



國立臺灣大學共同教育中心統計碩士學位學程

碩士論文

Master's Program in Statistics

Center for General Education

National Taiwan University

Master Thesis

抽樣方式及抽樣方法對於土地利用變遷模式之影響

—以桃園地區為例

The Effect of Sampling Method to Land Use Land Change

Modeling—the case study of Taoyuan area

王馨儀

Sing-Yi Wang

指導教授：林裕彬 博士

Advisor: Yu-Pin Lin, Ph.D.

中華民國一一二年二月

February, 2023

謝誌



回想碩士階段的三年間發生了許多的大大小小的事，改變我對於事物的看法與認知，尤其是 covid-19 的疫情深深影響了所有人的日常生活與我原本規畫好的人生藍圖。然而可以順利完成這本碩士論文，我需要感謝所有給予過我幫助的人。

首先最感謝的是林裕彬老師的指導，在老師的細心引導與層層修改之下，我才得以順利完成碩士論文。也十分感謝老師願意讓我在碩士的最後一年到日本交換留學，並協助我申請日本台灣交流協會的獎學金。此外，我也十分感謝口試委員給予我的建議，使本研究可以更加完善。

再來感謝在碩士期間給予我協助及指點人生迷津的大家，霈欣、雅晴、文欣、志源、瑤恩、馥容、冠妤、惠安、詹兆雯老師；研究室的同學們，中安、倩敏、安村、連博士等；統計碩士學程的同學、劉錡、妍凌助教、畢琳助教、鄭克聲老師、蔡政安老師；以及日本交換期間的指導老師牛垣先生、有澤先生。

最後感謝永遠最支持我的家人，使我可以在無後顧之憂的條件下順利完成碩士學位。雖然偶爾會吵架，但總是給予我最客觀建議的姊姊，以及擔任我和姐姐吵架時傳聲筒的記賢；總是把我放在第一順位的爸爸跟媽媽；不吝於給予我關愛與疼惜的外公外婆及其他的家族成員們；以及我的精神支柱腿腿、あいみょん、中村倫也。因為有你們給予的包容、關愛及慰藉，我才得以完成學業。

我會持續懷抱謙虛向學的心邁向人生接下來的旅程，揮別學生身分努力朝向獨當一面的社會人邁進。

馨儀 2022.02.15

摘要



土地利用變遷的動態過程中具有多重複雜性與不確定性。在眾多領域中皆為重要的研究議題，如：地理學、生態學及社會科學領域。隨著資料科學的進步，土地利用研究目前可處理的網格資料精度日趨精細，亦使得演算時間大幅增加。然而由於土地利用資料相較於一般二維資料具有空間自相關性，容易使得分類模型產生偏誤以至於無法準確代表該研究區的土地利用分布情形。因此如何有效控制資料間的空間自相關性以降低模擬及預測模型中的偏誤(bias)是建立模型中的首要課題。

本研究以桃園地區為研究區比較系統抽樣(systematic sampling)、隨機抽樣(random sampling)與分層隨機抽樣(stratified random sampling)在不同抽樣比例下應用邏輯斯迴歸(logistic regression)、廣義可加性模型(general Additive Model)、隨機森林(random forest)三種演算法對於 CLUE-s 模式的土地利用驗證結果之影響。推論結合各土地利用類別之最佳演算法的混合演算模式對於 CLUE-s 模式是否有過度擬合(overfitted)的問題。此外，本研究亦針對僅使用兩時期土地利用地圖之情形提出可完整校正及驗證模型之方法。

研究結果顯示，抽樣方式差異對於 CLUE-s 模式之模擬結果並無顯著影響，然而抽樣比例差異對於模型驗證則有顯著影響，且在 3 種演算法中皆以 100%抽樣模型為最佳；若採取 90%樣本於邏輯斯迴歸，60%於廣義可加性模型或是 90%於隨機森林之 CLUE-s 模式則為最具效率之抽樣比例。結合各土地利用 AUC 值最佳演算法之混合模型並無法提升 CLUE-s 模式土地利用配置精確度，反而影響其他土地利用配置產生誤差。此外，本研究之結果可適用僅有 1~2 時期之土地利用圖資情形，先以網格抽樣方式切割時期 1 資料集建立土地利用變遷模型，再以時期 2 圖資驗證，以確保模型預測結果與實際利用情形相符。

關鍵字：土地利用變遷模式、CLUE-s 模式、土地利用抽樣、三圖比較、品質因素、土地利用模型驗證

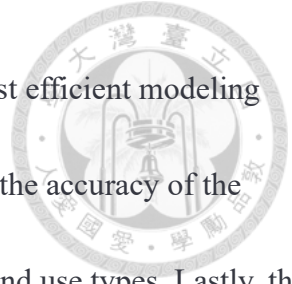
Abstract



The process of land use change contains multiple complexities and uncertainties. Therefore, it is viewed as an important issue for global environmental change. As data science has advanced, the resolution of raster data has become higher, but this also makes the workload heavier. Additionally, land use data have spatial autocorrelation, which can cause bias in classification algorithms and prevent the model from representing the real land use. Thus, it is crucial to efficiently control the autocorrelation in land use and land change research.

This study used Taoyuan as the research area to compare the differences between three sampling methods (systematic sampling, random sampling, and stratified random sampling) in different sampling ratios (100%, 90%, 80%, 70%, 60%, and 50%) with three classification algorithms (logistic regression, general additive model, and random forest) in CLUE-s model performance. Furthermore, it explored the overfitting problem in the mixed model which combined the best suitability of each land use type. Additionally, the research proposed a method that can completely calibrate and validate the model for the situation where only two periods of land use maps are used.

First, the results show that there is no significant difference between the sampling methods. It shows a significant difference in sampling ratios, with the 100% sampling model having the highest accuracy in all three algorithms. However, the 90% sample rate in logistic regression, the 60%



sample rate in GAM, and the 90% sample rate in random forest are the most efficient modeling methods in the Taoyuan area. Second, the mixed model could not improve the accuracy of the CLUE-s model; instead, the mixed model reduces the allocation of other land use types. Lastly, this research shows a stable and solid method of validating a LULC model by separating the period 1 land use data into a training dataset and a testing dataset with spatial sampling methods and validating the LULC model with the period 2 land use data. The procedure can ensure that the LULC model conforms to the real land use situation.

The results of this research could provide a viewpoint on the effect of grid sampling in land-use modeling and land-use land-change research.

Key Word: land-use modeling, land use land change, CLUE-s model, grid sampling, Three map comparison, Figure of Merit, LULC model validation

目錄



謝誌.....	i
摘要.....	ii
Abstract	iii
目錄.....	v
圖目錄.....	vii
表目錄.....	ix
壹、緒論.....	1
第一節、研究背景與動機.....	1
第二節、研究目的與問題.....	4
貳、文獻回顧.....	5
第一節、抽樣與土地利用變遷模式.....	5
第二節、CLUE-s 模式.....	8
第三節、模型校正與驗證 Calibration and Validation.....	9
參、研究方法與架構流程.....	11
第一節、研究區域.....	11
第二節、研究方法.....	14
(一) 抽樣設計.....	14
(二) 重複抽樣(permutation).....	17
第三節、研究架構流程.....	18
(一) CLUE-s 模式.....	20
(二) 二元分類演算法.....	22

(三) 土地變遷校正與驗證 Calibration and Validation.....	24
肆、研究結果與討論.....	28
第一節、研究結果.....	28
(一) 抽樣方式對空間自相關性的影響.....	28
(二) 抽樣方式對分類演算法結果的影響.....	32
(三) 抽樣方式對 CLUE-s 變遷模式結果.....	43
(四) 混合模型 Mix Model.....	54
第二節、討論.....	65
(一) 抽樣方式對於土地利用變遷模式的影響.....	65
(二) 兩時期土地利用地圖驗證模式比較.....	69
(三) 混合模型之過度擬合.....	71
(四) 替換演算法之土地利用配置結果之影響.....	73
伍、結論與建議.....	79
第一節、結論.....	79
第二節、後續應用與研究建議.....	81
參考文獻.....	83

圖目錄



圖 1 桃園地區 1995 年土地利用分布	12
圖 2 桃園地區 2007 年土地利用分布	12
圖 3 桃園地區 2014 年土地利用分布	13
圖 4 桃園地區土地利用隨機抽樣示意圖	15
圖 5 桃園地區土地利用系統抽樣示意圖	16
圖 6 桃園地區土地利用分層抽樣示意圖	17
圖 7 研究架構圖	18
圖 8 CLUE-s 模式之示意圖(修改自 Verburg et al. (2002)).....	20
圖 9 隨機抽樣之抽樣比例對 Moran's I 數值變化圖	28
圖 10 系統抽樣之抽樣比例對 Moran's I 數值變化圖	29
圖 11 分層隨機抽樣之抽樣比例對 Moran's I 數值變化圖.....	29
圖 12 農地 AUC 95%信賴區間分布圖	32
圖 13 林地 AUC 95%信賴區間分布圖	33
圖 14 草地 AUC 95%信賴區間分布圖	33
圖 15 建地 AUC 95%信賴區間分布圖	34
圖 16 水體 AUC 95%信賴區間分布圖	34
圖 17 埤塘 AUC 95%信賴區間分布圖	35
圖 18 荒地 AUC 95%信賴區間分布圖	35
圖 19 土地利用分歧散佈圖	43
圖 20 figure of merit 與 miss, wrong hit, false alarm 之散佈圖	47
圖 21 各模型的三圖比較數值與 100%模型比率	52
圖 22 各模型的土地利用分歧數值與 100%模型比率	52
圖 23 混合模型之土地利用分歧之正確配置盒狀圖	55



圖 24 混合模型之土地利用分歧之量分歧盒狀圖	56
圖 25 混合模型之土地利用分歧之配置分歧盒狀圖	56
圖 26 混合模型三圖比較之品質因素盒狀圖	59
圖 27 混合模型三圖比較之 Correct Rejection 盒狀圖	60
圖 28 混合模型三圖比較之 Hit 盒狀圖	61
圖 29 混合模型三圖比較之 False Alarm 盒狀圖	61
圖 30 混合模型三圖比較之 Miss 盒狀圖	62
圖 31 混合模型三圖比較之 Wrong Hit 盒狀圖	63
圖 32 各土地利用品質因素變化平行座標圖	72
圖 33 變更土地利用之土地利用分歧差異盒狀圖	75
圖 34 變更土地利用之三圖比較差異盒狀圖	77

表目錄



表格 1 抽樣方式與土地利用變遷模式相關文獻整理.....	5
表格 2 桃園地區土地利用資料.....	13
表格 3 TOC 矩陣表.....	24
表格 4 TOC 分類變數表.....	25
表格 5 三圖比較分類變數表.....	25
表格 6 土地利用變遷混淆矩陣.....	26
表格 7 土地利用變遷混淆機率矩陣.....	27
表格 8 不同抽樣方法及抽樣比例之 Moran's I 數值.....	31
表格 9 邏輯斯迴歸 AUC 雙因子變異數分析表.....	37
表格 10 邏輯斯迴歸事後檢定結果整理.....	38
表格 11 廣義可加性模式 AUC 雙因子變異數分析表.....	39
表格 12 廣義可加性模式事後檢定結果整理.....	40
表格 13 隨機森林 AUC 雙因子變異數分析表.....	41
表格 14 隨機森林事後檢定結果整理.....	42
表格 15 邏輯斯迴歸土地利用配置分歧多因子變異數分析表.....	44
表格 16 廣義可加性模式模型土地利用配置分歧多因子變異數分析表.....	45
表格 17 隨機森林模型土地利用配置分歧多因子變異數分析表.....	45
表格 18 邏輯斯迴歸三圖比較多因子變異數分析.....	48
表格 19 廣義可加性模式模型三圖比較多因子變異數分析.....	49
表格 20 隨機森林三圖比較多因子變異數分析.....	50
表格 21 各模型的三圖比較、土地利用分歧與 100%模型比率數值表.....	53
表格 22 混合模型配置表.....	54
表格 23 混合模型之土地利用分歧之 100%抽樣模型數值結果.....	57



表格 24 混合模型與隨機森林之土地利用分歧差異比較.....	58
表格 25 混合模型三圖比較與品質因素之 100%抽樣模型數值結果	63
表格 26 混合模型與隨機森林之三圖比較及品質因素差異比較.....	64
表格 27 土地利用空間資料代表性之土地利用分歧數值排行.....	66
表格 28 土地利用空間資料代表性之三圖比較與品質因素數值排行.....	67
表格 29 1995 至 2007 模型驗證數值比較	69
表格 30 土地利用變遷模式過度擬合程度.....	71
表格 31 水體、埤塘、荒地土地利用適宜度描述統計量.....	73
表格 32 混合模型差異驗證評估組合.....	74
表格 33 變更土地利用之土地利用分歧差異.....	76
表格 34 變更土地利用之三圖比較及品質因素差異.....	78

壹、緒論

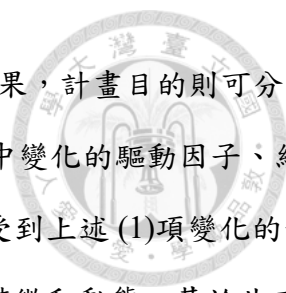


本研究旨在探討空間抽樣方式及抽樣比例對於土地利用變遷模式的影響，進而延伸討論混合土地利用分類模型的可行性。本章為緒論，第一節介紹研究背景與動機，第二節說明研究目的、研究問題與研究流程。

第一節、研究背景與動機

土地利用變遷為長久以來人類行為與自然環境及生態系統的互動而留下的軌跡，被視為全球環境變遷的重要議題之一。自 18 世紀第一次工業革命以來，人類開發利用自然資源的需求快速增加，大規模的土地利用開發造成地表覆蓋改變，進而影響生態系統結構的異動，也間接影響地球大氣氣候調節與水文系統循環的改變，連帶造成土地可被利用的條件有所改變。土地利用變遷與社會發展及生態環境息息相關，因此必須了解土地利用變遷機制，以利於制定相關政策及推估未來發展可能情境。

土地利用變遷主要為時間、空間、人類行為的交互作用下而產生的軌跡，此動態過程中具有多重複雜性與不確定性。在眾多領域中皆為重要的研究議題，如：地理學、生態學及社會科學領域。於 1990 年代，國際地圈-生物圈計畫(International Geosphere-Biosphere Programme, IGPB)與全球變遷人文科學面向計畫(International Human Dimension Programme, IHDP)共同發展跨領域研究計畫：土地利用與地表覆蓋變遷計畫(Land-Use and Land-Cover Change project, LUCC)。此項計畫有三個研究目的，(1)建立土地利用與地表覆蓋之動態資訊概要，(2)確立少數穩健原則，該原則為能更好地將當地居民之見解整合於科學預測中，(3)促進共同（或社區）模型的發展，使該模型能被廣泛使用。此計畫建立了目前土地利用變遷研究中的主要架構，經由實際的案例分析找出不同時空條件下的土地動態特徵；發展包含當地居民見解之土地利用及地表覆蓋變遷模型，藉由模型探討驅動因子間與土地利用或地表覆蓋的變遷關係。並透過此模型預測未來區域性或全球性的土地利用及地表覆蓋可能分布。全球土地計畫(Global Land Project, GLP)則接續了土地利用與地表覆蓋變遷計畫與全球變遷與陸域生態系



計畫(Global Change and Terrestrial Ecosystems Project, GCTE)之研究結果，計畫目的則可分為三項：(1) 識別人與環境耦合系統(coupled human-environment systems)中變化的驅動因子、結構及性質，並量化它們對耦合系統的影響(2) 評估生態系統服務如何受到上述(1)項變化的影響(3) 識別氣候變遷擾動下人與環境耦合系統中的脆弱性與永續性的特徵和動態。基於此三個研究目的，GLP 致力於發展氣候變遷下可計量分析並模擬預測的人類環境耦合系統模型，以利於評估系統變化對於全球或區域的影響。土地利用變遷研究為利用過往的土地使用情形結合問卷調查結果及可能的土地利用需求預測未來的土地利用可能模式。

CLUE-s(Conversion of Land Use and its Effect at Small regional extent)模式為 Verburg et al. (2002)於 Environmental Management 期刊中所提出之土地利用變遷模式模型，為改良 CLUE 模型中無法適應高精度網格資料之缺點。CLUE-s 模式為可結合不同二元分類演算法應建立土地利用變遷模型如：邏輯斯迴歸 (Logistic Regression, LR)、廣義可加性模型(Generalized Additive Model, GAM)、隨機森林(Random forest, RF)、人工類神經網路 (Artificial Neural Network, ANN)、支持向量機(Support Vector Machine, SVM)等。

由於土地利用資料相較於一般二維資料具有空間自相關性(spatial autocorrelation)，若是直接將研究區中的所有土地利用網格資料全部放入分類演算法中，將容易使得分類模型產生偏誤以至於無法準確代表該研究區的土地利用分布情形(Fitzpatrick-Lins, 1981)。基於上述空間資料的特性，土地利用變遷研究中經常使用抽樣方法以控制網格點間的空間自相關性及土地資料間的操作誤差，但是該以何種空間抽樣方式或該抽取多少比例目前的相關研究中並無確切提及。隨著資料科學的進步，目前可處理的網格資料精度日趨精細，使得演算時間大幅增加。就土地利用變遷研究而言，如何有效控制及運用資料間的空間自相關性以降低模擬及預測模型中的偏誤(bias)是建立模型中的首要議題。

本研究將比較系統抽樣(systematic sampling, SYS)、隨機抽樣(random sampling, SRS)與分層隨機抽樣(stratified random sampling, STS)在不同抽樣比例(90%、80%、70%、60%、50%)下應用邏輯斯迴歸(Logistic Regression, LR)、廣義可加性模型(General Additive Model, GAM)、



隨機森林(Random Forest, RF)等三種分類演算法對於 CLUE-s 模式的土地利用驗證結果之影響。基於抽樣方法及抽樣比例的研究結果進而延伸推論結合各土地利用類別之最佳演算法的混合演算模式對於 CLUE-s 模式是否有過度擬合(overfitted)的問題以及較有效率之桃園地區土地利用變遷模型的抽樣方法，以做為未來的土地利用變遷研究對於資料抽樣方式及抽樣比例的參考。此外，回顧過往文獻時發現部分土地利用變遷研究經常僅使用 1~2 時期土地利用作為研究資料，然而土地利用變遷為時間與空間的變化結果，若僅使用 1~2 時期土地利用建立土地利用變遷模式，在無法同時完成校正(calibration)與驗證(validation)之情形下，缺少了土地利用時間上的變化描述，而使得模型無法可能損失準確性，若進行長期土地利用變遷預測則容易與實際土地利用出入甚大。本研究亦將延伸抽樣方式與土地利用變遷模式的研究結果提出可在歷史土地利用資料有限的情形下完成校正及驗證程序，以確保模型的預測品質。

第二節、研究目的與問題



本研究旨在探討系統抽樣、隨機抽樣與分層隨機抽樣在抽樣比例為 100%、90%、80%、70%、60%、50% 下應用邏輯斯迴歸、廣義可加性模式、隨機森林等三種分類演算法對於 CLUE-s 模式的土地利用適宜度及土地利用變遷模式的驗證結果之影響，並推論針對不同土地利用套用該類別之最佳分類演算法之混合分類模式的可行性及是否存在過度擬合的問題。依據上述研究背景與動機，將本研究之目的歸納如下：

1. 探討空間抽樣差異及抽樣比例差異對於 TOC 曲線下面積 AUC (area under curve) 數值之影響：

就 3 種空間抽樣方式、6 種抽樣比例、3 種分類演算法，共 48 種組合。以桃園地區為研究區域，進行重複抽樣取後放回試驗 100 次。再以多因子變異數分析重複抽樣後的數據。

2. 比較不同條件之 CLUE-s 模式的土地利用變遷模型指標之差異：

應用土地利用配置分歧(Disagreement)與品質因素(Figure of Merit)，以多因子變異數分析(Anova)探討抽樣方式差異及抽樣比例差異對於土地利用變遷之檢定驗證之影響，以及較有效之桃園地區土地利用模型之抽樣比例。

3. 探討混合分類模式的可行性及過度擬合的問題：

以重複試驗 AUC 數值平均值最佳為標準，針對不同土地利用類別挑選該土地利用類別之最佳分類演算法。再以不同組合依序替換的方式驗證混合模型之 CLUE-s 模式土地利用變遷結果。

4. 提出適用於兩時期土地利用資料情形下之比較方法：

參考空間抽樣研究結果，比較抽樣模型與 100% 模型在時期 1 到時期 2 的土地利用驗證數值，建立可同時完成校正及驗證之程序

貳、文獻回顧

本章將依序回顧抽樣方式與土地利用變遷模式關係、CLUE-s 模式及檢定驗證方法之相關研究文獻。



第一節、抽樣與土地利用變遷模式

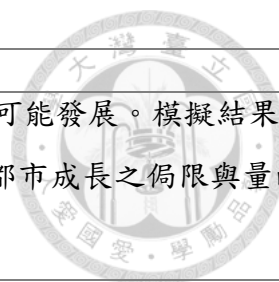
在結合統計模型的土地利用變遷模式研究中，土地利用網格資料的空間自相關性(spatial autocorrelation)一直是建模時的一大問題。由於一般線性回歸模型皆假設資料點為獨立隨機分佈且資料點間無相關性，但是土地利用資料為空間分佈且具有一定程度之自相關性。因此若是直接套用一般線性回歸模式則容易產生偏誤，導致模型預測之土地利用變遷機率過於高估或低估。過去的土地利用研究使用空間系統抽樣以固定間隔距離的方式隨機抽取樣本點以消弭網格與網格間的自相關性 (Overmars et al., 2003)。

