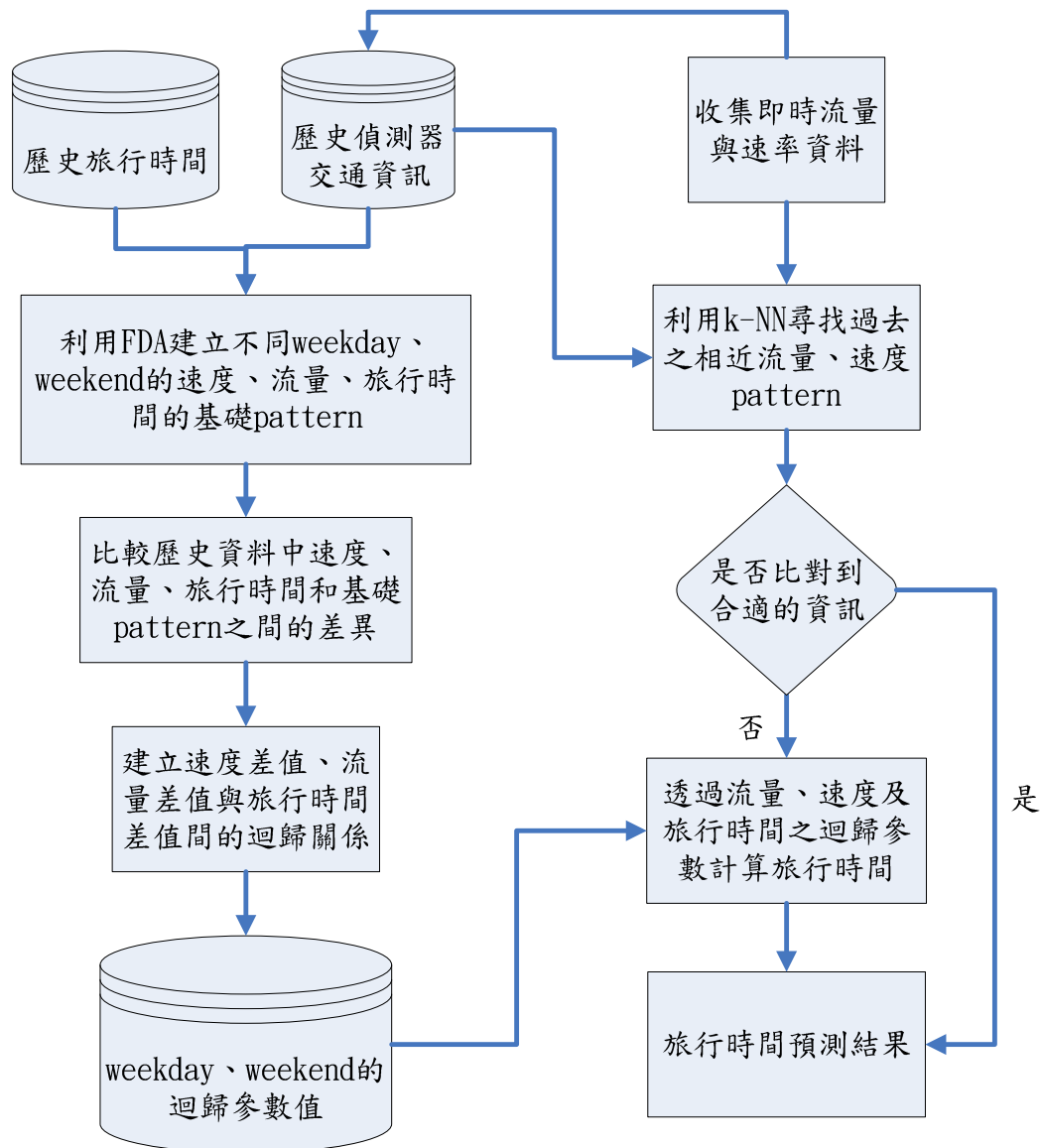
3.5.2 旅行時間預測模組(一般與事件)

