

國立交通大學
運輸科技與管理學系碩士班

碩士論文

利用灰色理論於偵測器遺失資料插補之研究



Using Grey Theory in the interpolation for
missing value of detector study

研究生：許程詠

指導教授：王晉元

中華民國一百年六月

利用灰色理論於偵測器遺失資料插補之研究

Using Grey Theory in the interpolation for missing value of detector
study

研究生：許程詠

Student : Cheng-Yung Hsu

指導教授：王晉元

Advisor : Jin-Yuan Wang

國立交通大學

運輸科技與管理學系

碩士論文



Submitted to Department of Transportation Technology and Management

College of Management

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Transportation Technology and Management

June 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年六月

利用灰色理論於偵測器遺失資料插補之研究

學生：許程詠 Cheng-Yung Hsu

指導教授：王晉元 Jin-Yuan Wang

摘要

本研究目的為了告知用路人準確的旅行時間預估資訊，最直接的資料蒐集來源為車輛偵測器(Vehicle Detector, VD)，其所回傳的資訊為流量(volume)、速度(speed)與佔有率(occupancy)等。但在實務上常會發生收集的資訊不完全的情況，若忽略偵測器遺漏的情況則會造成旅行時間預估模式發生問題。為了確保其模式預估的正確性，需針對資料遺失值作處理，本研究利用灰色理論不需要符合統計分配的優點，發展出有效的資料遺失值插補方法。在實證分析上以國道三號車輛偵測器為對象，利用灰預測法 GM(1,1)和最小循環式殘差修正法 (Minimum Recursive Residual GM(1,1), MRRGM(1,1)) 在不同資料遺失比例和不同遺漏情境 (任意時間段、尖峰時間段、離峰時間段) 下，比較兩種演算法的插補結果。

研究結果證實當遺失比例高和多重插補次數多時，以 MGGRM(1,1)其插補績效優於 GM(1,1)法。在建立插補模式時亦較其他插補理論簡單，且插補績效相當良好。



關鍵字：灰色理論、資料遺失值插補、MRRGM(1,1)

Using Grey Theory in the interpolation for missing value of detector study

Student: Cheng-Yung Hsu

Advisor: Dr. Jin-Yuan Wang

Department of Transportation Technology and Management
National Chiao Tung University

ABSTRACT

Purpose of this study to inform road users accurate estimates of travel time information, the most direct source of data collection for vehicle detectors (Vehicle Detector, VD), it returns the information to flow, speed and occupancy and so on. But often occurs in practice the information collected is not entirely the case, if the detector ignores the case of missing travel time prediction model will result in problems. In order to ensure the accuracy of their model estimates, the value of data loss need to be targeted for treatment, this study do not meet the statistical distribution of gray theory, the advantages of developing an effective value of the loss of data interpolation methods. In the empirical analysis on the National Highway No. 3 vehicle detectors for the object, using gray prediction method GM (1,1) and minimum cyclic residual correction method (Minimum Recursive Residual GM(1,1), MRRGM (1,1)) in different proportions and different missing data loss situations (arbitrarytime, peak time, off-peak time), the result of comparing the two interpolation algorithms.

The results confirmed that a high proportion of loss when the number of long and multiple interpolation to MGGRM (1,1) the interpolation performance is better than GM (1,1) method. Interpolation mode than in the establishment of other interpolation theory is simple and very good interpolation performance.

Keywords: Grey Model 、 interpolation missing value 、 MRRGM(1,1)

誌謝

首先，在這我要先感謝交通大學運管系的全體老師，因為有你們的教導，使我在短短這兩年的碩士生涯更添加了許多的知識。感謝指導教授王晉元老師在這兩年的期間諄諄教誨，不辭辛勞，也知道老師在背後總是為我們著想，在課業上，老師都會義不容辭的開導我們，老師就像黑夜裡的一盞明燈，指向我朝往正確的道路，這本論文的完成，除了自己的努力之外，還有老師在背後所無法衡量的奉獻，在此，再跟老師說一聲謝謝。

再來，還要感謝兩位口試委員黃寬丞老師與蘇昭銘老師，有了他們的寶貴評論與建議之下，使我的論文能夠更佳的完善且充實，在此也向兩位老師致上十二萬分的感謝。

另外，感謝一起進入研究所、一起成長、一起生活的夥伴們，小豬、阿胖、NoNo、馬代、凱開、QQ、鱷魚和所有碩二的同學們，雖然只有短短的兩年，但因為有了你們，使我的碩士生涯充滿了歡樂與色彩，我永遠不會忘記這兩年的點點滴滴，還有碩一的學弟妹們，阿幹、呂璇、老頭、佩慈，我們的感情，就好比家人一樣，雖然只有一年的相處時間，但以後卻可以成為一輩子的朋友。還有感謝每次都被我煩的彥佑、一姐、hoho、黛西、小惠、阿尼吉，在研究所期間對我的教導，讓我變的更強壯，去面對未來的困境。畢業，不是結束，而是另一段路的開始。

我要感謝我的家人 28 年來的照顧，給予我無限的關懷與愛，讓我有今天的成就，在求學過程中，每次到外地念書，都會讓你們擔心，但因為你們的擔心，使我更能警剔自己用功唸書。在這裡，我要大聲的跟全部的人說，我很得意，我出生在我的家庭，另外感謝我的女友，徐莎，雖然常常因為論文的事情而冷落了你，但因為有你的體諒，讓我更能加倍認真，當我遇到挫折時，你會適時的鼓勵我，當我徬徨無助的時候，你會偷偷在背後幫助我。

最後，我要感謝我認識的所有人，因為有你們，才会有今天的我。

許程詠謹誌

2011 年 6 月於新竹交大

目錄

中文摘要	i
英文摘要	ii
誌謝	iii
目錄	iv
表目錄	vi
圖目錄	vii
第一章、緒論	1
1.1 研究動機.....	1
1.2 研究目的.....	2
1.3 研究流程.....	2
第二章、文獻回顧	5
第三章、模式建立	9
3.1 灰色理論介紹.....	9
3.2 模式建立.....	11
3.2.1 蒐集交通資料.....	13
3.2.2 異常資料過濾模式.....	13
3.2.3 歷史資料庫的建立.....	14
3.2.4 灰關聯分析(GREY RELATION ANALYSIS; GRA).....	15
3.2.5 灰預測 GM(1,1)	16
3.2.6 最小循環式殘差修正(MINIMUM RECURSIVE RESIDUAL GM(1,1); MRRGM(1,1))	19
3.3 範例說明.....	23
3.4 誤差分析.....	27
第四章實證分析	28
4.1.1 資料蒐集.....	28
4.1.2 異常資料過濾.....	29
4.2 建立資料庫.....	29
4.3 實證方法.....	29
4.3.1 隨機時間段遺漏.....	30
4.3.1.1 任意時段遺漏.....	30
4.3.1.2 尖峰時段遺漏.....	31
4.3.3.3 離峰時間段遺漏.....	33

4.3.2 比對方式.....	34
4.3.3 插補次數.....	34
4.4 測試方式.....	36
4.5 測試範例.....	36
4.5.1 測試對象.....	36
4.6 測試結果.....	36
4.6.1 以 N3-47.98 為測試對象	38
4.6.2 以 N3-50.38 為測試對象	43
4.6.3 以 N3-51.55 為測試對象.....	49
4.6.4 以 N3-54 為測試對象.....	55
4.6.5 以 N3-55.34 為測試對象.....	61
第五章 結論與建議	69
5.1 結論.....	69
5.2 建議.....	69
參考文獻	71
簡歷	74



圖目錄

圖 1. 1 速度資料遺漏(VD-N3-N-47.980-M)	1
圖 1. 2 研究流程圖	4
圖 3. 1 灰色理論示意圖	9
圖 3. 2 插補模式流程圖	12
圖 3. 3 速率異常分佈圖(VD-N3-N-47.980-M)	14
圖 3. 4 計算流程圖	22
圖 3. 5 車輛偵測器(VD)資料遺失示意圖	23
圖 4. 1 均勻分配機率密度函數	30
圖 4. 2 任意時段遺漏起始點未超過最大遺失比例 50%示意圖	31
圖 4. 3 任意時段遺漏起始點超過最大遺失比例 50%示意圖	31
圖 4. 4 尖峰時間段遺漏起始點未超過最大遺失比例 50%示意圖	32
圖 4. 5 尖峰時間段遺漏起始點超過最大遺失比例 50%示意圖	32
圖 4. 6 離峰時間段遺漏起始點未超過最大遺失比例 50%示意圖	33
圖 4. 7 離峰時間段遺漏起始點超過最大遺失比例 50%示意圖	33
圖 4. 8 實證分析流程圖	35
圖 4. 9 不同遺失比例下插補誤差(N3-47.98 任意時段)	38
圖 4. 10 不同遺失比例下插補誤差(N3-47.98 尖峰)	39
圖 4. 11 不同遺失比例下插補誤差(N3-47.98 離峰)	40
圖 4. 12 不同插補次數下插補誤差 (N3-47.98 任意時段)	41
圖 4. 13 不同插補次數下插補誤差(N3-47.98 尖峰)	42
圖 4. 14 不同插補次數下插補誤差(N3-47.98 離峰)	43
圖 4. 15 不同遺失比例下插補誤差(N3-50.38 任意時段)	44
圖 4. 16 不同遺失比例下插補誤差(N3-50.38 尖峰)	45
圖 4. 17 不同遺失比例下插補誤差(N3-50.38 離峰)	46
圖 4. 18 不同插補次數下插補誤差(N3-50.38 整日)	47
圖 4. 19 不同插補次數下插補誤差(N3-50.38 尖峰)	48
圖 4. 20 不同插補次數下插補誤差(N3-50.38 離峰)	49
圖 4. 21 不同遺失比例下插補誤差(N3-51.55 任意時段)	50
圖 4. 22 不同遺失比例下插補誤差(N3-51.55 尖峰)	51

圖 4. 23 不同遺失比例下插補誤差(51.55 離峰).....	52
圖 4. 24 不同插補次數下插補誤差(N3-51.55 任意時段)	53
圖 4. 25 不同插補次數下插補誤差(N3-51.55 尖峰).....	54
圖 4. 26 不同插補次數下插補誤差(N3-51.55 離峰).....	55
圖 4. 27 不同遺失比例下插補誤差 (N3-54 任意時段)	56
圖 4. 28 不同遺失比例下插補誤差(N3-54 尖峰).....	57
圖 4. 29 不同遺失比例下插補誤差(N3-54 離峰).....	58
圖 4. 30 不同插補次數下插補誤差(N3-54 任意時段).....	59
圖 4. 31 不同插補次數下插補誤差(N3-54 尖峰)	60
圖 4. 32 不同插補次數下插補誤差(N3-54 離峰).....	61
圖 4. 33 不同遺失比例下插補誤差 (N3-55.34 任意時段)	62
圖 4. 34 不同遺失比例下插補誤差(N3-55.34 尖峰).....	63
圖 4. 35 不同遺失比例下插補誤差(N3-55.34 離峰).....	64
圖 4. 36 不同插補次數下插補誤差 (N3-55.34 任意時段)	65
圖 4. 37 不同插補次數下插補誤差(N3-55.34 尖峰).....	66
圖 4. 38 不同插補次數下插補誤差(N3-54.33 離峰)	67



表目錄

表 2. 1 插補遺失值文獻整理	7
表 2. 2 車輛偵測器(VD)資料表	13
表 2. 3 歷史資料庫欄位表	14
表 2. 4 GM(1, 1)計算值	25
表 2. 5 MRRGM(1, 1)計算值	26
表 2. 6 MAPE 之評估標準	27
表 2. 7 龍潭到樹林收費站北上路段間偵測器之資料缺漏率	28
表 2. 8 歷史資料庫欄位表	29
表 2. 9 車輛偵測器(VD)任意時段最佳插補次數	68
表 2. 10 車輛偵測器(VD)尖峰時段最佳插補次數	68
表 2. 11 車輛偵測器(VD)離峰時段最佳插補次數	68



第一章、緒論

1.1 研究動機

先進旅行者資訊系統(Advance Traveler Information System, ATIS)是智慧型運輸系統(Intelligent Transportation System, ITS)的子系統之一。近年來政府投資了許多經費和資源於 ATIS 中的發展，其功能除了彙整所有交通訊息之路交通資訊的提供和行車路徑導引。其目的為提供民眾在「行」方資訊，並透過道路相關設施所回傳的路況，提供用路人即時或預測包括何時出發、行駛哪條路徑、到達時間為何，以協助用路人於資下做決策。而在 ATIS 所提供的功能中，以旅行時間預測是大家所預測可減少用路人的旅行時間，並間接促進經濟發展。

為了告知用路人準確的旅行時間預估資訊，首要作業是先收集料，最直接來源為車輛偵測器(Vehicle Detector, VD)，其所回傳的數量(volume)、速率(speed)與佔有率(occupancy)等。但在實務上常會訊不完全的情況，如面臨天候因素、偵測器故障、通訊傳輸斷線、異常或斷電等，而造成偵測器每日所傳回資料在不同時間段或是連失(missing)。經由實際蒐集 2009/04/05~2010/04/05 期間國道三號北到樹林收費站間某一偵測器所回傳的歷史資料，其回傳的頻率為4其中各偵測器遺漏值情況不盡相同，經由資料分析結果如圖 1.1 所(VD-N3-N-47.980-M)於時間點 16:00~18:40 顯示所回傳的資訊的確形。

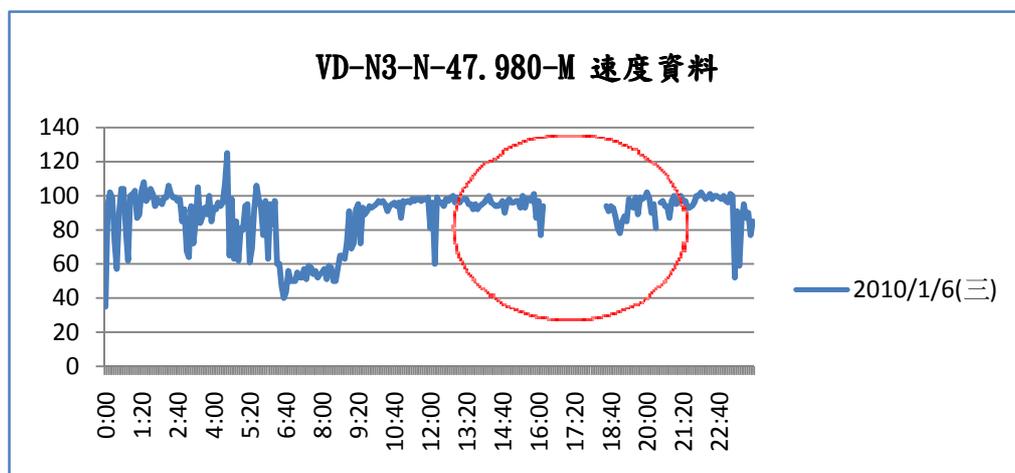


圖 1. 1 速度資料遺漏(VD-N3-N-47.980-M)

若忽略偵測器遺漏的情況則會造成旅行時間預估模式發生問題，為確保其模式預估的正確性，當務之急必先解決資料遺失的問題，因此針對遺失資料發展插補演算法是相當重要的課題。

而在遺失資料的處理上，最早是由 Little and Rubin[1987]提出將含有遺失值的資料項目(欄位或紀錄)直接刪除，僅利用剩餘的完整資料(complete data)進行分析。若所蒐集的資料量很大時，僅遺失一、兩筆資料則對後續推估沒有影響；倘若遺失比例過多的情況下，則所推估出來的結果會發生問題。

目前已有多位學者提出遺失值插補的相關研究，如 Gold[2001]事先假設所有資料必先滿足某一統計分配下才可插補，因此需事先檢定大量歷史資料後，再依最大值(EM)法、最大概似法(MLE)等方法針對遺失值做插補。在資料探勘方面，有許多插補方法被提出，如 Shen and Chen[2003]關聯性法則(association rules)、Lee and Mong[1976]叢集分析與 Pedreira and Parente[1995]機械式學習等，其概念皆是應用資料探勘的工具用於大量歷史資料中，期許能找出有利於遺失值的插補方式，進而正確的推估其遺失值。

但在實際應用上並非所蒐集的歷史資料都是大量的，有時僅能取得少數的歷史資料，若遇到歷史資料量少的情況下，其資料無法滿足統計分配而導致其方法失效。因此，本研究利用灰色理論中灰預測可用於少量資料預測的優點，發展出更有效的遺失資料插補模式。

1.2 研究目的

本文研究目的主要為利用車輛偵測器(VD)之資料，發展出一套演算法，針對資料遺失值部份進行插補，並就其插補績效與其它演算法相比較。

1.3 研究流程

本研究之研究流程如圖 1.2 研究流程圖所示，其詳細步驟說明如下：

(1) 確認研究問題

決定實際研究對象以及欲進行插補的資料，並根據研究背景與目的將問題清楚描述與界定。

(2) 相關文獻回顧

針對遺失資料插補演算法之相關研究進行文獻蒐集與回顧，以插補類型不同區分，如速度、流量、佔有率等。並分析灰色理論於資料遺失插補之適用性，將其優缺點加以歸納，進而整理出本研究之研究方向。

(3)建立插補模式

利用灰色理論等方法建立遺失資料插補模式。

(4)資料處理

蒐集車輛偵測器資料，經由資料過濾處理後，將非交通狀況所造成的異常值(suspicious data)刪除，如速率異常、流量異常等；並建立歷史資料庫。

(5)實證分析

本研究為了驗證插補後資料精確度與可用性，利用車輛偵測器(VD)歷史資料發展出遺失資料插補模式，並在不同之資料遺漏比例下測試其模式的精確度。如先尋找一組沒有遺失的原始資料，並隨機刪除不同比例之資料，再利用本模式進行遺失資料插補。

藉由計算插補後的遺失值與原本資料之平均絕對誤差(Mean Absolute Percentage Error, MAPE)，來驗證此模式是否適用於現實狀況，並分析結果以作為修正模式之依據。

(6)結論與建議

將本研究之過程與結果提出結論與建議，並提出後續可以進行研究的方向。

以下為本論文研究流程：

第二章、文獻回顧

一般在處理資料遺漏值的方式有很多種，可分為資料探堪和統計方法兩大類。在資料探勘方面如關聯法則(Association Rules)、叢集分析(Clustering Analysis)、機械式學習(Machine Learning)與分類迴歸樹(Classification and Regression Tree)等方法，找出有關資料遺失值的資訊。然而，統計插補法最早由Rubin[1987]所提出，如平均值插補法(Mean Imputation)、迴歸插補法(Regression Imputation)等，後續的學者常搭配其他演算法和統計法來插補遺失資料值，其文獻說明如下所示：

Gold[2001]利用期望最大值演算法(Expectation Maximization Algorithm, EM)和 Kernel 迴歸方法插補遺失值。EM 插補法是針對有母數中遺失值的處理方法，主要分為「E(Expectation)」和「M(Maximization)」兩個步驟，計算方式和最大概似法(MLE)方法相同，只差別在求解概似函數的不同，計算方式是透過迭代的方法求取其參數估計量。但在實際應用上以常態分配(Normal Distribution)來描述該偵測器資料屬於那一群集的機率密度，而此機率函數用來取代群集的距離函數。在計算資料分群後，再利用因子分析(Factor up)和直線式插補(Straight-line interpolation)兩種插補方式，結合多項式迴歸與 Kernel 迴歸兩種迴歸法針對 VD 五分鐘、十分鐘的流量和占有率進行填補。結果發現多項式迴歸和因子分析的插補績效優於核迴歸法和直線式插補法。

Chen[2001]使用兩步驟插補遺漏值。先以自組織映射圖(Self-Organising Map, SOM)輸入車輛偵測器(以 15 分鐘為單位)之資料(速率、流量與佔有率)作為交通型態的判別，然後再使用 ARIMA 和另兩種類神經網路 MLP 和 RBFN 來插補遺漏值，其方式為取遺漏值發生時段前的 4 筆歷史資料(一小時資料)進行流量的插補。結果顯示 SOM/ARIMA 插補績效優於 SOM/MLP 與 SOM/RBFN。

Hung and Zhu[2002]以擬最鄰近法(pseudo nearest neighbor)進行資料插補，先以擬相似(pseudo-similarity)法找出各完整數列和不完整數列的相同元素值和遺失值(Null)，並以權重法比較兩兩數列的比值大小，然後再以最鄰近參考數列之對應數值進行不完整資料之插補。實驗假定資料都符合常態分配並用 k-mean 將資料分群，接著再和其他三種方法做比較，結果顯示擬最鄰近法之插補績效較好。

不過在實際應用上並非所有的交通資料都符合常態分配，因此 Hung and Lee[2004]提出改善以灰關聯擬最鄰近法來插補遺漏值，並先利用灰關聯分析找出與遺失的屬性值關聯大的附近屬性值，然後再以 k-最鄰近(k nearest neighbors, k-NN)法進行插補。實驗結果證明，灰關聯擬最鄰近法插補績效優於多值插補法(Multiple Value Imputation)與平均值替代法(Mean Substitution)。

Wen *et al.*[2005]沿用 Hung and Lee[2004]和 Hung and Zhu[2002]的概念使用灰關聯擬最鄰近法(grey-based pseudo nearest neighbor approach)和灰色時間序列法(grey time-series model)，針對車輛偵測器在時間上所產生之遺漏值進行資料的填補。其次使用遞迴式類神經網路(Recurrent Neural Network, RNN)進行高速公路旅行時間的預測。結果顯示，當遺漏值產生率介於 50% ~ 80% 時，誤差約為 20%，顯示填補完後的資料用於建立旅行時間預測模式亦有不錯的績效。

Chen *et al.*[2006]以五種插補遺失值的方法進行比較分析，包括歷史平均法(HIST)係取同樣資料的前一天(time of day)和前一週(time of week)資料插補；時間差補法(TPI)係取遺失值發生的前一時段與後一時段的平均值做插補；空間差補法(SPI)係利用發生遺漏值之偵測器的上下游偵測器插補；混合插補法(Hybrid)係將歷史資料經過指數平滑法處理後，再以時間序列插補遺失值；類神經網路(ANN)模仿人類神經運作方式，藉由不斷學習過程插補遺失值的預測模式。結果顯示以混合模式的績效最好。

Zang and Liu[2009]使用最小二乘支持向量機(Least Squares Support Vector Machines, LS-SVMs)，在時間和空間的架構下預測市區道路遺漏的車流量資料，此方法為非線性迴歸模式，具有快速收斂、高精準度和運算過程簡單等特性。以皮爾森積差相關係數分群(Pearson Product Moment correlation coefficient)對 VD 歷史資料分群，再利用期望最大值法(expectation maximization, EM)與資料擴增法(data augmentation, DA)法做比較，結果發現 LS-SVMs 在單筆或是連續遺漏資料插補績效優於 EM 和 DA 法。