Cheng 與 Masser(2003)則認為分層隨機抽樣(stratified random sampling)與空間系統抽樣(systematic spatial sampling)皆可降低因空間自相關性而產生的誤差。

Ahmed 等人(2014)則認為系統抽樣較能有效降低空間上的相關性但可能忽略封閉區域的重要資訊；然而隨機抽樣則較能有效代表母體但對於消弭空間自相關性的效率較差。

表格 1 抽樣方式與土地利用變遷模式相關文獻整理

作者(年分)	抽樣方法	大綱
Ahmed et al. (2014)	採用空間隨機抽樣以極小化空間自相關性(總樣本數的 30%)，由於抽樣後之資料數仍具有代表性，因此不使用總樣本數進行模擬。	應用 clue-s 模式結合邏輯斯迴歸模擬孟加拉首都達卡在 1988 年~1999 年間的都市化過程再以 1999 年~2005 年間土地利用作為驗證。結果指出驅動因子足以解釋達卡地區之動態變化，且實際土地利用資料與都市成長模型結果類似。
Puertas et al. (2014)	由於空間系統抽樣可以減少空間相依與空間自相關性，因此使用空間系統抽	應用馬可夫鍊、細胞自動機模型及邏輯斯迴歸模擬智利聖地牙哥首都之都市動態變化，並以 1975 年~2010 年之資料建立模型，並預



作者(年分)	抽樣方法	大綱
	樣在時間 s 下依照土地利用網格 i 到 j 抽取 20%、30%、40%之樣本點。	測 2030 年~2045 年之可能發展。模擬結果指出綜合模型成功解釋都市成長之侷限與量的變化。
Li and Huang (2015)	考慮到研究區之大小及研究工作量，使用六邊形網格系統抽樣，再依照 ERI 方法計算每個網格之生態景觀危險指數。	結合 clue-s 模式與生態景觀危險指數方法分析 Luanhe 流域未來土地利用變遷之特徵，並模擬三種情境下 2010 年~2030 年之土地利用變遷熱區。研究發現 Luanhe 流域之生態景觀指數呈現負向趨勢，然而其中數個區域存在高度風險。
Waiyasusri et al. (2016)	空間系統抽樣一般而言適用於減少空間相依度，然而空間隨機抽樣更能有效代表整個母體，但無法有效減少空間相依度。由於土地利用為不均勻分布於研究區域中，因此以空間分層抽樣為抽樣方法	應用 GIS 遙測影像與 Clue-s 模式分析泰國 Changwat Uthaithani 地區 Huai Thap Salao 分水嶺 1988 年至 2007 間的土地立變遷之空間型態。研究結果發現 1988 - 2007 年間林地面積減少 7.36%，農地增加 10.5%，並以 CLUE-S 預測未來 20 年(至 2027 年)之土地利用，假設在無空間限政策之條件下，Huai Thap Salao 分水嶺之去森林化之趨勢將持續，且農地及建地比例增加。
Wagner and Waske (2016)	為了避免空間自相關性，使用空間分層抽樣從二元資料網格中抽取 20%之單位像素為樣本。	以邏輯斯迴歸與水文變數分析中等規模印度流域之土地利用變遷。研究發現加入水文變數於模型中，可提升整體模型的準確度，使得 ROC 曲線下面積增加 11%。
Shafizadeh-Moghadam et al. (2017)	為了避免過度擬合與減少程式運算時間，利用空間分層抽樣抽取母體之 50% 資料點作為訓練資料。	比較人工類神經網路(artificial neural networks, ANN)、支援向量回歸(support vector regression, SVR)、隨機森林(random forest, RF)、分類決策樹(classification and regression trees, CART)、邏輯斯迴歸(logistic



作者(年分)	抽樣方法	大綱
		regression, LR)及多元適應雲型回歸 (multivariate adaptive regression splines, MARS)等六種機器學習方法應用於細胞自動機之 Tehran Metropolitan 地區的土地利用變遷模式差異。研究結果顯示不同機器學習模型之土地利用變遷模式有顯而易見之差異。
Peng et al. (2020)	為了減少空間自相關性的影響，以空間分層抽樣抽取研究區中 2%之母體網格作為訓練資料。	應用隨機森林於 CLUE-s 模式以模擬 Wuhan 地區濕地之空間動態，並預測 2015 年至 2040 年於 3 種情境下之土地利用變遷。結果指出 3 種情境中下，高度開發地區的天然濕地、稻田將逐年減少。
Peng et al. (2021)	為了減少空間自相關性對邏輯斯迴歸造成的不確定性誤差，採用空間隨機抽樣產生訓練樣本建立回歸模型。	結合邏輯斯迴歸、馬可夫鍊及 CLUE-s 模式建立 Wuhan 地區 2015 年至 2035 年之土地利用變遷模型。研究結果顯示該模型與實際土地利用有良好的一致性，且 2015 年至 2040 年 Wuhan 地區中心之建地擴張程度顯著，農地與人工渠道面積明顯減少，濕地及其他綠地則大致維持相同比例。
Mazloun et al. (2021)	使用空間分層抽樣為抽樣方法，並維持樣本點間之最小距離為 900 公尺以減少空間自相關性。	應用隨機森林於陸地衛星影像資料分析 1985 年至 2016 年間 Zayandehrood 次流域之土地利用變遷情形。衛星影像分類結果顯示，1985 年至 2016 年間住宅地增加一倍，且以牧場為主之工業區亦增加。另以類神經網路與馬可夫鍊預測 2036 年之土地利用結果工業區及住宅區持續擴張，此結果可證明伊朗境內之牧場逐漸被開發為其他土地利用。

第二節、CLUE-s 模式



CLUE-s(Conversion of Land Use and its Effect at Small regional extent)模式為 Verburg et al. (2002)於 Environmental Management 提出之土地利用變遷模型。CLUE-s 模式的前身為 Veldkamp and Fresco (1996)於生態建模(Ecological Modeling)所提出的 CLUE 模式。CLUE-s 為改良 CLUE 模式無法適應網格精度較高之區域，使得模型預測能力較為不足之缺點，發展成為針對小區域且可適應高網格精度之土地利用變遷模式。Verburg 等人(2002)運用 CLUE-s 模式結合邏輯斯迴歸(logistic regression)分析菲律賓的 Sibuyan 島及馬來西亞的 Klang-Langat 集水區的土地利用變遷情形。研究結果也支持 CLUE-s 模式能有效模擬這兩個區域的土地使用情形，並提出預測土地利用分佈以作為政府規劃政策使用。

CLUE-s 模式發展之初為利用邏輯斯迴歸為演算法，計算土地利用變遷機率以建立土地利用適宜度。隨著土地利用變遷研究及演算法之發展，許多研究將 CLUE-s 模式結合線性、非線性及機器學習等演算法以適應不同區域的土地利用分佈情形。Lin 等人(2011)以邏輯斯迴歸(logistic regression)、自動邏輯斯迴歸(auto-logistic regression)與人工類神經網路(artificial neural network, ANN)等三種演算法結合 CLUE-s 模式預測新店地區的寶僑集水區的土地利用變遷並比較三種演算法之結果。研究結果認為人工類神經網路搭配 CLUE-s 模式對於研究區的土地利用變遷預測較佳，且指出以人工類神經網路與自動邏輯斯迴歸的土地利用模型較能有效消弭土地利用資料間的空間自相關性。

Ahmed 等人(2014)以隨機森林(random forest)作為分類演算法應用 CLUE-s 模式分析孟加拉的首都達卡(Dhaka)快速都市化的驅動因子。研究結果認為結合隨機森林與 CLUE-s 的土地利用變遷模型足以解釋達卡在 1988 年至 1999 年期間的土地利用變遷方式，並指出 CLUE-s 模式相較於其他基於主體模型(agent-based model)更能完整描述達卡地區的複雜土地利用行為與變遷過程。Peng 等人(2020)以隨機森林作為分類演算法於 CLUE-s 模式模擬 Wuhan 地區

濕地之土地利用變化，預測 2015 年至 2040 年於 3 種不同情境下土地利用。研究結果認為在所有情境中，隨著開發程度增加，Wuhan 地區的天然濕地、稻田將逐年減少。

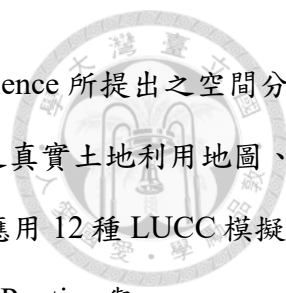


第三節、模型校正與驗證 Calibration and Validation

檢定與驗證為土地利用變遷模式研究中最為重要的一環。作為比較研究結果良好與否及優劣之比較標準，長久以來持續備受重視。目前與土地利用變遷研究有關的主要檢定驗證指標有 ROC 曲線與曲線下面積(AUC)、kappa 數值，而近年來 Pontius 則提出 TOC 曲線、土地利用配置分歧(disagreement)及品質因素(Figure of merit)等概念更為細緻比較土地利用變遷模式之預測結果。

ROC(Receiver operating characteristics)接收者操作特性為第二次世界大戰時被運用來分析雷達訊號，之後則被應用於訊號偵測理論(signal detection theory)。目前也經常被應用於分析機器學習(machine learning)及心理物理學(psychophysics)等領域。Pontius 與 Schneider 認為土地利用變遷研究逐漸發展出許多不同的研究方法及工具，因此需要一個可以做為比較及分析的驗證方法。他們認為 ROC 曲線讓土地覆蓋變化研究者得以驗證模型於特定區域的能力，並同時維持特變化的自由度。而 TOC(Total operating characteristics)曲線為 Pontius 與 Si (2014) 提出，為改良 ROC 過度簡化資訊不利於土地利用研究分析之缺點，透過變更繪製方式的 TOC 曲線完整保留 ROC 曲線之性質並增加完整的土地利用配置結果。

Kappa 指數為土地利用變遷領域中，廣泛用來比較模型配置優劣的指標之一。透過與隨機分配之網格配置比較，計算出土地利用變遷模式的 kappa 數值。Pontius 與 Millones (2011)所提出的論文則認為，kappa 指數以隨機配置為基線進行比較，但於地圖上的隨機配置較不穩定、不可靠；且 kappa 指數的計算及定義較為複雜，不利於廣泛分析及討論。因此 Pontius 與 Millones 提出量分歧(quantity disagreement)與量分歧(quantity disagreement)以取代 kappa 指數，作為評估土地利用配置結果的指標。



三圖比較(three map)為 Pontius 等人(2011)於 The Annals of Regional Science 所提出之空間分布驗證方法。驗證方法為(1)時期 0 之真實土地利用地圖、(2)時期 n 之真實土地利用地圖、(3) 時期 n 之預測土地利用地圖比較此三張地圖之空間分布差異。且應用 12 種 LUCC 模擬結果為例，證明三圖比較可以進一步區分土地利用變遷模型中的誤差。Pontius 與 Peethambaram 以越南 Cho Don 區為研究區域，應用三圖比較分析地圖解析度對於土地利用變遷模式之誤差影響，其中亦提出品質因素(Figure of Merit)作為模型驗證之指標。研究結果指出三圖比較可清楚分辨模型預測變化及觀察變化差異，且有利於動態模擬之比較。Varga 等人(2019)則根據 Pontius 等人(2008、2011、2018)提出之三圖比較與 Figure of Merit 概念，清楚定義出適合土地利用變遷模式的品質因素(FOM)之四個元素及 FOM 之計算方式。

參、研究方法與架構流程



本研究認為不同抽樣方式、抽樣比例及演算法對於土地利用變遷模式應有明顯之影響。以下依序介紹研究區域及研究方法。

第一節、研究區域

為驗證抽樣方法及抽樣比例對於土地利用變遷模式的影響，因此在研究區的挑選上需考量：研究區的網格數多寡、土地利用變化穩定並具有完整土地利用地理圖資系統。桃園地區位於台灣本島的西北部，總面積為 1220.9540 平方公里，於 1960 年代初期開始進入工業化及都市化發展，區域內有桃園國際機場、石門水庫、高速公路及台灣高鐵等國家重要交通及水利設施，臨近台北都會區且近年人口社會增加率逐年上升，為土地利用穩定變遷之區域，符合本研究所需的研究區域條件。

本研究以桃園地區西元 1995 年(圖 1)，以及西元 2007 年(圖 2)之土地利用資料建立土地利用適宜度地圖，再以相同地區西元 2014 年(圖 3)之土地利用資料驗證土地利用變遷模式。本研究使用之地圖資料為內政部國土測繪中心所提供之國土利用現況調查之成果，但由於西元 1995 年、西元 2007 年及西元 2014 年之調查方法及分類標準不盡相同，故本研究依照研究需求將此三時期之土地利用重新分類為：農地、林地、草地、建地、水體、埤塘、荒地等六類。

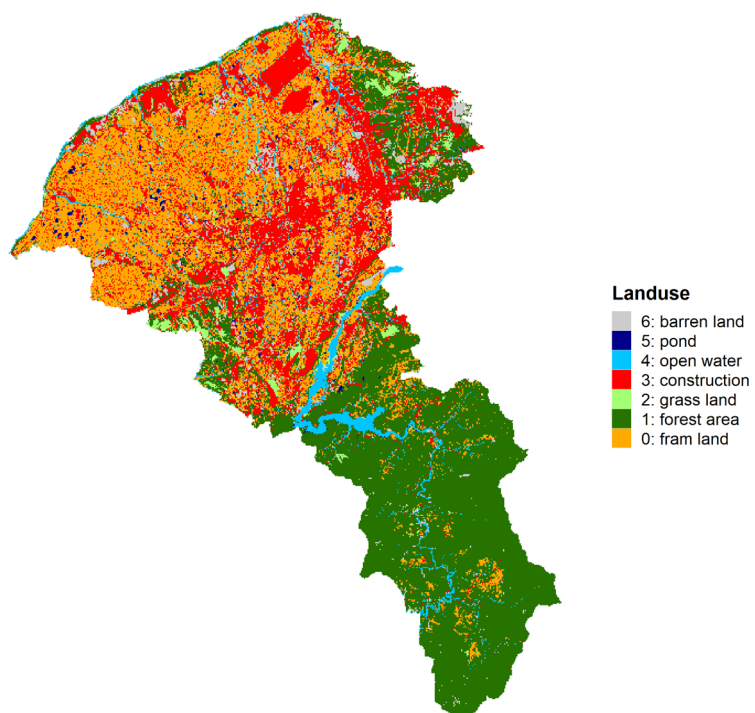


圖 1 桃園地區 1995 年土地利用分布

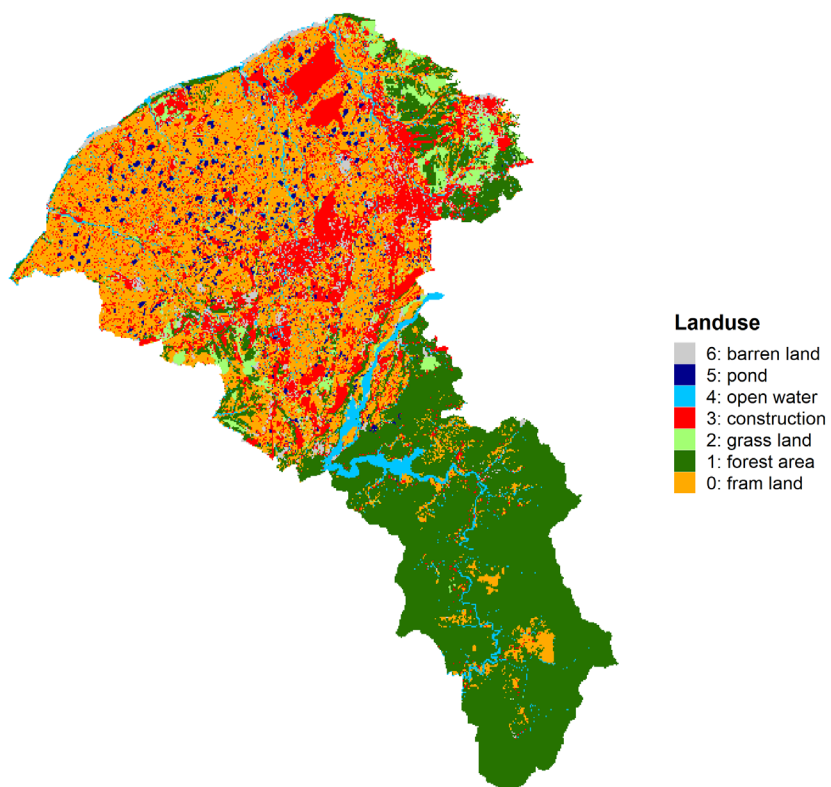


圖 2 桃園地區 2007 年土地利用分布

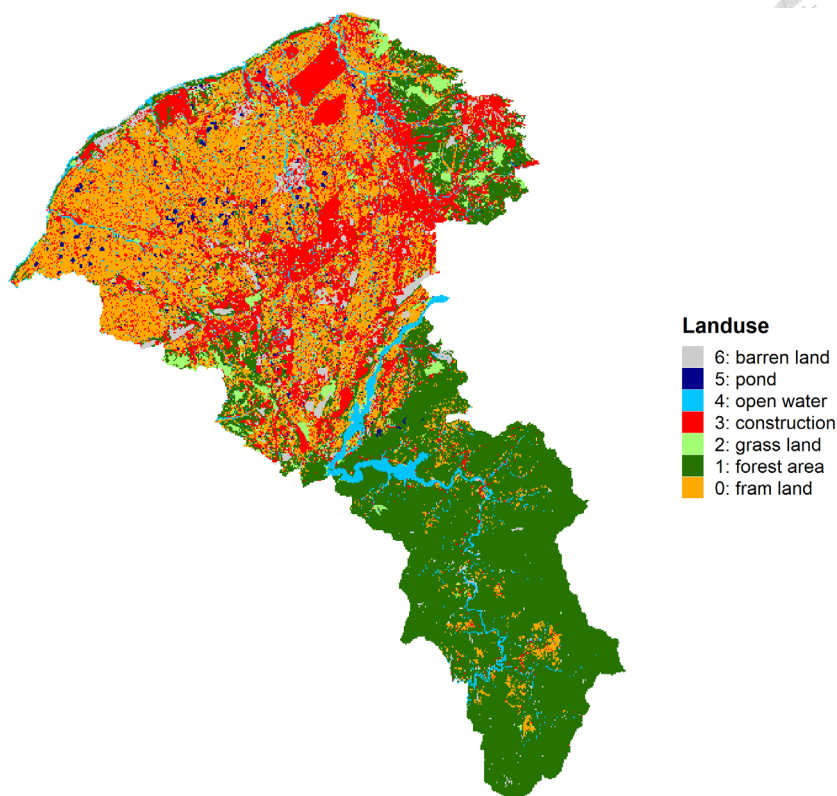


圖 3 桃園地區 2014 年土地利用分布

桃園地區之研究範圍為 1220.9540km²，總網格數為 120398 格，網格大小為 100m*100m 之方格。桃園地區的土地利用根據地表覆蓋景觀之差異可分為 0 至 6 類，依序代表農地(farm land)、林地(forest area)、草地(grass land)、建地(construction)、水體(open water)、埤塘(pond)、荒地(barren land)，詳細土地利用網格資料及比例如表格 2 所示。

表格 2 桃園地區土地利用資料

年份	0: 農地	1: 林地	2: 草地	3: 建地	4: 水體	5: 埤塘	6: 荒地	總網格數
1995	32994	43988	2742	28064	5806	1214	5590	120398(格)
	27.40%	36.54%	2.28%	23.31%	4.82%	1.01%	4.64%	100%
2007	40276	41703	4587	21657	5691	2684	3800	120398(格)
	33.45%	34.64%	3.81%	17.99%	4.73%	2.23%	3.16%	100%
2014	33749	44132	4474	26022	5691	1411	4919	120398(格)
	28.03%	36.66%	3.72%	21.61%	4.73%	1.17%	4.09%	100%

第二節、研究方法



(一)抽樣設計

參考文獻回顧的結果，隨機抽樣、系統抽樣及分層抽樣等三種抽樣方式為土地利用變遷研究中最頻繁使用的三種抽樣方式。考量桃園地區的土地利用網格數較多，以及研究易於編寫程式，本研究使用 R 語言的 sf 套件(Pebesma, 2018)對桃園地區的網格點進行隨機抽樣(random sampling)、系統抽樣(systematic sampling)、分層隨機抽樣(stratified random sampling)等三種抽樣方式。以下將分別詳述每種抽樣方式的抽樣方法。