一般旅行時間預測模組與事件旅行時間預測模組基本概念與方法是類似的，因此本小節將一般及事件旅行時間預測模組合併說明。唯透過事件旅行時間預測模組預測出之旅行時間由於準確度之考量，僅供內部參考之用，不會發佈。

旅行時間預測模組的目的在於從即時的交通資訊預測由該時間點從起點出發至訖點的旅行時間為何，本研究將利用即時資訊找出該時間點下交通資訊的特徵，並利用該特徵找出擁有類似特徵的歷史資料，進而得到滿足該特徵下的實際日期及時間點。將該日期與時間點資料代入歷史時間資料庫後，得到這些日期的歷史旅行時間平均值，即得到各時間點下，路徑上平均的或各日的歷史旅行時間解。經過 KNN 法可求得即時資訊的旅行時間解，然而道路上之交通情況多變，由於本研究使用之 KNN 法為加入門檻值改良後的 KNN 法，此方法精準度較高，但缺點為不一定可以比對的到資料。因此本研究團體為能持續得到旅行時間預測值，將加入迴歸模式進行預測。

本研究首先透過歷史資料分析，取得各參數（包含旅行時間、車速以及流量）之基礎型態。並將此基礎型態進行分群，進而分析每日變化、有無事件與基礎型態之差異，並求取各分群之參數間關係（如旅行時間與偵測器速度、流量間的關係）。

於旅行時間預測階段，將會藉由即時車輛偵測器資訊，進行特徵辨識以便將此即時資訊歸類於相對應群組中。接著使用該相對應群組中的參數關係來進行旅行時間預測，此預測流程如圖 3.5-3 所示。



[資料來源：本研究整理]

圖 3.5- 3 預估模組流程架構圖

本研究希望可以利用即時的交通資料找出歷史資料和即時資料之間的關係以預測旅行時間，因此將以KNN做為資料分析的方法，先找出那一些歷史資料與即時資料相類似，而進行下一步的研究。

一、KNN 模式

由於不同的交通狀況會對應不同的旅行時間，利用本模式利用加入門檻值後之 *KNN* 模式，找出與即時交通狀況類似的歷史資訊，再由該歷史資訊的旅行時間預測目前的旅行時間為何，以下將介紹 *KNN* 模式。

學者 Robinson 和 Polak(2005)提出如果要建立一個準確的 *KNN* 模式，首先要先決定以下這四個要素：1、決定模式中的特徵向量。2、決定判斷分群的樣本數量。3、利用加權法減少 *KNN* 模式的誤差。4、設定距離量度(distance metric)。以下將針對這四點進行一步說明。

1. 決定模式中的特徵向量。

KNN 的目標是在找尋最接近即時資訊的歷史資訊，本研究利用偵測器即時可收集到的資訊(流量、速率)做為特徵向量，由這些特徵進行比對，找出與即時資訊相類似的歷史資訊。

2. 利用加權法減少 *KNN* 模式的誤差。

由於本研究將同時比對流量與速率資訊的差異，由於這兩個變數的單位不一致，所以需要再設定權重以減少因單位差異而造成的誤差。本研究希望在尖鋒時刻下仍具有相當的準確率，因此以尖鋒時刻通過各偵測器之平均流量與平均速率做為權重設定的判斷方式。

3. 設定距離量度(distance metric)及門檻值。

將即時資訊與各分群資訊的差異加總，即可得到樣本與各分群的特徵差異距離，而距離量度為判斷即時資料與那一組特徵資料相類似的一種判斷指標，當距離量度小於某一個門檻值，則判斷該即時資訊與某一組歷史資料類似，也正如此，可能會造成所有的距離量度皆在門檻值外，造成比對不到資料的情況。然而距離量度的判斷目前並沒有一個很好的準則來對各種資料進行距離量度的判斷，在面對不一樣的環境條件，分析人員必須自行建立一套準則來做處理，因此隨著不同的特徵分群，將會採用不同的距離量度以判斷即時資料與那一個歷史資料類似。

4. 決定判斷類似樣本的數量。

本研究將比對每半個小時下每五分鐘通過各偵測器的流量變化和平均速度變化，再與歷史資料進行比對，比對出該半小時的交通狀況較近似於那些歷史資料，再從這些歷史資料中挑選最接近的 k 筆資料進行旅行時間預測。