Li Qu *et al.*[2009]針對市區道路車流量，提出以機率主成份分析(Probabilistic Principal Component Analysis, PPCA)進行插補，先以主成份分析法(PCA)分析歷史資料，再以最大概似法(MLE)進行遺漏插補。並利用最鄰近歷史資料(Nearest Historical)和歷史平均值(Mean Historical)，再以迴歸分析進行預測比較，結果發現 PPCA/MLE 插補績效比歷史資料和迴歸、歷史平均值和迴歸插補績效來得好。

張堂賢與黃宏仁[2008]建立即時資料插補技術，並利用傅立葉轉換和中央極限定理來平滑車輛偵測器的時間曲線，以降低交通號誌的躁動。然後利用卡爾曼濾波器和基因演算法發展車輛偵測器遺漏資料插補技術。因卡爾曼濾波器可考量資料遺漏前期資料走勢，可提高插補準確度，結果可獲得良好之插補績效。

吳健生與廖梓淋[2009]於資料填補分析時，採用華德法與 k-mean 法將資料進行兩階段分群，再以倒傳遞類神經網路加以填補，最後以雪山隧道車輛偵測器所得流量、速度及佔有率三項資料進行實證分析，以速率填補績效最佳。其作法為先分群而後再填補偵測器遺漏資料之兩步驟方式，因此能將同質性高之車流狀態資料匯集在一起分別處理，而能獲得良好的插補績效。

汪進財與邱孟佑 [2009]使用兩步驟來插補資料遺漏值，先以群集分析法對車輛偵測器歷史資料做交通狀態分類處理，並參考 Wong and Chiu[2009]基本車流理論之服務水準來分類；再根據某偵測器所代表之次路段結合整體路段 ETC 之旅行時間來建立整體迴歸關係，並依照不同遺漏時窗多寡來選用不同插補方法，如歷史均值法、移動平均法(MA)、分類迴歸樹(CART)。實驗結果顯示當連續遺漏時窗數超過 10 筆時採用 CART 法可獲得插補最佳績效。

本研究將插補遺失值之文獻相關整理如下表 2.1：

表 2. 1 插補遺失值文獻整理

作者	研究方法
Gold[2001]	EM 演算法、多項式迴歸、Kernel 迴歸
Chen[2001]	SOM、MLP、RBFNN
Hung and Zhu[2002]	擬最鄰近法、權重法
Hung and Lee[2004]	灰關聯分析、KNN
Wen et al.[2005]	灰關聯擬最鄰近法、RNN
Chen et al.[2006]	歷史平均法(HIST)、時間差補法(TPI)、空間差補法(SPI)、混合插補法(Hybrid)、類神經網路(ANN)
Zang and Liu[2009]	皮爾森積差相關係數、LS-SVMs、EM、DA
Li Qu et al.[2009]	PPCA、MLE
張堂賢與黃宏仁[2008]	傅立葉轉換、卡爾曼濾波器和基因演算法
吳健生與廖梓淋[2009]	華德法與 k-mean 法、倒傳遞類神經網路(BPANN)
汪進財與邱孟佑 [2009]	群集分析、分類迴歸樹(CART)

資料來源：【本研究整理】

綜合上述相關文獻回顧得知，插補方式是採用車輛偵測器所回傳的歷史資料，並針對時間段的遺漏值做插補。其方法主要分為兩步驟處理：

首先利用 EM、SOM、群集分析與 k-mean 等方法找出同質性高的資料；再依照資料型態的不同插補資料遺漏值，如迴歸方法、MLE、KNN、類神經網路(ANN)與基因(GA)演算法等。

然而在利用 MLE、EM、多項式迴歸、Kernel 迴歸與 LS-SVMs 插補時，前提條件為歷史資料要先符合統計分配與不偏性的要求，才可獲得良好的插補績效。但文獻上並未提及歷史資料不符合統計分配的情況下應該如何處理，且在實務上常存在著資料不符合統計分配的情況。而 KNN 和 ANN 的使用上則需要設定門檻值，並以多次試誤的方式才能得到適當的解。

綜合上述缺點，Hung and Lee[2004]和 Wen[2005]提出灰關聯方法比對其歷史資料，其優點為基於灰色理論，並可引入灰關聯度作為相鄰資料的相似度衡量的概念，與歐氏距離所延伸 Minkowski distance 相似度衡量方法雷同，其限制為當

資料型態不同時則需採用不同維度的距離尺度來衡量，因此在大量歷史資料比對下會增加計算的複雜度。而 Zhang[2007]提出灰關聯法在比對距離尺度時其計算方式簡單，可先將不同資料型態做正規化動作後再做運算，則可簡化計算的複雜度。

本研究採取 Hung and Lee[2004]和 Wen[2005]所提出灰關聯的優點來做歷史資料比對的方式，所比對為車輛偵測器的歷史速度資料，其資料型態為相同的，因此可省略正規化步驟直接運算。

Wen[2005]將含有遺失值之資料採用灰關聯擬最鄰近方法來尋找歷史資料，當灰關聯度越大時表示可比對的歷史資料有較少的不完整數列，並以關聯度最大之數列資料取代其含有遺失值之數列資料，其缺點為歷史資料不一定吻合於原資料，且在歷史資料量少的情況下，可能會比對到重複的歷史資料值。

Hung[2004]先利用灰關聯分析找出與遺失屬性值關聯大的附近屬性值，接著利用 KNN 法進行遺失值插補，找出 k 筆相似紀錄以推估遺失值，其缺點為 k 值選取會影響到遺失值的誤差，因此需要測試多筆 k 值才能選出誤差最小的 k 值。

有鑑於此，本研究採用灰色理論中不需符合統計分配之優點，採用灰關聯法來比對歷史資料，並利用灰預測少量數據預測的優點，可在歷史資料不足的情況下，針對現有遺失值進行插補。但其缺點為在長時間預測下，當歷史數據資料呈現不規律時，其預測結果較易產生波動，因此預測誤差也會變大，本研究提出改善方法以提昇灰預測的準確度。

綜言之，本文將利用灰色理論發展插補法，針對車輛偵測器(VD)的速度作插補資料之遺失值。第三章將說明灰色理論原理並建立灰色理論資料插補模式；第四章說明如何處理原始資料與實證分析；第五章結論與建議。

第三章、模式建立

本研究以灰色理論為基礎，針對資料遺失值部份進行插補，並利用灰關聯分析法(GRA, Grey Relation Analysis)簡化資料複雜性之缺點，從原本的歷史資料中找出相似的歷史數列並計算其灰關聯度，找出含有遺失值數列與歷史數列間相似性的特徵。然後再利用灰預測方式有效處理遺失資料插補問題。

此外，本研究提出之修改型的灰預測方法，能有效解決灰色理論的缺點。

3.1 灰色理論介紹

灰色理論係由鄧聚龍[1982]所提出之訊息區間分析概念，該理論將訊息分為白色、黑色、灰色等，白色為系統內資料完全且為大樣本，可用統計方法分析，例如迴歸分析或時間序列等方法加以預測；黑色則為系統內資料不知且無法蒐集到任何樣本，必須利用偏好性理論進行分析，例如透過問卷調查建立模式來預測；若系統內可知某部份特性且可蒐集到少量樣本，則該訊息稱為灰色，可用灰色理論中灰色模型來預測。其理論係結合數學、系統與力學等觀點所發展出的灰色模型。主要是針對在系統模式不明確、資訊不完整之情況下，進行模式的關聯性分析(Relation Analysis)與模式建構(Construct Model)，並藉由預測(Prediction)及決策(Decision)之方法來探討並了解其系統。灰色理論的應用極為廣泛，能對多變量資料輸入、離散資料與不完整數據做有效的處理，而其理論主要由灰關聯分析、灰生成、灰建模與灰預測等方法所組成，其方法如下：

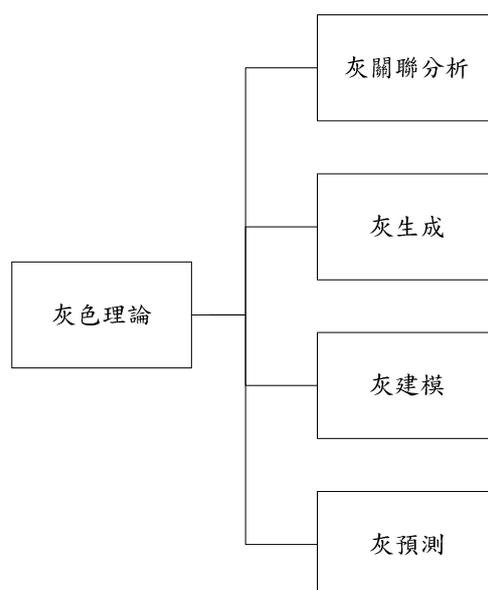


圖 3. 1 灰色理論示意圖

(1) 灰關聯分析(Grey Relational Analysis)

灰關聯分析法(GRA)是一種相對性的排序觀念，用來衡量離散數列間的相關程度，並根據數列間曲線幾何形狀的相似程度來判斷其關聯性，越相近者其關聯度越大，反之則小。由於灰關聯分析具備不需要大量數據即可進行「多因素、同時比較」的優點，其具體作法可區分為局部性關聯分析和整體性關聯分析兩類，前者適用於有參考數列時之關聯性比較，反之當參考數列不存在時則可利用後者來進行數列間比較，而本研究因有不完整資料(含有遺失值的資料)作為參考數列，因此可採用局部性灰關聯分析。

(2) 灰生成(grey generating)

灰生成為補充訊息的數據處理技巧，常透過數據整理來尋找其數據變化的規律性，其方法將原始數據依序做累加(累減)的動作，以降低原始數據雜亂程度，讓數據呈現有規律的分佈型態，得以尋找其被掩蓋的潛在規則和特徵。

(3) 灰建模(Grey Model Construction)

利用灰生成數據可建立一組灰色差分方程(Grey difference equation)與灰色擬微分方程(Pseudo differential equation)之模式，稱之為灰色模式 GM(h,n)，其中 G 為 grey、M 為 model、h 為微分方程式階數、n 為所輸入的變數；其中又以一階和單變數 GM(1,1)模型最常使用，本研究採用 GM(1,1)為建模的基礎。

(4) 灰預測(Grey Prediction)

灰預測以 GM(1,1)模型為基礎並對已知數據的變化進行預測，不同於統計迴歸分析需大量數據與典型分配之限制，其優點在於少量數據($n > 4$)即可預測未來數據的變化。此模型可應用於數列預測、異常值預測、拓撲預測等。

在現實世界裡原始資料蒐集和傳輸過程中會發生資料缺漏或異常問題，為確保資料可靠度，一般方式為將原始資料過濾後含有遺失值的紀錄和屬性刪除，利用剩餘完整資料來做分析，此種方式容易造成分析上的偏誤，導致其結果不佳。因此為了尋求資料的完整性，有必要插補資料中的遺失值。

而為彌補上述問題，本研究採取灰色理論中灰關聯分析(GRA)的優點建立遺失資料插補模式，並針對灰色理論不足之處提出有效的解決方案。由鄧聚龍[1982]研究中提出灰預測的缺點，當原始數據呈現劇烈變化時，利用灰預測 GM(1,1)模型所預測出來的精確度誤差較大，因此已超出合理的預測範圍。

而最早的改善法由劉思峰[1991]利用殘差數列建立 GM(1,1)模型，對原來的模型進行修正，以提高預測精確度。之後有多位學者提出改善方法如灰色馬可夫(Grey Markov)殘差修正法、灰色基因(Grey GA)殘差修正法、灰色類神經法(Grey BP)等，其目的主要改善灰預測的缺點。

因此本研究所插補方法利用灰預測模型 GM(1,1)為基礎，並針對其缺點提出改善。

3.2 模式建立

本研究演算流程分為兩部份，分為歷史資料庫的建立和遺失資料插補模式。

在歷史資料庫的建立上，需先蒐集一段時間車輛偵測器(VD)的原始資料，而所蒐集到的資料通常無法直接使用，大部分所取得資料是非常雜亂且不完整的，如受到通訊斷線、收集設備異常等問題。因此需將原始資料透過異常資料過濾模式的處理，才能建立歷史資料庫。若是直接使用原始資料，將會影響到後續插補的績效。

而在含有遺失值的資料處理上，本研究透過灰色理論所發展的插補模式作插補，其說明如下：

1. 先透過灰關聯分析(GRA)法去尋找歷史資料庫，並找出與含有遺失值的樣本，有相同特性的歷史資料。
2. 將尋找出的歷史資料透過灰預測 GM(1,1)方法插補出資料遺失值。

此外為了改善灰預測 GM(1,1)模式之缺點，本研究將改良灰預測 GM(1,1)模式以提昇插補準確度，其計算內容於 3.3 節中會詳述。

以下為插補模式流程說明：

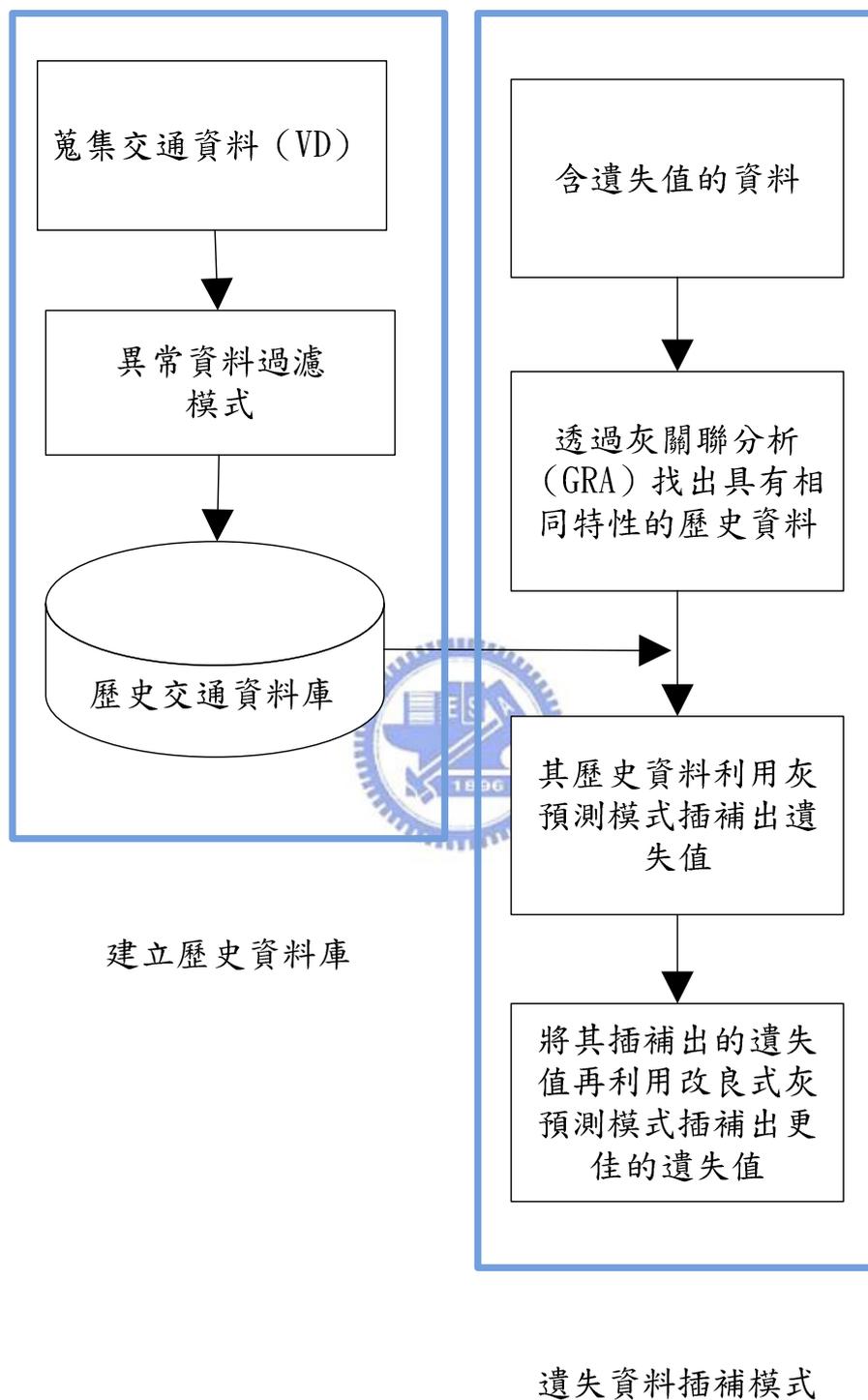


圖 3. 2 插補模式流程圖

3.2.1 蒐集交通資料

本研究資料來源為車輛偵測器(VD)所回傳的資料，蒐集頻率為 5 分鐘/次。資料欄位說明如下：

表 2. 2 車輛偵測器(VD)資料表

編號	欄位	型態	長度	欄位說明
1	ID	nvarchar	50	偵測器編號
2	Detectiondatetime	datetime	50	日期
3	LineType	nvarchar	50	車道種類
4	LineNumber	nvarchar	50	車道數
5	Direction	nvarchar	50	方向
6	Line1_smallcar_flow	nvarchar	50	車道 1 小客車流量
7	Line1_smallcar_avg_speed	nvarchar	50	車道 1 小客車速度
...
48	Total_flow	nvarchar	50	總流量
49	avg_speed	nvarchar	50	平均速度
50	avg_occ	nvarchar	50	平均佔有率

3.2.2 異常資料過濾模式

本研究採用蒐集一段時間內的速度資料，經研究整理後，發現有異常資料發生的情形，如速度突然的增加和降低。由圖 3.3 可知車輛偵測器(VD)於 09:10 所蒐集到的速度為 20 公里，而從原始資料知 09:00 到 10:00 所蒐集到的速度資料介於 80 公里到 100 公里左右，因此不可能會有一筆突然降低的速度資料。此外在 20:00 到 21:00 的速度資料介於 90 公里到 100 公里左右，而 20:15 的速度為 40 公里，由常理判斷不可能會有一筆突然降低的速度資料。此外，有一筆速度資料在 17:00 於 5 分鐘內突然增加到 130 公里，之後在 17:05 時馬上回到正常速限範圍 110 公里內，因此可視為此筆資料為異常的情形。

倘若忽略上述問題，而直接採用所蒐集的速度資料則會影響後續插補遺失資料的準確率。因此本研究有必要發展資料過濾模式，其評判準則如下：

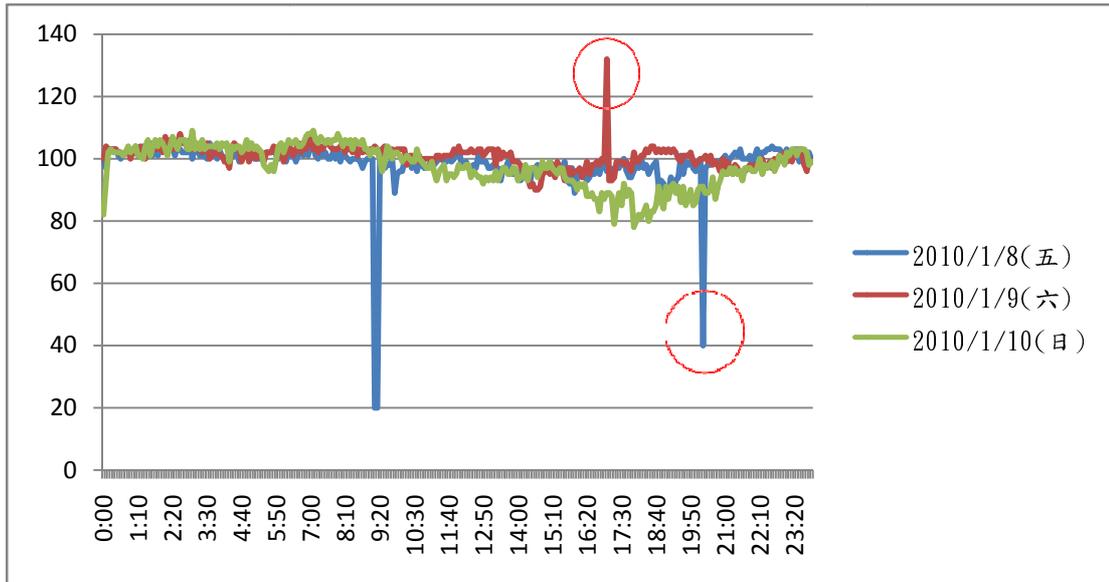


圖 3. 3 速率異常分佈圖(VD-N3-N-47. 980-M)

過濾條件：

本研究以資料逐筆過濾方式，將資料前後間的差值有劇烈變化料過濾的條件，依序拿三筆資料出來，分別為資料編號 $i-1, i, i+1$ 資料是比較基準點，而第 i 筆資料需與第 $i-1$ 筆資料及第 $i+1$ 筆資料變化均超過 15% 時，將第 i 筆資料視為異常資料，則刪除此筆資料。

然後在刪除第 i 筆資料後，只剩第 $i-1$ 筆和第 $i+1$ 筆資料，因 $i+2$ 筆資料，則此時有三筆資料比較，可比較是否第 $i+1$ 筆資料與其他兩筆資料，若有則刪除此筆資料。倘若均無超過 15%，則此時為比較基準點，可在納入後續一筆資料逐一判斷。

本研究所蒐集的資料屬於大樣本性質，依照統計學原理當樣本參數估計值的抽樣分配趨近於常態。而常態分配中大部份的觀測值正負三倍標準差的範圍內，因此落在三倍標準差外的值視為離群值。

3.2.3 歷史資料庫的建立

由所蒐集到交通資料透過 3.2.2 節異常資料過濾模式，將過濾後的交通資料建立歷史資料庫，其資料欄位說明如下：

表 2. 3 歷史資料庫欄位表

編號	欄位	型態	長度	欄位說明
1	ID	nvarchar	50	VD 編號
2	detectdatetime	datetime	8	收集日期
3	weekday	int	4	星期別
4	Hour	int	4	小時

編號	欄位	型態	長度	欄位說明
5	min	int	4	分鐘
6	Avg_speed	int	4	平均速度
7	Total_flow	int	4	平均流量
8	Avg_occ	int	4	平均佔有率

3.2.4 灰關聯分析(Grey Relation Analysis; GRA)

本研究將單筆含有遺失值的資料稱為參考數列，而此參考數列內含有多個參考數值，因此可利用灰關聯分析法從歷史交通資料庫裡，比對相同時段歷史數列內的數值資料，並計算各參考數值與歷史數值的灰關聯係數值。