1.隨機抽樣

本研究使用球面抽樣(Sphere Point Picking)的方式在研究區中隨機抽取網格。此抽樣方式假設研究區中的所有網格點皆以球座標系(Spherical Coordinates)描述，並服從均勻分布(uniform distribution)，使得研究區的每一網格被抽中的機率均為相同。抽樣方式則可描述為下列數學式， U, V 分別服從均勻分布(0,1)區間中的兩個隨機變數，由此可定義出 θ 為式 1， ϕ 為式 2。

$$\theta = 2\pi u \quad \text{式 1}$$

$$\phi = \cos^{-1}(2v - 1) \quad \text{式 2}$$

已知樣本點均勻地分布於球表面 S^2 上，則可對立體角(Solid Angle)取導數為式 3。

$$d\Omega = \sin \phi \, d\theta d\phi = -d\theta d(\cos \phi) \quad \text{式 3}$$

同時，極角(polar angle)可描述為式 4，將式 2 及 $P_v = 1$ 帶入式 4 中即可得出極角 P_ϕ 為式 5。

再將 $du = \sin \phi \, d\phi$ 帶入式 5，則可得出抽樣點的座標式 6。

$$P_\phi \, d\phi = P_v \left| \frac{dv}{d\phi} \right| d\phi \quad \text{式 4}$$

$$P_{\phi} = \frac{1}{2} \sin \phi$$

$$x = \sqrt{1 - u^2} \cos \theta$$

$$y = \sqrt{1 - u^2} \sin \theta$$

$$z = u$$



式 5

式 6

其中 $\theta \in [0, 2\pi)$ ， $u \in [-1, 1]$ 並均勻分布於球表面 S^2 上。

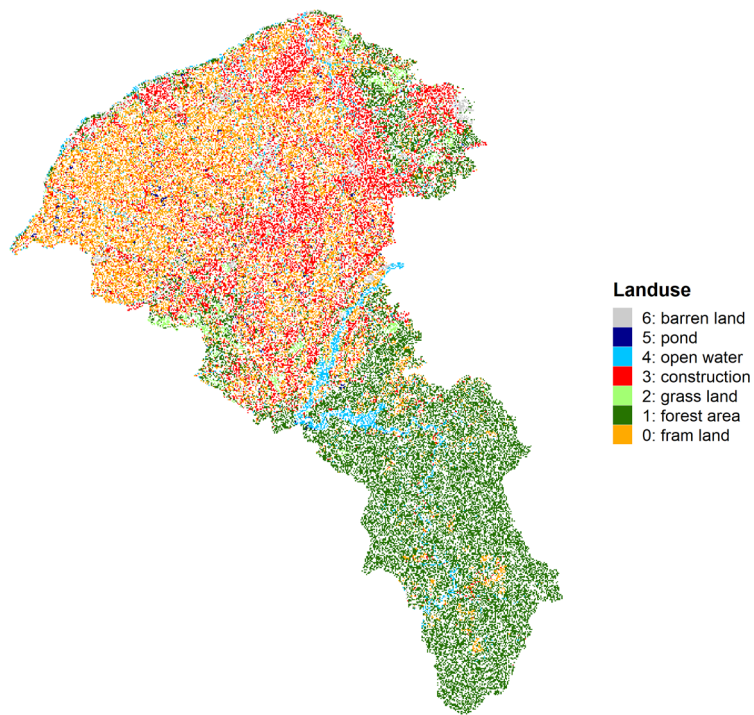


圖 4 桃園地區土地利用隨機抽樣示意圖



2.系統抽樣

系統抽樣為固定間隔抽取樣本的抽樣方法。首先於研究區以隨機抽樣選取 1 個起始點，以起始點為中心間隔一定距離選取網格。由於系統抽樣受限於距離限制，故實際抽取網格樣本個數近似於設定抽樣數。本研究中因土地利用資料皆處理為 100m*100m 之單位方格，因此抽樣上皆以方格中心計算樣本距離。

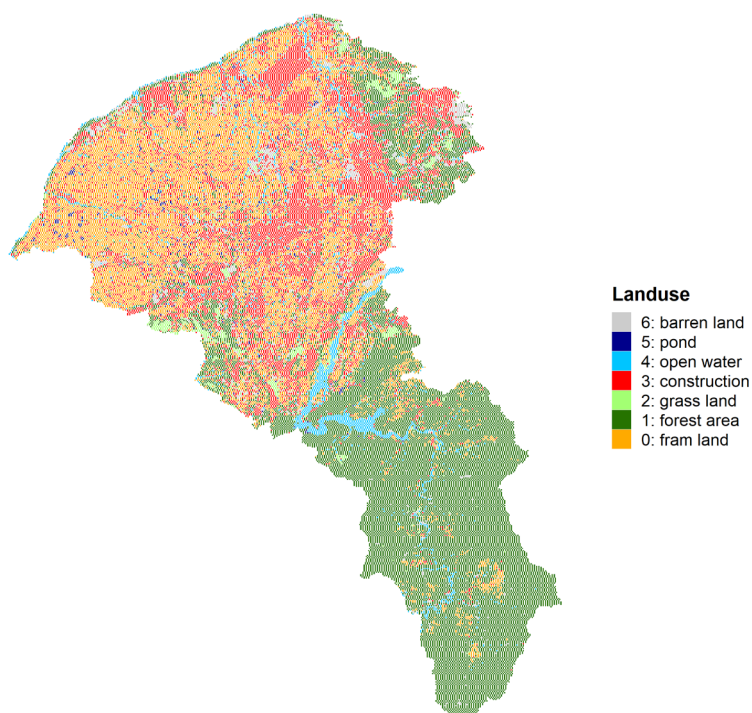


圖 5 桃園地區土地利用系統抽樣示意圖

3.分層隨機抽樣

分層隨機抽樣為將桃園地區中的土地利用網格依照土地利用類別不同分為 0~6 類，每一類別皆以隨機抽樣抽取相同比例的樣本網格點，目的為保持樣本內的土地利用類別比例與母體比例相同。詳細抽樣流程為: (1)先計算第 K 類別的土地利用網格佔整體土地利用的比例為 $k\%$ ；(2)決定抽樣個數 n ；(3)以隨機抽樣的方式抽取第 K 土地利用類別中 $n \cdot k\%$ 個土地利用樣本點。

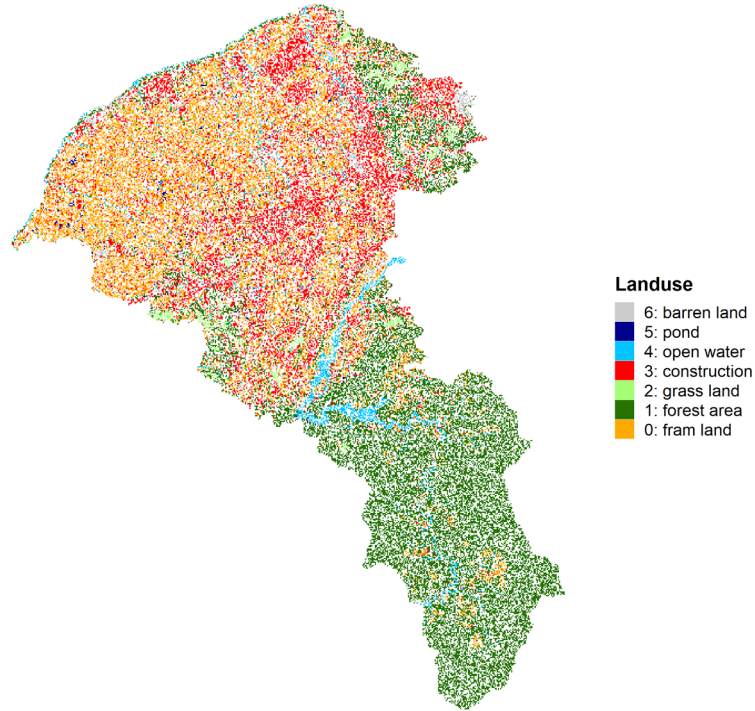


圖 6 桃園地區土地利用分層抽樣示意圖

(二)重複抽樣(permutation)

為了驗證抽樣方式與抽樣比例的對於及土地利用變遷模式的影響，本研究針對桃園地區的樣本空間網格點進行隨機取後放回的抽樣，重複抽樣 100 次。每次抽樣皆生成一次土地利用適宜度，計算 TOC 曲線下面積 AUC 值。完成 100 次循環後，並依式 7、式 8 計算該分類模型的 AUC 的平均值與標準差，再將 100 個 AUC 平均值及標準差分別由小到大排列，取第 2.5 百分位數 $P_{2.5}$ 與第 97.5 百分位數 $P_{97.5}$ 為 95%信賴區間(95% confidence interval)的下界及上界。

$$\overline{AUC} = \frac{1}{n} \sum_{i=1}^n AUC_i \quad \text{式 7}$$

$$AUC_{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (AUC_i - \overline{AUC})^2 \quad \text{式 8}$$



第三節、研究架構流程

研究流程上，本研究先由統計模型與土地利用變遷模式的關係，對演算法及 CLUE-模式，回顧現有文獻對於土地利用變遷模式的應用及探討，並確立以抽樣方式及抽樣比例對於土地利用變遷模式的影響為本文研究之主軸。再將研究問題對應 CLUE-s 模式演算階段，分為土地適宜度及土地利用配置兩種不同面向後，在這兩個研究階段構面下，各自進行重複抽樣並以對應檢驗工具驗證，最後進行混合模型對於土地利用變遷模式的應用及檢核是否可能過度擬合(overfitting)，研究流程則可歸納如圖 7 研究架構圖：

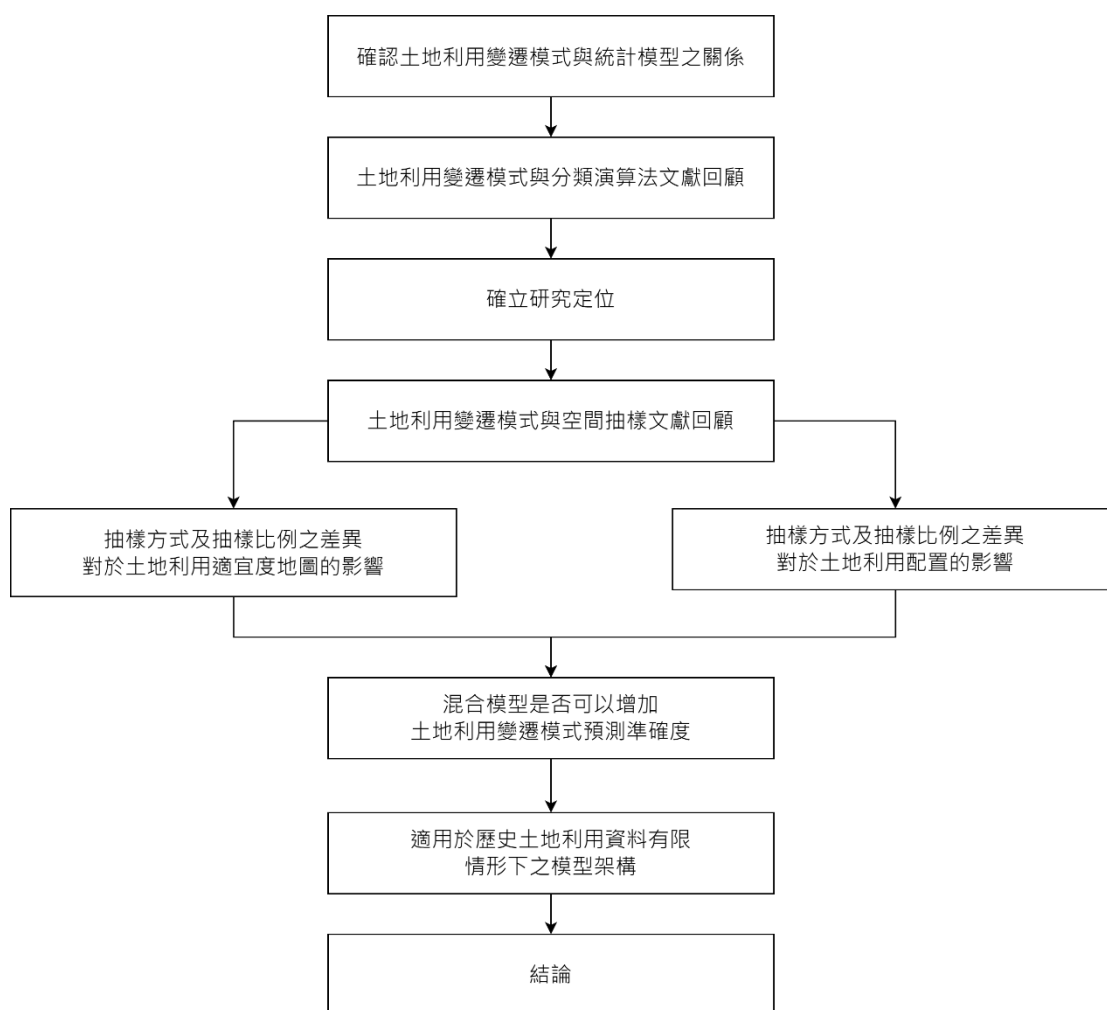


圖 7 研究架構圖



大綱則分述如下：

1. 抽樣方法與土地利用變遷研究、CLUE-s 模式、曲線下面積 AUC 之應用、土地利用檢定驗證方法等文獻回顧。
2. 資料處理：整理桃園地區 1995 年、2007 年、2014 年土地利用資料並分類。
3. 重複抽樣取後放回：以系統抽樣、隨機抽樣與分層抽樣；抽樣比例 100%、90%、80%、70%、60%、50%；邏輯斯迴歸、廣義可加性模式、隨機森林演算法，共 48 種實驗組合以桃園地區為研究區，每組進行拔靴法重複抽樣試驗 100 次。
4. CLUE-s 模式依照 48 組實驗組合分析 2007-2014 年之土地利用變遷模型結果，並計算各組之土地利用配置分歧及品質因素。
5. 依照 AUC 結果，針對不同土地利用類別挑選該土地利用類別之最佳分類演算法。再以依序替換分類演算法之方式進行 CLUE-s 模式混合模型分析。
6. 運用多因子變異數分析上述實驗結果之數據

詳細步驟如下：

本研究首先蒐集抽樣方式及抽樣比例與土地利用變遷模式之相關文獻，並同時整理桃園地區土地利用資料以作為 CLUE-s 模式之輸入資料。以 3 種空間抽樣、6 種抽樣比例、3 種分類演算法，總共 48 種組合進行桃園地區重複試驗。再以 CLUE-s 模式進行 48 組桃園地區土地利用變遷分析。最後根據 AUC 比較之結果，針對不同土地利用類別挑選該土地利用類別之最佳分類演算法。以迭代之方式進行 CLUE-s 模式之混合模型驗證。最後針對實驗結果加以探討。此外挑選次佳的抽樣模型與 100%模型比較時期一到時期二的土地利用驗證數值，提出可在有限圖資下完成校正及驗證之模型架構。



(一)CLUE-s 模式

CLUE-s 模式(Verburg et al., 2002)如圖 8，為成熟的土地利用變遷模型，此模式藉由建構過去的土地利用變化模型，推估未來情境的土地利用需求，模擬將來可能的土地利用分布方式。CLUE-s 模式整體架構主要分為非空間分析與空間分析，其中本研究中非空間分析為使用馬可夫鏈 (markov chain) 推估不同時期的土地利用需求搭配土地利用轉換類型、限制區域、土地利用適宜度空間分布機率，根據總機率大小對土地利用需求進行空間分配，透過多次迭代出目標年之土地利用網格分配。

馬可夫鏈(Markov chain)假設土地利用的轉變為隨機過程 (stochastic process)，意即土地利用的轉變與過去的轉變過程無關，僅與轉變當下的狀態有關。因此可以利用當前的土地利用面積藉由馬可夫鏈(Markov chain)轉移矩陣推導出每一時期不同土地利用的需求面積。

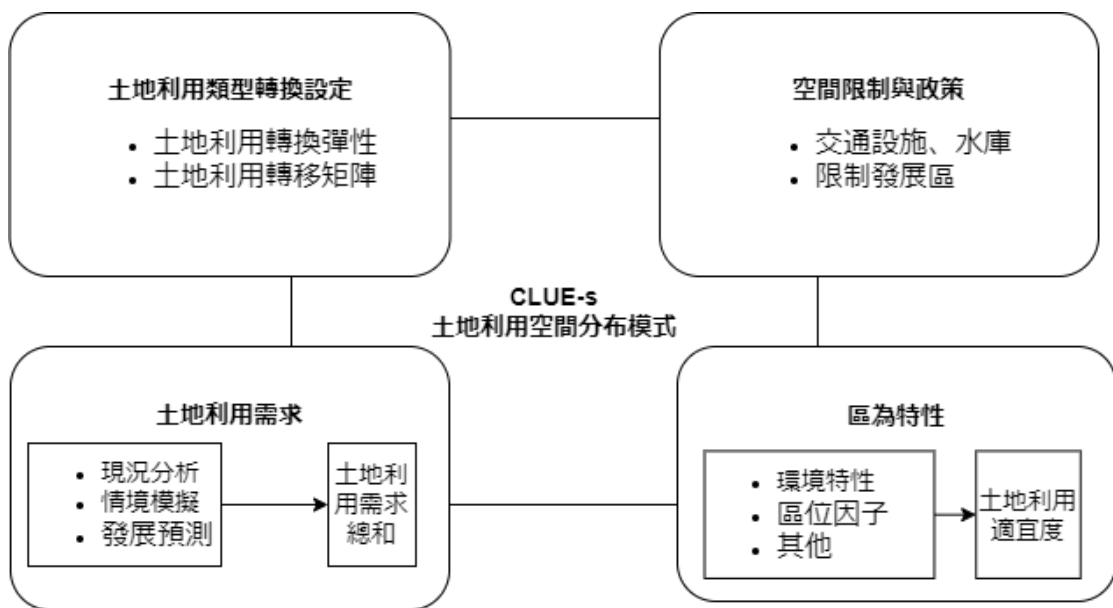


圖 8 CLUE-s 模式之示意圖(修改自 Verburg et al. (2002))

土地利用轉換類型為描述土地利用間可轉換的可能性，介於 0 到 1 之間；接近 1 表示轉移彈性越低，接近 0 則表示轉移彈性越高。土地利用在低開發程度時可以在許多類型間互相轉換，如林地可轉移為耕地與荒草地，耕地可轉為荒草地，荒草地可轉為耕地。本研究以土地利用

轉移彈性矩陣紀錄不同之土地利用變遷機制：(1)為林地擴張為耕地之機制；(2)為林地採伐後成荒地；(3)為耕地棄耕後成為荒地；(4)為荒地重新耕種後為耕地。

CLUE-s 模式中的空間分包含了空間政策及限制、區位特性。空間政策與開發限制會直接限制土地利用變遷的可能區域與範圍，因此 CLUE-s 模式中將無法開發或改變土地利用的區塊（如國家公園、水體等區域）獨立，以避免影響模擬結果。在本研究桃園地區中，桃園國際機場為國家重要交通建設，其利用與開發受政策限制為非一般土地利用變遷模式可以描述，故建立土地利用適宜度及分析驗證中將桃園國際機場之網格點從資料中獨立。

區位特性為不同區域的土地利用依據時空條件的不同具有不同的土地利用變化方式，在 CLUE-s 模式中以土地利用適宜度（land use suitability map）描述該地的區位特性。為了建立桃園地區土地利用適宜度的，本研究透過發放問卷整理出 11 種可能影響桃園地區土地利用改變的驅動因子作為二元分類模型的自變項，並以基年土地利用作為依變項，以邏輯斯迴歸、廣義可加性模型及隨機森林建立土地利用適宜度模型。



(二)二元分類演算法

邏輯斯迴歸 (Logistic Regression)、廣義可加性模式 (Generalized Additive Model, GAM)，以及隨機森林 (Random Forest, RF) 三種不同的分類模型為土地利用變遷研究中經常被使用的三種模型。

1. 邏輯斯迴歸 (Logistic Regression, LR)

邏輯斯迴歸為常見的二元分類模型，為給定驅動因子條件下可以推估每個網格點之土地利用轉變機率，進而產生適宜度模型 (Lin et al. 2011; Chan et al. 2015)。

邏輯斯迴歸可以描述為在給定 $X_1 = x_1, \dots, X_k = x_k$ 的情況下， $Y = 1$ 的機率為

$$Pr(Y = 1 | X_1 = x_1, \dots, X_k = x_k) = \frac{\exp(\sum_{j=1}^k \beta_j x_j)}{1 + \exp(\sum_{j=1}^k \beta_j x_j)}$$

在土地利用適宜度模型中， $Pr(Y = 1)$ 為網格中的土地利用適宜性機率，共有 k 個驅動因子， x_i 為網格中的第 i 個驅動因子， β_0 為推估常數， β_i 為邏輯斯迴歸中第 i 個驅動因子之係數。

本研究中使用 R 語言內建之 `lm()` 函式進行邏輯斯迴歸計算。

2. 廣義可加性模型 (General Additive Model, GAM)

廣義可加性模型不同於邏輯斯迴歸，使用函數取代邏輯斯迴歸係數，描述自變量與應變量間的關係。因此廣義可加性模型相較邏輯斯迴歸模型更能完整捕捉自變量與應變量間的非線性關係。廣義可加性模型方程組可描述如下式：

$$E(Y | X_1 = x_1, \dots, X_k = x_k) = \beta_0 + f_1(x_1) + \dots + f_k(x_k)$$

當依變項 Y 為二元變數時，廣義可加性模型方程組可以改寫為：

$$Pr(Y = 1 | X_1 = x_1, \dots, X_k = x_k) = \frac{\exp(\beta_0 + f_1(x_1) + \dots + f_k(x_k))}{1 + \exp(\beta_0 + f_1(x_1) + \dots + f_k(x_k))}$$