在分類及分群之後，本研究目標將即時資訊和歷史資訊相比對，找尋即時資訊和那一些歷史資訊相類似，而找尋方式為找尋歷史與即時資訊的差平方加總在距離量度內時，則判斷該分群的歷史資訊與即時資訊類似，因此可以該天的歷史資訊來預測交通狀況，而判斷的模式建立如下：

$$tss = \sum_{i=1}^T \sum_{j=1}^L [w_q (t_{rij} - t_{mij})^2] \quad (3.5-4)$$

t ：VD 間之旅行時間

L ：比對範圍的路段數量，以偵測器劃分路段。

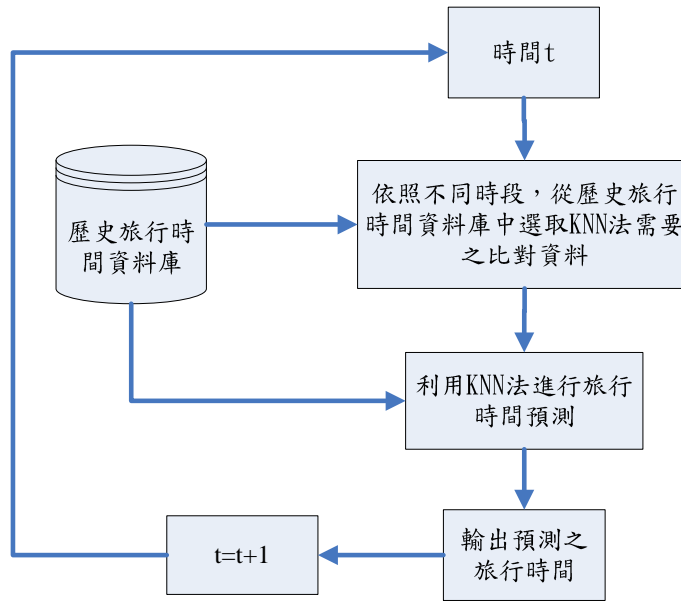
T ：比對的時間長度，本研究將比對一個小時的交通變化。

rij ：即時各時間(i)、各偵測器間(j)之交通資訊。

mij ：歷史各時間(i)、各偵測器間(j)之交通資訊。

w_q ：權重。

在找到類似的分群特徵後，本研究將收集歷史資料中滿足該特徵下實際日期，並代入歷史時間資料庫後，得到這些日期的歷史旅行時間平均值，即可預測各時間點之旅行時間。如圖 3.5-4 所示



[資料來源：本研究整理]

圖 3.5- 4 旅行時間預估模組

二、迴歸模式構建

1. 簡單線性迴歸 (Simple Linear Regression Model)

簡單線性迴歸模式中僅包含兩個變數 (Variables)，且此兩變數間呈現線性之關係，其模式型態如下：

$$Y = \alpha + \beta X + \varepsilon \quad (3.5-5)$$

Y 為依變數 (Dependent variable)。

X 為自變數 (Independent variable)。

α 為迴歸參數 (Regression parameters)，表示截距

β 為迴歸參數 (Regression parameters)，表示直線斜率。

ε 為誤差項。

線性迴歸模式通常可用最小平方法 (Method of Least Squares) 來進行線性迴歸分析，得到 α 與 β 之迴歸參數校估數值，畫出一條符合觀察資料的直線。

2. 多元線性迴歸模式(Multiple Linear Regression Model)

多元線性迴歸又稱為複迴歸，其主要是延伸簡單線性迴歸之觀念，但是在自變數的部分，為探討兩個以上的自變數與應變數間之關係，其模式型態如下式所示：

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3.5-6)$$

其變數與簡單線性模式大同小異，只是自變數的數目由 1 個變成 n 個。多元迴歸分析目的是確認有關變數之間的關係，利用最小平方法建立多元線性迴歸方程，其主要有如下用途：