將灰關聯係數值加總後取平均可得到各歷史數列的灰關聯度值，則灰關聯度值越高者，表示該筆歷史資料與含有遺失值的資料最相近。

因此可考慮含有遺失值的資料為參考數列：

$$X_0 = [x_0(1), x_0(2), \dots, x_0(s), NA_0(s+1), \dots, NA_0(n)], s < n \quad (1.1)$$

以 $NA_0(i)$ 為遺失值，且 $i = s+1, s+2, s+3, \dots, n$ 。

而歷史數列資料為：

$$X_j = [x_j(1), x_j(2), \dots, x_j(s), \dots, x_j(d)], j=1,2,\dots,m \text{ 且 } s \leq d \quad (1.2)$$

因此可先從歷史資料庫取相同時段之歷史數列，並取各歷史數列前 S 個數值，以式(1.1)為單筆參考數列，以式(1.2)為歷史數列資料。

其計算步驟如下：

- (1) 求其參考數列 X_0 和各歷史數列資料 X_j 的差值，並找出差值數列最大值和最小值。

則差值數列為

$$\{|x_0(1) - x_j(1)|, |x_0(2) - x_j(2)|, \dots, |x_0(k) - x_j(k)|\} \\ ; k = 1,2,3, \dots, s, j = 1,2, \dots, m$$

因此可找出差值數列內

$$\text{最小差值 } \min_j \min_k |x_0(k) - x_j(k)|, k = 1,2,3, \dots, s, j=1,2, \dots, m \quad (1.3)$$

$$\text{最大差值 } \max_j \max_k |x_0(k) - x_j(k)|, k = 1,2,3, \dots, s, j=1,2, \dots, m \quad (1.4)$$

(2) 計算數值間的灰關聯係數

令 X_0 與 X_j 於數列中各數值的灰關聯係數為 $\zeta_{0j}(k)$ ，其式子如下：

$$\zeta_{0j}(k) = \frac{\min_j \min_k |x_0(k) - x_j(k)| + \rho \max_j \max_k |x_0(k) - x_j(k)|}{|x_0(k) - x_j(k)| + \max_j \max_k |x_0(k) - x_j(k)|} \quad (1.5)$$

其中 $\rho \in [0,1]$ 。

ρ 表為辨識係數，定義為背景值與待測值之間的對比，其大小可依據實際的需要作適當的調整。吳漢雄[1996]所提出之數學證明中可得知，辨識係數 ρ 只會改變相對數值的大小，並不會造成相關因子對目標序列的高低，而影響灰關聯的排序；因此一般皆假設為0.5。

由灰關聯係數計算可知參考數列與歷史數列內的各數值其相關程度，本研究需採用關聯度最大之歷史數列以利後續預測之用，因此將取各歷史數列的灰關聯係數值加總後取其平均值。

(3) 計算灰關聯度(Grey Relational Grade ; GRG)

$$\gamma_{0j} = \frac{1}{d} \sum_{i=1}^d \zeta_{0j}(k), \quad j = 1, 2, \dots, m \quad (1.6)$$

， d 表各歷史數列中其數值的個數。

經過灰關聯度計算後，依數值大小作排序，得知歷史數列與參考數列間之相似程度。取出灰關聯度最大之歷史數列後，則可帶入3.2.5節灰預測GM(1,1)模式，即可預測出參考數列內其遺失值資料。

3.2.5 灰預測 GM(1,1)

本研究係以灰預測GM(1,1)為基礎，目的為了預測遺失值的未來情況，主要優點為所需要的原始數據不多，數據大於四筆即可。在構建灰預測GM(1,1)的數列必須為一相同符號的數列，若有正負值符號不一致時，在灰生成時則會發生模式無法計算之課題。以下說明其預測方式：

由3.2.4節找到灰關聯度最大之歷史數列

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)) = \sum_{i=1}^k x^{(0)}(i), \quad i = 1, 2, 3, \dots, n \quad (2.1)$$

(1) 灰生成(Grey generating)

先對歷史數列 $X^{(0)}$ 做一次累加生成(Accumulated Generating Operation, AGO)，

可得累加生成序列，其中 $x^{(0)}(1) = x^{(1)}(0)$

$$X^{(1)} = \left(-, x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n) \right) = \sum_{i=1}^k x^{(1)}(i) \quad (2.2)$$

其中 $k = 1, 2, \dots, n$

(2) 灰建模(Grey model construction)生成

由式(2.2)可構建灰預測 GM(1,1)之一階單變量的微分方程式

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b, \quad a, b \in R \quad (2.3)$$

其中 t 為系統之自變數， a 為發展係數， b 為灰色控制變數，因此 a 與 b 為模式之待定參數； $X^{(1)}$ 為背景值， $\frac{dX^{(1)}}{dt}$ 為一階導數。

在灰預測的計算過程上屬於時間間隔相等($\Delta t = 1$)，因此可視為近似離散數列之差方方程式。

$$\frac{dx^{(1)}(t)}{dt} \approx \frac{\Delta x^{(1)}(t)}{\Delta t} = x^{(1)}(t + \Delta t) - x^{(1)}(t) = x^{(1)}(t + 1) - x^{(1)}(t) = x^{(0)}(t + 1) \quad (2.4)$$

(3) 再利用均值計算生成 $X^{(1)}$ 背景值

$X^{(1)} = \sum_{i=1}^k x^{(0)}(i)$ ， $k = 1, 2, \dots, n$ ，令 $Z^{(1)}$ 為 $X^{(1)}$ 緊鄰的均值生成數列，則

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)) \quad (2.5)$$

其中 $z^{(1)}(k) = \alpha x^{(1)}(k) + (1 - \alpha)x^{(1)}(k - 1)$ ， $k = 2, 3, \dots, n$ ，本研究取 $\alpha = 0.5$

$$\text{則 } Z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k - 1) \approx x^{(1)}(t) \quad (2.6)$$

由式(2.4)和(2.6)可將 GM(1,1)灰微分方程 $\frac{dX^{(1)}}{dt} + aX^{(1)} = b$

$$\text{轉化成 } X^{(0)}(k + 1) + aZ^{(1)}(k) = b \quad (2.7)$$

將其移項 $X^{(0)}(k + 1) = -aZ^{(1)}(k + 1) + b$

且藉由式(2.2)和(2.5)得到以下線性組合：

可轉化成矩陣形式：

$$y_n = \begin{bmatrix} X^{(1)}(2) \\ X^{(1)}(3) \\ \dots \\ X^{(1)}(n) \end{bmatrix}, B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \dots & \dots \\ -Z^{(1)}(n) & 1 \end{bmatrix} \quad (2.8)$$

因為 $Y = B\hat{a}$ ，因此可利用最小平方法求解係數向量 $\hat{a} = (B^T B)^{-1} B^T y_n = \begin{bmatrix} a \\ b \end{bmatrix}$ 。

(4) 將所求出係數向量 \hat{a} 所內含係數值 a 和 b 分別代入灰色微分方程，可得到灰色累加 (AGO) 預測模型為 $\hat{X}^{(1)}(k+1)$ ，其表示式子如下：

$$\hat{X}^{(1)}(k+1) = \left[X^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (2.9)$$

(5) 還原成 $\hat{X}^{(0)}(k)$ 預測式，以逆累加生成 (Inverse Accumulated Generating Operation, IAGO) 方式，可得到：

$$\hat{X}^{(0)}(k) = \left[X^{(0)}(1) - \frac{b}{a} \right] (1 - e^{-a}) e^{-a(k-1)}, \text{ 其中 } k = 1, 2, 3, \dots, n$$

$$\text{或 } \hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) \quad (2.10)$$

因此可得預測數列

$$\hat{X}^{(0)}(k) = \left(\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \dots, \hat{x}^{(0)}(n) \right), k=1, 2, 3, \dots, n \quad (2.11)$$

以上步驟為灰預測 GM(1,1) 模型，取 $k=1, 2, \dots, n$ 可計算得到預測值；然而由鄧聚龍 [1982] 所提出灰色預測模式為一平滑的指數曲線，對於隨機性較大的數據預測結果較差，且其預測精度易受數列波動的影響，會降低模型的預測精度。

為了提高 GM(1,1) 模式之預測精確度，以往作法考慮原始值與預測值的關係，來重新修正預測模式。而本研究採用劉思峰 [1991] 所提出殘差 (Residual) 的概念，利用原始值減去預測值即可得到差異值。則其殘差式如下：

$$\underline{\varepsilon}^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k) \quad (2.12)$$

可藉由式 (2.1) 與 (2.11) 得到殘差數列

$$\varepsilon^{(0)}(k) = \left(-, \underline{\varepsilon}^{(0)}(2), \underline{\varepsilon}^{(0)}(3), \dots, \underline{\varepsilon}^{(0)}(n) \right); k = 2, 3, \dots, n \quad (2.13)$$

由式 (2.13) 將其殘差數列 $\varepsilon^{(0)}(k)$ 做一次累加生成 (AGO)，可得累加生成殘差數列

$$\varepsilon^{(1)}(k) = \left(-\varepsilon^{(1)}(3), \varepsilon^{(1)}(4), \dots, \varepsilon^{(1)}(n) \right) = \sum_{i=3}^k \varepsilon^{(1)}(i); k = 3, \dots, n \quad (2.13)$$

將式(2.13)計算累加生成殘差數列後，如同灰預測 GM(1,1)模式構建方式，由(2.3)到(2.10)之計算，可重新建立 GM(1,1)殘差修正模式。

$$\hat{\varepsilon}^{(0)}(k) = \left(\varepsilon^{(0)}(2) - \frac{b_{\varepsilon}}{a_{\varepsilon}} \right) (1 - e^{a_{\varepsilon}}) e^{-a_{\varepsilon}(k-1)}, k=2,3, \dots, n \quad (2.14)$$

而溫裕宏[1997]與 Hsu[2003]提出在建構 GM(1,1)殘差修正模式時，因受正負號之影響，使構建模式之殘差數列具連續相同符號之殘差數據並不一定超過四筆，因此可能違背灰預測 GM(1,1)模式之構建條件，有時會面臨到無法運算之困難。

目前已有多位學者利用灰預測 GM(1,1)為模式基礎，提出不同的殘差修正 GM 模型來改善灰預測的缺點，如溫裕宏[1997]與卓訓榮[2005]提出灰色馬可夫(Grey Markov)殘差修正法、Lee[2010]提出灰色基因(Grey Genetic Algorithm)殘差修正法與 Erda[2010]所提出的灰色傅立葉 (Grey Fourier) 殘差修正法，其目的為了有效解決殘差修正模式，因殘差值的正負號影響而使模式無法運算，導致無法有效提昇灰預測的準確度。

本研究係以灰預測 GM(1,1)為模式基礎，並利用劉思峰[1991]殘差概念建立新的修正模型，期許可對 GM(1,1)模式進行修正，以提高模式預測的準確度。

以下為其修正模式原理說明：

3.2.6 最小循環式殘差修正(Minimum Recursive Residual GM(1,1);

MRRGM(1,1))

由 3.2.5 節可知灰預測 GM(1,1)修正法，都以殘差數列為基礎所發展的演算法，而 GM(1,1)殘差修正模式無法構建主要因為殘差值的正負號是隨機產生的，因此有時會遇到模式無法運算的問題導致無法提昇灰預測 GM(1,1)的準確度。

有鑑於此，本研究為了解決殘差數列所產生的問題，因此將灰預測 GM(1,1)後所得到的殘差值都取絕對值。

$$\varepsilon^{(0)}(k) = (|\varepsilon^{(0)}(2)|, |\varepsilon^{(0)}(3)|, \dots, |\varepsilon^{(0)}(n)|); k = 2, 3, \dots, n \quad (3.1)$$

式(3.1)可解決模式構建的問題，由張廷政[2009]提出有些預測值與原始值的差值過大，因此所得到的殘差絕對值再帶入 GM(1,1)殘差模式修正時，會讓其模式所得到的預測值誤差更大，反而降低了預測的準確度。

而本研究考量到此問題，因此從殘差數列中取絕對值最小殘差值，其目的為了在 GM(1,1)預測模式修正時，可由差值最小者為基礎修正原預測模型，因此可減少預測值與原始值差異過大的情況，而影響到模式的預測。其方法如下：

(1)取絕對值最小殘差值

$$\min_k |\varepsilon^{(0)}(k)| \quad (3.2)$$

(2) 將 $\min_k |\varepsilon^{(0)}(k)|$ 加入式(2.11)之 $\hat{X}^{(0)}$ 預測數列，可得到一組新的預測數列

$$y^{(0)}(k) = \hat{X}^{(0)} + \min_k |\varepsilon^{(0)}(k)| \quad (3.3)$$

(3) 再對數列 $y^{(0)}(k)$ 做一次累加生成(AGO)可得 $y^{(1)}(k)$ 。

$$y^{(1)} = (y^{(1)}(1), y^{(1)}(2), y^{(1)}(3), \dots, y^{(1)}(n)) = \sum_{i=1}^k y^{(0)}(i) \quad (3.4)$$

$k = 1, 2, \dots, n$

利用均值計算生成 $y^{(1)}$ 背景值。

(4) 同 3.2.4 節灰建模生成計算可得到 a 和 b 值，並代入灰預測 GM(1,1)模式，因此可得到一組新的 AGO 預測模型 $\hat{y}^{(1)}(k+1)$ 。

$$\hat{y}^{(1)}(k+1) = \left[y^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (3.5)$$

(5) 利用逆累加生成 (IAGO) 還原 $\hat{y}^{(0)}(k+1)$ ，可得：

$$\hat{y}^{(0)}(k+1) = \left[y^{(0)}(1) - \frac{b}{a} \right] (1 - e^a) e^{-a(k-1)}; k = 2, 3, \dots, n \quad (3.6)$$

$$\text{或} \quad \hat{y}^{(0)}(k) = \hat{y}^{(1)}(k+1) - \hat{y}^{(1)}(k) \quad (3.7)$$

(6) 則其殘差為

$$\varepsilon_y^{(0)}(k) = y^{(0)}(k) - \hat{y}^{(0)}(k) \quad (3.8)$$

殘差數列為

$$\varepsilon_y^{(0)}(k) = (|\varepsilon_y^{(0)}(2)|, |\varepsilon_y^{(0)}(3)|, \dots, |\varepsilon_y^{(0)}(k)|); k = 2, 3, \dots, n \quad (3.9)$$

(7) 同式(3.2)並重複上述步驟可得到 $\min_k |\varepsilon_y^{(0)}(k)|$ 的最小殘差值。

(8)則當所有殘差 $\left|\varepsilon_y^{(0)}(k)\right| \cong 0$ 時,即可停止最小循環式殘差修正法 MRRGM(1,1)。

(9)將修正後的 MRRGM(1,1)模式,即可預測 $k=n+1$ 筆預測值。

本研究整合灰關聯分析(GRA)、灰預測 GM(1,1)和最小循環式殘差修正 MRRGM(1,1)等方法發展遺失資料插補模式,其計算流程圖 3.4 如下:



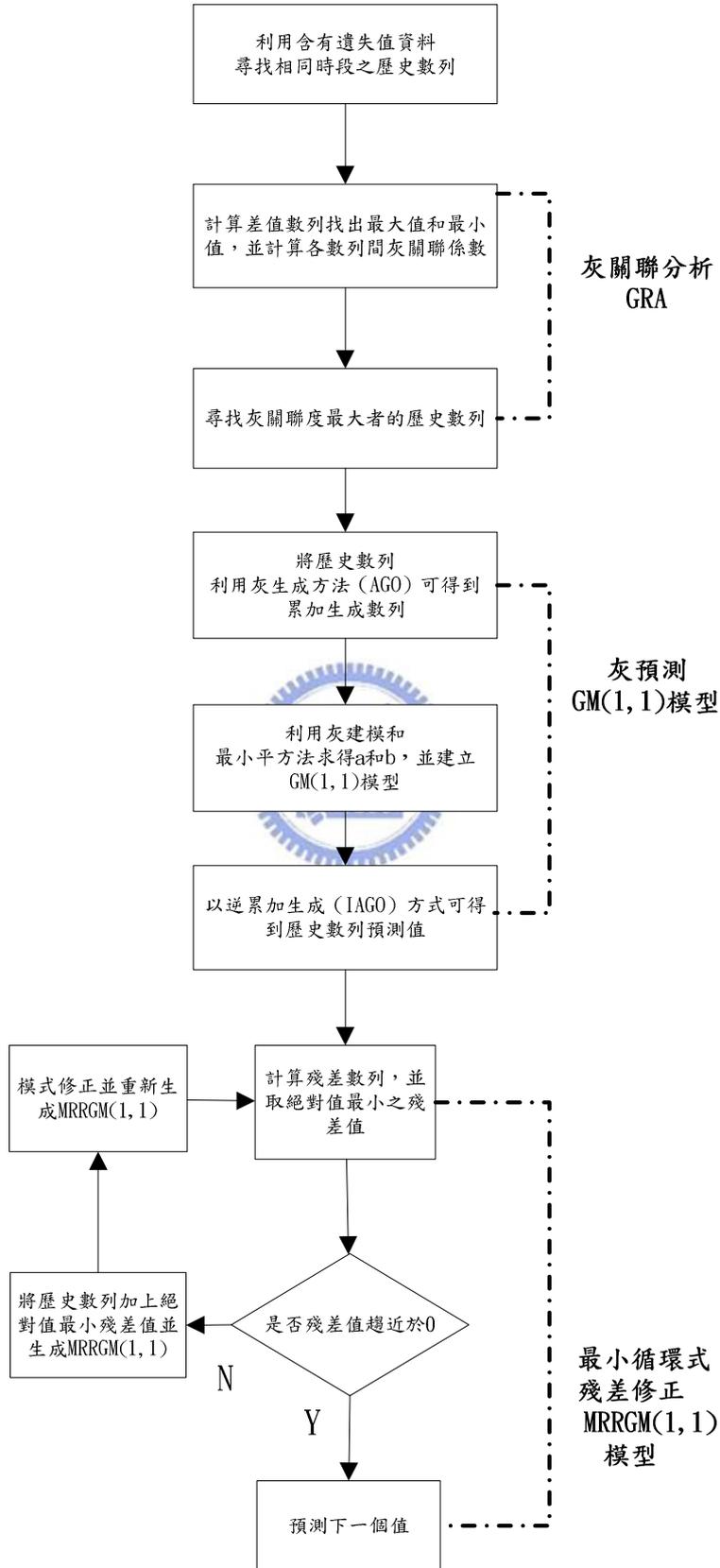


圖 3. 4 計算流程圖

3.3 範例說明

本研究以單一偵測器為例，每五分鐘回傳一筆資訊，如流量、速率、佔有率等，本範例以速率為探討對象。則此偵測器每一小時內有 12 筆資料，因此一天有 288 筆，並以歷史交通資料庫以供本演算法比對，其說明如下：

假設某一偵測器在 2010/01/25(一)由時間點 00:00 到 01:00 該偵測器所回傳不完整數列為 $X_0=(103,108,109,116,106,90,95,98,*,*,*,\dots,*)$ ，*表遺失值，則含有遺失值之不完整數列有 8 筆數值。而時間點 00:35 所回傳的數值為 98，在之後時間點沒有回傳資料回來，如圖 3.5 所示：

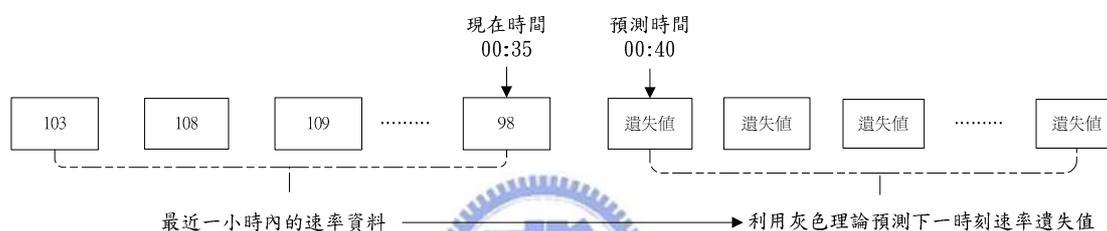


圖 3.5 車輛偵測器(VD)資料遺失示意圖

其計算步驟如下：

- (1) 先從歷史資料庫尋找同時段的歷史數列，再從其歷史數列中取其前 8 筆數值，因此可找到 6 筆歷史數列。

$$X_1=(109,90,96,104,99,101,111,93)$$

$$X_2=(116,109,105,100,90,89,94,101)$$

$$X_3=(100,108,108,97,100,107,106,107)$$

$$X_4=(110,109,113,108,109,113,104,100)$$

$$X_5=(112,113,118,104,103,94,93,106)$$

$$X_6=(104,105,100,91,96,94,95,86)$$

利用灰關聯分析法，從這 6 筆歷史數列中尋找一筆與含有遺失值資料 X_0 相似度最接近的歷史數列。

$$X_0=(103,108,109,116,106,90,95,98)$$

- (2) 分別計算差值序列為 $\{\|x_0(k) - x_j(k)\|\}$ ，並找出最小值 $\min_j \min_k \|x_0(k) - x_j(k)\|$ 、最大值 $\max_j \max_k \|x_0(k) - x_j(k)\|$ 。

差值數列

$$\{\|x_0(k) - x_1(k)\|\} = (6, 18, 13, 12, 7, 11, 16, 5)$$

$$\{\|x_0(k) - x_2(k)\|\} = (13, 1, 4, 16, 16, 1, 1, 3)$$

$$\{\|x_0(k) - x_3(k)\|\} = (3, 0, 1, 19, 6, 17, 11, 9)$$

$$\{\|x_0(k) - x_4(k)\|\} = (7, 1, 4, 8, 3, 23, 9, 2)$$

$$\{\|x_0(k) - x_5(k)\|\} = (9, 5, 9, 12, 3, 4, 2, 8)$$

$$\{\|x_0(k) - x_6(k)\|\} = (1, 3, 9, 25, 10, 4, 0, 12)$$

其中 $k=1,2,3,\dots,8$, $j=1,2,3,\dots,6$

而最大值和最小值為

$$\max_j \max_k \|x_0(k) - x_j(k)\| = 25$$

$$\min_j \min_k \|x_0(k) - x_j(k)\| = 0$$

$$(3) \text{ 計算灰關聯係數 } \zeta_{0j}(k) = \frac{\min_j \min_k |x_0(k) - x_j(k)| + \rho \max_j \max_k |x_0(k) - x_j(k)|}{|x_0(k) - x_j(k)| + \max_j \max_k |x_0(k) - x_j(k)|}; \rho = 0.5$$