在土地利用適宜度模型中， $Pr(Y = 1)$ 為網格中的土地利用適宜性機率，共有 k 個驅動因子， x_i 為網格中的第 i 個驅動因子， β_0 為推估常數， $f_i(x_k)$ 為廣義可加性模型中第 i 個驅動因子輸出機率。本研究使用 R 語言中 GAM 套件中的 `gam()` 函式建立廣義可加性模型。

3.隨機森林(Random Forest, RF)

隨機森林為常見的機器學習二元分類演算法，透過拔靴法(bootstrap aggregating/bagging)重複抽樣建立隨機子空間(random subspace)，建構多個決策樹再採取多數決的方式預測分類結果(Liaw and Wiener 2002)。由於隨機森林為透過數個隨機子空間建立多棵決策樹，每棵決策樹再生成多節點，每一節點中再篩選出數個驅動因子分類該節點，最後輸出個別決策樹的平均預測值，因此較能適應高維度模型且較不易產生過度擬合(overfitting)之情形。但隨機森林缺點為需要耗費大量運算時間與記憶體空間，故運算效率較為緩慢。本研究為了提升隨機森林之運算效率，使用 R 語言中的 ranger 套件中的 `ranger()` 函式進行隨機森林運算。

`ranger`(Wright & Ziegler, 2017)為以 C++編寫之 R 語言套件，其預設參數皆與 R 語言中經常使用的 `RandomForest()` 函式相同，但運算效率較 `RandomForest()` 函式快速，且演算結果與 `RandomForest()` 函式相同。



(三)土地變遷校正與驗證 Calibration and Validation

土地利用變遷研究著重於模型對於土地利用改變的預測能力，本研究使用 Total operating characteristic(TOC)、曲線下面積(area under curve, AUC)為評估二元分類模型之指標，衡量適宜性模型的分類結果；另外使用三圖比較(three-map comparison) 及土地利用配置分歧(disagreement)評估土地利用變遷模型的配置預測結果。

1.Total operating characteristic 曲線(TOC)

TOC 曲線為坐標圖式的分析工具，橫軸為分類器分類布林變數為真的數值，縱軸為分類器分類布林變數為真且真值亦為真的數值。TOC 曲線上的每一點皆為一次分類器的分類結果，分類上，若預測值大於等於閾值，則布林變數分類為真；反之，若預測值小於閾值則布林變數分類不為真。依照分類結果及真值分佈結果，可以概括為以下四種情形，如表格 3 TOC 矩陣表所示，表格 4 TOC 分類變數表則為各個變數之邏輯定義。

表格 3 TOC 矩陣表

		參考值 (Reference)		診斷值總計
		存在 Presence	不存在 Absence	Diagnosed Total
診斷值 Diagnosis	存在 Presence	擊中 (Hits, H)	誤報 (False Alarms, F)	H+F
	不存在 Absence	遺漏 (Misses, M)	正確拒絕 (Correct Rejections, C)	N+C
參考值總計 Reference Total		H+M=P	F+C=Q	P+Q



表格 4 TOC 分類變數表

變數	解釋
Hit	分類為真且真值為真
False Alarm	分類為真但真值不為真
Misses	分類不為真但真值為真
Correct Rejection	分類不為真且真值不為真

2. 曲線下面積(Area Under Curve, AUC)

AUC 為 TOC 曲線下的面積，其值介於 0 到 1 之間。AUC 為評估分類器的重要指標，可檢驗比較不同模型的表現。若 AUC 值越接近 1，表示模型的分類效率越高；反之，越接近 0 分類效率越差。其中，當 TOC 曲線為一對角線時 AUC 值為 0.5，代表模型可辨別為真的機率為 50%，與隨機分配相同故模型無鑑別率。若 TOC 曲線愈往左上角則表示模型敏感度高，誤判率低，且 AUC 值愈接近 1。

3. 三圖比較(Three-map comparison)與品質因素(Figure of Merit, FoM)

為了比較土地利用變遷模型預測結果與實際變遷機制是否一致，本研究使用三圖比較法(Three-map comparison)(Pontius et al., 2004)進行驗證。三圖比較法為將(1)時期 1 的土地利用參考圖、(2)時期 2 的土地利用參考圖以及(3)時期 2 的土地利用預測圖套疊在一起進行比較。依據預測變化及真實變化的狀況不同可以概括為表格 5。

表格 5 三圖比較分類變數表

變數	解釋
Hit(H)	土地利用有改變，且實際類別與預測類別相同
Wrong Hit(WH)	土地利用有改變，但實際類別與預測類別不同
False Alarm(FA)	土地利用無改變，但預測改變
Misses(M)	土地利用有改變，但預測無改變
Correct Rejection(CR)	土地利用無改變，且預測無改變

透過三圖比較後的列聯表，進而定義出品質因素(figure of merit) (Pontius et al., 2011)。品質因素可定義為式 9，用來評估土地利用變遷模型與實際土地利用變遷機制是否一致之指標。品質因素介於 0 到 1 之間，若品質因素為 0 表示實際變化與預測變化之間沒有交集；反之，若品質因素為 1，則表示實際變化與預測變化之間完全交集。

$$FoM = \frac{H}{H + WH + FA + M} \quad \text{式 9}$$

4. 土地利用配置不確定性

基於土地利用變遷模式預測分配之結果，本研究使用 Pontius and Millones(2011)提出之分歧分析方法(disagreement approach)量化土地利用變遷模型的不確定性。Pontius and Millones(2011)以土地利用變遷模型預測之配置結果與實際土地利用空間分布差異建立混淆矩陣(confusion matrix)，如表格 6，再以式 10 計算出混淆矩陣中每一元素的分布機率，如表格 7。

$$P_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^J n_{ij}} \right) \left(\frac{N_i}{\sum_{j=1}^J N_i} \right) \quad \text{式 10}$$

表格 6 土地利用變遷混淆矩陣

		參考值(Reference)				總和(Total)
		$i = 1$	$i = 2$	...	$i = J$	
比較值 (comparasion)	$i = 1$	n_{11}	n_{12}	...	n_{1J}	$N_1 = \sum_{j=1}^J n_{1j}$
	$i = 2$	n_{21}	n_{22}	...	n_{2J}	$N_2 = \sum_{j=1}^J n_{2j}$
	...					
	$i = J$	n_{J1}	n_{J2}	...	n_{JJ}	$N_J = \sum_{j=1}^J n_{Jj}$

表格 7 土地利用變遷混淆機率矩陣

	參考值(Reference)					比較總和 (Comparison Total)
	$i = 1$	$i = 2$	...	$i = J$		
比較值 (Comparison)	$i = 1$	p_{11}	p_{12}	...	p_{1J}	$\sum_{j=1}^J p_{1j}$
	$i = 2$	p_{21}	p_{22}	...	p_{2J}	$\sum_{j=1}^J p_{2j}$
	...	...	...	...	...	...
	$i = J$	p_{J1}	p_{J2}	...	p_{JJ}	$\sum_{j=1}^J p_{Jj}$
參考總和 (Reference Total)		$\sum_{j=1}^J p_{i1}$	$\sum_{j=1}^J p_{i2}$	...	$\sum_{j=1}^J p_{iJ}$	1

從表格 7 土地利用變遷混淆機率矩陣中可以定義出土地利用的不確定性；分別為總分歧 (total disagreement, TD) 式 16、量分歧 (quantity disagreement, QD) 式 12 與配置分歧 (allocation disagreement, AD) 式 14。透過區分量分歧與配置分歧，可以比較土地利用變遷模式中的誤差來源。若是量分歧大於配置分歧，代表此模型中對於為未來土地利用需求設定較為不足；若是配置分歧大於量分歧，則代表土地利用驅動因子可能不足以完整描述該區域之土地利用變化。

$$q_g = \left| \sum_{i=1}^J p_{ig} - \sum_{j=1}^J p_{gj} \right| \quad \text{式 11}$$

$$QD = \frac{\sum_{g=1}^J q_g}{2} \quad \text{式 12}$$

$$a_g = 2 \min \left| \left(\sum_{i=1}^J p_{ig} \right) - p_{gg}, \left(\sum_{j=1}^J p_{gj} \right) - p_{gg} \right| \quad \text{式 13}$$

$$AD = \frac{\sum_{g=1}^J a_g}{2} \quad \text{式 14}$$

$$C = \sum_{j=1}^J p_{jj} \quad \text{式 15}$$

$$TD = 1 - C = QD + AD \quad \text{式 16}$$

肆、研究結果與討論



本研究係以桃園地區為研究區域，應用隨機抽樣(SRS)、系統抽樣(SYS)以及分層隨機抽樣(STS)，與固定抽樣比例為 90%、80%、70%、60%、50%，搭配 CLUE-s 模式結合邏輯斯迴歸(LR)、廣義可加性模式(GAM)及隨機森林(RF)，所得之結果為研究母體。分析不同抽樣方法下之 TOC、曲線下面積(AUC)、土地利用配置分歧與品質因素(FoM)等指標之差異。

第一節、研究結果

(一)抽樣方式對空間自相關性的影響

本研究以 Moran's I 數值評估桃園地區的土地利用的空間自相關性。Moran's I 數值介於-1 到 1 之間，-1 代表負空間自相關，0 為無空間自相關，1 則為完全空間自相關。

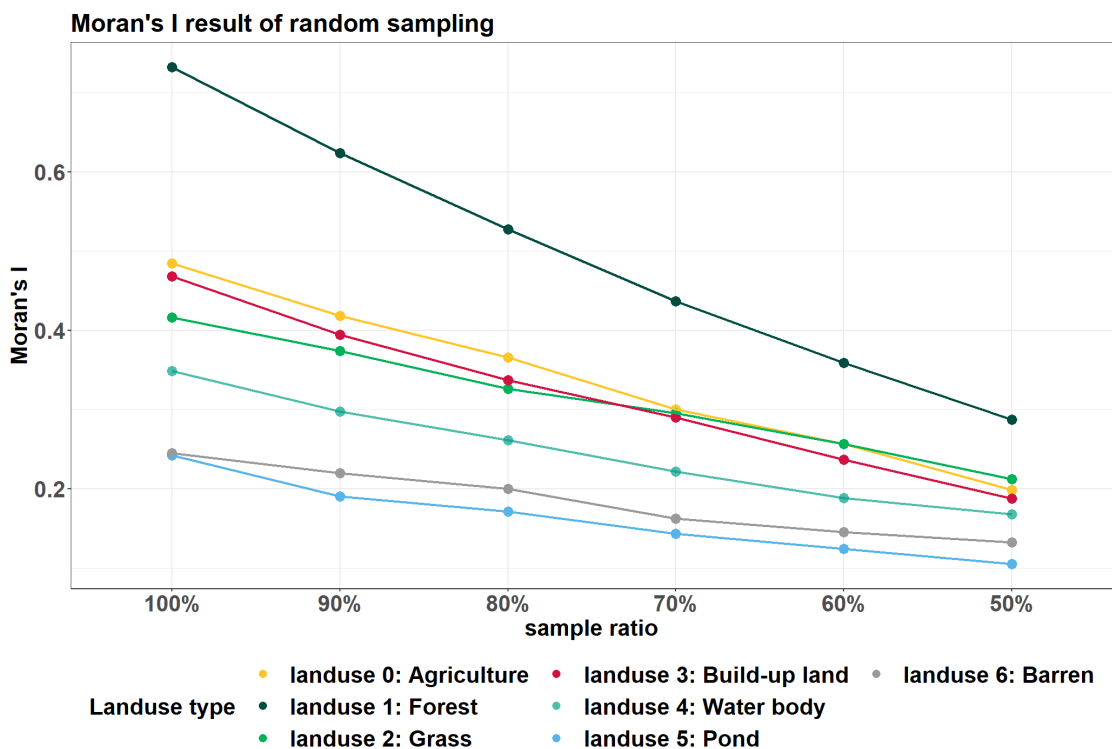


圖 9 隨機抽樣之抽樣比例對 Moran's I 數值變化圖

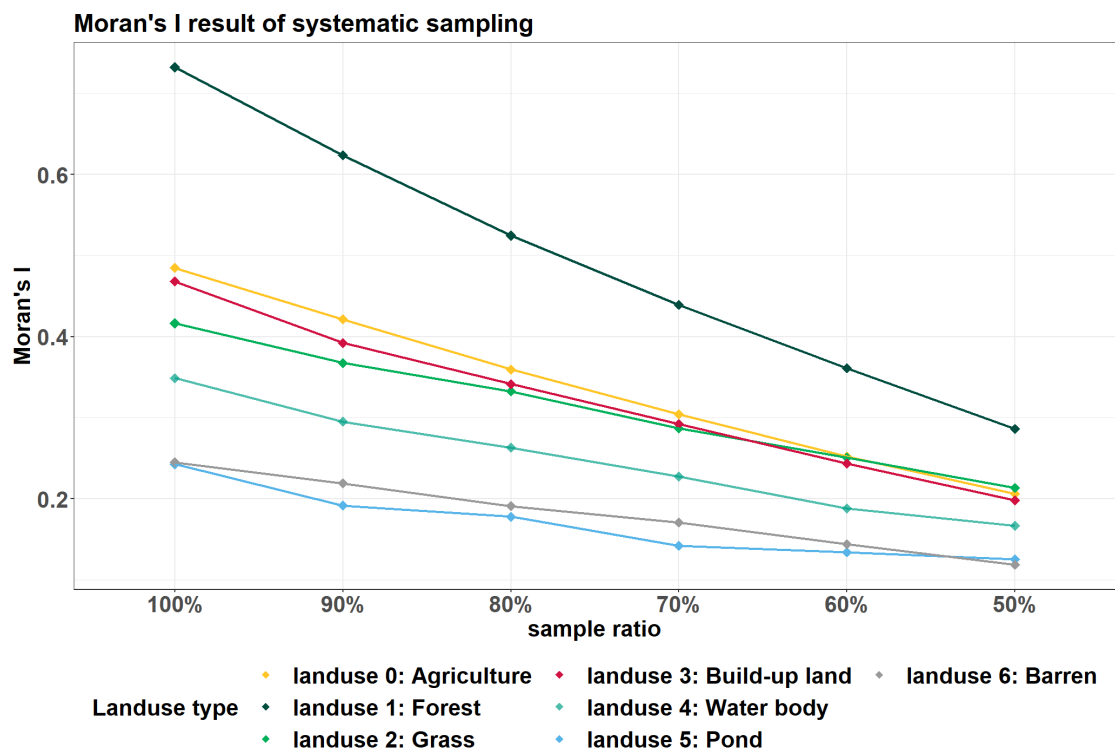


圖 10 系統抽樣之抽樣比例對 Moran's I 數值變化圖

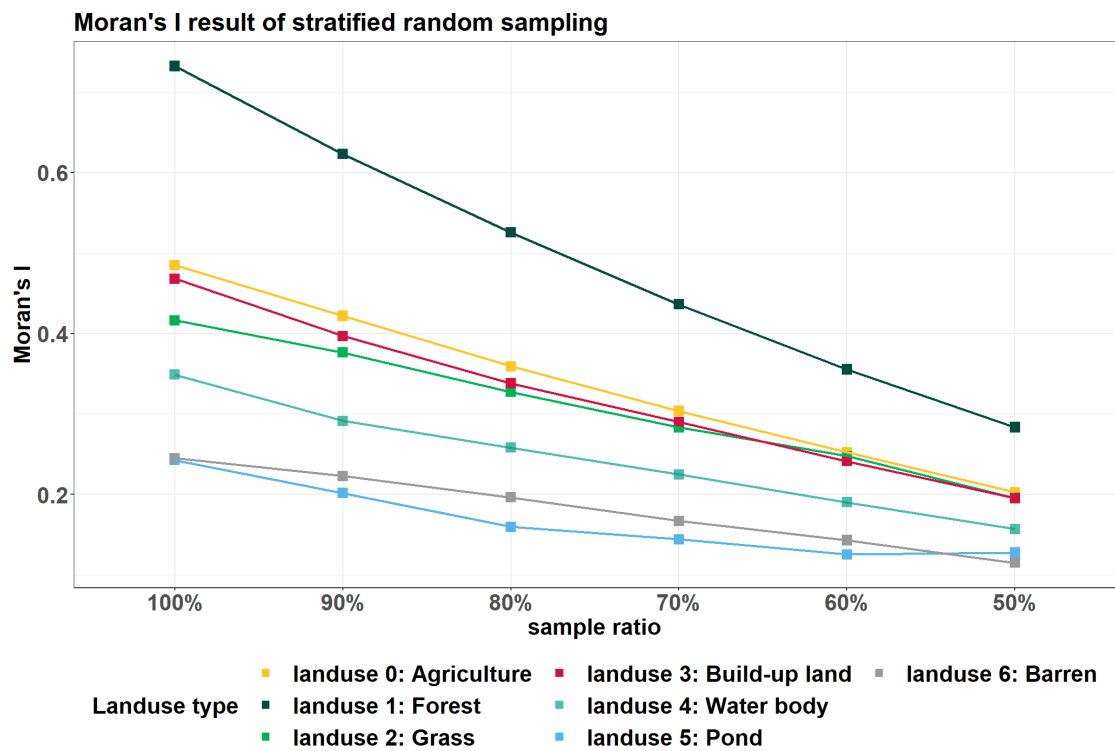


圖 11 分層隨機抽樣之抽樣比例對 Moran's I 數值變化圖



圖 9、圖 10 及圖 11 分別為桃園地區의 各個土地利用類別在固定抽樣比例經由隨機抽樣、系統抽樣、分層隨機抽樣後的 Moran's I 數值的變化圖。縱軸為 Moran's I 數值，橫軸則為抽樣比例。由於 Moran's I 數值為評估空間分布是否聚集的指標，因此當土地利用網格佔比較高則容易呈現分布較為聚集，Moran's I 數值較大之趨勢。在桃園地區中則以林地、農地、建地等三種土地利用類別佔比較高，故此三類別之 Moran's I 數值在圖 9、圖 10 及圖 11 皆明顯較其他土地利用類別大。此外，水體¹因為空間分布的特性較其他土地利用類別較為聚集，故呈現土地利用佔比較小，但 Moran's I 數值較大之趨勢。綜觀而言，由此三圖的趨勢大致呈現林地下降幅度最大，農地、建地、草地次之，水體第五，埤塘與荒地最小的趨勢。由此趨勢可明顯看出，抽樣方式確實可以有效降低空間自相關性，但抽樣方式間並無明顯差異。表格 8 則為不同抽樣方式及抽樣比例之 Moran's I 數值列聯表。由此表可知 100%抽樣(全樣本)下各土地利用類別的 Moran's I 數值約為 50%抽樣時的 Moran's I 數值的 2 倍。Moran's I 值依照土地利用類別不同及抽樣大小不同介於 0.1053 到 0.6238，其中又以林地的差異最大介於 0.2835 到 0.6229，埤塘的差異最小介於 0.1053 到 0.1906，此一趨勢則可明顯看出 Moran's I 數值與土地利用的網格數多寡相關，與上述趨勢相符。

¹ 本研究中的水體為包含湖泊、河川、水庫等較大範圍之水。由於埤塘為桃園地區的地理景觀特色，故分類上與水體區分以利於研究操作。

表格 8 不同抽樣方法及抽樣比例之 Moran's I 數值

抽樣比例	抽樣方式	0:農地	1:林地	2:草地	3:建地	4:水體	5:埤塘	6:荒地
100%		0.4848	0.7324	0.4164	0.4681	0.3488	0.2426	0.2450
90%	SRS	0.4181	0.6238	0.3738	0.3946	0.2974	0.1906	0.2200
	SYS	0.4214	0.6236	0.3675	0.3922	0.2951	0.1916	0.2191
	STS	0.4219	0.6229	0.3764	0.3969	0.2914	0.2016	0.2227
	全距	0.0038	0.0009	0.0088	0.0047	0.0060	0.0110	0.0037
80%	SRS	0.3660	0.5273	0.3261	0.3372	0.2616	0.1715	0.2002
	SYS	0.3594	0.5246	0.3325	0.3415	0.2631	0.1778	0.1910
	STS	0.3593	0.5256	0.3273	0.3382	0.2583	0.1598	0.1963
	全距	0.0067	0.0027	0.0064	0.0044	0.0047	0.0180	0.0092
70%	SRS	0.3004	0.4371	0.2952	0.2904	0.2218	0.1433	0.1626
	SYS	0.3042	0.4392	0.2868	0.2919	0.2275	0.1422	0.1708
	STS	0.3037	0.4362	0.2835	0.2899	0.2249	0.1444	0.1671
	全距	0.0038	0.0031	0.0118	0.0020	0.0057	0.0022	0.0082
60%	SRS	0.2570	0.3588	0.2565	0.2371	0.1882	0.1243	0.1455
	SYS	0.2520	0.3613	0.2511	0.2433	0.1881	0.1342	0.1438
	STS	0.2523	0.3554	0.2477	0.2414	0.1903	0.1255	0.1431
	全距	0.0050	0.0059	0.0089	0.0062	0.0022	0.0099	0.0025
50%	SRS	0.1989	0.2876	0.2121	0.1875	0.1681	0.1053	0.1325
	SYS	0.2059	0.2859	0.2131	0.1981	0.1667	0.1255	0.1188
	STS	0.2027	0.2835	0.1956	0.1954	0.1570	0.1279	0.1151
	全距	0.0070	0.0042	0.0176	0.0106	0.0111	0.0226	0.0174