1. 較容易測得的應變數 X_i 來推算較難預測的因變數 Y 。
2. 描述某一現象時，考慮與某些因素的數量關係。在此用來建立模式來探討各變數之影響力。

多重線性迴歸分析除了求取參數係數值外，還必須確定所得到的參數係數值是否「有效」，在求解迴歸方程前，線性迴歸模型只是一種假設，因此還要進行統計檢定，以確保所得之參數及該迴歸式是否具有說明價值。

3.5.3 旅行資料分析

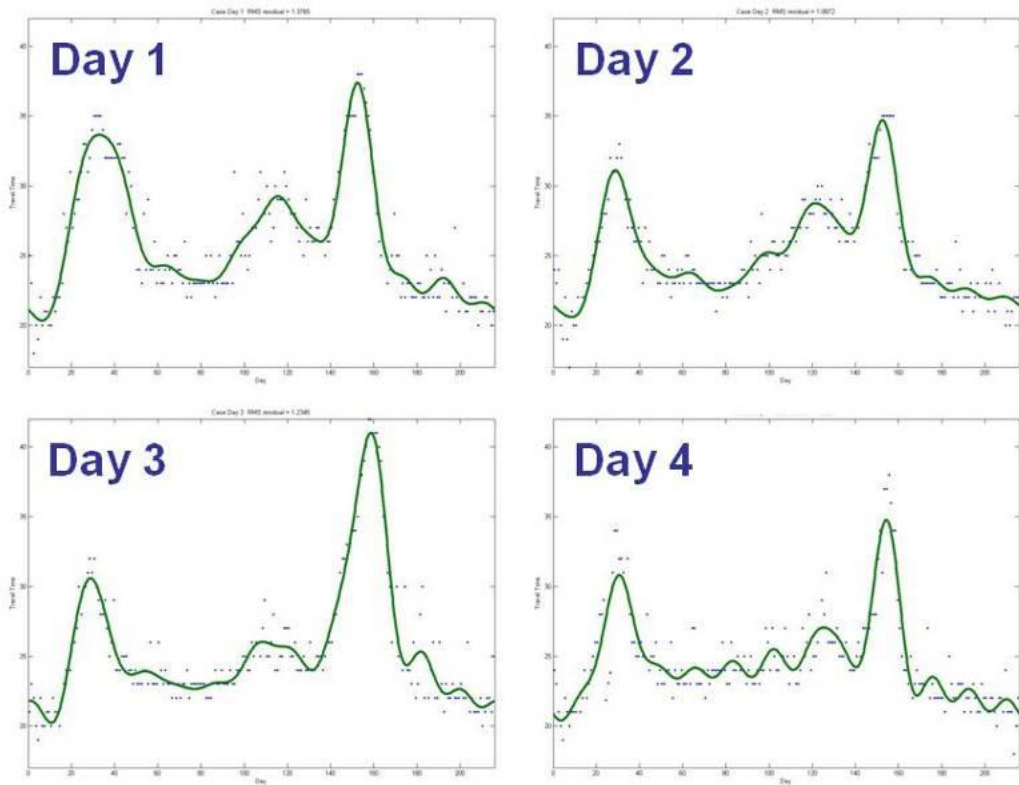
在傳統統計分析中，常會將一組多維度的資料看成多變量(multivariate)資料來進行資料處理，在這種情況下，以多變量統計學中的方法論來進行資料分析；然而，此類型的資料的維度大多是屬於有限(finite)的，在處理可能具有無限(infinite)維度的資料時，多變量統計學中的一些方法論可能就無法完全滿足這類型的資料型態，造成在分析上的顯著性不足或是效果不如預期。

所謂無限維度的資料型態，最簡單的例子就是隨著時間觀察某一事件發生的頻率，由於時間在某種意義上可以看成是無限維度的一個空間，因此這類型的資料都可以看成是一種無限維度的資料型態。本團隊參考 Ramsay 和 Silverman (1997)提出的函數資料分析(Functional Data Analysis, FDA)來處理這類型的資料。此類資料通常稱作函數資料(functional data)，大多以函數(function)或是以圖形(curve)的形式呈