則其所計算灰關聯係數如下：

$$\{\zeta_{01}\} = (0.676, 0.410, 0.490, 0.510, 0.641, 0.532, 0.439, 0.714)$$

$$\{\zeta_{02}\} = (0.490, 0.926, 0.758, 0.439, 0.439, 0.926, 0.926, 0.806)$$

$$\{\zeta_{03}\} = (0.806, 1.000, 0.926, 0.397, 0.676, 0.424, 0.532, 0.581)$$

$$\{\zeta_{04}\} = (0.641, 0.926, 0.758, 0.610, 0.806, 0.352, 0.581, 0.862)$$

$$\{\zeta_{05}\} = (0.581, 0.714, 0.581, 0.510, 0.806, 0.758, 0.862, 0.610)$$

$$\{\zeta_{06}\} = (0.926, 0.806, 0.581, 0.333, 0.556, 0.758, 1.000, 0.510)$$

; 其中 $j=1,2,3,\dots,6$

$$(4) \text{ 接著計算灰關聯度 } \gamma_{kj} = \frac{1}{d} \sum_{i=1}^d \zeta_{kj}(t); d=8$$

$\gamma_{01} = 0.57; \gamma_{02} = 0.64; \gamma_{03} = 0.60; \gamma_{04} = 0.65; \gamma_{05} = 0.64; \gamma_{06} = 0.72$ ，則以 γ_{06} 為關聯度最大，表示與原始數列相關聯程度最大。

因此可用第六筆歷史數列經由灰預測 GM(1,1) 模式預測遺失值，再利用 MRRGM(1,1) 模型修正去預測所有時間點的遺失值，以下以預測 00:40 遺失值為例：

(5) 計算灰生成

利用 $X_6 = (104, 105, 100, 91, 96, 94, 95, 86)$ 共八筆原始數值，欲預測第九筆遺失值。

則歷史數列為 $(104, 105, 100, 91, 96, 94, 95, 86)$

AGO 累加生成： $(-, 209, 309, 400, 496, 590, 685, 771)$

均值生成： $Z_X^{(1)}(k+1) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k+1)$

$$= (-, 156.5, 259, 354.5, 448, 543, 637.5, 728)$$

(6) 計算灰建模 GM(1,1)，並求得 a 和 b

$$y_n = \begin{bmatrix} 209 \\ 309 \\ 400 \\ 496 \\ 590 \\ 685 \\ 771 \end{bmatrix}, B = \begin{bmatrix} -156.5 & 1 \\ -259 & 1 \\ -354.5 & 1 \\ -448 & 1 \\ -543 & 1 \\ -637.5 & 1 \\ -728 & 1 \end{bmatrix}, B^T B = \begin{bmatrix} 1649186.75 & -3126.5 \\ -3126.5 & 7 \end{bmatrix}$$

$$(B^T B)^{-1} = \begin{bmatrix} 0 & 0.0018 \\ 0.0018 & 0.9321 \end{bmatrix}$$

可由最小平方方法求得 $\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T y_n$ 則 $\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.02413 \\ 106.0619 \end{bmatrix}$

(7) 以逆累加生成方式 (IAGO) 求得 $\hat{X}^{(0)}(k)$

$$\text{可得 AGO 累加生成函數 } \hat{X}^{(1)}(k+1) = \left[104 - \frac{106.0619}{0.02413} \right] e^{-0.02143k} + \frac{106.0619}{0.02413}$$

則還原成 IAGO 逆累加生成函數

$$\hat{X}^{(0)}(k) = \left[104 - \frac{106.0619}{0.02413} \right] (1 - e^{-0.02413}) e^{-0.02413(k-1)} \text{ 可得到 } \hat{X}^{(0)}(k) \text{ 之預測值}$$

表 2. 4 GM(1,1) 計算值

K	1	2	3	4	5	6	7	8
$x^{(0)}(k)$	104	105	100	91	96	94	95	86
$\hat{x}^{(0)}(k)$	104	102.3134	99.8745	97.4936	95.1695	92.9009	90.6863	88.5245
$\varepsilon^{(0)}(k)$	0	2.6866	0.1255	-6.4936	0.8305	1.0991	4.3137	-2.5245

(8) 計算殘差 $\varepsilon^{(0)}(k) = X^{(0)}(k) - \hat{X}^{(0)}(k)$

(9) 判斷是否各 $\varepsilon^{(0)}(k)$ 是否近似於零，若是則不需要做循環式殘差修正，可預測遺失值。若否則需進行循環式殘差修正模型 MRRGM(1,1)，其修正步驟如下：

(10) 取模型預測誤差最小者

$$\min |\varepsilon^{(0)}(k)| = 0.1255$$

(11) $\hat{x}^{(0)}(k) + \min_k |\varepsilon^{(0)}(k)|$ 則可得到一筆新數列 $y^{(0)}$

$$(104.1255, 102.4389, 100, 97.6191, 95.295, 93.0264, 90.8118, 88.65)$$

(12) 生成循環式殘差修正模型 MRRGM(1,1)

$$y^{(0)}(k) = \hat{x}^{(0)}(k) + \min_k |\varepsilon^{(0)}(k)|; k=1 \sim 8, \text{ 其計算步驟如下:}$$

(13) 計算灰生成

$y^{(0)}$ 數列為：

$$(104.1255, 102.4389, 100, 97.6191, 95.295, 93.0264, 90.8118, 88.65)$$

AGO 生成：

(-, 206.5644, 306.5644, 404.1835, 499.4785, 592.5049, 683.3167, 771.9667)
 均值生成： $Z_y^{(1)}(k) = 0.5y^{(1)}(k) + 0.5y^{(1)}(k-1)$
 (-, 155.3450, 256.5644, 355.3740, 451.8310, 545.9917, 637.9108, 727.6417)

(14) 計算灰建模 MRRGM(1,1)，並求得 a 和 b

$$y_n = \begin{bmatrix} 206.5644 \\ 306.5644 \\ 404.1835 \\ 499.4785 \\ 592.5049 \\ 683.3167 \\ 771.9667 \end{bmatrix}, B = \begin{bmatrix} -155.3450 & 1 \\ -256.5644 & 1 \\ -355.3740 & 1 \\ -451.8310 & 1 \\ -545.9917 & 1 \\ -637.9108 & 1 \\ -727.6417 & 1 \end{bmatrix} B^T B = \begin{bmatrix} 1654898.8105 & -3130.6585 \\ -3130.6585 & 7 \end{bmatrix}$$

$$(B^T B)^{-1} = \begin{bmatrix} 0 & 0.0018 \\ 0.0018 & 0.9280 \end{bmatrix}$$

$$\text{由最小平方法求得 } \hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T y_n \text{ 則 } \hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.024094 \\ 106.18162 \end{bmatrix}$$

(15) 利用 AGO 累加生成函數，還原成 IAGO 逆累加生成函數

$$\hat{y}^{(0)}(k) = \left[104.1255 - \frac{106.18162}{0.024094} \right] (1 - e^{0.024094}) e^{-0.024094(k-1)}$$

可得到 $\hat{y}^{(0)}(k)$ 之預測值。

表 2. 5 MRRGM(1, 1) 計算值

K	1	2	3	4	5	6	7	8
$y^{(0)}(k)$	104.1255	102.4389	100	97.6191	95.295	93.0264	90.8118	88.65
$\hat{y}^{(0)}(k)$	104.1255	102.4338	99.9953	97.6148	95.2910	93.0225	90.8080	88.6462
$\varepsilon^{(0)}(k)$	----	0.0051	0.0047	0.0043	0.004	0.0039	0.0038	0.0038

當所有 $\varepsilon^{(0)}(k)$ 皆趨近於零，即可停止循環式殘差修正模型 MRRGM(1,1)，則可用此模式來預測第九筆資料遺失值，即 $k=9$ 帶入 $\hat{y}^{(0)}(k)$ 可得到 86.5359 取四捨五入即為 87。

3.4 誤差分析

由於預測值與實際值間可能存在誤差，為進行下一點之預測，預測數據經前一小節之步驟進行生成及建模後，須進一步檢驗其預測結果之準確性與適用性。

本研究利用平均絕對值誤差百分比 (Mean Absolute Percent Error, MAPE) 做為檢驗插補準確度績效的指標，因 MAPE 值較不易受到實際值與預測值的單位和大小影響，可以客觀比較插補值與實際值之間的差異程度。其定義如式(4.1)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (4.1)$$

其中 x_i 為實際值， \hat{x}_i 為預測值， n 為預測個數

而平均絕對誤差百分比(MAPE)主要係衡量模式中未被解釋部分之百分比。MAPE 值愈小，表示模式的預測能力愈強，模式預測結果與歷史資料吻合精確度愈大。根據 Lewis[1982]所提出 MAPE 之評估標準，當 MAPE 值小於 10%時，表示預測能力為高精準插補。

表 2. 6 MAPE 之評估標準

MAPE	等級
< 10%	高精準插補
10~20%	優良插補
20~50%	合理插補
>50%	不正確插補

第四章實證分析

本章目的為了驗證本研究所提出的插補法是否有良好的績效，因此需要設計實驗的測試方法，以驗證本演算法是否有良好的插補效果和適用的情境。

以國道三號高速公路之車輛偵測器作為測試對象，並透過 4.1.1 節資料蒐集取得歷史資料、4.1.2 節異常資料過濾、4.2 節建立資料庫等，並由 4.3 節實證分析設計不同遺漏之比例，根據時間插補設計原則，依偵測器所回傳的資料頻率，因此所可分別測試兩種演算法在不同情境下插補準確度，以下針對各節加以說明。

4.1.1 資料蒐集

本研究蒐集 2009/04/05~2010/04/05 期間國道三號北上龍潭收費站 46 公里處到樹林收費站 72 公里處之間 VD 資料作為實證分析之對象，全長為 26 公里，該路段共有 15 個偵測器。然而在這些偵測器中，如表 4.1-1 中所示存在著各偵測器資料遺漏的問題。

為了確保歷史資料的完整性，本研究採用缺漏率小於 10% 之偵測器資料，因此只測試 VD 編號 N3-47.98、N3-50.38、N3-51.555、N3-54、N3-55.34，共 5 個偵測器所蒐集到的資料，其餘偵測器資料則捨棄不用。

而這 5 個偵測器所蒐集資料內容包含 VD 編號、日期時間、總流量、平均速度、佔有率等資料，其蒐集頻率為 5 分鐘/次，每小時共有 12 筆資料，每日資料量為 288 筆，總蒐集天數為 365 天，因此每一偵測器應有 105,120 筆資料，來建立歷史資料庫。

表 2.7 龍潭到樹林收費站北上路段間偵測器之資料缺漏率

個數	1	2	3	4	5	6
VD 編號	N3-46.5	N3-47.98	N3-50.38	N3-51.555	N3-52.5	N3-54
缺漏率	11.25%	8.90%	8.70%	9.43%	13.90%	8.83%
遺漏數	11826	9355	9147	9911	14613	9280
個數	7	8	9	10	11	12
VD 編號	N3-55.34	N3-60.492	N3-61.99	N3-62.925	N3-64.000	N3-65.996
缺漏率	8.93%	18.68%	15.98%	18.52%	15.08%	15.24%
遺漏數	9391	19642	16801	19475	15848	16017
個數	13	14	15			
VD 編號	N3-67.990	N3-69.510	N3-70.470			
缺漏率	14.51%	14.99%	14.72%			
遺漏數	15251	15755	15472			

4.1.2 異常資料過濾

在建立歷史資料時，將所蒐集缺漏率小於 10% 的 5 個偵測器資料，透過本模式異常資料過濾模式，以資料前後間的差值有劇烈變化者作為異常資料過濾的條件，因此本研究假定車輛偵測器 (VD) 於時間點 00:00 時所回傳的第一筆速率資料為比較基準點。

而本研究分析 2010 年 1 月份 VD 變化程度，15 % 為三倍標準差為條件下，因此由 3.2.2 節過濾條件，可將異常資料過濾掉。

4.2 建立資料庫

本研究利用偵測器每五分鐘回傳資料建立歷史資料庫，所設計歷史資料庫欄位說明如下：

表 2. 8 歷史資料庫欄位表

編號	欄位	型態	長度	欄位說明
1	ID	nvarchar	50	VD 編號
2	detectdatetime	datetime	8	收集日期
3	weekday	int	4	星期別
4	Hour	int	4	小時
5	min	int	4	分鐘
6	Avg_speed	int	4	平均速度
7	Total_flow	int	4	平均流量
8	Avg_occ	int	4	平均佔有率

4.3 實證方法

本研究透過實際驗證下，了解本插補法在不同情境下其插補結果是否有良好的績效。

以蒐集 2011/04/10 與 2011/04/11 兩天完整連續且不中斷的資料作為測試的基礎，並將這兩天完整資料採用固定比例刪除方式 5%、10%、15%、...、50%，因此共有 10 種資料遺漏值作為測試範例，由測試範例比較找到本演算法在不同情境下的插補精確度。

為了比較兩種演算法 MRRGM(1, 1) 和 GM(1, 1) 插補法的差異，本研究採用時間插補原則，其插補頻率為每 5 分鐘/筆。

本研究將所蒐集到兩天完整不中斷的資料，利用 4.3.1 節隨機時間段遺漏產生本研究將所蒐集到兩天完整部中斷的資料，利用 4.3.1 節隨機時間段遺漏產生不同的遺漏方式。然後再將所建立的歷史資料庫，由 4.3.2 節說明比對歷史資料。此外，本研究採用多重插補法則以驗證在多次插補的情況下，其插補次數

是否會影響插補的結果，於 4.3.3 節說明其插補次數。

4.3.1 隨機時間段遺漏

本研究在隨機時間段遺漏先假定每個時間點所遺漏的機率是相等的，並利用均勻分配產生隨機亂數，其機率密度函數如圖 4.1 所示：

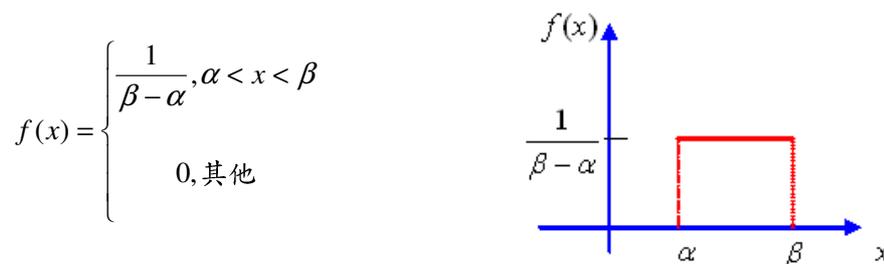


圖 4.1 均勻分配機率密度函數

則 x 為在區間 (α, β) 上均勻分布的隨機變數，以 $x \sim U(\alpha, \beta)$ 表示，其機率密度函數可表示成 $f(x) = \frac{1}{\beta - \alpha}$ 。

本研究採用上述統計理論，將隨機時間段遺漏方式分為整日、尖峰時段與離峰時段，且依照偵測器所回傳的頻率作為資料遺漏方式，其說明如下：

4.3.1.1 任意時段遺漏

將偵測器每日於 00:00 所傳回第 1 筆資料到偵測器於 23:55 回傳第 288 筆資料，可表示成 $x \sim U(1, 288)$ ，依照不同遺失比例產生隨機時間段遺漏的起始點，其說明如下所示：

- (1) 若遺漏起始點 $<$ 最大遺失比例（本研究為 50%）時，則產生遺失連續資料落在當日（無跨日）的範圍，其說明如下圖 4.2 所示：

若遺失起始點為第20筆，且遺失比例45%(130個)

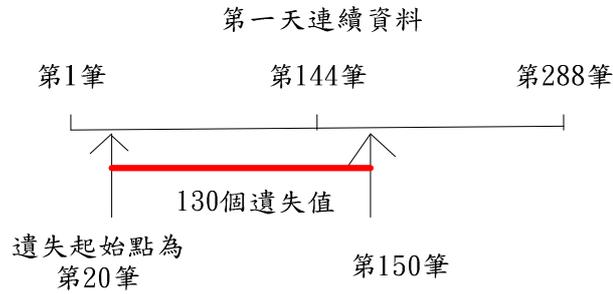


圖 4. 2 任意時段遺漏起始點未超過最大遺失比例 50%示意圖

(2)遺漏起始點 \geq 最大遺失比例（本研究為 50%）時，則會產生遺失連續資料超過當日（有跨日）的範圍，其說明如下圖 4.3 所示：

若遺失起始點為第188筆，且遺失比例45%(130個)



圖 4. 3 任意時段遺漏起始點超過最大遺失比例 50%示意圖

本研究主要目的為探討連續時間遺漏方式，因此為了解決圖 4.3 所產生遺漏起始點的問題而影響到遺失資料比例正確性，所設計的計算方法如下列式子，以確保隨機資料遺失比例的完整性。

$$\text{隨機資料遺失比例} = \frac{\text{遺失值個數} \times 2}{\text{兩天連續完整資料}}$$

4.3.1.2 尖峰時段遺漏

將偵測器於每日尖峰時段早上 07:00~09:00 和下午 17:00~19:00 所傳回 48 筆資料，可表示成 $x \sim U(1,48)$ ，依照不同遺失比例產生隨機時間段遺漏的起始

點，其說明如下所示：

- (1) 若遺漏起始點 < 最大遺失比例（本研究為 50%）時，則產生遺失連續資料落在當日（無跨日）的範圍，其說明如下圖 4.4 所示：

若遺失起始點為第2筆，且遺失比例45%(22個)



圖 4. 4 尖峰時間段遺漏起始點未超過最大遺失比例 50%示意圖

- (2) 遺漏起始點 \geq 最大遺失比例（本研究為 50%）時，則會產生遺失連續資料超過當日（有跨日）的範圍，其說明如下圖 4.5 所示：

若遺失起始點為第28筆，且遺失比例45%(22個)

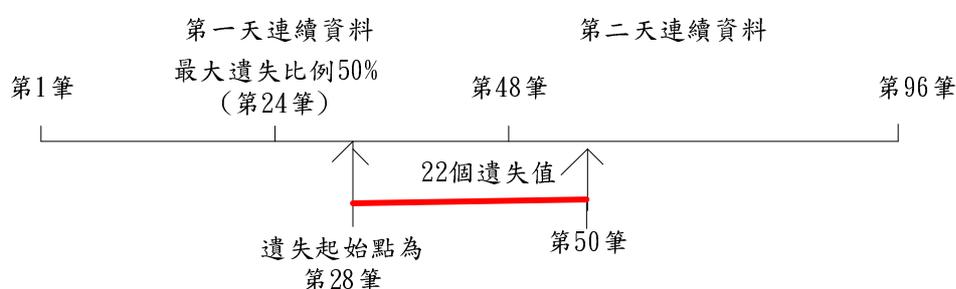


圖 4. 5 尖峰時間段遺漏起始點超過最大遺失比例 50%示意圖

本研究主要目的為探討連續時間遺漏方式，因此為了解決圖 4.3- 4 所產生遺漏起始點的問題而影響到遺失資料比例正確性，所設計的計算方法如下列式子，以確保隨機資料遺失比例的完整性。

$$\text{隨機資料遺失比例} = \frac{\text{遺失值個數} \times 2}{\text{兩天連續完整資料}}$$

4.3.3.3 離峰時間段遺漏

將偵測器於每日非尖峰時段所傳回的 240 筆資料，可表示成 $x \sim U(1,240)$ ，依照不同遺失比例產生隨機時間段遺漏的起始點，其說明如下所示：

- (1) 若遺漏起始點 < 最大遺失比例（本研究為 50%）時，則產生遺失連續資料落在當日（無跨日）的範圍，其說明如下圖 4.6 所示：

若遺失起始點為第 2 筆，且遺失比例 45% (108 個)



圖 4. 6 離峰時間段遺漏起始點未超過最大遺失比例 50% 示意圖

- (2) 遺漏起始點 \geq 最大遺失比例（本研究為 50%）時，則會產生遺失連續資料超過當日（有跨日）的範圍，其說明如下圖 4.7 所示：

若遺失起始點為第 188 筆，且遺失比例 45% (108 個)

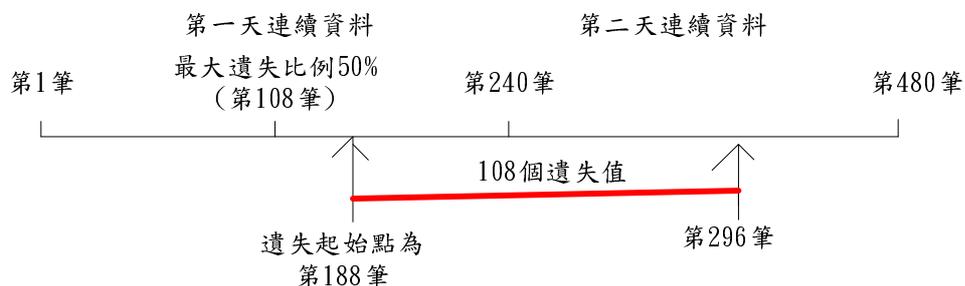


圖 4. 7 離峰時間段遺漏起始點超過最大遺失比例 50% 示意圖

4.3.2 比對方式

建立歷史資料庫後，其目的是為了將含有遺失值的參考數列，利用灰關聯法比對歷史資料，而找到相同特性的歷史資料，並透過本演算法插補遺失值。

4.3.3 插補次數

本研究採用 Rubin[1978]提出了多重插補法(multiple imputation method)的概念，其優點為多重插補代表重複地隨機從模型中抽取出來，結合完整資料的推論會較有效。