(二)抽樣方式對分類演算法結果的影響

前一節為抽樣方式對土地利用資料 Moran’s I 數值的研究結果，本節則延續以隨機抽樣(SRS)、系統抽樣(SYS)及分層隨機抽樣(STS)在固定抽樣比例為 90%、80%、70%、60%、50%並分別抽樣 100 次建立3×5組重複抽樣樣本，再以邏輯斯迴歸(LR)、線性可加性模式(GAM)及隨機森林(RF)依據不同土地利用類別進行分類，紀錄每次分類後的 TOC 曲線下的 AUC 值，共3×5×3組重複試驗樣本，每組各 100 個 AUC 值，總共 4500 個 AUC 值為分析資料。依照土地利用類別，每組重複試驗樣本之 95%信賴區間如圖 12 到圖 18 所示。

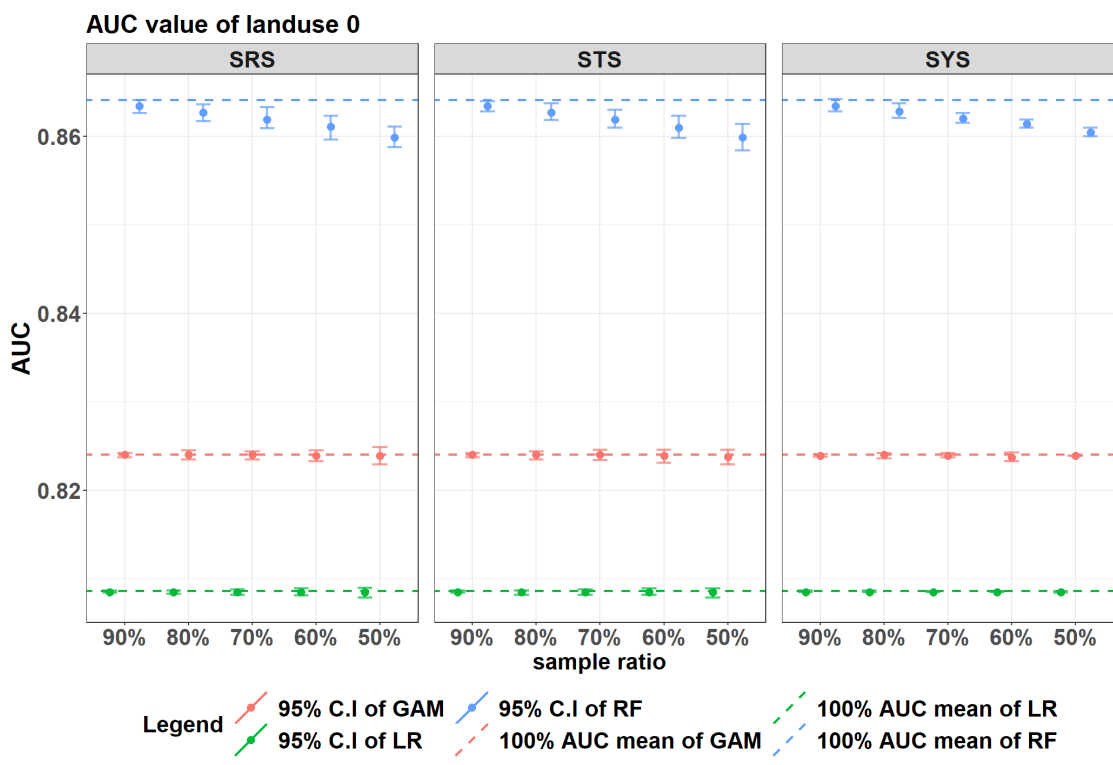


圖 12 農地 AUC 95%信賴區間分布圖

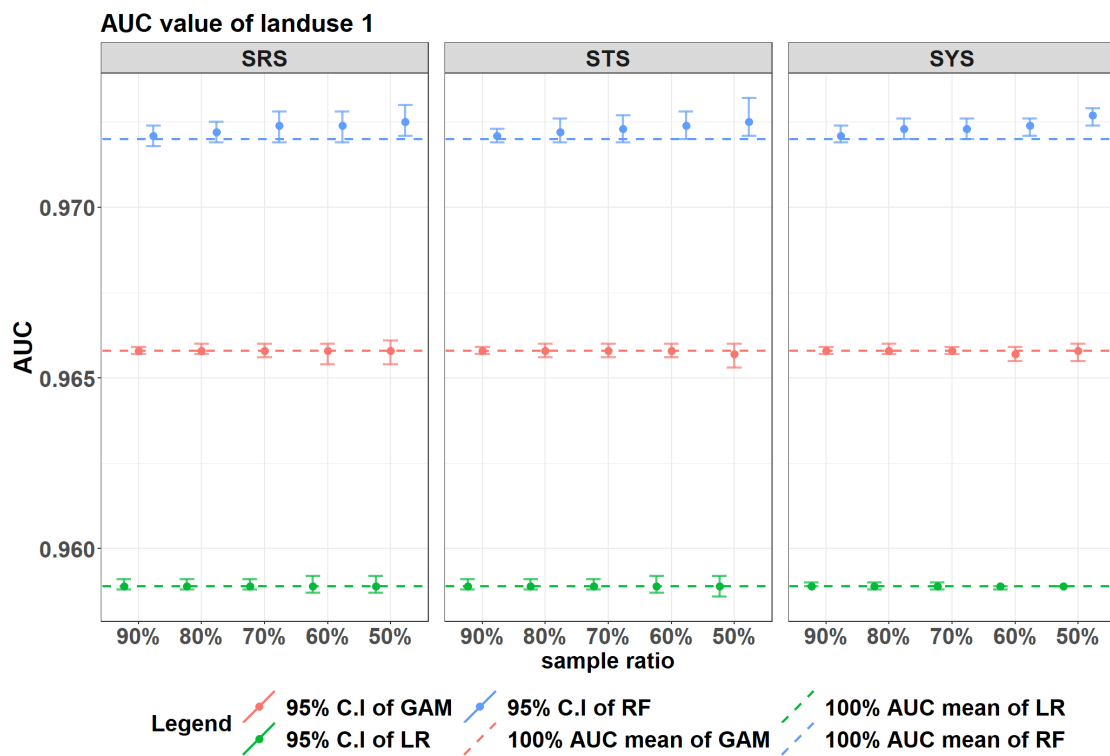


圖 13 林地 AUC 95%信賴區間分布圖

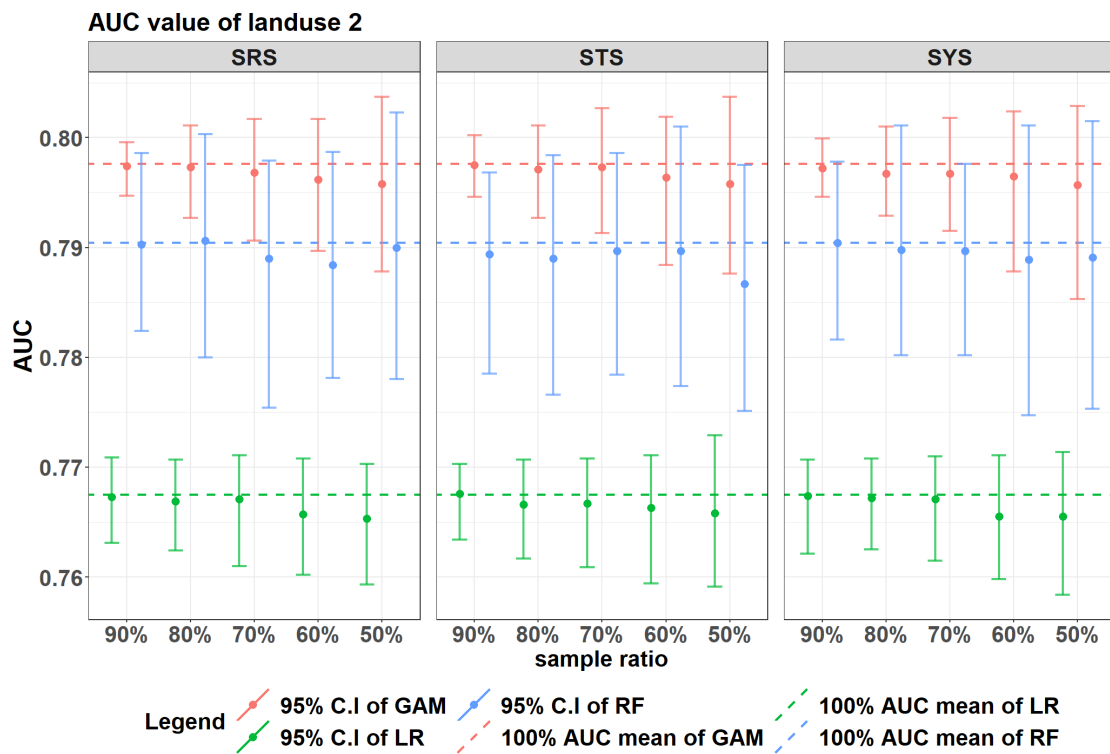


圖 14 草地 AUC 95%信賴區間分布圖

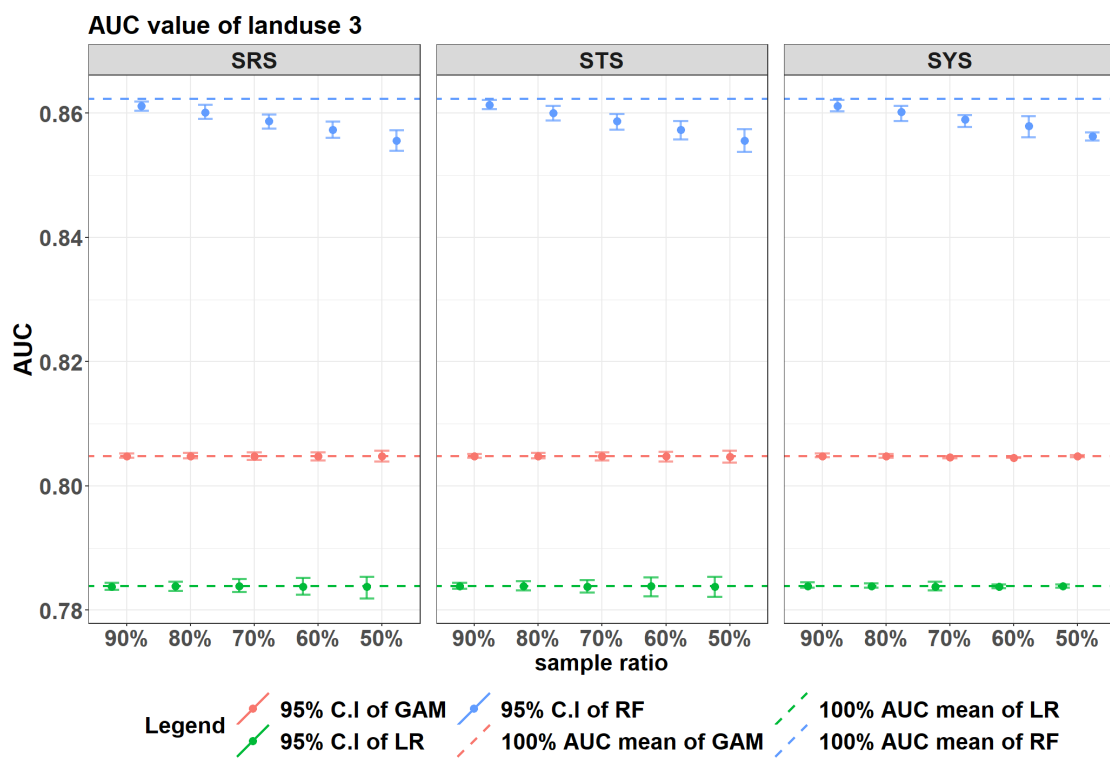


圖 15 建地 AUC 95%信賴區間分布圖

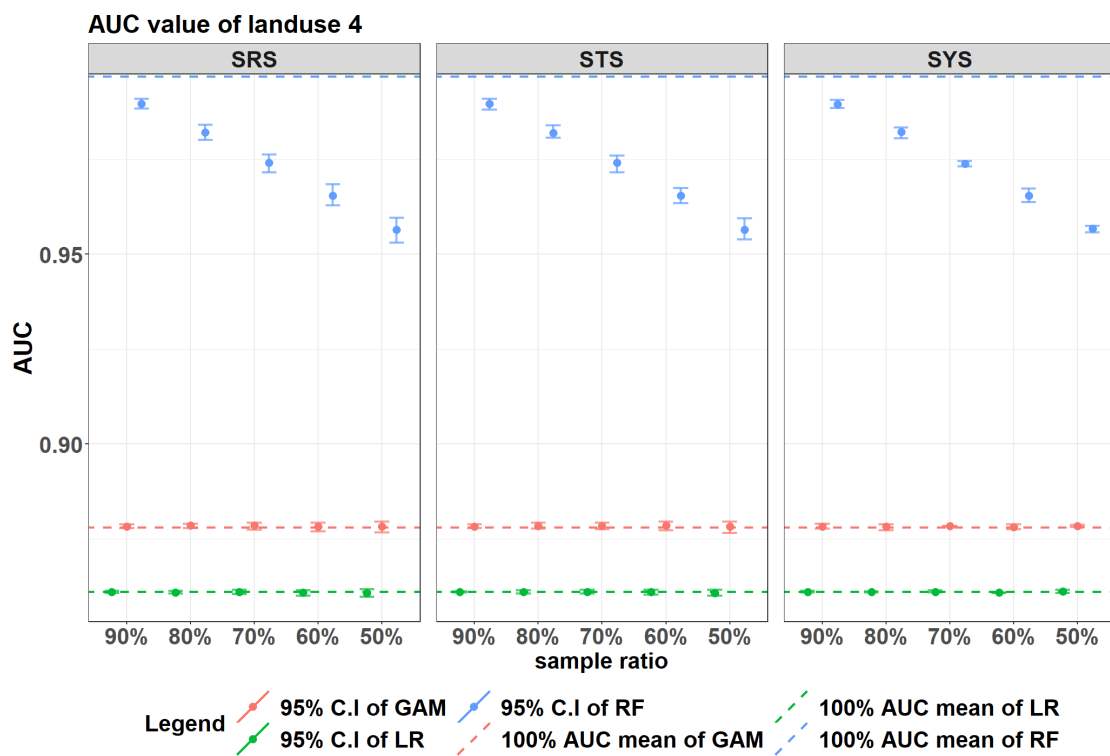


圖 16 水體 AUC 95%信賴區間分布圖

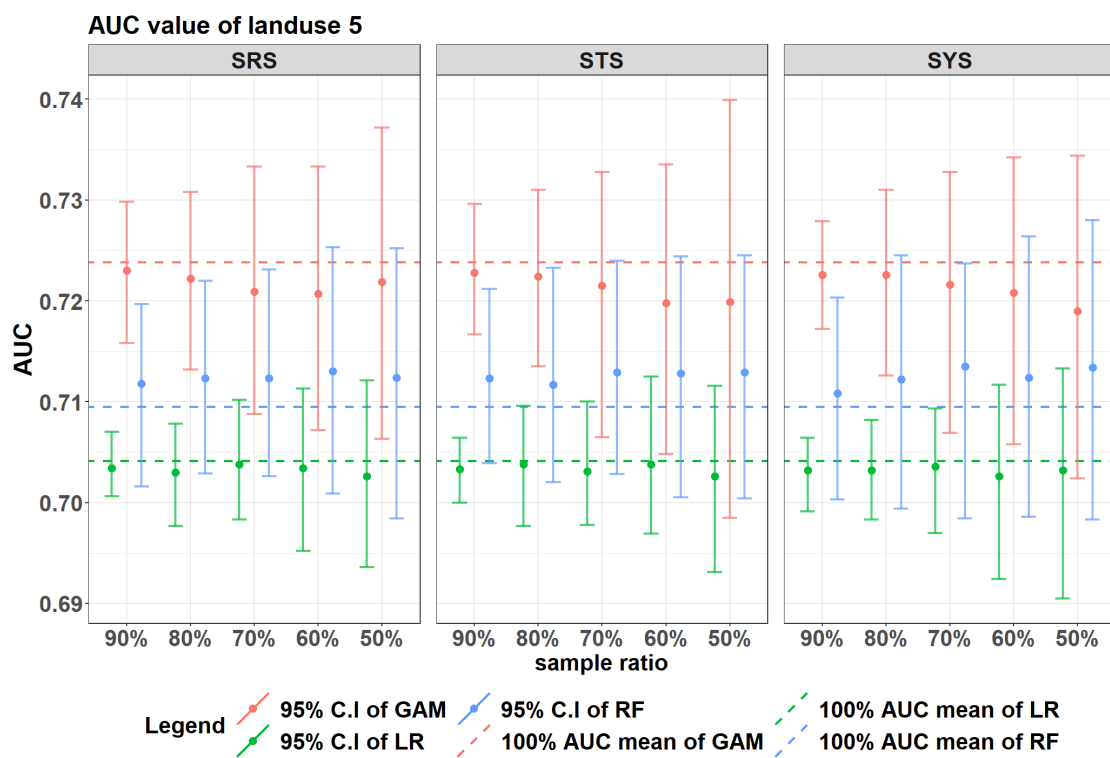


圖 17 埤塘 AUC 95%信賴區間分布圖

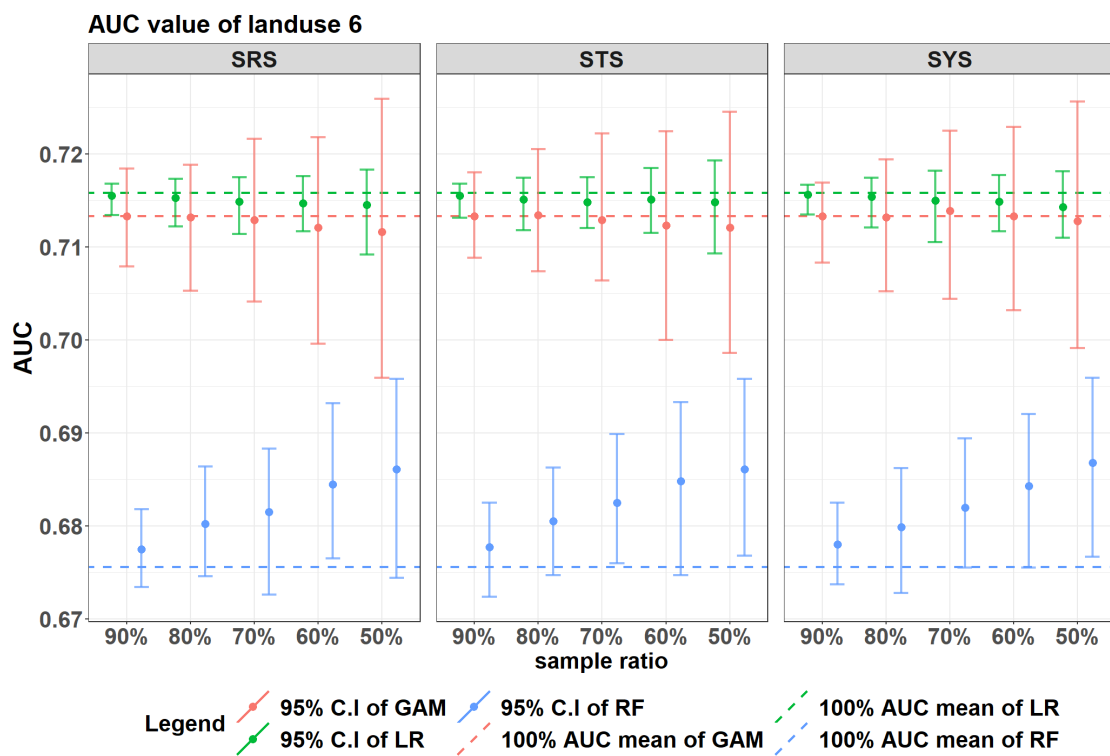


圖 18 荒地 AUC 95%信賴區間分布圖

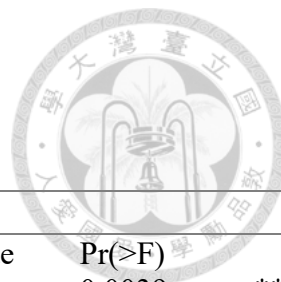


圖的縱軸為 AUC 值，橫軸為抽樣比例。由於分類演算法的特性不同，故先討論邏輯斯迴歸 (LR)與廣義可加性模式(GAM)。從圖 12 至圖 18 可以清楚看出，在空間隨機抽樣(SRS)、空間分層抽樣(STS)、空間系統抽樣(STS)及抽樣比例 90%、80%、70%、60%、50%下 LR 與 GAM 的 95%信賴區間(confidence interval, C.I)皆包含 100%AUC 平均值，且信賴區間範圍隨著抽樣比例下降而逐漸變大。其中，當土地利用類別為 2：草地、5：埤塘、6：荒地時，由於上述土地利用類別的網格點比例較少，草地 2.28%、埤塘 1.01%、荒地 4.64%，使得此三類別的信賴區間範圍明顯較其他土地利用類別大，但仍明顯隨著抽樣比例下降而逐漸變大。