現，因此本團隊認為此方法論將是個可以應用於後續資料分析的一個有利的工具。

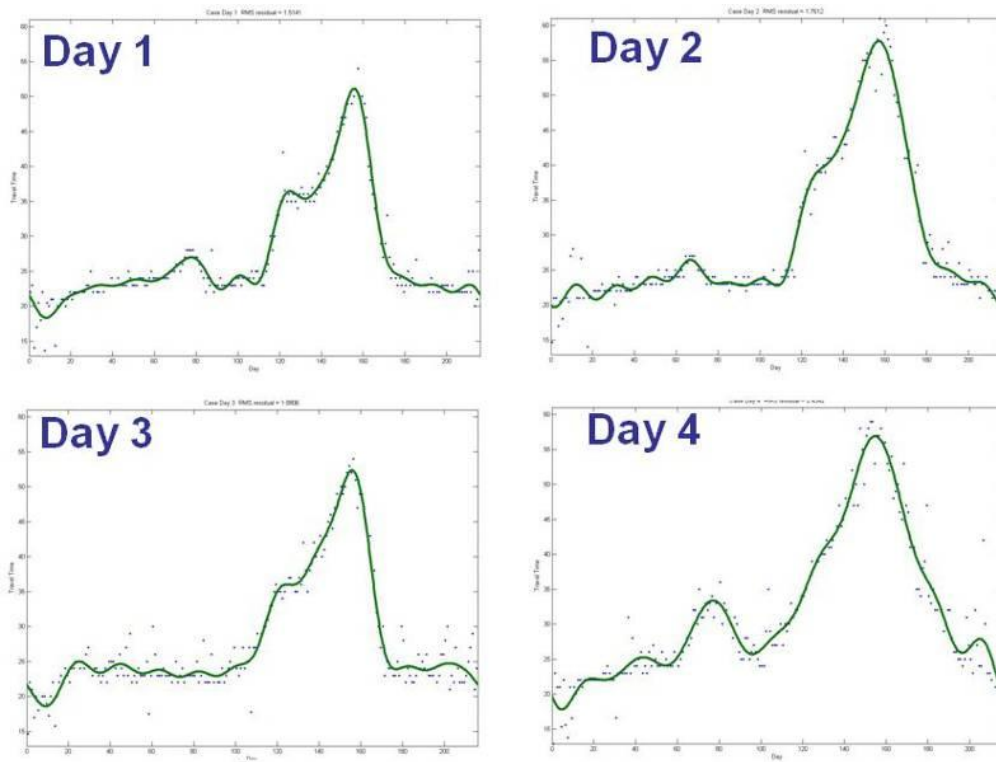
時間的預測主要根據速度、流量、佔有率的資訊來進行，由於旅行時間容易受到天氣狀況、事故意外發生、道路施工所影響，預測過程更顯得複雜。為建立合理的預測模式，首先必須將旅行時間資料進行分析，尋找資料之特性。

下圖 3.5-5 與 3.5-6 分別是將同一條路徑在平日(weekday)與假日(weekend)各挑選四天旅行時間資料，並利用函數資料分析(functional data analysis, FDA)所描繪出之結果。可以發現平日與假日之間旅行時間的分佈有相當的差異。



[資料來源：本研究整理]