因此在實驗設計上，需要測試不同插補次數和不同資料遺漏比例下，所插補的資料與原始資料的平均絕對誤差百分比(MAPE)。

將插補次數分為 1 次、10 次、100 次、200 次、300 次、400 次、500 次，來驗證本演算法在不同情境下插補之準確度。

其實證分析流程如圖 4.8：



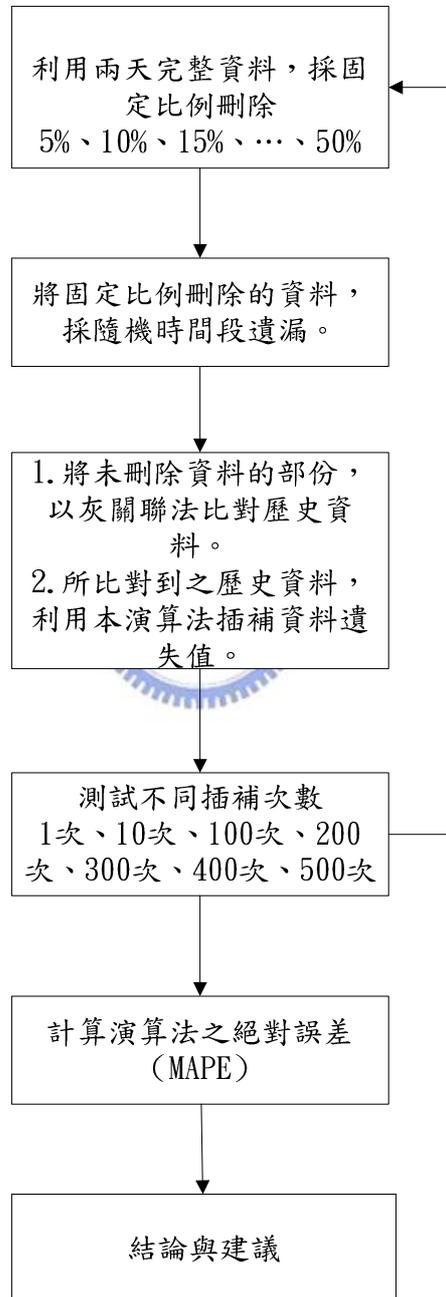


圖 4. 8 實證分析流程圖

4.4 測試方式

比較兩種插補法 GM (1, 1)和 MRRGM(1, 1)在不同時段（任意、尖峰、離峰）插補設計下是否有差異，以多重插補法原則，可分為兩個方式來探討：

- (1)在插補次數固定的條件下，比較不同遺失比例下兩演算法之插補績效。
- (2)在遺失比例固定的條件下，比較不同插補次數下兩演算法之插補績效。

4.5 測試範例

利用固定比例刪除完整資料 5%、10%、…、50%，共有 10 種情境，來比較兩種插補法的優劣。

4.5.1 測試對象

本研究採用缺漏率小於 10%之偵測器資料，因此由表 4.1 所示只測試 5 個偵測器（N3-47.98、N3-50.38、N3-51.555、N3-54、N3-55.34），並將其蒐集到的歷史資料用來插補資料遺失值。



4.6 測試結果

其假設檢定如下：

由測試結果可知兩插補法的插補績效，接著利用統計方法檢定兩插補法的插補績效是否存在差異。先以 F 檢定判斷兩母體變異數是否相等之檢定，再利用 t 檢定兩母體平均數差異檢定，以檢定兩插補績效是否存在差異。

其檢定假設如下：

1. 兩母體變異數是否相等檢定假設

$$\begin{cases} H_0: \sigma_{MRRGM(1,1)}^2 = \sigma_{GM(1,1)}^2 \\ H_1: \sigma_{MRRGM(1,1)}^2 \neq \sigma_{GM(1,1)}^2 \end{cases}$$

$\sigma_{MRRGM(1,1)}^2$ 以 MRRGM(1,1)為插補法，其插補績效的母體變異數。

$\sigma_{GM(1,1)}^2$ 以 GM(1,1)為插補法，其插補績效的母體變異數。

$$\text{檢定統計量為 } F = \frac{S_{MRRGM(1,1)}^2}{S_{GM(1,1)}^2}$$

其中
$$S_{MRRGM(1,1)}^2 = \frac{\sum (X_{MRRGM(1,1)} - \bar{X}_{MRRGM(1,1)})^2}{n_{MRRGM(1,1)} - 1}$$

$$S_{GM(1,1)}^2 = \frac{\sum (X_{GM(1,1)} - \bar{X}_{GM(1,1)})^2}{n_{GM(1,1)} - 1}$$

拒絕域

$$C = \left\{ F \mid F < F_{1-\frac{\alpha}{2}}(n_{MRRGM(1,1)} - 1, n_{GM(1,1)} - 1) \text{ 或 } F > F_{\frac{\alpha}{2}}(n_{MRRGM(1,1)} - 1, n_{GM(1,1)} - 1) \right\}$$

2. 兩母體平均數差異檢定假設

(1) 在變異數相等情況下之檢定假設：

$$\begin{cases} H_0: \mu_{MRRGM(1,1)} = \mu_{GM(1,1)} \\ H_1: \mu_{MRRGM(1,1)} \neq \mu_{GM(1,1)} \end{cases}$$

其中

$\mu_{MRRGM(1,1)}$ 以 MRRGM(1,1) 為插補法，其插補績效的平均值。

$\mu_{GM(1,1)}$ 以 GM(1,1) 為插補法，其插補績效的平均值。

$$\text{檢定統計量 } t = \frac{(\bar{X}_{MRRGM(1,1)} - \bar{X}_{GM(1,1)}) - (\mu_{MRRGM(1,1)} - \mu_{GM(1,1)})}{S_p \sqrt{\frac{1}{n_{MRRGM(1,1)}} + \frac{1}{n_{GM(1,1)}}}}$$

$$S_p^2 = \frac{(n_{MRRGM(1,1)} - 1)S_{MRRGM(1,1)}^2 + (n_{GM(1,1)} - 1)S_{GM(1,1)}^2}{n_{MRRGM(1,1)} + n_{GM(1,1)}}$$

拒絕域

$$C = \left\{ t \mid t < -t_{\frac{\alpha}{2}}(n_{MRRGM(1,1)} + n_{GM(1,1)} - 2) \text{ 或 } t > t_{\frac{\alpha}{2}}(n_{MRRGM(1,1)} + n_{GM(1,1)} - 2) \right\}$$

(2) 在變異數不相等情況下之檢定假設：

$$\begin{cases} H_0: \mu_{MRRGM(1,1)} = \mu_{GM(1,1)} \\ H_1: \mu_{MRRGM(1,1)} \neq \mu_{GM(1,1)} \end{cases}$$

其中

$\mu_{MRRGM(1,1)}$ 以 MRRGM(1,1) 為插補法，其插補績效的平均值。

$\mu_{GM(1,1)}$ 以 GM(1,1) 為插補法，其插補績效的平均值。

$$\text{檢定統計量 } t = \frac{(\bar{X}_{\text{MRRGM}(1,1)} - \bar{X}_{\text{GM}(1,1)}) - (\mu_{\text{MRRGM}(1,1)} - \mu_{\text{GM}(1,1)})}{\sqrt{\frac{S_{\text{MRRGM}(1,1)}^2}{n_{\text{MRRGM}(1,1)}} + \frac{S_{\text{GM}(1,1)}^2}{n_{\text{GM}(1,1)}}}}$$

拒絕域

$$C = \left\{ t \mid t < -t_{\frac{\alpha}{2}}(\gamma) \text{ 或 } t > t_{\frac{\alpha}{2}}(\gamma) \right\}$$

$$\text{其中自由度}(\gamma)\text{為, } \gamma = \frac{\left(\frac{S_{\text{MRRGM}(1,1)}^2}{n_{\text{MRRGM}(1,1)}} + \frac{S_{\text{GM}(1,1)}^2}{n_{\text{GM}(1,1)}} \right)^2}{\left(\frac{S_{\text{MRRGM}(1,1)}^2}{n_{\text{MRRGM}(1,1)}} \right)^2 + \left(\frac{S_{\text{GM}(1,1)}^2}{n_{\text{GM}(1,1)}} \right)^2}$$

$$\frac{n_{\text{MRRGM}(1,1)} - 1}{n_{\text{MRRGM}(1,1)} - 1} + \frac{n_{\text{GM}(1,1)} - 1}{n_{\text{GM}(1,1)} - 1}$$

4.6.1 以 N3-47.98 為測試對象

1. 比較不同遺失比例下插補誤差之探討

(1) 任一時段

將插補次數固定，比較在不同遺失比例下其插補誤差，根據圖 4.9 得知在不同遺失資料比例下，其插補誤差皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1, 1) 標準差 6.050358，GM(1, 1) 標準差 6.61517，則 MRRGM(1, 1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

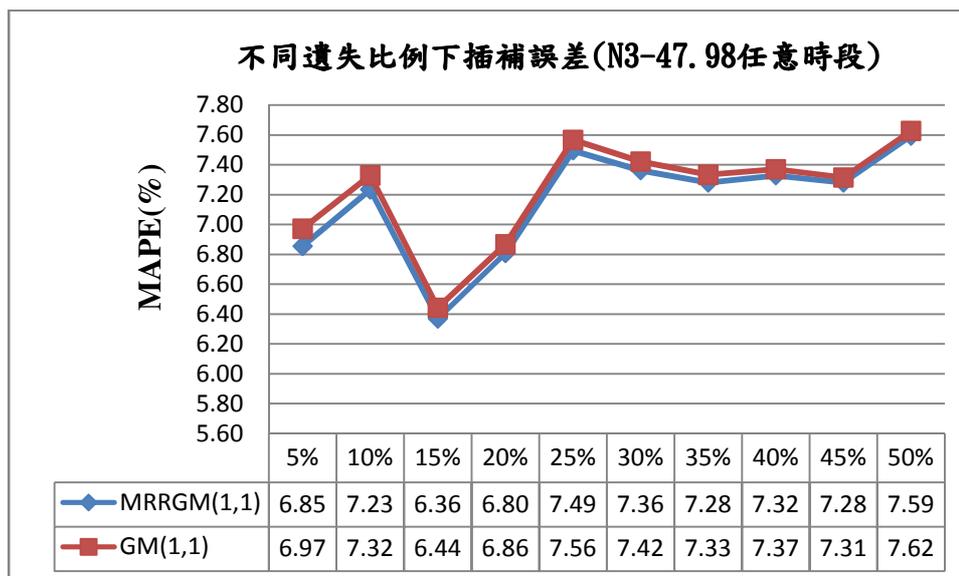


圖 4. 9 不同遺失比例下插補誤差(N3-47.98 任意時段)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$

之下不拒絕 H_0 (P-value=0.46339)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定。在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.350435)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果可說明在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

將插補次數固定，比較不同遺失比例其插補誤差，根據圖 4.10 得知當資料遺失比例小於等於 40% 時，其插補誤差百分比小於 10% 屬於高精準插補。而遺失比例大於 40% 時，其插補誤差介於 10%~20% 屬於優良插補。

由測試結果得知 MRRGM(1,1) 標準差 6.315061，GM(1,1) 標準差 6.440074，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

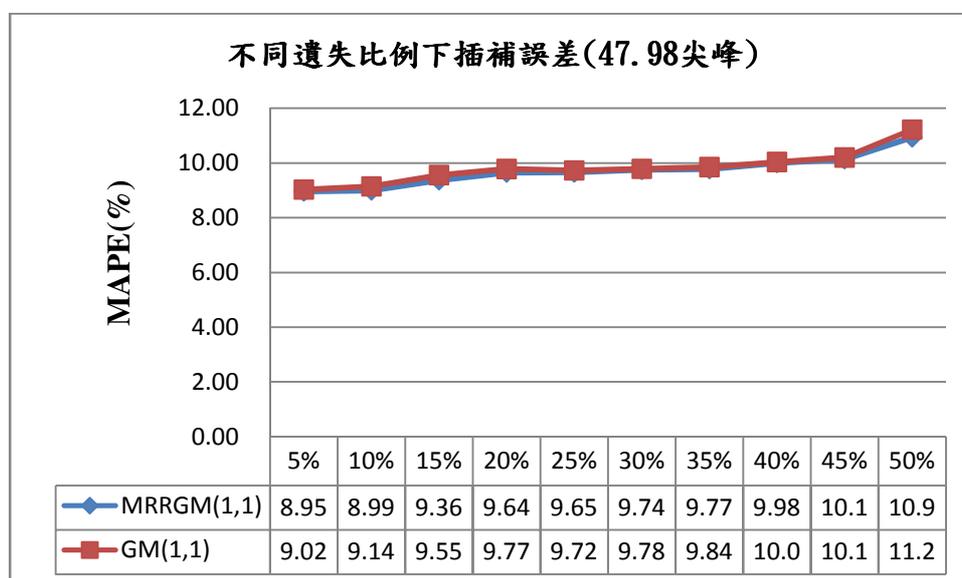


圖 4. 10 不同遺失比例下插補誤差(N3-47.98 尖峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.434665)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定。在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.338804)，GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果可說明在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

將插補次數固定，比較不同遺失比例其插補誤差，根據圖 4.11 得知當資料遺失比例小於等於 40% 時，其插補誤差百分比小於 10% 屬於高精準插補。而遺失比例大於等於 40% 時，其插補誤差介於 10%~20% 屬於優良插補。

由測試結果得知 MRRGM(1, 1)標準差 6.232187，GM(1, 1)標準差 6.354498，則 MRRGM(1, 1)的標準差較小，所插補資料值和遺失資料值的變異程度較小。

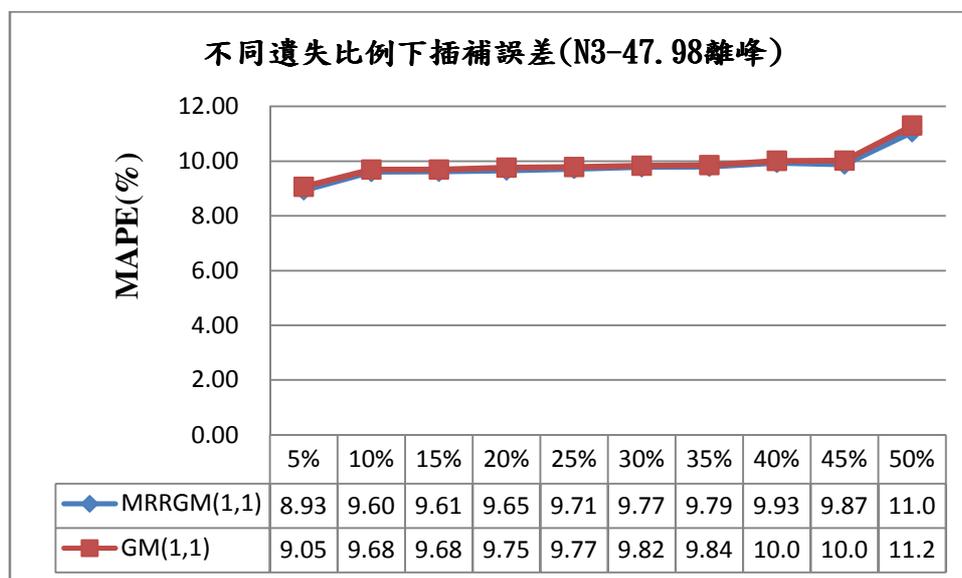


圖 4. 11 不同遺失比例下插補誤差(N3-47.98 離峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.456609)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定。在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.346397)，GM(1, 1)插補之 MAPE 值不大於 MRRGM(1, 1)插補之 MAPE 值。結果可說明在遺失比例不同的情況下 MRRGM (1, 1)插補法之績效與 GM(1, 1)插補法無顯著的差異。

2. 比較不同插補次數下插補誤差之探討

(1) 任意時段

在任一時段下比較不同插補次數的插補誤差，根據圖 4.12 可知 MRRGM (1, 1) 的插補績效在不同插補次數下皆優於 GM(1, 1)，且隨著插補次數的增加，其插補誤差有下降的趨勢。

由 Lewis[1982]評估標準得知，在不同插補次數下其插補誤差百分比(MAPE)皆小於 10%，則這兩種演算法皆屬於高精準的插補。且隨著插補次數的增加，其誤差百分比有逐漸下降的趨勢，由圖 4.12 可知插補次數 500 次其誤差百分比最小。

由實驗結果得知 MRRGM(1, 1)標準差 6.050358，GM(1, 1)標準差 6.61517，則 MRRGM(1, 1)的標準差較小，所插補資料值和遺失資料值的變異程度較小。

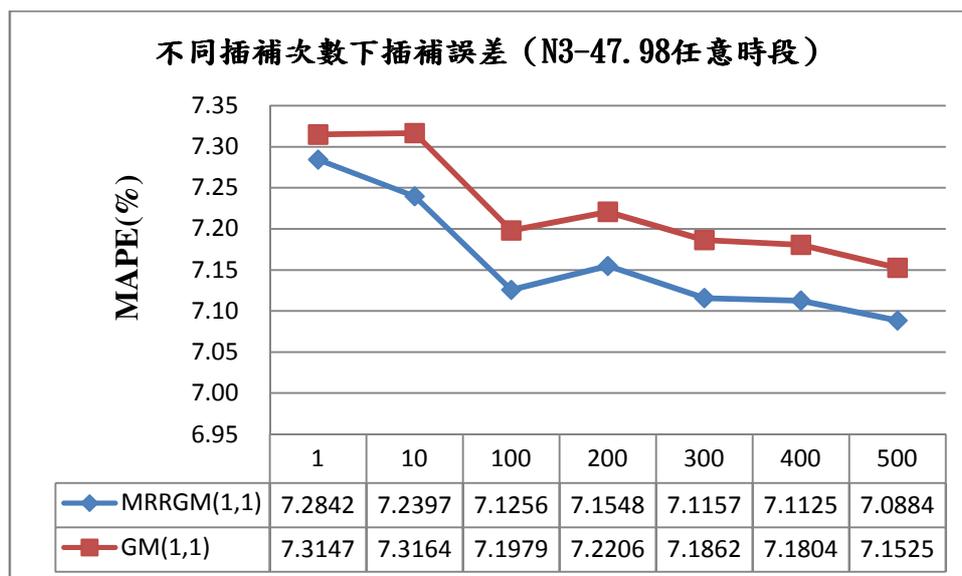


圖 4. 12 不同插補次數下插補誤差 (N3-47.98 任意時段)

依照插補次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value=0.39724)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.055751)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時間段

在尖峰時段下比較不同插補次數的插補誤差，根據圖 4.13 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優於 GM(1,1)，且隨著插補次數的增加，其插補誤差有下降的趨勢。由 Lewis[1982] 評估標準得知，在不同插補次數下其插補誤差小於 10%，則此兩種演算法屬於高精準的插補，由圖 4.13 知插補次數 10 次其誤差百分比最小。

由測試結果得知 MRRGM(1,1) 標準差 6.315061，GM(1,1) 標準差 6.440074，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

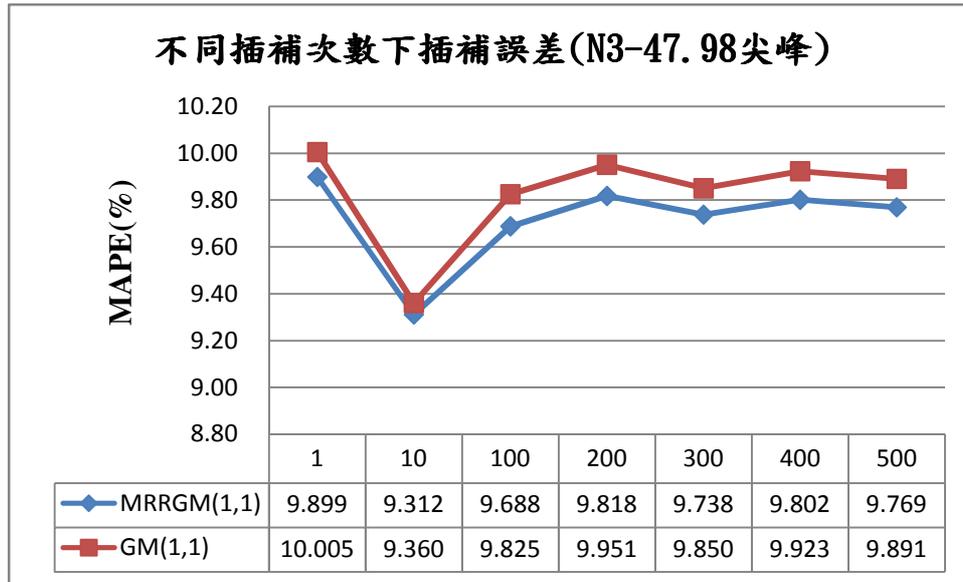


圖 4. 13 不同插補次數下插補誤差(N3-47.98 尖峰)

依照插補次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.387011)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.163119)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

在離峰時段下比較不同插補次數的插補誤差，根據圖 4.14 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優於 GM(1,1)，且隨著插補次數的增加，其插補誤差有下降的趨勢。由 Lewis[1982] 評估標準知，在不同插補次數下其插補誤差小於 10%，則此兩種演算法屬於高精準的插補，由圖 4.14 知插補次數 10 次其誤差百分比最小。

由實驗結果得知 MRRGM(1,1) 標準差 6.232187，GM(1,1) 標準差 6.354498，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

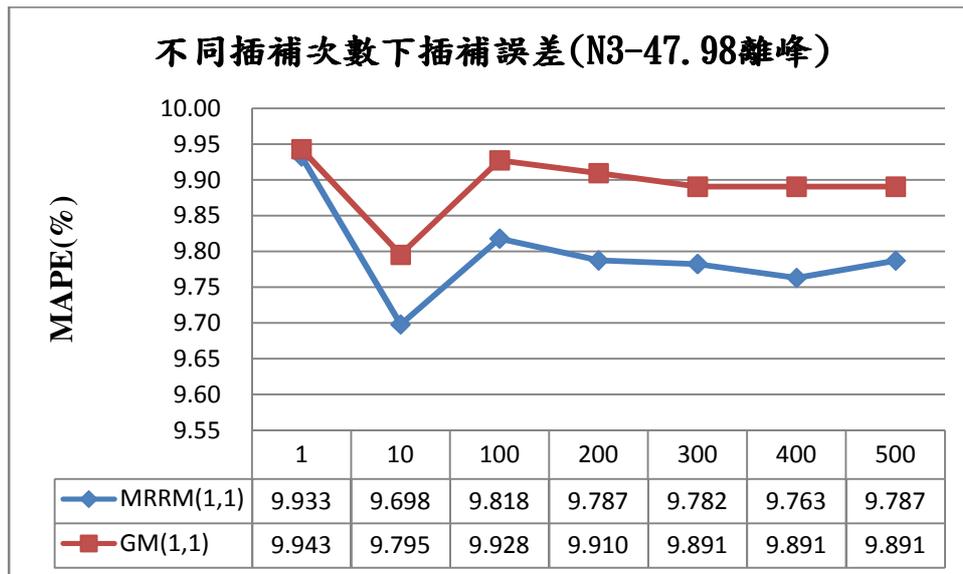


圖 4. 14 不同插補次數下插補誤差(N3-47.98 離峰)

依照插補次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.175782)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.005445)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

4.6.2 以 N3-50.38 為測試對象

1. 比較不同遺失比例下插補誤差之探討

(1) 任意時段

將插補次數固定，比較任意時段在不同遺失比例下其插補誤差，根據圖 4.15 得知這兩種演算法在不同遺失資料比例下，其插補誤差皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 5.250522，GM(1,1) 標準差 5.326293，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