RF 呈現與 LR 及 GAM 明顯不同的趨勢。在土地利用類別為 0：農地、3：建地、4：水體此三類別中，不分抽樣方式及抽樣比例 RF 的 95%信賴區間並未完全包含 100%AUC 值平均值，且 AUC 平均值隨著抽樣比例增加而上升並以 100%AUC 值平均值最佳，此趨勢在土地利用類別為 4:水體時最為明顯。當土地利用類別為 1：林地、6：荒地時，信賴區間同樣未完全包含 100%AUC 值平均值，但 AUC 值平均值以 50%為最佳，隨著抽樣比例下降而增加。當土地利用類別 2：草地、5：埤塘時，信賴區間皆包含 100%AUC 平均值，且信賴區間範圍隨著抽樣比例下降而逐漸變大。

以下將分小節針對不同分類模型及土地利用類別探討抽樣方式對於 AUC 值的影響。分析方式為雙因子變異數分析，討論在相同分類模型及土地利用類別²下，抽樣方式及抽樣比例對於 AUC 值之影響。

²註:土地利用分類為 0：農地、1：林地、2：草地、3：建地、4：水體、5：埤塘、6：荒地



1.邏輯斯迴歸(LR)

表格 9 邏輯斯迴歸 AUC 雙因子變異數分析表

農地:landuse 0						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	2.6030E-07	1.3015E-07	5.57	0.0039	**
ratio	4	3.7609E-07	9.4023E-08	4.03	0.0030	**
Residuals	1493	3.4859E-05	2.3348E-08			
林地:landuse 1						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	3.5390E-07	1.7695E-07	20.58	<0.0001	***
ratio	4	8.9170E-08	2.2293E-08	2.59	0.0350	*
Residuals	1493	1.2836E-05	8.5973E-09			
草地:landuse 2						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	6.3820E-06	3.1910E-06	0.37	0.6910	
ratio	4	7.8641E-04	1.9660E-04	22.78	<0.0001	***
Residuals	1493	1.2885E-02	8.6305E-06			
建地:landuse 3						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	2.2662E-08	1.1331E-08	0.04	0.9621	
ratio	4	7.3298E-07	1.8324E-07	0.62	0.6453	
Residuals	1493	4.3834E-04	2.9360E-07			
水體:landuse 4						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	2.9777E-06	1.4888E-06	17.80	<0.0001	***
ratio	4	5.3319E-06	1.3330E-06	15.93	<0.0001	***
Residuals	1493	1.2489E-04	8.3653E-08			
埤塘:landuse 5						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	5.7478E-06	2.8739E-06	0.20	0.8196	
ratio	4	6.8584E-05	1.7146E-05	1.19	0.3146	
Residuals	1493	2.1565E-02	1.4444E-05			
荒地:landuse 6						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.8743E-06	9.3713E-07	0.33	0.7191	
ratio	4	1.6215E-04	4.0539E-05	14.27	<0.0001	***
Residuals	1493	4.2419E-03	2.8412E-06			

本小節為討論以邏輯斯迴歸(以下稱 LR)為分類模型之 ANOVA 分析結果，從表格 9 中可以看出，在土地利用為 0:農地、1:林地、4:水體等三類別中，抽樣方式的 p-value 小於顯著水準(p-value=0.05)，顯示不同抽樣方式對此三類別的 AUC 平均值在統計上有顯著影響；另外在土地利用為 0:農地、1:林地、2:草地、4:水體、6:荒地等五類別中，抽樣比例的 p-value 小



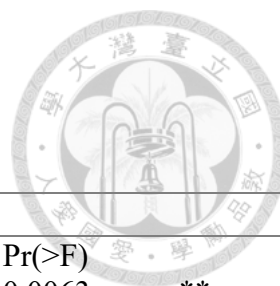
於顯著水準($p\text{-value}=0.05$)，顯示不同抽樣比例對此五類別的 AUC 平均值在統計上有顯著影響。

表格 10 邏輯斯迴歸事後檢定結果整理

邏輯斯迴歸(LR)				
土地利用	抽樣方式		抽樣比例	
	groups	AUC	groups	AUC
0:農地	SYS	0.8085	90%	0.8085
1:林地	STS	0.9589	null	
	SRS	0.9589		
2:草地	null		90%	0.7674
			70%	0.7670
			80%	0.7669
3:建地	null		null	
4:水體	SYS	0.8611	70%	0.8611
			90%	0.8611
			80%	0.8610
5:埤塘	null		null	
6:荒地	null		90%	0.7155

註: 表格中僅列出平均數最佳的組別，null 值則代表經事後檢定後不同組別間差異未達顯著

再從表格 10 邏輯斯迴歸事後檢定結果整理則可以知道各土地利用對於抽樣方式及抽樣比例組別平均 AUC 表現的差異。事後檢定採用 Scheffé Test 比較不同抽樣方式及抽樣比例之 AUC 平均數差異。由 ANOVA 報表已知抽樣方式在農地及水體中有顯著差異，再由事後檢定結果則可知，林地中隨機抽樣及分層抽樣的平均 AUC 數值較高，農地中系統抽樣的平均 AUC 數值較高。抽樣比例的事後檢定則呈現農地、草地、水體、荒地的 90% 平均 AUC 數值較高；草地、水體的 70%、80% 平均 AUC 數值較高。由此可知在邏輯斯迴歸下，抽樣方式差異對 AUC 數值的影響較不顯著，然而抽樣比例高的組別的影響較為明顯。



2.廣義可加性模式(GAM)

表格 11 廣義可加性模式 AUC 雙因子變異數分析表

農地:landuse 0						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	9.3900E-07	4.6955E-07	5.08	0.0063	**
ratio	4	4.0640E-06	1.0159E-06	10.99	<0.0001	***
Residuals	1493	1.3805E-04	9.2460E-08			
林地:landuse 1						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	4.8110E-07	2.4055E-07	13.00	<0.0001	***
ratio	4	1.4522E-06	3.6305E-07	19.62	<0.0001	***
Residuals	1493	2.7630E-05	1.8510E-08			
草地:landuse 2						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.3700E-05	6.8420E-06	0.71	0.4899	
ratio	4	4.7740E-04	1.1935E-04	12.45	<0.0001	***
Residuals	1493	1.4311E-02	9.5850E-06			
建地:landuse 3						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.2200E-06	6.0984E-07	6.37	0.0018	**
ratio	4	2.6760E-06	6.6901E-07	6.99	<0.0001	***
Residuals	1493	1.4292E-04	9.5730E-08			
水體:landuse 4						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.5000E-06	7.4760E-07	3.03	0.0485	*
ratio	4	2.7300E-06	6.8248E-07	2.77	0.0261	*
Residuals	1493	3.6802E-04	2.4650E-07			
埤塘:landuse 5						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	6.5000E-05	3.2410E-05	0.74	0.4782	
ratio	4	1.5450E-03	3.8631E-04	8.80	<0.0001	***
Residuals	1493	6.5562E-02	4.3910E-05			
荒地:landuse 6						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.2200E-04	6.0974E-05	2.47	0.0853	
ratio	4	3.1700E-04	7.9226E-05	3.20	0.0125	*
Residuals	1493	3.6918E-02	2.4728E-05			

本小節為討論以廣義可加性模式(以下稱 GAM)為分類模型之結果，依照土地利用類別 ANOVA 分析結果如表格 2 所示。從表格 8 中可以看出，在土地利用為 0:農地、1:林地、3:建地、4:水體等四類別中，抽樣方式的 p-value 小於顯著水準(p-value=0.05)，顯示不同抽樣方式對此四類別的 AUC 平均值在統計上有顯著影響；另外在所有土地利用類別中，抽樣比



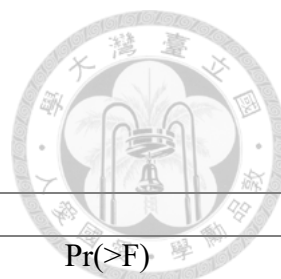
例的 p-value 皆小於顯著水準($p\text{-value}=0.05$)，顯示在分類模型為 GAM 的條件下不同抽樣比例對所有土地利用類別的 AUC 平均值在統計上有顯著影響。

表格 12 廣義可加性模式事後檢定結果整理

廣義可加性模式(GAM)				
土地利用	抽樣方式		抽樣比例	
	groups	AUC	groups	AUC
0:農地	SRS	0.8239	90%	0.8240
			80%	0.8240
			70%	0.8239
1:林地	STS	0.9658	90%	0.9658
	SRS	0.9658	80%	0.9658
2:草地	null		90%	0.7974
3:建地	SRS	0.8048	80%	0.8048
			90%	0.8048
4:水體	STS	0.8784	null	
5:埤塘	null		90%	0.7228
			80%	0.7224
6:荒地	null		null	

註: 表格中僅列出平均數最佳的組別，null 值則代表經事後檢定後不同組別間差異未達顯著

再從表格 12 廣義可加性模式事後檢定結果整理則可以知道各土地利用對於抽樣方式及抽樣比例組別平均 AUC 表現的差異。由 ANOVA 報表已知抽樣方式在農地、林地、建地、水體中有顯著差異，再由事後檢定結果則可知，隨機抽樣及分層抽樣的平均數值較高，抽樣比例的事後檢定則呈現農地、林地、草地、建地、埤塘的 90% 平均 AUC 數值較高。由此可知在廣義可加性模式下，系統抽樣的樣本的 AUC 數值較差，但抽樣比例大的組別的 AUC 數值仍然較佳。



3.隨機森林(RF)

表格 13 隨機森林 AUC 雙因子變異數分析表

農地:landuse 0						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.9109E-05	9.5545E-06	37.68	<0.0001	***
ratio	4	2.0350E-03	5.0874E-04	2006.30	<0.0001	***
Residuals	1493	3.7858E-04	2.5357E-07			
林地:landuse 1						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	6.0448E-07	3.0224E-07	7.90	0.0004	***
ratio	4	3.4296E-05	8.5741E-06	224.17	<0.0001	***
Residuals	1493	5.7103E-05	3.8247E-08			
草地:landuse 2						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	1.6886E-04	8.4428E-05	2.68	0.0689	
ratio	4	4.0370E-04	1.0093E-04	3.20	0.0124	*
Residuals	1493	4.7024E-02	3.1497E-05			
建地:landuse 3						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	3.6657E-05	1.8328E-05	35.09	<0.0001	***
ratio	4	5.4011E-03	1.3503E-03	2585.15	<0.0001	***
Residuals	1493	7.7982E-04	5.2232E-07			
水體:landuse 4						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	9.9471E-07	4.9736E-07	0.44	0.6453	
ratio	4	2.0522E-01	5.1304E-02	45203.65	<0.0001	***
Residuals	1493	1.6945E-03	1.1350E-06			
埤塘:landuse 5						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	7.9303E-06	3.9651E-06	0.08	0.9244	
ratio	4	3.7967E-04	9.4918E-05	1.88	0.1113	
Residuals	1493	7.5337E-02	5.0460E-05			
荒地:landuse 6						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
sample	2	3.1781E-05	1.5891E-05	1.01	0.3637	
ratio	4	1.3966E-02	3.4915E-03	222.36	<0.0001	***
Residuals	1493	2.3443E-02	1.5702E-05			

本小節為討論以隨機森林(以下稱 RF)為分類模型之結果，依照土地利用類別 ANOVA 分析結果如表格 3 所示。從表格 13 中可以看出，在土地利用為 0:農地、1:林地、3:建地等三類別中，抽樣方式的 p-value 小於顯著水準(p-value=0.05)，顯示不同抽樣方式對此三類別的 AUC 平均值在統計上有顯著影響；另外在 0:農地、1:林地、2:草地、3:建地、4:水體、6:荒地等六



種土地利用類別中，抽樣比例的 p-value 皆小於顯著水準($p\text{-value}=0.05$)，顯示在分類模型為 RF 的條件下不同抽樣比例對 AUC 平均值在統計上有顯著影響。

表格 14 隨機森林事後檢定結果整理

隨機森林(RF)				
土地利用	抽樣方式		抽樣比例	
	groups	AUC	groups	AUC
0:農地	SYS	0.8620	90%	0.8634
1:林地	SYS	0.9724	50%	0.9726
2:草地	null		90%	0.7900
3:建地	SYS	0.8589	90%	0.8612
4:水體	null		90%	0.9896
5:埤塘	null		null	
6:荒地	null		50%	0.6863

註: 表格中僅列出平均數最佳的組別，null 值則代表經事後檢定後不同組別間差異未達顯著

從表格 14 隨機森林事後檢定結果整理則可以知道各土地利用對於抽樣方式及抽樣比例組別平均 AUC 表現的差異。由 ANOVA 報表已知抽樣方式在農地、林建地中有顯著差異，再由事後檢定結果則可知，系統抽樣的平均數值較高，抽樣比例的事後檢定則呈現農地、草地、建地、水體的 90% 平均 AUC 數值較高，林地、荒地的 50% 平均 AUC 數值較高。由此可知在隨機森林下，系統抽樣的樣本的 AUC 數值較佳，90% 與 50% 抽樣比例對 AUC 數值因土地利用類別不同而有差異。



(三)抽樣方式對 CLUE-s 變遷模式結果

本節將討論隨機抽樣(SRS)、系統抽樣(SYS)以及分層隨機抽樣(STS)，與固定抽樣比例為 90%、80%、70%、60%、50%應用 CLUE-s 模式結合邏輯斯迴歸(LR)、廣義可加性模式(GAM)及隨機森林(RF)的土地利用配置差異。本節將以多因子變異數分析 48 組 CLUE-s 模型之土地利用配置結果。

1.土地利用配置分歧

由前一節的研究結果知道，在不同的分類演算法下，抽樣方式對於 AUC 數值的影響隨著土地利用類別不同而有差異，但抽樣比例則呈現抽樣比例較高的組別普遍呈現 AUC 平均值較高的趨勢。AUC 數值為評估模型分類準確度的指標，數值越高則分類結果越為準確。在此認知之下，可以預期抽樣比例較高的模型的土地利用配置分歧的結果可能較好。

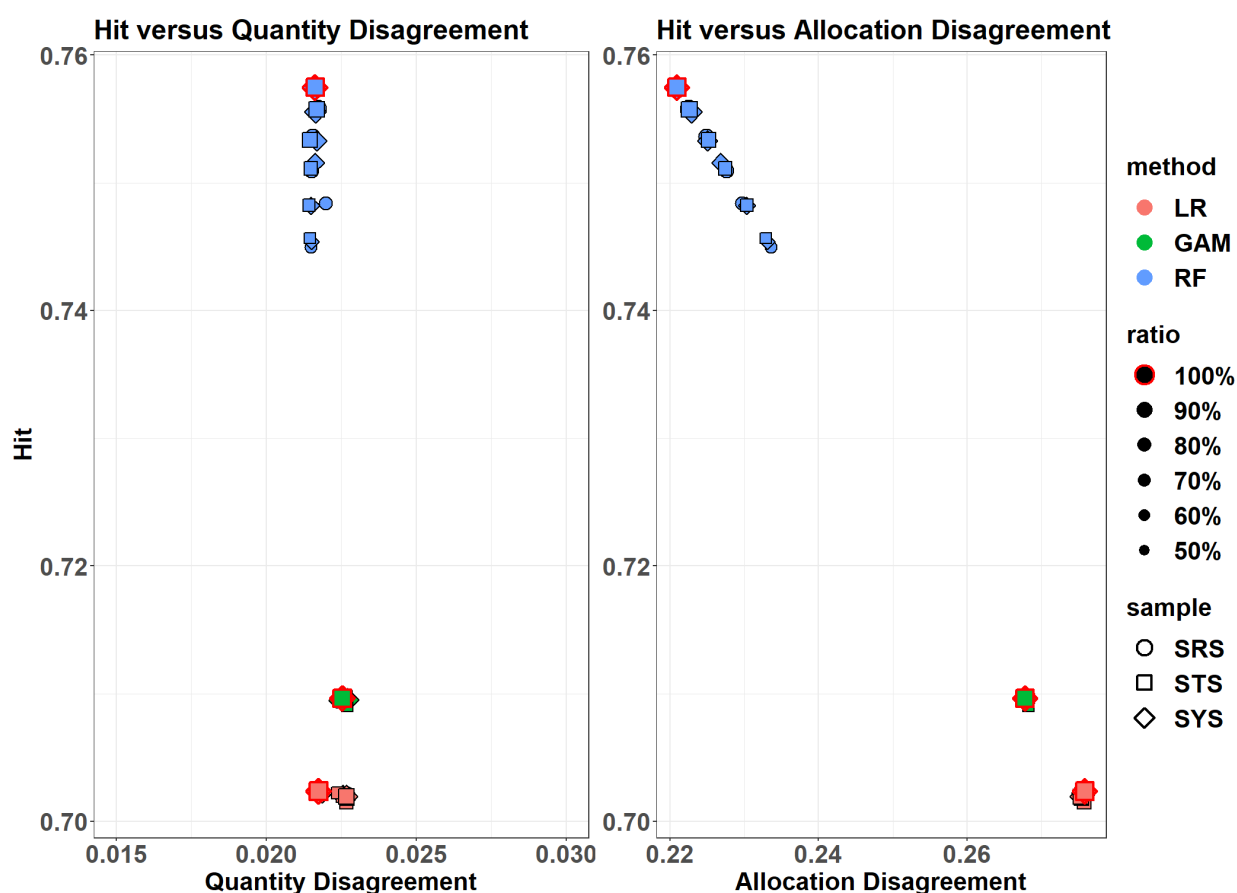


圖 19 土地利用分歧散佈圖

圖 19 土地利用分歧散佈圖，此圖呈現不同分類演算法、抽樣方式與抽樣比例之模型的土地利用配置結果差異，其中縱軸為正確配置數值，橫軸為量分歧與配置分歧數值，粉紅色為邏輯斯迴歸、綠色為廣義可加性模式、藍色為隨機森林。從圖 19 看出隨機森林的數值隨著抽樣比例而有明顯變動，邏輯斯迴歸及廣義可加性模式則無明顯變動。以下將透過變異數分析各土地利用配置分歧數值於統計上之意義。以下為邏輯斯迴歸、廣義可加性模式及隨機森林的土地利用配置分歧之多因子變異數分析結果。控制因子 ratio 代表抽樣比例、sample 則為抽樣方式。

表格 15 邏輯斯迴歸土地利用配置分歧多因子變異數分析表

正確配置 Correct (C)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	5.0377E-07	1.0075E-07	2.29	0.1238
sample	2	3.4410E-08	1.7204E-08	0.39	0.6860
Residuals	10	4.3953E-07	4.3953E-08		
量分歧 Quantity Disagreement (QD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	1.8201E-06	3.6403E-07	13.38	0.0004
sample	2	1.0594E-07	5.2970E-08	1.95	0.1932
Residuals	10	2.7216E-07	2.7220E-08		
配置分歧 Allocation Disagreement (AD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	5.9278E-07	1.1856E-07	3.90	0.0318
sample	2	2.6840E-08	1.3420E-08	0.44	0.6547
Residuals	10	3.0360E-07	3.0360E-08		

由表格 15 可知，ratio 的 p-value 在量分歧與配置分歧中皆小於顯著水準(p-value=0.05)，故可拒絕不同抽樣比例之量分歧與配置分歧的平均為相同之原假設，接受至少有一組抽樣比例之量分歧與配置分歧數值不同；意即，對於以邏輯斯迴歸為分類方法的土地利用變遷模式，抽樣比例之差異會影響量分歧與配置分歧的數值。



表格 16 廣義可加性模式模型土地利用配置分歧多因子變異數分析表

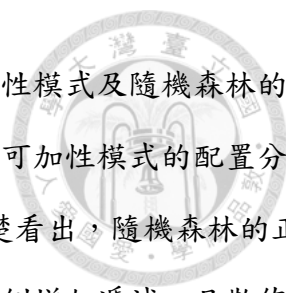
正確配置 Correct (C)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	2.9754E-07	5.9508E-08	2.71	0.0844
sample	2	1.6890E-08	8.4470E-09	0.38	0.6903
Residuals	10	2.1958E-07	2.1958E-08		
量分歧 Quantity Disagreement (QD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	4.0140E-08	8.0280E-09	0.61	0.6936
sample	2	6.8300E-09	3.4130E-09	0.26	0.7759
Residuals	10	1.3114E-07	1.3114E-08		
配置分歧 Allocation Disagreement (AD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	3.5308E-07	7.0616E-08	4.30	0.0239
sample	2	4.0100E-09	2.0050E-09	0.12	0.8864
Residuals	10	1.6431E-07	1.6431E-08		

由表格 16 可知，配置分歧中的 ratio 之 p-value 小於顯著水準(p-value=0.05)，故可推論抽樣比例差異對於以廣義可加性模式為分類方法的土地利用變遷模式，抽樣比例之差異會影響配置分歧的數值。