圖 3.5- 5 圖 平日旅行時間趨勢線

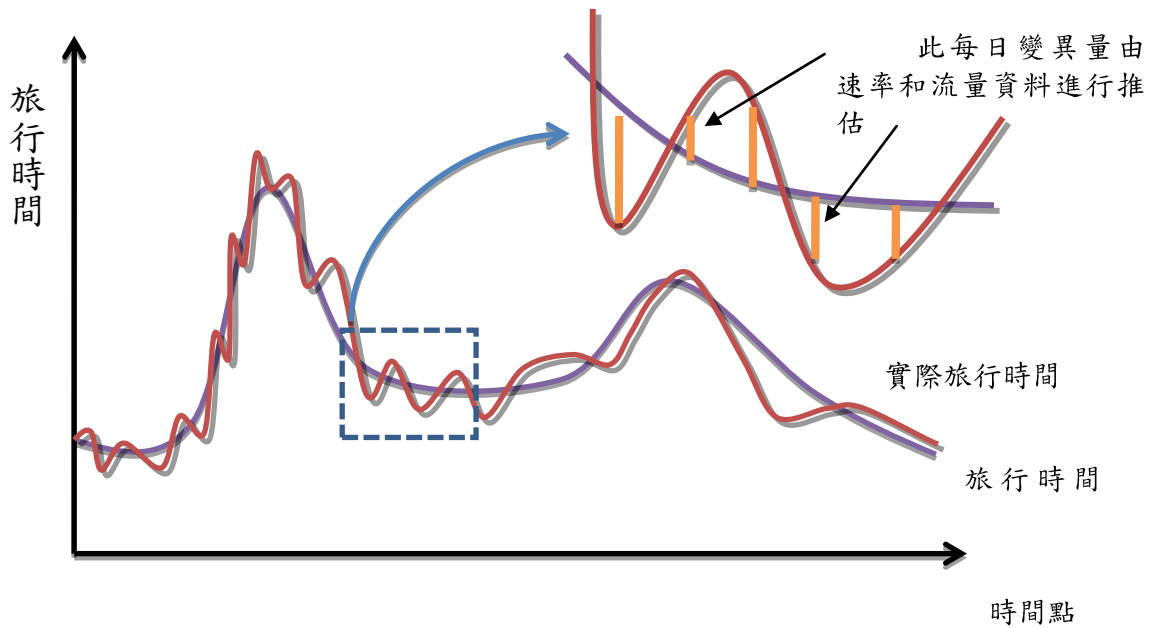


[資料來源：本研究整理]

圖 3.5- 6 假日旅行時間趨勢線

由以上結果發現，假日和平日有不同的交通狀態，因此本模式在不同的分群下，進行旅行時間預測。

參考 Yu 等(2008)的文獻，可以找出即時交通資訊(速率、流量)的變化量先將旅行時間分群，找該分群中旅行時間的基底，以各參數變化量與時間變化量關係建立迴歸模式。模式概念如下圖：



[資料來源：本研究整理]