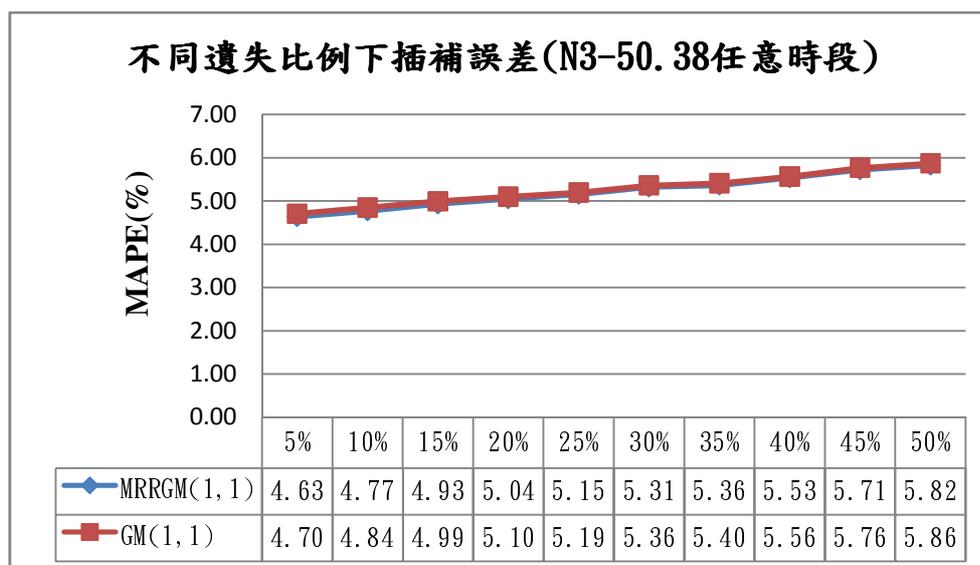


圖 4. 15 不同遺失比例下插補誤差(N3-50. 38 任意時段)

而依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value=0.465159)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.385476)，所示 GM(1, 1) 插補之 MAPE 值不大於 MRRGM(1, 1) 插補之 MAPE 值。可說明結果為在遺失比例不同的情況下 MRRGM(1, 1) 插補法之績效與 GM(1, 1) 插補法無顯著的差異。

(2) 尖峰時段

將插補次數固定，比較在不同遺失比例下其插補誤差，根據圖 4.16 得知這兩種演算法在不同遺失資料比例下，其插補誤差皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1, 1) 標準差 3.825625，GM(1, 1) 標準差 3.998382，則 MRRGM(1, 1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

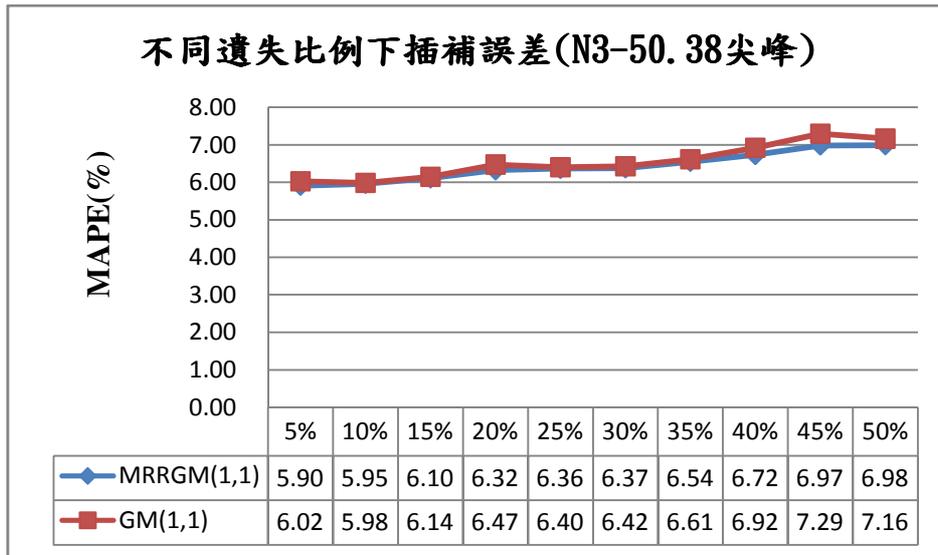


圖 4. 16 不同遺失比例下插補誤差(N3-50.38 尖峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.309067)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.270242)，所示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。可說明結果為在遺失比例不同的情況下 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

將插補次數固定，比較在不同遺失比例下其插補誤差，根據圖 4.17 得知這兩種演算法在不同遺失資料比例下，當遺失比例小於等於 30% 時，其插補誤差小於 10%，屬於高精準的插補。當遺失比例大於 30% 時，其插補誤差大於 10%，屬於優良插補。

由測試結果得知 MRRGM(1,1) 標準差 6.010745，GM(1,1) 標準差 6.587435，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

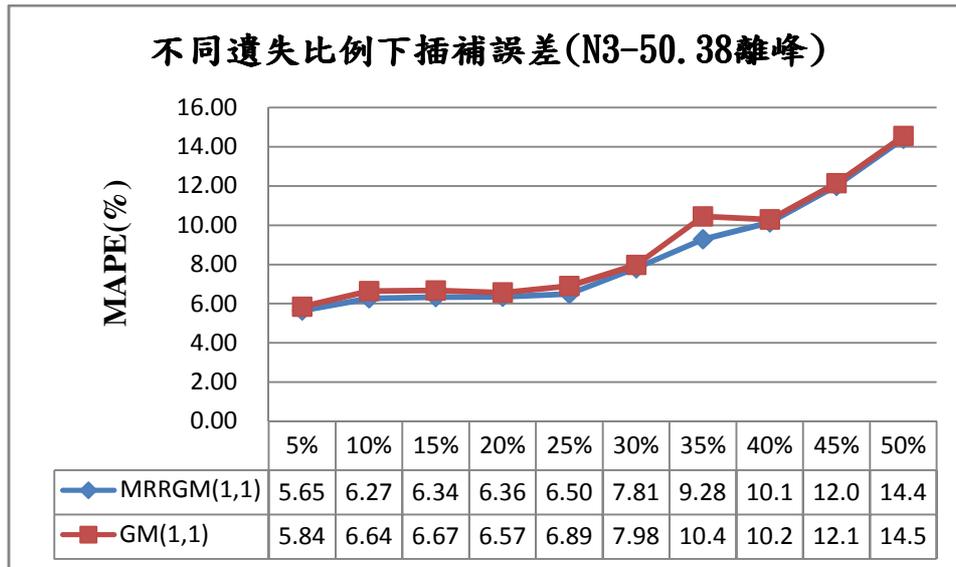


圖 4. 17 不同遺失比例下插補誤差(N3-50.38 離峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.494179)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.403002)，所示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。可說明結果為在遺失比例不同的情況下 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

2. 比較不同插補次數下插補誤差之探討

(1) 任意時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.18 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，插補次數 10 次時，插補誤差最小。由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 5.250522，GM(1,1) 標準差 5.326293，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

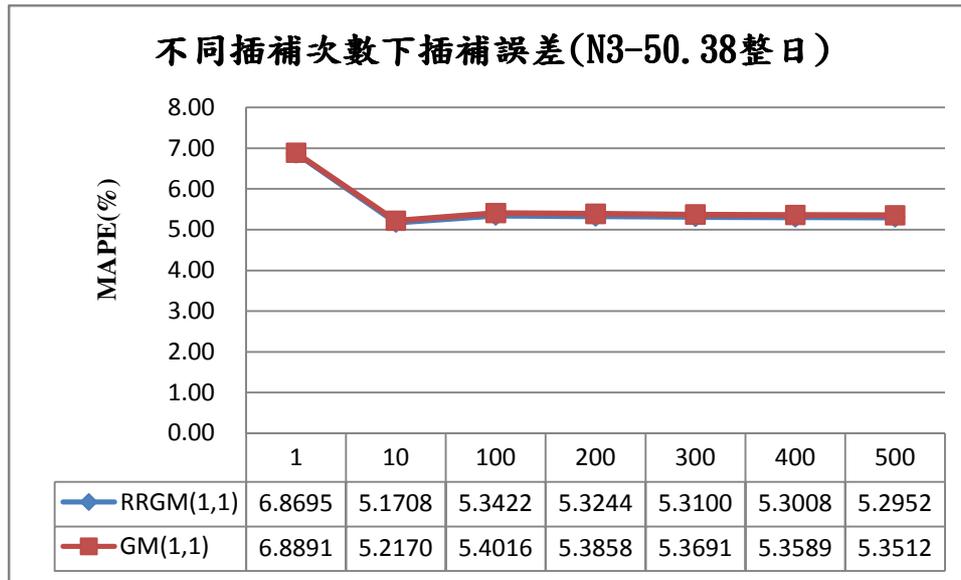


圖 4. 18 不同插補次數下插補誤差(N3-50.38 整日)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value=0.421525)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.297643)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.19 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 10 次時，插補誤差最小。由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 3.825625，GM(1,1) 標準差 3.998382，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

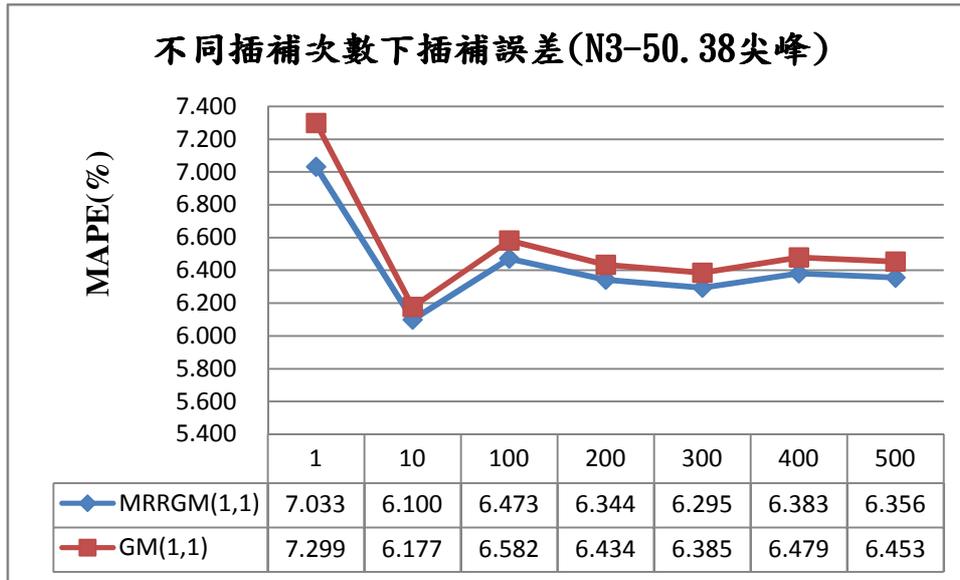


圖 4. 19 不同插補次數下插補誤差(N3-50.38 尖峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.319716)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.254654)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.20 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 100 次時，插補誤差最小。由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由實驗結果得知 MRRGM(1,1) 標準差 6.010745，GM(1,1) 標準差 6.587435，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

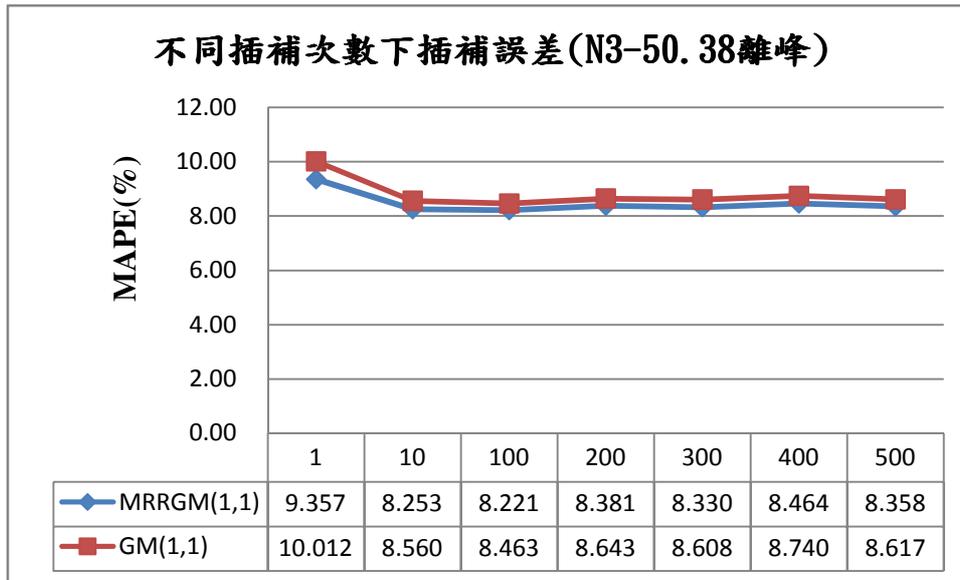


圖 4. 20 不同插補次數下插補誤差(N3-50.38 離峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.235268)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.110624)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

4.6.3 以 N3-51.55 為測試對象

1. 比較不同遺失比例下插補誤差之探討

(1) 任意時段

將插補次數固定，而比較任意時段在不同遺失比例下的插補誤差，根據圖 4.21 得知兩種演算法在資料遺失比例在 5% 與 10% 的情況下其插補誤差小於 10%，則此兩種演算法屬於高精準的插補。而資料遺失比例由 15% 到遺失比例 45% 情況下，其插補誤差百分比小於 20%，此兩種演算法屬於優良的插補。而資料遺失比例為 50% 的情況下，其誤差百分比大於 20%，此兩種演算法屬於合理的插補。

由測試結果得知 MRRGM(1,1) 標準差 4.980772，GM(1,1) 標準差 5.253874，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

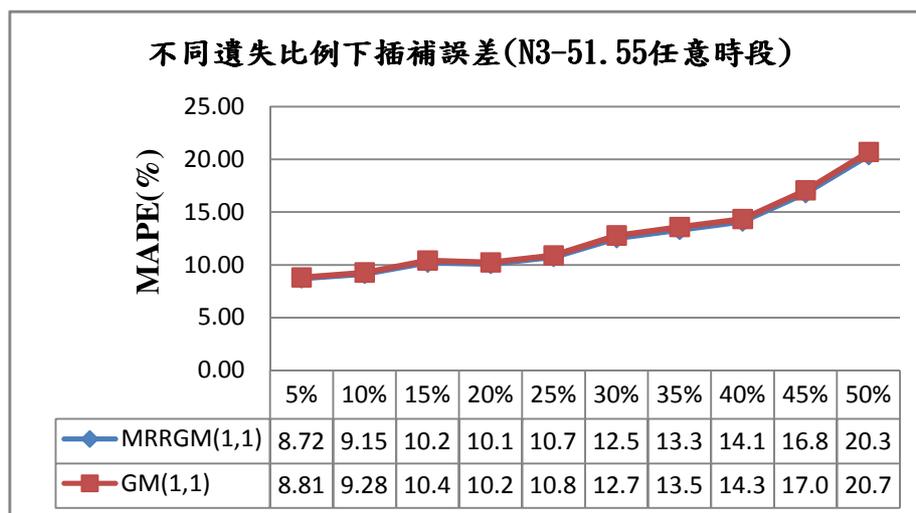


圖 4. 21 不同遺失比例下插補誤差(N3-51.55 任意時段)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value=0.480124)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.452919)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

將插補次數固定，比較尖峰時段在不同遺失比例下其插補誤差，根據圖 4.22 得知這兩種演算法在不同遺失資料比例下，當遺失比例小於等於 30% 時，則其插補誤差小於 10%，屬於高精準的插補。當遺失比例大於 30% 時，則其插補誤差介於 10%~20%，屬於優良插補。

由實驗結果得知 MRRGM(1,1) 標準差 6.845259，GM(1,1) 標準差 7.475309，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

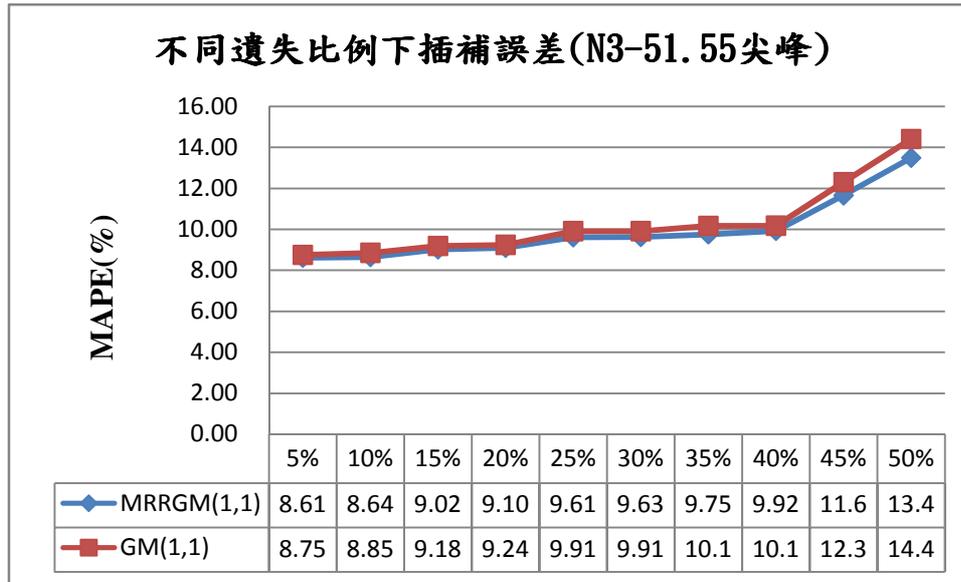


圖 4. 22 不同遺失比例下插補誤差(N3-51.55 尖峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.33063)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.321454)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

將插補次數固定，比較尖峰時段在不同遺失比例下其插補誤差，根據圖 4.23 得知這兩種演算法在不同遺失資料比例下，其插補誤差小於 10%，屬於高精準插補。

由實驗結果得知 MRRGM(1,1) 標準差 5.305978，GM(1,1) 標準差 5.538281，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

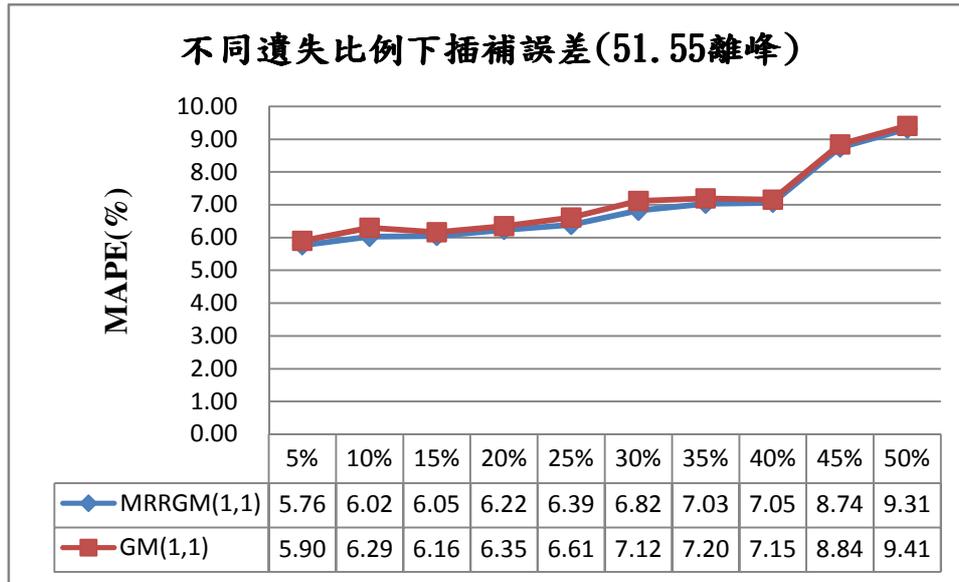


圖 4. 23 不同遺失比例下插補誤差(51.55 離峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.47204)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.380077)，所示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異

2. 比較不同插補次數下插補誤差之探討

(1) 任意時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.24 知 MRRGM (1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，插補次數 10 次時，插補誤差最小。而插補次數 1 次時插補誤差最大，由 Lewis[1982] 評估標準得知，所有資料在不同的插補次數下，其插補誤差百分比皆小於 20%，因此這兩種演算法皆屬於合理的插補。

由測試結果得知 MRRGM(1,1) 標準差 4.980772，GM(1,1) 標準差 5.253874，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

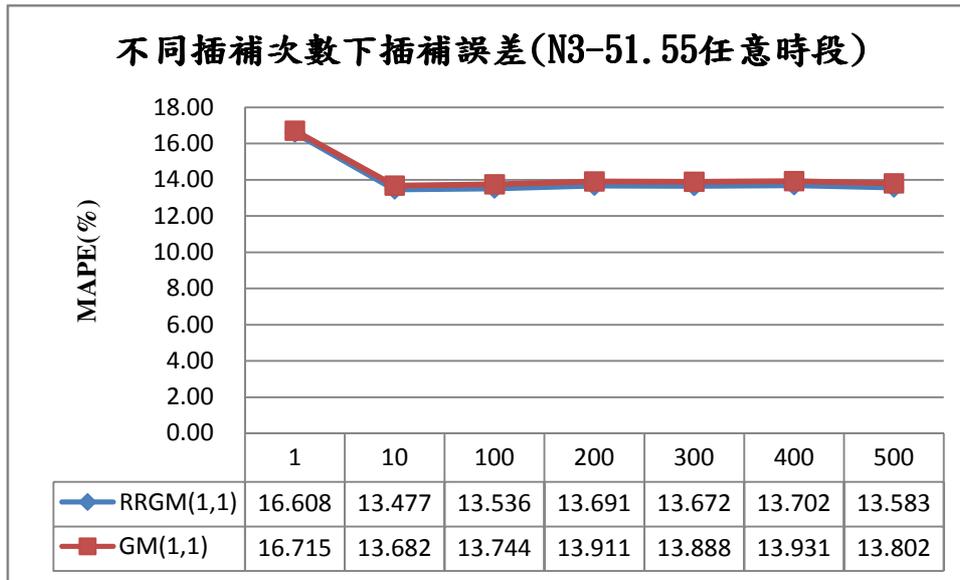


圖 4. 24 不同插補次數下插補誤差(N3-51.55 任意時段)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value=0.465705)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.371372)，表示 GM(1,1) 插補之 MAPE 值不大於 RRGM(1,1) 插補之 MAPE 值。所以 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

固定遺失比例下，比較不同插補次數下的插補誤差，根據圖 4.25 可知 MRRGM (1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 400 次時插補誤差最小。插補次數 1 次時插補誤差最大，由 Lewis[1982] 評估標準得知插補次數 400 次，其誤差百分比最小屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 4.980772，GM(1,1) 標準差 5.253874，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