表格 17 隨機森林模型土地利用配置分歧多因子變異數分析表

正確配置 Correct (C)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	2.9971E-04	5.9943E-05	1100.57	<0.0001
sample	2	1.7000E-08	8.0000E-09	0.15	0.8607
Residuals	10	5.4500E-07	5.4000E-08		
量分歧 Quantity Disagreement (QD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	8.1000E-08	1.6000E-08	0.97	0.4806
sample	2	4.2000E-08	2.1000E-08	1.25	0.3270
Residuals	10	1.6800E-07	1.7000E-08		
配置分歧 Allocation Disagreement (AD)					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	3.0650E-04	6.1299E-05	655.96	<0.0001
sample	2	9.0000E-09	5.0000E-09	0.05	0.9517
Residuals	10	9.3400E-07	9.3000E-08		

由表格 17 可知，正確配置、配置分歧中的 ratio 之 p-value 小於顯著水準(p-value=0.05)，故可推論抽樣比例差異對於以隨機森林為分類方法的土地利用變遷模式，抽樣比例之差異會影響正確配置及配置分歧的數值。



綜合以上研究結果，抽樣方式之差異並不影響邏輯斯迴歸、廣義可加性模式及隨機森林的土地利用分歧結果。但抽樣比例之差異對於邏輯斯迴歸的量分歧、廣義可加性模式的配置分歧與隨機森林的正確配置、配置分歧數值皆有顯著影響。由圖 19 可清楚看出，隨機森林的正確配置數值明顯隨著抽樣比例增加遞增，配置分歧數值則隨著抽樣比例增加遞減，且散佈方式為以抽樣比例為群之分布，此結果與多因子變異數分析結果相吻合。



2.土地利用變遷模式品質因素

三圖比較為比較實際土地利用轉變方式與模型預測土地利用轉變方式之結果是否一致，以檢視土地利用變遷模式是否完整描述該地的土地利用變遷過程。本節將以多因子變異數分析比較 48 組土地利用變遷模型之品質因素(Figure of Merit)及其他數值之平均值在統計上之意義。48 組土地利用變遷模式之品質因素(Figure of Merit)及其他數值的散佈圖為圖 20 縱軸為 Hit 值，橫軸依序為 Miss、Wrong Hit 與 False Alarm 值，其中以紅色外框標示為各分類演算法之 100%抽樣時之樣本點。

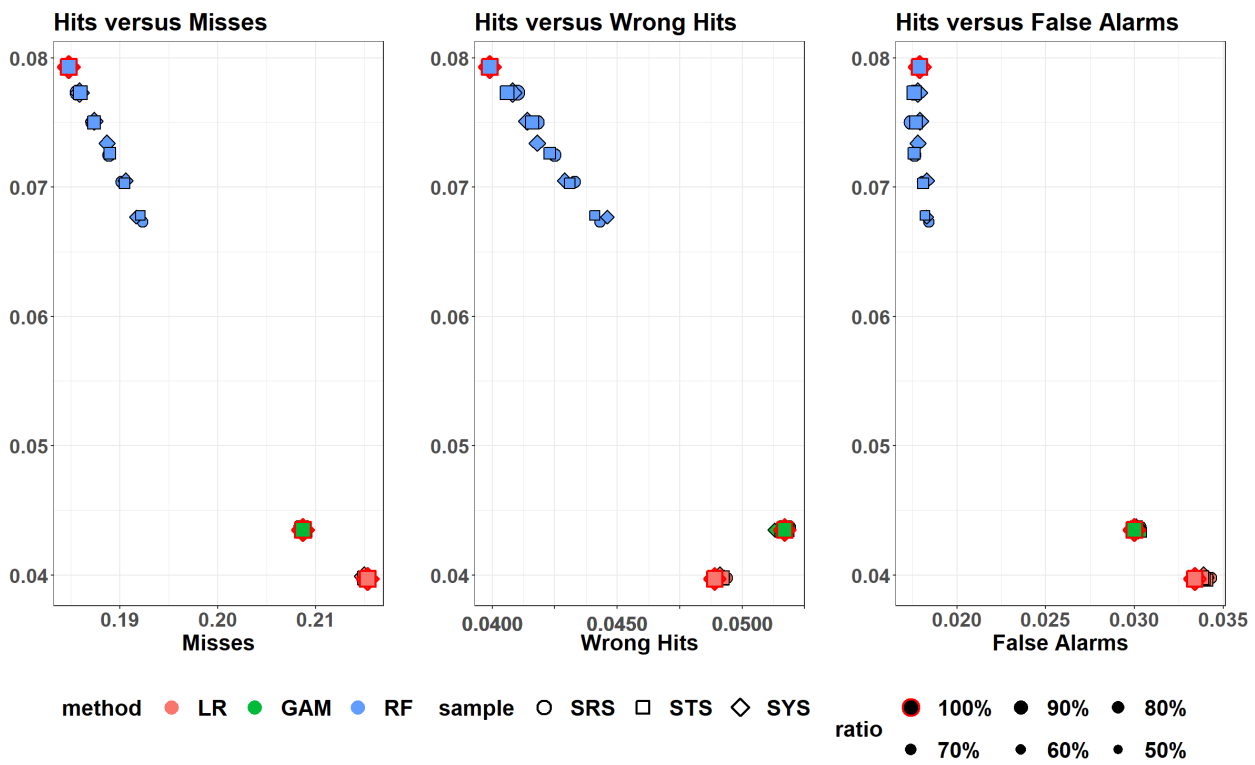


圖 20 figure of merit 與 miss, wrong hit, false alarm 之散佈圖

由圖 20 看出 LR 與 GAM 的樣本點不分抽樣比例與抽樣方式分布得較為聚集，RF 的樣本點則呈現以抽樣比例為群的方式聚集分布。又 48 組土地利用變遷模式中以 RF 的 100%抽樣模型的 Hit 值最大，Miss、Wrong Hit 與 False Alarm 值為最小，使得 Figure of Merit 值為最大。



(1) 邏輯斯迴歸(LR)

本小節為討論以邏輯斯迴歸(以下稱 LR)為分類模型之三圖比較 ANOVA 分析結果如表格 18 所示。

表格 18 邏輯斯迴歸三圖比較多因子變異數分析

Correct rejection					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	5.8278.E-07	1.1656.E-07	4.56	0.0199 *
sample	2	3.1110.E-08	1.5556.E-08	0.61	0.5630
Residuals	10	2.5556.E-07	2.5556.E-08		
False alarm					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	6.9111.E-07	1.3822.E-07	5.34	0.0120 *
sample	2	4.1110.E-08	2.0556.E-08	0.79	0.4786
Residuals	10	2.5889.E-07	2.5889.E-08		
Hit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	3.7778.E-08	7.5556.E-09	0.72	0.6259
sample	2	1.4444.E-08	7.2222.E-09	0.68	0.5266
Residuals	10	1.0556.E-07	1.0556.E-08		
Miss					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	3.0278.E-07	6.0556.E-08	2.57	0.0956
sample	2	7.1111.E-08	3.5556.E-08	1.51	0.2674
Residuals	10	2.3556.E-07	2.3556.E-08		
Wrong hit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	1.9833.E-07	3.9667.E-08	2.59	0.0942
sample	2	5.3333.E-08	2.6667.E-08	1.74	0.2248
Residuals	10	1.5333.E-07	1.5333.E-08		

由此表，ratio 與 sample 的 p-value 在所有類別中皆大於顯著水準($p\text{-value}=0.05$)，故可知抽樣比例及抽樣方式對於三圖比較之數值平均無顯著影響。多因子變異數分析結果與散佈圖中 LR 之樣本點分布情形一致，抽樣方式及抽樣比例對於三圖比較之數值並無差異。



(2) 廣義可加性模式(GAM)

本小節為討論廣義可加性模式(GAM)為分類模型之三圖比較 ANOVA 結果。

表格 19 廣義可加性模式模型三圖比較多因子變異數分析

Correct rejection					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	1.5611.E-07	3.1222.E-08	3.16	0.0574
sample	2	8.1111.E-08	4.0556.E-08	4.10	0.0501
Residuals	10	9.8889.E-08	9.8890.E-09		
False alarm					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	1.4000.E-07	2.8000.E-08	2.80	0.0779
sample	2	4.0000.E-08	2.0000.E-08	2.00	0.1859
Residuals	10	1.0000.E-07	1.0000.E-08		
Hit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	8.0000.E-08	1.6000.E-08	1.02	0.4547
sample	2	2.3333.E-08	1.1667.E-08	0.74	0.4995
Residuals	10	1.5667.E-07	1.5667.E-08		
Miss					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	3.2667.E-07	6.5333.E-08	2.93	0.0699
sample	2	1.3000.E-07	6.5000.E-08	2.91	0.1009
Residuals	10	2.2333.E-07	2.2333.E-08		
Wrong hit					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ratio	5	1.0667.E-07	2.1333.E-08	1.94	0.1744
sample	2	8.3333.E-08	4.1667.E-08	3.79	0.0596
Residuals	10	1.1000.E-07	1.1000.E-08		

由此表，ratio 與 sample 的 p-value 皆大於顯著水準(p-value=0.05)，可知抽樣方式對於 False Alarm、Hit、Wrong hit 與 Miss 之數值無顯著影響，抽樣比例對於 GAM 的三圖比較之數值無顯著影響。多因子變異數分析結果與散佈圖中 GAM 之樣本點分布情形一致，抽樣方式及抽樣比例對於 False Alarm、Hit、Wrong hit 與 Miss 之數值並無差異。



(3) 隨機森林(RF)

本小節為討論以隨機森林 (以下稱 RF) 為分類模型之三圖比較 ANOVA 分析結果如表格 20 所示。

表格 20 隨機森林三圖比較多因子變異數分析

Correct rejection						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
ratio	5	1.1428.E-06	2.2856.E-07	12.32	0.0005	***
sample	2	1.3444.E-07	6.7222.E-08	3.62	0.0656	
Residuals	10	1.8556.E-07	1.8556.E-08			
False alarm						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
ratio	5	1.1828.E-06	2.3656.E-07	18.84	0.0001	***
sample	2	1.0111.E-07	5.0556.E-08	4.03	0.0522	
Residuals	10	1.2556.E-07	1.2556.E-08			
Hit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
ratio	5	2.8481.E-04	5.6962.E-05	1241.31	<0.0001	***
sample	2	1.9400.E-07	9.7000.E-08	2.12	0.1709	
Residuals	10	4.5900.E-07	4.6000.E-08			
Miss						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
ratio	5	1.1364.E-04	2.2728.E-05	670.65	<0.0001	***
sample	2	5.4000.E-08	2.7200.E-08	0.80	0.4748	
Residuals	10	3.3900.E-07	3.3900.E-08			
Wrong hit						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
ratio	5	3.8111.E-05	7.6222.E-06	175.00	<0.0001	***
sample	2	1.9100.E-07	9.5600.E-08	2.19	0.1622	
Residuals	10	4.3600.E-07	4.3600.E-08			

由表格 20，ratio 的 p-value 在各類數值中皆小於顯著水準($p\text{-value}=0.05$)可知抽樣比例對於 RF 的三圖比較數值有顯著影響。多因子變異數分析的結果大致與 RF 的三圖比較數值散佈圖趨勢一致，呈現以抽樣比例為群的集中分布趨勢。



3.最有效之抽樣比例大小

由前述土地利用配置分歧與土地利用變遷模式品質因素的研究結果可以知道抽樣方式對於土地利用變遷模式土地利用分歧及品質因素無顯著影響，但抽樣比例的影響十分顯著，且同時又以 100%抽樣時的模型表現最為優異。然而考慮多數土地利用變遷研究經常以抽樣減少模型的運算時間，本小節將討論桃園地區的最有效之抽樣比例大小。

圖 21、圖 22 為邏輯斯迴歸、廣義可加性模式及隨機森林的土地利用變遷模型分數與 100% 模型的土地利用變遷模型分數比率對抽樣比例的折線圖。由於抽樣方式對三圖比較並無顯著，故採用相同演算法相同抽樣比例之平均值作為比較基準。圖的橫軸為抽樣比例，縱軸為與 100%模型數值的比較結果，在圖 21 中實線為三圖比較中的 hit 與 correct rejection 的加總，虛線則為 false alarm、wrong hit 與 miss 數值的加總，代表模型中的總誤差；在圖 22 中，實線為土地利用分歧中的正確配置，虛線則為總分歧。由圖 21 及圖 22 可以觀察出，隨著抽樣比例下降，模型的正確數值遞減，同時誤差數值遞增，此趨勢在隨機森林中最为明顯。

表格 21 則為各模型的三圖比較、土地利用分歧與 100%模型比率數值表，可知在邏輯斯迴歸中當抽樣比例由 100%減少為 90%時，H+CR 下降 0.04%，FA+WH+M 上升 0.11%，正確配置下降 0.05%，總分歧上升 0.13%，為僅次於 100%模型的次佳抽樣比例；在廣義可加性模式中當抽樣比例由 100%減少為 60%時，H+CR 下降 0.01%，FA+WH+M 上升 0.01%，正確配置下降 0.01%，總分歧上升 0.03%，為僅次於 100%模型的次佳抽樣比例；在隨機森林中當抽樣比例由 100%減少為 90%時，H+CR 下降 0.23%，FA+WH+M 上升 0.71%，正確配置下降 0.24%，總分歧上升 0.74%，為僅次於 100%模型的次佳抽樣比例。

若考慮樣本比例與模型演算時間，次佳抽樣比例可大幅減少模型的計算時間，且同時維持一定準確性及不具有過多誤差。

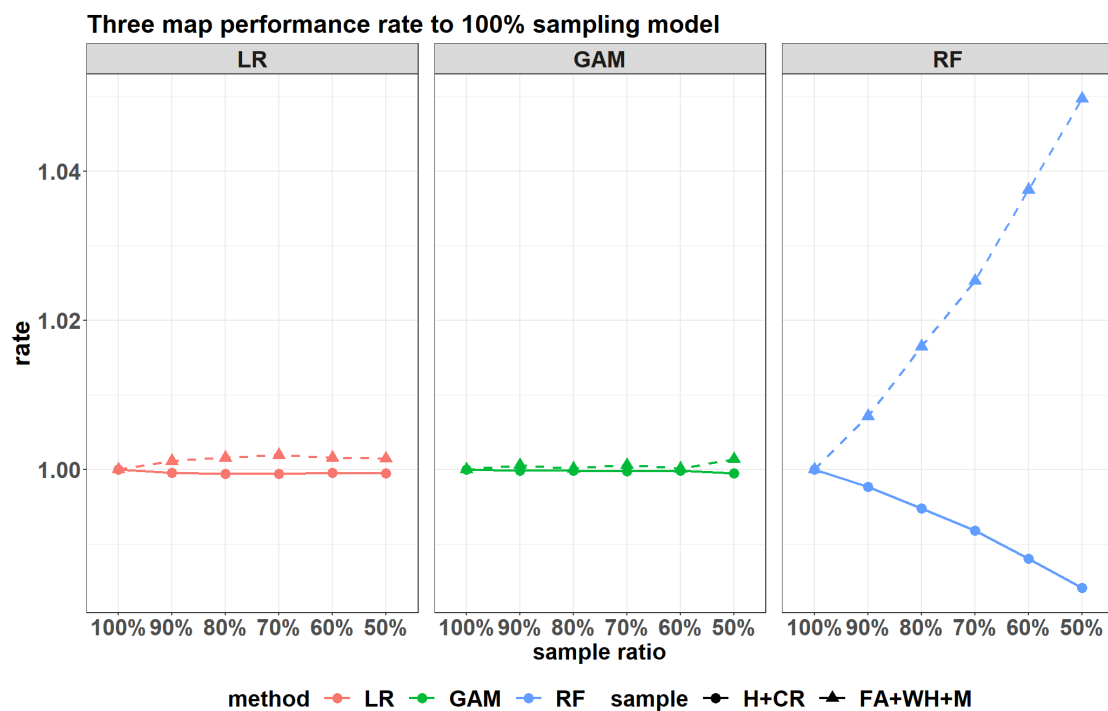


圖 21 各模型的三圖比較數值與 100%模型比率

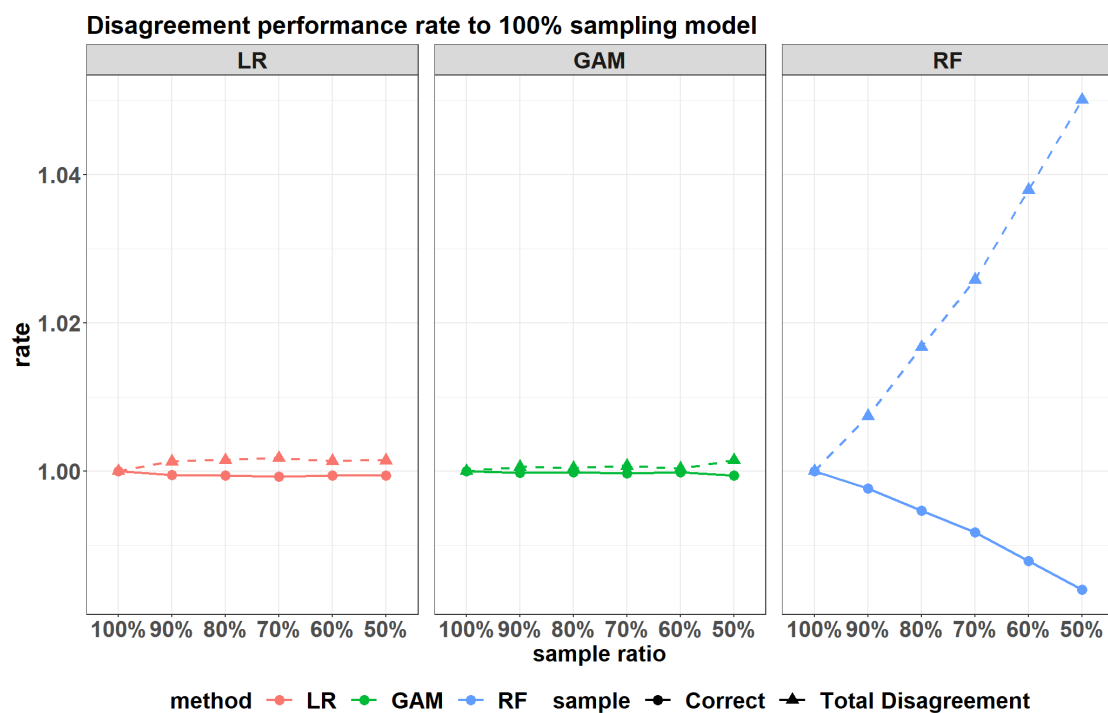


圖 22 各模型的土地利用分歧數值與 100%模型比率

表格 21 各模型的三圖比較、土地利用分歧與 100%模型比率數值表

method	ratio	三圖比較		土地利用分歧	
		H+CR	FA+WH+M	正確配置(C)	總分歧(TD)
LR	100%	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)
	90%	0.9996 (-0.04%)	1.0011 (0.11%)	0.9995 (-0.05%)	1.0013 (0.13%)
	80%	0.9994 (-0.06%)	1.0016 (0.16%)	0.9994 (-0.06%)	1.0015 (0.15%)
	70%	0.9994 (-0.06%)	1.0019 (0.19%)	0.9993 (-0.07%)	1.0017 (0.17%)
	60%	0.9995 (-0.05%)	1.0016 (0.16%)	0.9994 (-0.06%)	1.0013 (0.13%)
	50%	0.9995 (-0.05%)	1.0015 (0.15%)	0.9994 (-0.06%)	1.0014 (0.14%)
	100%	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)
	90%	0.9999 (-0.01%)	1.0006 (0.06%)	0.9998 (-0.02%)	1.0005 (0.05%)
GAM	80%	0.9999 (-0.01%)	1.0001 (0.01%)	0.9998 (-0.02%)	1.0004 (0.04%)
	70%	0.9998 (-0.02%)	1.0006 (0.06%)	0.9997 (-0.03%)	1.0007 (0.07%)
	60%	0.9999 (-0.01%)	1.0001 (0.01%)	0.9999 (-0.01%)	1.0003 (0.03%)
	50%	0.9995 (-0.05%)	1.0014 (0.14%)	0.9994 (-0.06%)	1.0015 (0.15%)
	100%	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)	1.0000 (0.00%)
	90%	0.9977 (-0.23%)	1.0071 (0.71%)	0.9976 (-0.24%)	1.0074 (0.74%)
	80%	0.9948 (-0.52%)	1.0165 (1.65%)	0.9946 (-0.54%)	1.0167 (1.67%)
	70%	0.9918 (-0.82%)	1.0253 (2.53%)	0.9917 (-0.83%)	1.0258 (2.58%)
RF	60%	0.9880 (-1.20%)	1.0375 (3.75%)	0.9879 (-1.21%)	1.0379 (3.79%)
	50%	0.9841 (-1.59%)	1.0497 (4.97%)	0.9840 (-1.60%)	1.0501 (5.01%)



(四)混合模型 Mix Model

本節將討論結合各土地利用類別最佳之分類演算法之混合模型的土地利用配置分歧與三圖比較並與抽樣方式對 CLUE-s 變遷模式結果之結果進行比較。首先，基於抽樣方式對分類演算法結果的影響之研究結果，依照不同土地利用類別的 AUC 平均值之大小選擇各抽樣方式及抽樣比例下最佳分類演算法，並以依序代換之方式將混合模式之模型組合整理如表格 22 所示。