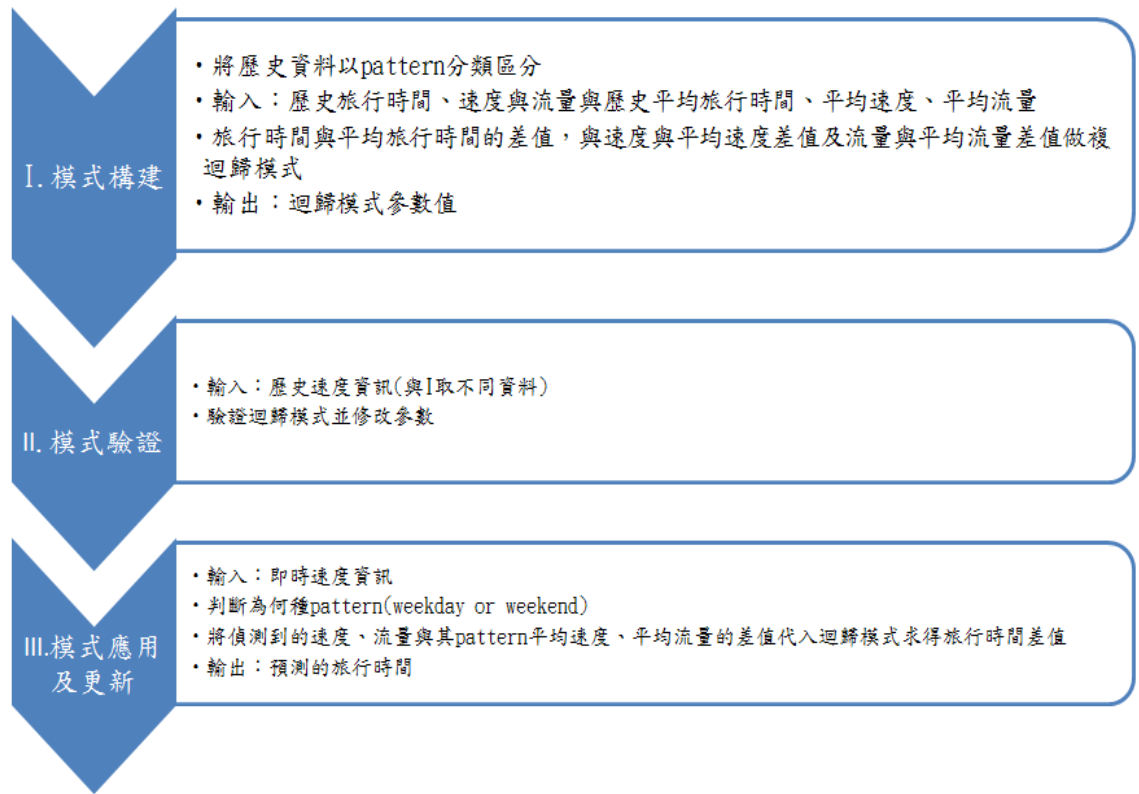
圖 3.5-7 推估概念示意圖

首先是基底曲線(Base curve)的找法，因為同樣分群有相似的參數分布，因此以訓練資料中各參數的日平均曲線來當作各交通參數的基底曲線。模式建立時，先取得資料庫的歷史資訊，以部份當作訓練資料用以校估參數，以每日的實際旅行時間曲線，找出其基底時間曲線，再以該訓練資料中各偵測器所收集的每日的流量與速度曲線找出其基底的流量與速度曲線。實際旅行時間是利用收集各收費站各車的到達時間資訊，因此可得到兩兩收費站間各車輛的實際旅行時間，而即時流量和速率是利用高速公路局各偵測器所偵測到的速度及流量資訊。本模式希望可以找出流量、速率和實際旅行時間參數其基底曲線差異量的關係，即流量、速度與其基底曲線間的差異量，對於旅行時間和其基底平均時間差異的關係。迴歸模式如下：

$$\Delta T^p = \sum_{i=1}^n \alpha_i^p \Delta v_i^p + \sum_{i=1}^n \beta_i^p \Delta f_i^p \quad (3.5-7)$$

其中 ΔT^p 為旅行時間與旅行時間基底值的差異量， Δv_i^p 為速度與速度基底值的差異量， Δf 為流量與流量基底值的差異量， i 為此模式所建構的收費站間的偵測器編號， n 為其間的偵測器數量， p 為分群數量。

流程圖如下圖所示



[資料來源：本研究整理]

圖 3.5- 8 迴歸模式架構圖

3.5.4 事故偵測模式

本研究針對非重現性擁塞部份發展事故偵測模式，由於全國路況中心資料庫已收集所有路況。因此本模式連線到全國路況中心資料庫，並設定事件篩選條件與類型，以取得發生在目標路段上之事故資訊。所採用之篩選規則如下表所示：

表 3.5- 1 事件篩選類型說明

編號	篩選類型	事件說明
1	區域內容	國道3號
2	事件類型	道路施工、交通管制、災變、其他、號誌故障、交通障礙、事故、阻塞（速率小於40km）
3	經度座標	120.987625~121.645861
4	緯度座標	24.757432 ~ 25.07285
5	起迄點	10.9 ~ 100.7

6	方向	北上(N)、南下(S)、雙向
---	----	----------------

事故偵測模式運作方式先連線到全國路況中心資料庫，收尋事件發生的區域是否為國道三號高速公路，並依照所發生的事件查詢是否為所定義的事件類型。若是，再次搜尋事件其發生的起迄點是否為新竹系統與汐止系統之間，依序再收尋發生事件路段的方向。

當推估出路徑旅行時間後，透過事故偵測模式查詢，若是全國路況資料庫查詢到路徑上有無事件發生，本模式定義有事件發生為“1”、無事件發生為“0”，有事件發生時則將推估出之路徑旅行時間存放至事件旅行時間資料庫，並發佈事件資訊，此時仍會透過事件旅行時間預測模組預測旅行時間，但此旅行時間由於誤差較大，因此僅供內部參考，不會公佈；若事件偵測模組沒有偵測到事件發生，將推估出之路徑旅行時間存放至一般旅行時間資料庫，並採用旅行時間預測模組來預測旅行時間。