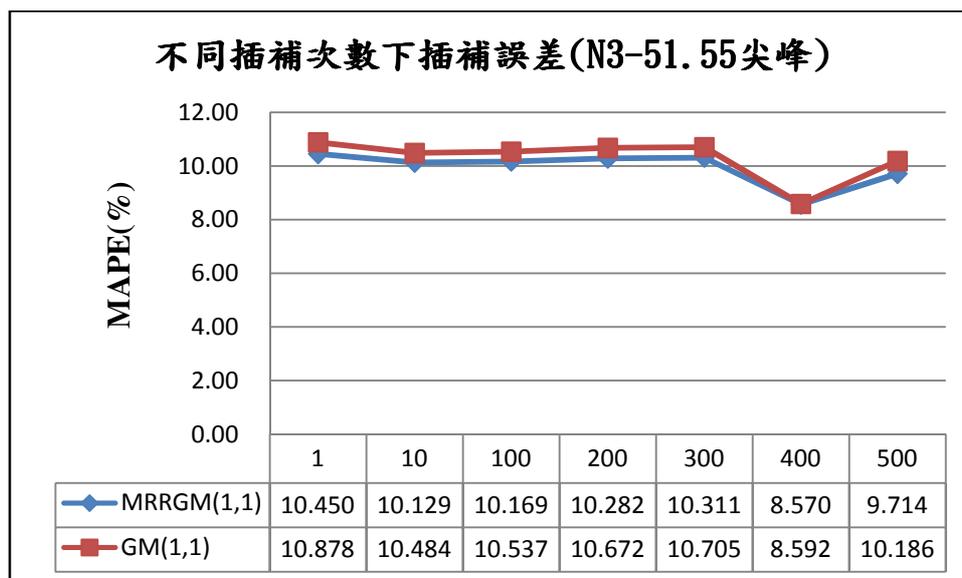


圖 4. 25 不同插補次數下插補誤差(N3-51.55 尖峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.33311)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.191792)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

固定遺失比例下，比較不同插補次數下的插補誤差，根據圖 4.26 知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 500 次時插補誤差最小，插補次數 1 次時插補誤差最大。由 Lewis[1982] 評估標準得知，所有資料在不同的插補次數下，其插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 5.305978，GM(1,1) 標準差 5.538281，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

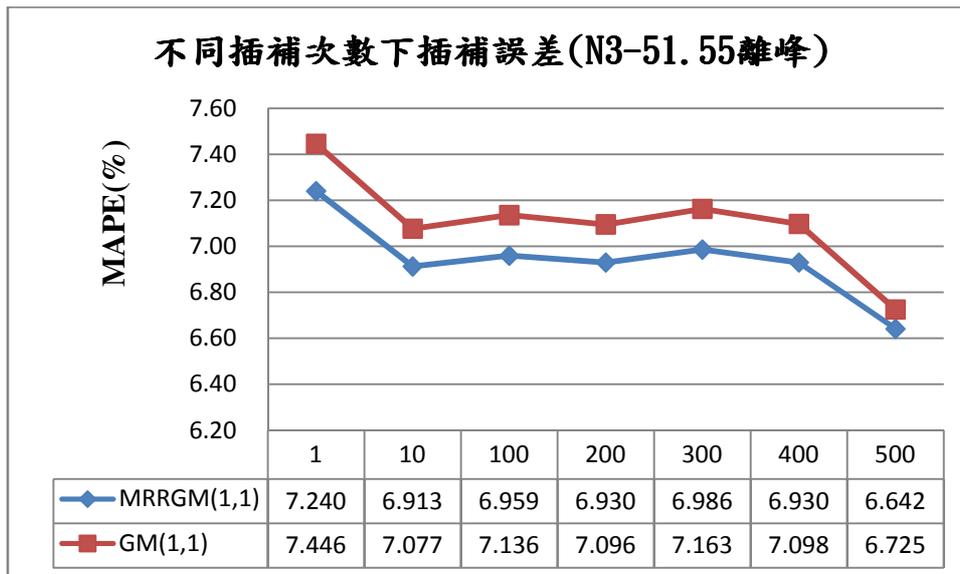


圖 4. 26 不同插補次數下插補誤差(N3-51.55 離峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.686802)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.070284)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

4.6.4 以 N3-54 為測試對象

1. 比較不同遺失比例下插補誤差之探討

(1) 任意時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4.27 得知兩種演算法在不同遺失資料比例下，由 Lewis[1982] 評估標準得知，其插補誤差百分比小於 20%，因此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 10.33656，GM(1,1) 標準差 10.51764，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

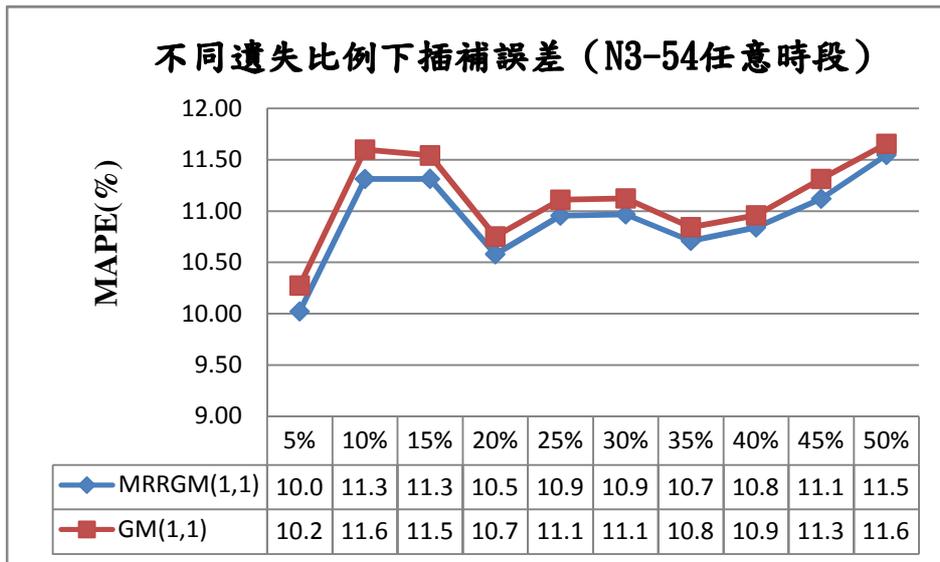


圖 4. 27 不同遺失比例下插補誤差 (N3-54 任意時段)

依照遺失比例的不同，可檢定兩模式之優劣。根據檢定檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value=0.489494)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.18351)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4.28 得知兩種演算法在不同遺失資料比例下，其插補誤差百分比小於 20%，因此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 12.74446，GM(1,1) 標準差 12.91936，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

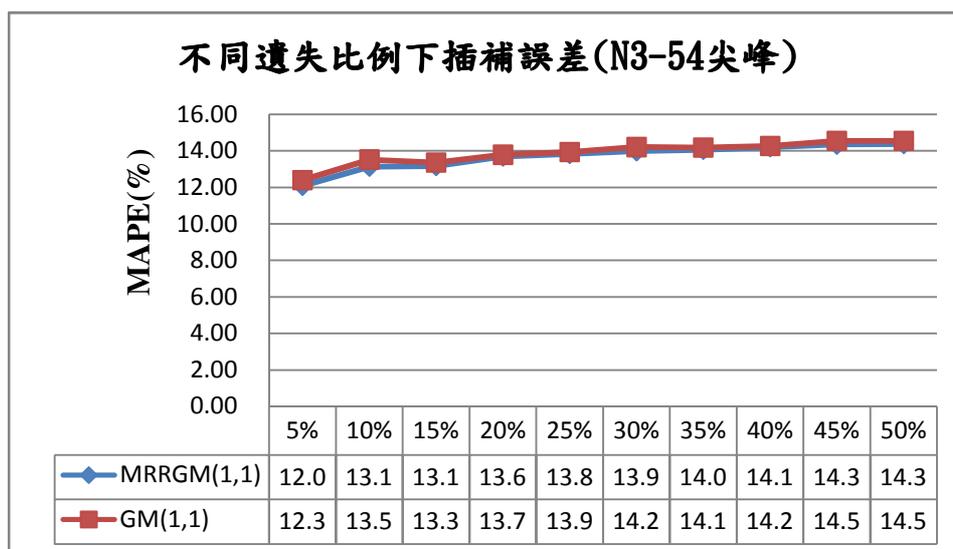


圖 4. 28 不同遺失比例下插補誤差(N3-54 尖峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據檢定檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.40194)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.268434)，則 GM(1, 1) 插補之 MAPE 值不大於 MRRGM(1, 1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1, 1) 插補法之績效與 GM(1, 1) 插補法無顯著的差異。

(3) 離峰時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4. 29 得知兩種演算法在不同遺失資料比例下，其插補誤差百分比小於 20%，因此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1, 1) 標準差 11.25183，GM(1, 1) 標準差 11.60636，則 MRRGM(1, 1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

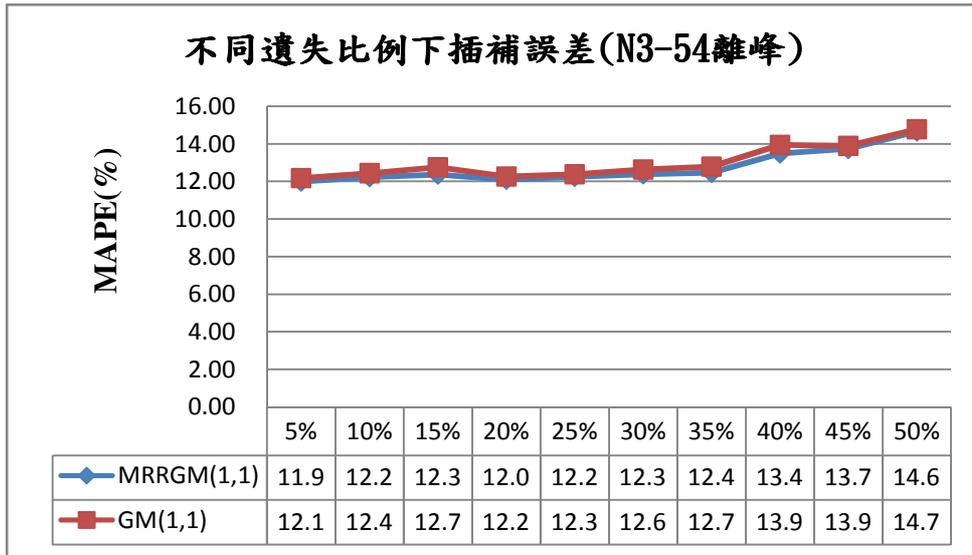


圖 4. 29 不同遺失比例下插補誤差(N3-54 離峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.497423)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.275206)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

2. 比較不同插補次數下插補誤差之探討

(1) 任意時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.30 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 100 次時插補誤差最小。而插補次數 1 次時插補誤差最大，由 Lewis[1982] 評估標準得知，所有資料在不同插補次數下的插補誤差百分比皆小於 20%，因此這兩種演算法皆屬於合理的插補。

由測試結果得知 MRRGM(1,1) 標準差 10.33656，GM(1,1) 標準差 10.51764，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

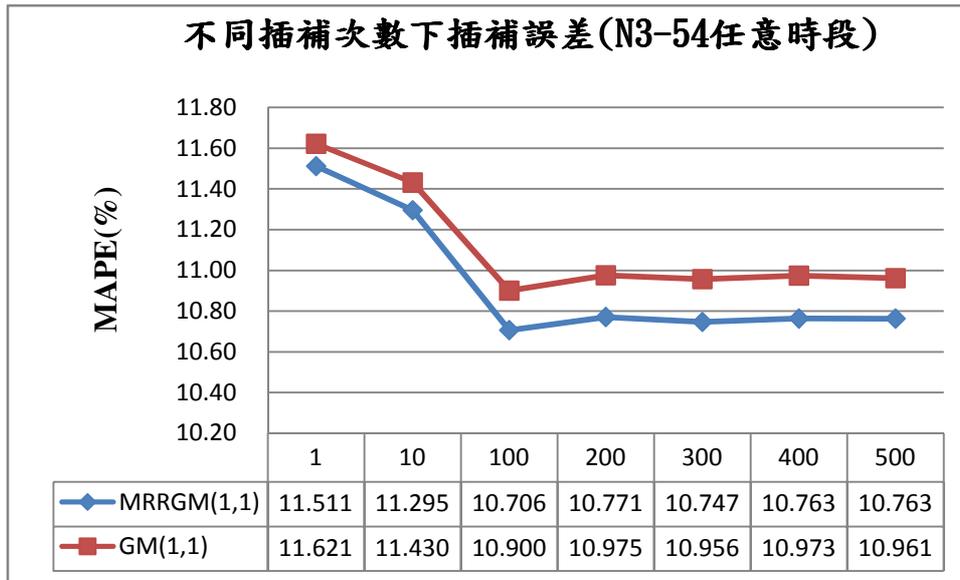


圖 4. 30 不同插補次數下插補誤差(N3-54 任意時段)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value=0.380811)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.146622)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.31 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 10 次時插補誤差最小。由 Lewis[1982] 評估標準得知，所有資料在不同插補次數下的插補誤差百分比皆小於 20%，因此這兩種演算法皆屬於合理的插補。

由測試結果得知 MRRGM(1,1) 標準差 12.74446，GM(1,1) 標準差 12.91936，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

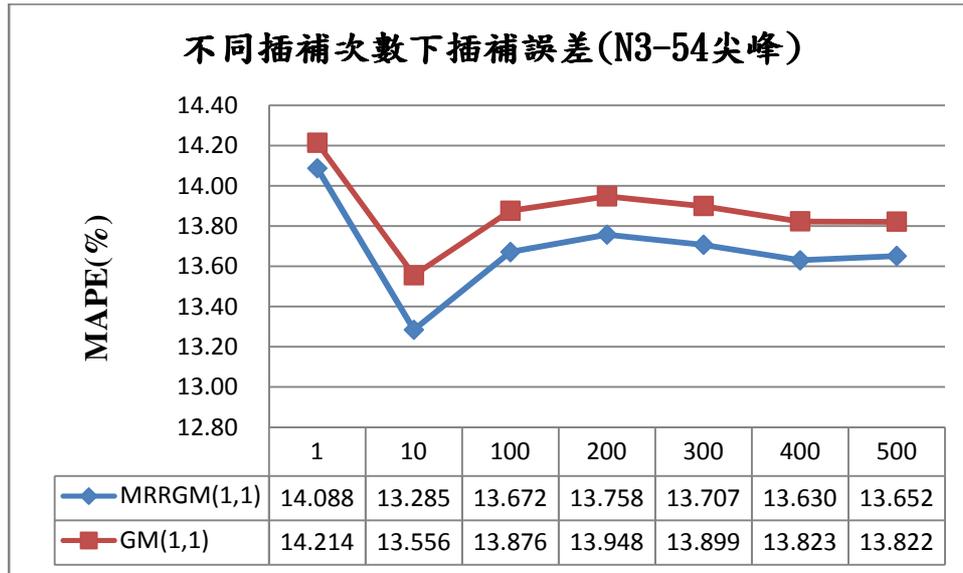


圖 4. 31 不同插補次數下插補誤差(N3-54 尖峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.329969)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.060931)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.32 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 500 次時插補誤差最小。由 Lewis[1982] 評估標準得知，所有資料在不同插補次數下的插補誤差百分比皆小於 20%，因此這兩種演算法皆屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 11.25183，GM(1,1) 標準差 11.60636，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

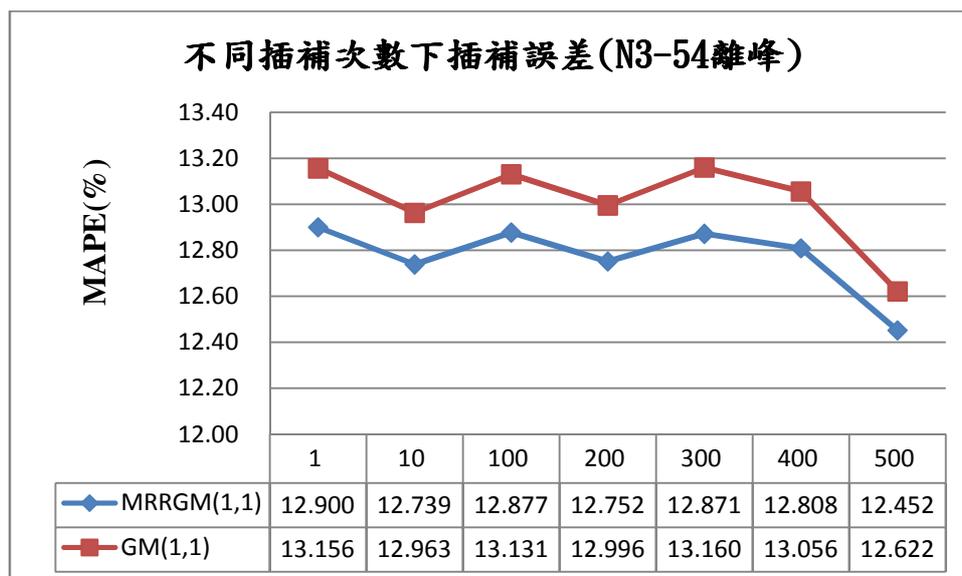


圖 4. 32 不同插補次數下插補誤差(N3-54 離峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.318552)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，拒絕 H_1 (P-Value= 0.01131)，表示 GM(1,1) 插補之 MAPE 值大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效優於 GM(1,1) 插補法。

4.6.5 以 N3-55.34 為測試對象

1. 比較不同遺失比例下插補誤差之探討

(1) 任意時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4.33 得知兩種演算法在不同遺失資料比例下，由 Lewis[1982] 評估標準得知，由資料遺失比例 5% 到遺失比例 45% 情況下，其插補誤差百分比小於 10%，這兩種演算法屬於高精準的插補。而資料遺失比例 50% 的情況下，其插補誤差百分比小於 20%，則此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 6.050358，GM(1,1) 標準差 6.61517，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

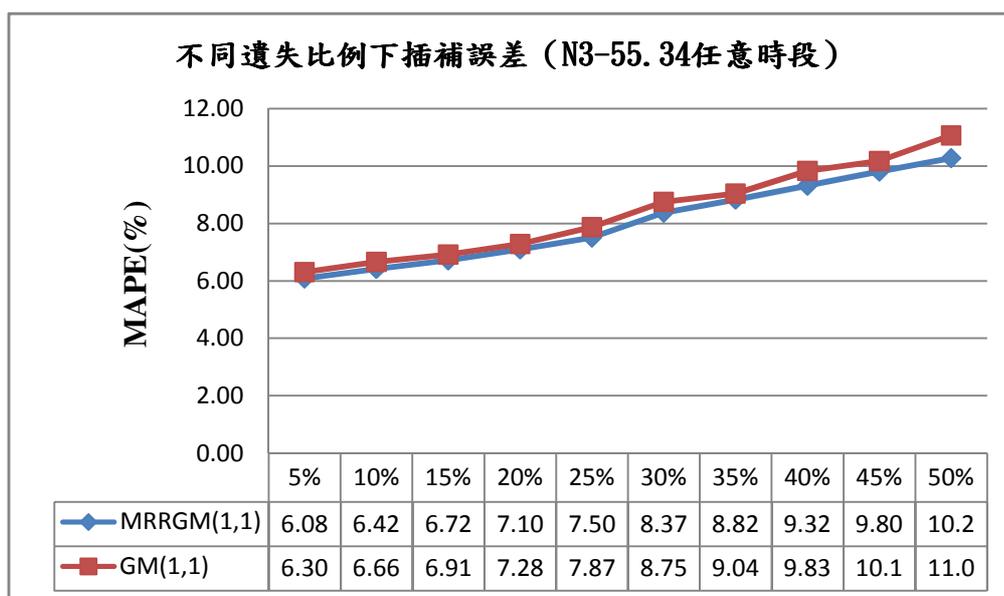


圖 4. 33 不同遺失比例下插補誤差 (N3-55.34 任意時段)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value=0.391167)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value=0.312054)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(2) 尖峰時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4.34 得知兩種演算法在不同遺失資料比例下，資料遺失比例 5%，其插補誤差小於 10%，屬於高精準插補。而遺失比例大於 5% 時，其插補誤差百分比小於 20%，則此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 8.24639，GM(1,1) 標準差 8.968251，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

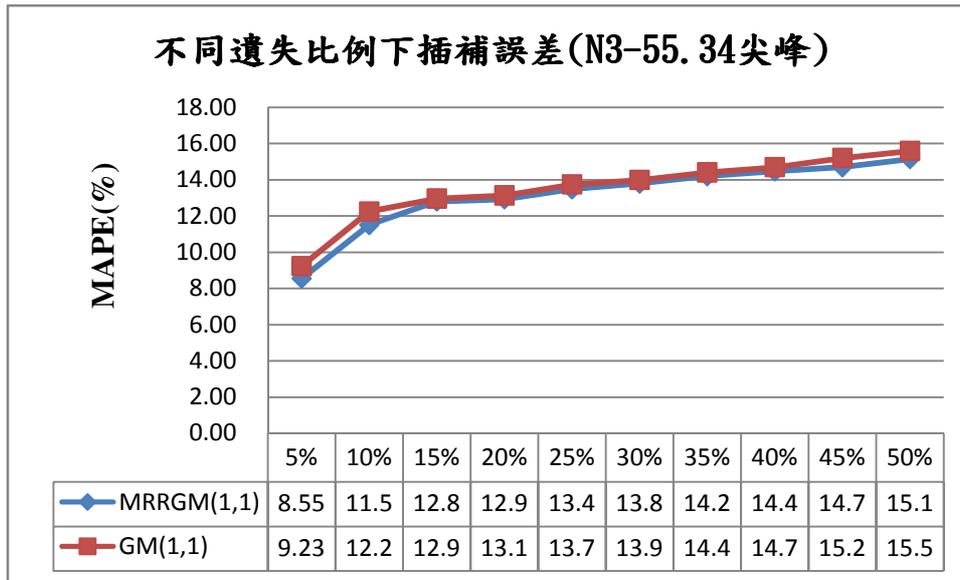


圖 4. 34 不同遺失比例下插補誤差(N3-55.34 尖峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.432411)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.336793)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

將插補次數固定，而比較不同遺失比例下的插補誤差，根據圖 4.35 得知兩種演算法在不同遺失資料比例下，當資料遺失比例小於 45%，其插補誤差小於 10%，屬於高精準的插補。而遺失比例大於等於 45% 時，其插補誤差百分比大於 10%，其插補誤差百分比小於 20%，則此兩種演算法屬於優良的插補。

由測試結果得知 MRRGM(1,1) 標準差 4.845049，GM(1,1) 標準差 5.538233，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

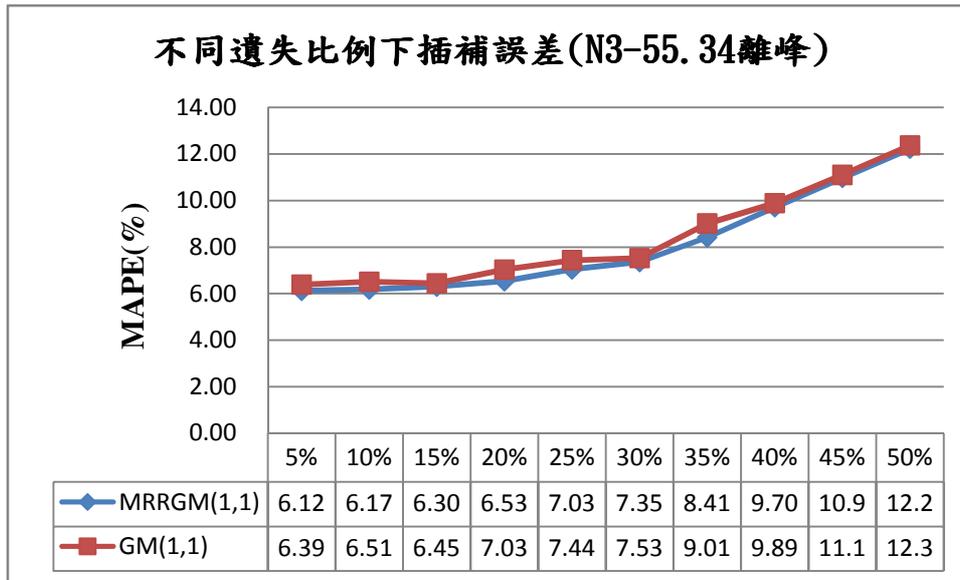


圖 4. 35 不同遺失比例下插補誤差(N3-55.34 離峰)

依照遺失比例的不同，可檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下不拒絕 H_0 (P-value= 0.467474)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，不拒絕 H_0 (P-Value= 0.383048)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

2. 比較不同插補次數下插補誤差之探討

(1) 任意時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.36 可知 MRRGM (1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 100 次時插補誤差最小，而插補次數 10 次時插補誤差最大。由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 6.050358，GM(1,1) 標準差 6.61517，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

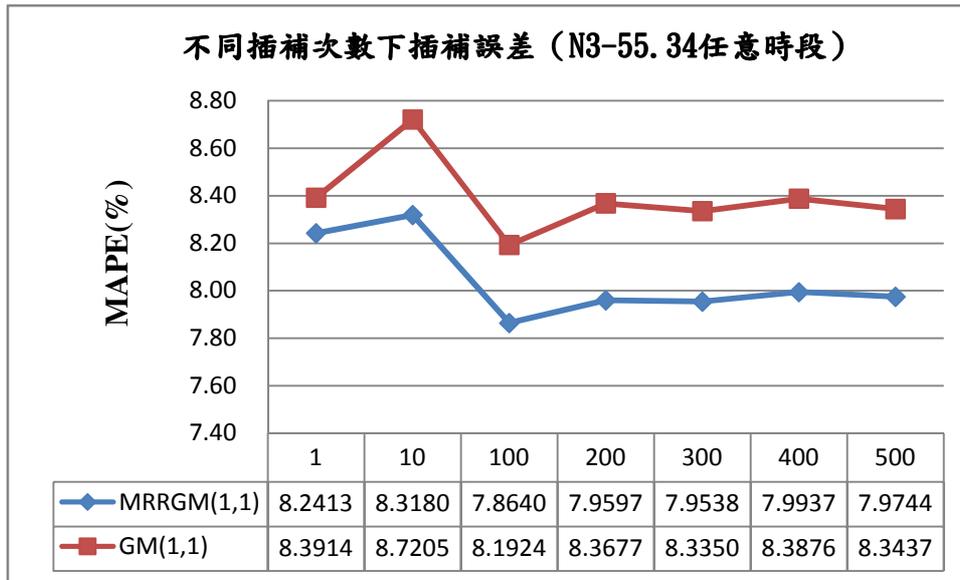


圖 4. 36 不同插補次數下插補誤差 (N3-55.34 任意時段)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 ($P\text{-value}=0.456177$)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，接受 H_1 ($P\text{-Value}=0.00094$)，表示 GM(1,1) 插補之 MAPE 值大於 MRRGM(1,1) 插補之 MAPE 值。所以 MRRGM(1,1) 插補法之績效與 GM(1,1) 插補法有顯著的差異。

(2) 尖峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.37 可知 MRRGM(1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 10 次時插補誤差最小。而插補次數 1 次時插補誤差最大，由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 8.24639，GM(1,1) 標準差 8.968251，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

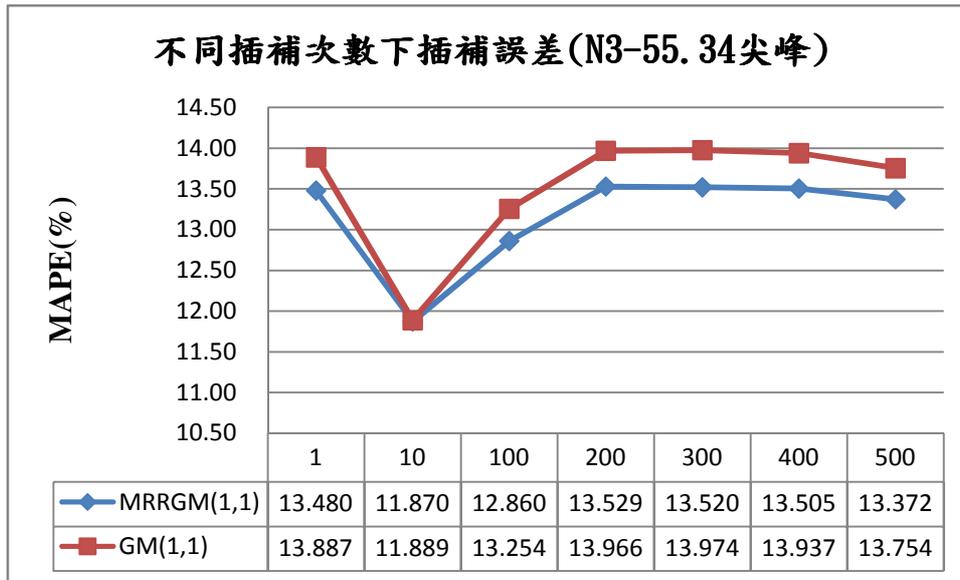


圖 4. 37 不同插補次數下插補誤差(N3-55.34 尖峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.309281)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，接受 H_0 (P-Value= 0.175189)，表示 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

(3) 離峰時段

固定遺失比例下，而比較不同插補次數下的插補誤差，根據圖 4.38 可知 MRRGM (1,1) 的插補績效在不同插補次數下皆優 GM(1,1)，由插補次數 400 次時插補誤差最小。而插補次數 1 次時插補誤差最大，由 Lewis[1982] 評估標準得知，所有資料不同插補次數下的插補誤差百分比皆小於 10%，因此這兩種演算法皆屬於高精準的插補。

由測試結果得知 MRRGM(1,1) 標準差 4.845049，GM(1,1) 標準差 5.538233，則 MRRGM(1,1) 的標準差較小，所插補資料值和遺失資料值的變異程度較小。

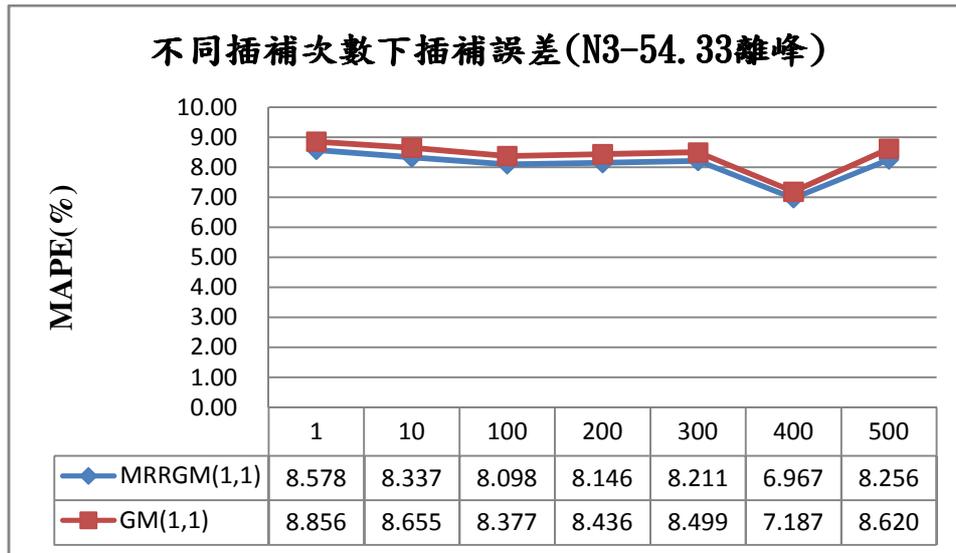


圖 4. 38 不同插補次數下插補誤差(N3-54.33 離峰)

依照測試次數的不同，檢定兩模式之優劣。根據統計檢定結果，在 $\alpha=0.05$ 之下，不拒絕 H_0 (P-value= 0.446774)，表示兩母體變異數無顯著的差異。再作兩母體平均數差異檢定，變異數相等情況下之 t 檢定，在 $\alpha=0.05$ 之下，由單尾檢定得知，接受 H_0 (P-Value= 0.163516)，則 GM(1,1) 插補之 MAPE 值不大於 MRRGM(1,1) 插補之 MAPE 值。結果為在遺失比例不同的情況下 MRRGM (1,1) 插補法之績效與 GM(1,1) 插補法無顯著的差異。

小結

本研究為了探討兩種插補法何種插補績效較好，在不同遺失比例的情況下，以 MRRGM(1,1) 插補績效優於 GM(1,1)。

1. 在不同遺失比例下其插補誤差

- (1) 比較整日時段插補績效，以車輛偵測器 N3-47.98、N3-50.38、N3-55.34，其插補誤差小於 10%，屬於高精準插補。而 N3-51.55、N3-54 其插補誤差介於 10%~20%，屬於優良插補。
- (2) 比較尖峰時段插補績效，以車輛偵測器 N3-50.38，其插補誤差小於 10%，屬於高精準插補。而 N3-47.98、N3-51.55、N3.54、N3-55.34 其插補誤差介於 10%~20%，屬於優良插補。
- (3) 比較離峰時段插補績效，以車輛偵測器 N3-47.98、N3-50.38、N3-51.55、N3.54、N3-55.34 其插補誤差介於 10%~20%，屬於優良的插補。

另外本研究比較不同插補次數的情況下，以 MRRGM(1,1) 插補績效優於

GM(1,1)。由其插補結果可發現隨著插補次數的增加，會逐漸降低插補誤差。

2. 在不同插補次數下其插補誤差

(1)比較整日時段插補績效，以車輛偵測器 N3-47.98、N3-50.38、N3-55.34，其插補誤差小於 10%，屬於高精準的插補。而 N3-51.55、N3-54，插補誤差介於 10%~20%，屬於優良插補。而整日時段各車輛偵測器最佳插補次數如下：

表 2. 9 車輛偵測器(VD)任意時段最佳插補次數

偵測器編號	N3-47.98	N3-50.38	N3-51.55	N3-54	N3-55.34
最佳插補次數	500 次	10 次	10 次	100 次	100 次

(2)比較尖峰時段插補績效，以車輛偵測器 N3-47.98、N3-50.38，其插補誤差小於 10%，屬於高精準的插補。而 N3-51.55、N3-54、N3-55.34，插補誤差介於 10%~20%，屬於優良插補。而尖峰時段各車輛偵測器最佳插補次數如下：

表 2. 10 車輛偵測器(VD)尖峰時段最佳插補次數

偵測器編號	N3-47.98	N3-50.38	N3-51.55	N3-54	N3-55.34
最佳插補次數	10 次	10 次	400 次	10 次	10 次

(3)比較離峰時段插補績效，以車輛偵測器 N3-47.98、N3-50.38、N3-51.55、N3-55.34，其插補誤差小於 10%，屬於高精準的插補。而 N3-54，插補誤差介於 10%~20%，屬於優良插補。而離峰時段各車輛偵測器最佳插補次數如下：

表 2. 11 車輛偵測器(VD)離峰時段最佳插補次數

偵測器編號	N3-47.98	N3-50.38	N3-51.55	N3-54	N3-55.34
最佳插補次數	10 次	100 次	500 次	500 次	400 次

第五章 結論與建議

本研究之目的以灰色理論為基礎，發展出新的插補法可針對資料遺失值進行插補，因此以不同車輛偵測器(VD)資料作為插補驗證之用，而根據前述各章節之分析與探討，本研究提出下列結論與建議：

5.1 結論

本研究之驗證方法藉由兩種方式為不同的遺失比例和不同的插補次數，來探討何種演算法的插補績效較好。

1. 在不同遺失比例下的情況下，以車輛偵測器 N3-47.98、N3-50.38、N3-55.34 其插補誤差小於 10%，因此可說明 MRRGM(1, 1) 與 GM(1, 1) 屬於高精準的插補。儘管在統計檢定得知此兩者插補法並無顯著的差異，但可從數據上得知本插補法 MRRGM(1, 1) 插補績效優於 GM(1, 1) 法。
2. 在不同的插補次數的情況下，隨著插補次數的增加，其誤差百分比(MAPE) 有逐漸下降的趨勢，結果說明多次插補法優於單次插補法。此外，由於兩種演算法其平均絕對誤差 (MAPE) 差距甚小，由統計檢定得知，只有車輛偵測器 N3-54 有顯著的差異，而其餘車輛偵測器均無顯著的差異，但可從數據上得知本插補法 MRRGM(1, 1) 插補績效優於 GM(1, 1) 法。
3. 當遺失比例高和插補次數多時，本插補法 MGGRM(1, 1) 其插補績效可優於 GM(1, 1) 法。

5.2 建議

1. 在資料遺失值插補法可分為「空間插補」和「時間插補」，而本研究所採用的是時間插補原則，來插補單一車輛偵測器(VD)的遺失值。根據 Chen[2002] 文獻指出待插補偵測器與鄰近偵測器間具有直接相關性，故可選擇相鄰上、下游偵測器之資料，作為插補資料的輸入值。因此未來本研究可針對空間插補的概念，倘若有某一偵測器遺失大量的資料，則可利用鄰近偵測器的資料來插補遺失值。
2. 本研究所採用歷史比對方式為「局部性」的灰關聯法(GRA)，以單因子比較速度的相似程度下，去尋找關聯度最大的歷史數列。而灰關聯分析法有「多

因子、同時比較」的優點，常用於偏好性理論的研究如 AHP，未來可建議後續研究者考慮比較多因子如速度和流量，期許可尋找到關聯度更佳的歷史數列，以作為後續插補法之用。

3. 本研究利用灰關聯法(GRA)和 MRRGM(1,1)法發展插補演算法，由實驗結果得知優於灰預測 GM(1,1)的方法。而在插補文獻中利用灰關聯法(GRA)和其他演算法如類神經(RNN)、KNN 法，來做遺失值資料插補。因此在未來應用上建議搭配 MRRGM(1,1)和其他演算法，可比較其插補績效是否優於灰關聯法(GRA)和 MRRGM(1,1)法。
4. 本研究利用殘差(Residual)概念，以 GM(1,1)灰預測為基礎發展新的插補法。因此在殘差的計算上存在著許多課題，不同的計算方式會影響到插補的結果，建議後續研究者可針對殘差的計算方式加以研究，以期許能得到更好的插補結果。
5. 本研究在多次插補設計上採用試誤方式尋找最小的插補誤差，而結果顯示並非插補次數越多的情況下，其誤差最小。未來可探討多次插補的合理範圍，以期許找到最小的插補誤差值。
6. 本研究設計不同遺失比例為 5%、10%、...、50%，而由結論得知當遺失比例增加和插補次數增加時，本演算法 MRRGM(1, 1)可得到較佳的插補績效。因此，在未來的研究可探討遺失比例 5%、10%、...、100%的情況下，是否本插補法其插補績效優於 GM(1, 1)。

參考文獻

1. Chen, D., Muller, S.G., Mussone, L. and Montgomey, F., "A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting," *Neural Computing & Applications*, pp. 277-286, 2001
2. Deng, J. L., "Control problems of grey systems," *System and Control Letters*, Vol. 1, No. 5, pp. 288-294, 1982
3. Gold, D. L., Turner, S. M., Gajewski, B. J. and Spiegelman, C., "Imputing Missing Values In ITS Data Archives For Intervals Under 5 Minutes," *Transportation Research Board 80th Annual Meeting January 7-11, Washington, D.C, 2001.*
4. Huang, X. L. and Zhu, Q. M., "A Pseudo-nearest-neighbor Approach For Missing Data Recovery On Gaussian Random Data Sets," *Pattern Recognition Letters*, Vol.23, pp. 1613-1622, 2002.
5. Huang, C. C. and Lee, H. M., "A Grey-based Nearest Neighbor Approach For Missing Attribute Value Prediction," *Applied Intelligent*, Vol.20, No.3, pp.239-252, 2004.
6. Jou, Y. J.; Lee, T. T.; Lan, C.L.; Hsu, C. H., "The implementation of Markov bias corrected grey system in freeway travel time prediction" *Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, Volume :1, IEEE 2003*
7. Mei Chen, Jingxin Xia, Rongfang Liu, "Developing a Strategy for Imputing Missing Traffic Volume Data", *JOURNAL OF THE TRANSPORTATION 121 RESEARCH FORUM*, Vol.45, No.3, pp. 57-76, 2006.
8. Qu, L., Li, L., Zhang, Y. and Hu, J., "PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach", *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, Vol. 10, No. 3, pp.512-522, 2009.
9. Wen, Y. H., Lee, T. T. and Cho, H. T., "Missing Data Treatment And Data Fusion Toward Travel Time Estimation For ATIS," *Journal of the Eastern Asia Society for Transportation Studies*, Vol.6, pp. 2546-2560, 2005.
10. Yang Zhang, Yuncai Liu, "Missing Traffic Flow Data Prediction using Least Squares Support Vector Machines in Urban Arterial Streets", *IEEE Symposium on Computational Intelligence and Data Mining*, pp. 76-83, 2009.
11. Zhang, Y., Liu, Y., "Missing Traffic Flow Data Prediction using Least Squares Support Vector Machines in Urban Arterial Streets", *IEEE Symposium on Computational Intelligence and Data Mining*, pp. 76-83, 2009.

12. Little, R. J. A., and Rubin, D. B., "Statistical Analysis with Missing Data", 2nd edition, John Wiley and Sons, New York, USA, 2002.
13. Shen, J. J. , and Chen, M. T., "A Recycle Technidue of Association Rule for Missing Value Completion," Proceeding, 17th International Conference on Advanced Information Networking and Applications, Xi'an, China, pp.638-641, 2003.
14. Lee, R. C. T., Slagle, J. R., and Mong, C. T., "Application of Clustering to Estimate Missing Data and Improve Data Integrity," Processedings, 2nd international conference on Software engineering, San Francisco, CA, USA, pp.539-544(1976).
15. Pedreira, C. E., and Parente, E., "Neural Networks with Missing Values Attributes", Proceedings, IEEE International Conference on Neural Networks. Vol. 6, pp.3021-3023, 1995.
16. Hsu, C. C. , Chen, C.Y. , "Aplications of improved grey prediction model for power demand forecasting" , Energy Conversion and Management 44,2241-2249,2003
17. Lee, Y. S. , Tong, I. L. , "Forecasting energy consumption using a grey model improved by incorporating genetic programming" , Energy Conversion and Management 52,147-152,2010
18. Chengqi Z. , Xiaofeng Z. , Jilian Z. , Yongsong Q. Z. , "GBKII: An Imputation Method for Missing Values* " , Springer-Verlag Berlin Heidelberg, 1080-1087, 2007
19. Caruana, R. , "A Non-parametric EM-style algorithm for Imputing Missing Value" . Artificial Intelligence and Statistics, January 2001.
20. Erdal K., Baris U., Okyay K., "Grey system theory-based models in time series prediction" . Expert Systems with Applications 37 ,1784–1789, 2010.
21. Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P., "Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems," Proceedings of the 82th Annual Meeting of Transportation Research Board, 2002.
22. 張廷政, 李碩建與林育政, 2009, "GM(1, 1) 循環式殘插修正法", 灰色系統期刊, 第二卷, 頁14-22
23. 陳冠竹, 2004, "高速公路旅行時間灰色預測模式之研究", 國立台灣大學土木工程學研究所碩士論文。
24. 曾憲雄等人, 2008, "資料探勘 Data Mining", 旗標出版股份有限公司。
25. 黃承傳, 戴輝煌, 2008, "兩岸三地主要樞紐港口相對競爭力之分析", 運輸學刊, 第二十卷, 第一期, 頁1-38。
26. 邱妍菁, 2006, "高速公路短期交通資訊之灰預測模型", 私立逢甲大學交通工程與管理學系碩士班碩士論文。

27. 溫裕宏,1997,“航空運量預測與航空網路設計之研究—應用灰色理論”,國立交通大學運輸科技與管理研究所
28. 卓訓榮,林國顯,張瓊文,藍健綸, 96年12月,“小型車高速公路旅行時間預測模式—應用灰色理論”,中華民國運輸學會學術論文國際研討會。
29. 卓訓榮,藍健綸,林春馨,林國顯,楊幼文,蔡瑞鉉, 94年11月,「灰色理論於高速公路旅行時間之應用」,中華民國運輸學會第20屆學術論文國際研討會
30. 吳健生與廖梓淋,「利用資料填補概念探討車輛偵測器佈設間距」,運輸學刊,第二十二卷,第三期,頁307-326,99年9月。
31. 張堂賢與黃宏仁,「車輛偵測器資料漏失之在線插補技術研究」,運輸學刊,第二十卷,第四期,頁377-404,97年12月。
32. 劉思峰與謝乃明,「灰色系統理論及其應用」,科學出版社,1991
33. 吳漢雄、鄧聚龍與溫坤禮,「灰色分析入門」,高立圖書有限公司,1996
34. 溫坤禮等,「灰預測原理與應用」,全華科技圖書股份有限公司,2002



簡歷



姓 名：許程詠

出生地：嘉義縣

出生日期：民國72年10月28日

聯絡地址：嘉義縣民雄鄉東榮村16鄰文化路26之52號

電子信箱：millersh@yaho.com.tw



學歷：

民國100年7月 國立交通大學運輸科技與管理學系碩士班畢業

民國96年6月 私立東海大學工業工程與經營資訊學系學士班畢業

民國92年6月 嘉義縣私立協同高中畢業