表格 22 混合模型配置表

抽樣方法	抽樣比例	模型	土地利用類別						
			0: 農地	1: 林地	2: 草地	3: 建地	4: 水體	5: 埤塘	6: 荒地
SRS STS SYS		RF	RF	RF	RF	RF	RF	RF	RF
	50%	model1	RF	RF	GAM	RF	RF	RF	RF
	60%	model2	RF	RF	RF	RF	RF	RF	LR
	70%	model3	RF	RF	RF	RF	RF	GAM	RF
	80%	model4	RF	RF	GAM	RF	RF	GAM	RF
	90%	model5	RF	RF	GAM	RF	RF	RF	LR
	100%	model6	RF	RF	RF	RF	RF	GAM	LR
		model7	RF	RF	GAM	RF	RF	GAM	LR

由表格 22 可知各土地利用類別之最佳分類演算法依照土地利用順序為: RF、RF、GAM、RF、RF、GAM、LR。另外 model1、model2、model3 為 C_1^3 之組合，model4、model5、model6 為 C_2^3 之組合，model7 則為 C_3^3 之模型組合。以下將分兩小節分別比較混合模型與隨機森林之的土地利用配置分歧與三圖比較之差異。



1.混合模型之土地利用分歧

混合模型之土地利用變遷模式之土地利用配置分歧結果如圖 23~圖 25 示。圖的橫軸為不同的混合土地利用變遷模型，縱軸為數值，散佈點由小到大分別標示抽樣比例之大小，其中以紅色外框標示之散佈點為 100%抽樣時土地利用變遷模式之數值，方形代表隨機抽樣、菱形為分層抽樣、三角形則為系統抽樣。從土地利用配置分歧的研究結果可知，土地利用配置分歧中正確配置及分歧的數值會隨著抽樣比例增加而變動，但量分歧則幾乎不隨抽樣比例而變動。

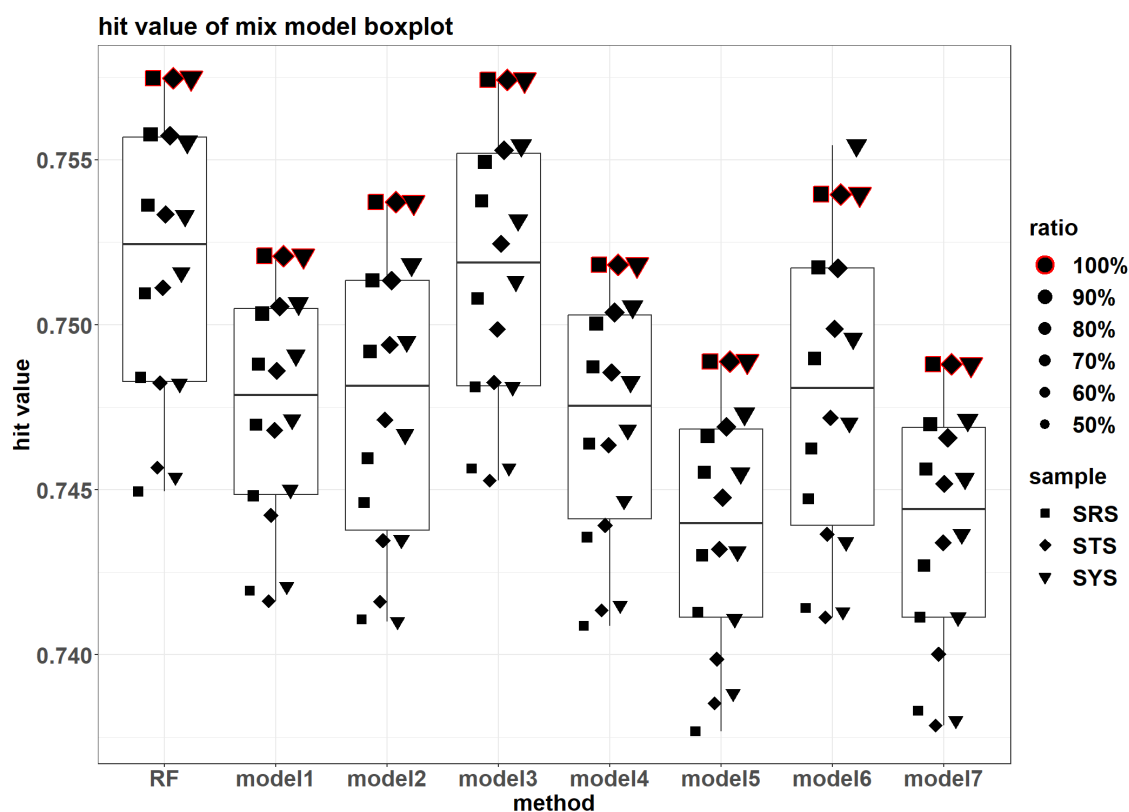


圖 23 混合模型之土地利用分歧之正確配置盒狀圖

由圖 23 可知各混合模型的正確配置數值皆隨著抽樣比例增加而遞增，並以 RF 與 model3 的正確配置數值分布皆明顯高於其他混合模型的正確配置數值。其中盒狀圖的分布呈現 RF 與 model3 相似、model1 與 model4 相似、model2 與 model6、model5 與 model7 等四組盒形圖分佈較為相近。

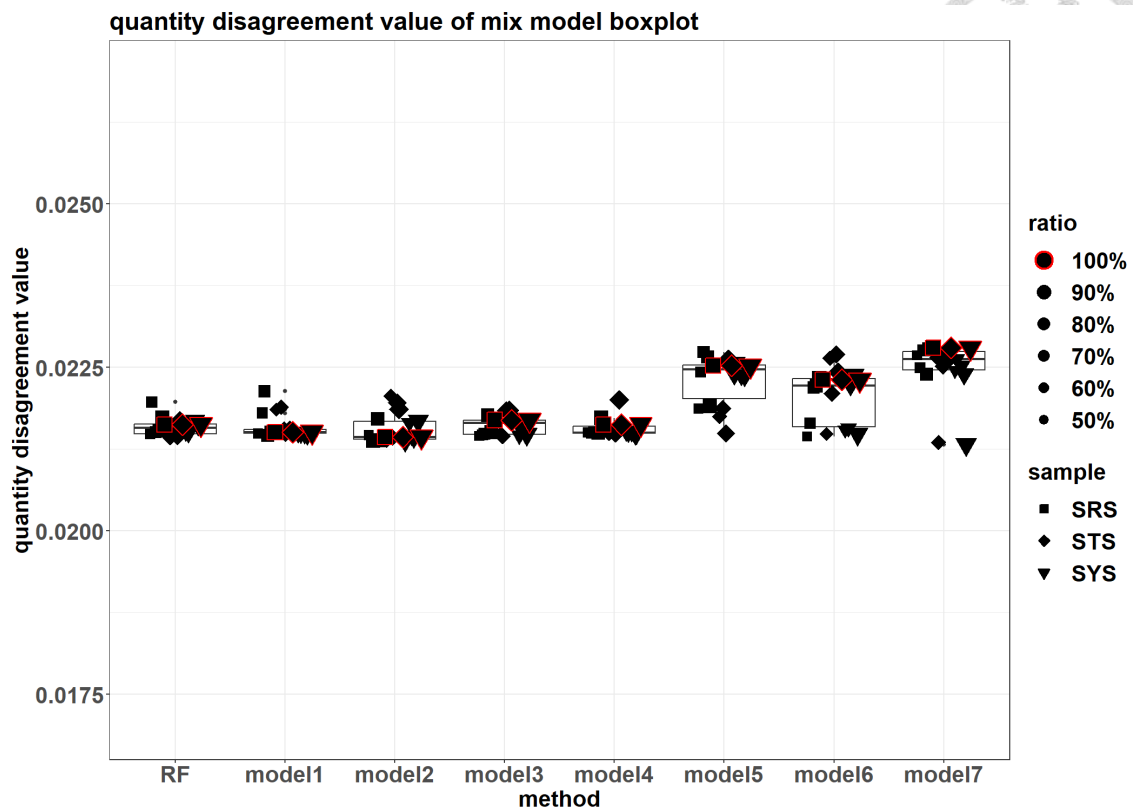


圖 24 混合模型之土地利用分歧之量分歧盒狀圖

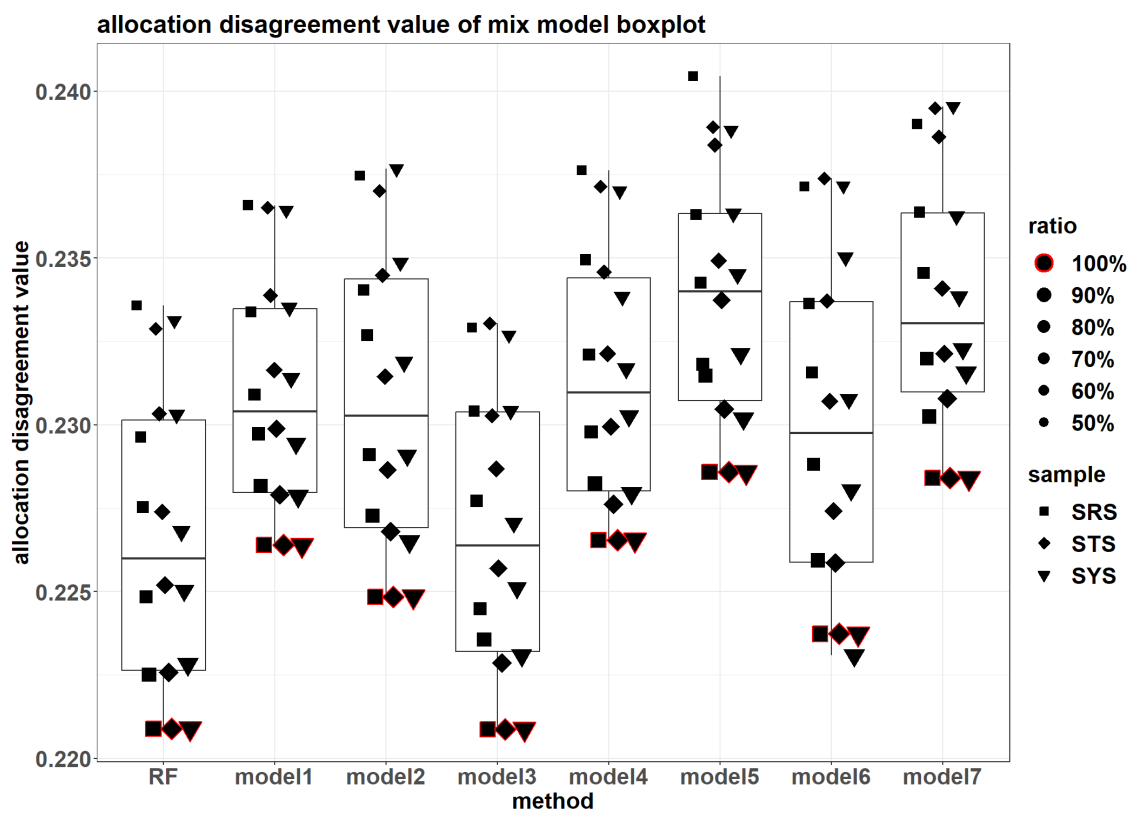


圖 25 混合模型之土地利用分歧之配置分歧盒狀圖



由圖 24 混合模型之土地利用分歧之量分歧盒狀圖知各混合模型的量分歧數值無明顯分布趨勢。盒狀圖的分布呈現 RF、model1、model2、model3 與 model4 的數值分布較為集中，另 model6、model5 與 model7 的數值分布則呈現較為發散的情況。

由圖 25 可知各混合模型的配置分歧數值皆隨著抽樣比例增加而遞減，並以 RF 與 model3 的配置分歧數值分布皆明顯小於其他混合模型的配置分歧數值。其中盒形圖分佈則與正確配置呈現相同的分組方式，RF 與 model3、model1 與 model4、model2 與 model6、model5 與 model7 等四組分佈。

混合模型之土地利用配置分歧之 100%抽樣模型數值結果如表格 23 所示。該表格為依照正確配置由大到小、配置分歧由小至大、量分歧由小至大排序而成。其中，以隨機森林 100%抽樣為最佳，完整替換之混合模型 model7 為最差。

表格 23 混合模型之土地利用分歧之 100%抽樣模型數值結果

排序	模型類別	抽樣方式	抽樣比例	正確配置 Hit	量分歧 QD	配置分歧 AD
1	RF	100%	100%	0.7575	0.0216	0.2209
2	model3	100%	100%	0.7574	0.0217	0.2209
10	model6	100%	100%	0.7540	0.0223	0.2237
12	model2	100%	100%	0.7537	0.0214	0.2248
18	model1	100%	100%	0.7521	0.0215	0.2264
20	model4	100%	100%	0.7518	0.0216	0.2265
44	model5	100%	100%	0.7489	0.0225	0.2286
46	model7	100%	100%	0.7488	0.0228	0.2284

各混合模型與隨機森林之土地利用分歧則如表格 24 所示。依照數值差異由小到大排列依序為 model3、model6、model2、model1、model4、model5、model7，且從替換之土地利用類別中可以看出對於土地利用分歧之影響由小到大依序為 Landuse 5:埤塘< Landuse 6:荒地< Landuse 2:草地。

表格 24 混合模型與隨機森林之土地利用分歧差異比較



模型類別	替換土地利用	正確配置 (比例)	量分歧 (比例)	配置分歧 (比例)
model3	LU5	-0.0001 (-0.01%)	0.0001 (0.46%)	0.0000 (0.00%)
model6	LU5+LU6	-0.0035 (-0.46%)	0.0007 (3.24%)	0.0028 (1.27%)
model2	LU6	-0.0038 (-0.50%)	-0.0002 (-0.93%)	0.0039 (1.77%)
model1	LU2	-0.0054 (-0.71%)	-0.0001 (-0.46%)	0.0055 (2.49%)
model4	LU2+LU5	-0.0057 (-0.75%)	0.0000 (0.00%)	0.0056 (2.54%)
model5	LU2+LU6	-0.0086 (-1.14%)	0.0009 (4.17%)	0.0077 (3.49%)
model7	LU2+LU5+LU6	-0.0087 (-1.15%)	0.0012 (5.56%)	0.0075 (3.40%)



2.混合模型之三圖比較與品質因素

混合模型之品質因素數值比較如圖 26 所示，混合模型之三圖比較數值盒型圖如圖 27 至圖 31 所示。圖的橫軸為不同的混合土地利用變遷模型，縱軸為數值，散佈點由小到大分別標示為抽樣比例之大小，其中以紅色外框標示之散佈點為 100%抽樣時土地利用變遷模式之數值，方形代表隨機抽樣、菱形為分層抽樣、三角形則為系統抽樣。從抽樣方式對 CLUE-s 變遷模式結果的研究結果可知，以隨機森林為分類演算法之土地利用變遷模式之品質因素會隨著抽樣比例增加而變動。

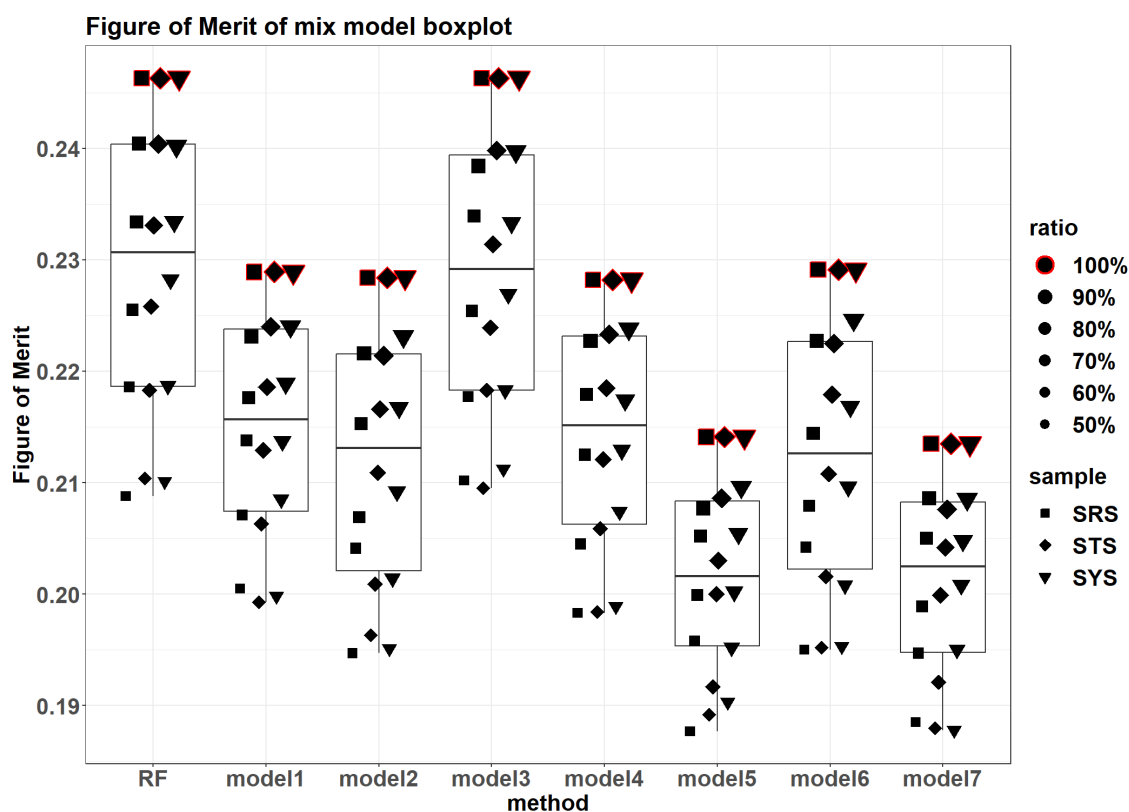


圖 26 混合模型三圖比較之品質因素盒狀圖

由圖 26 混合模型三圖比較之品質因素盒狀圖分佈情形可得出與混合模式土地利用分歧之正確配置相似結果，品質因素盒狀圖大致可分為 RF 與 model3、model1 與 model4、model2 與 model6、model5 與 model7 等四組分佈，但品質因素盒狀圖中，model5、model7 之分佈相較



於其他混合模式較為集中。另外，品質因素亦呈現數值隨著抽樣比例增加而遞增，並以 RF 與 model3 之 100%抽樣為最佳。

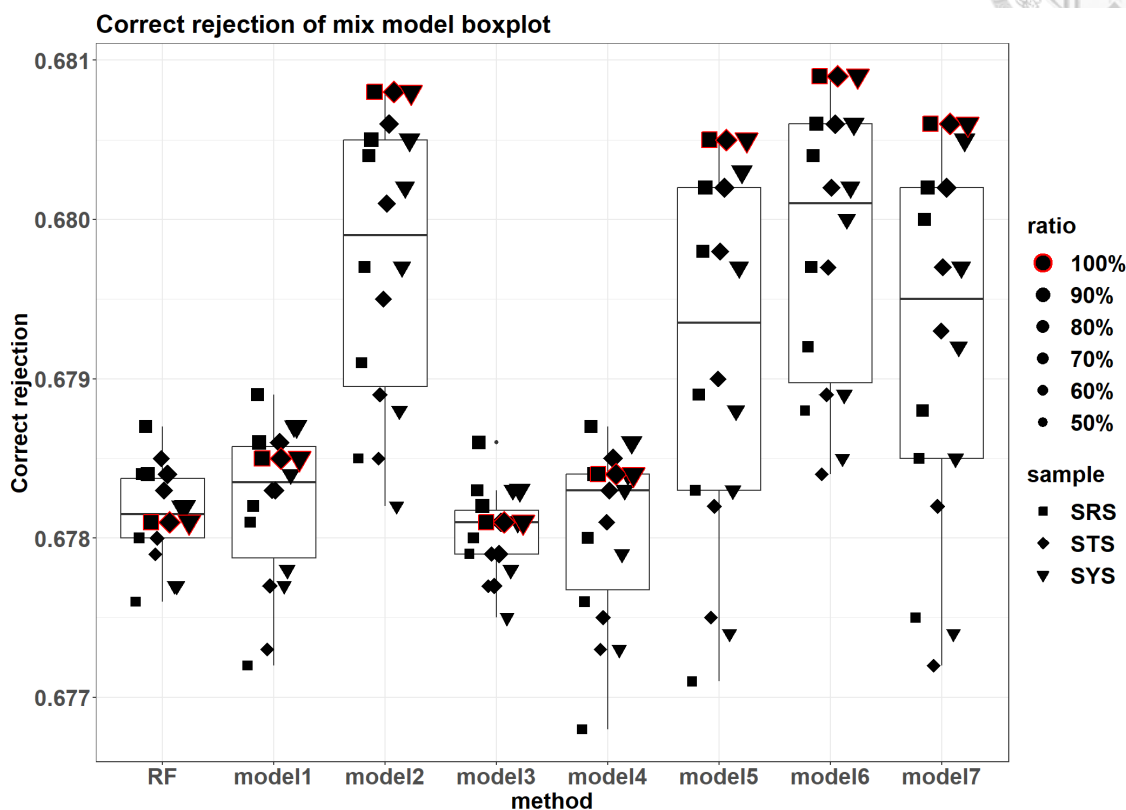


圖 27 混合模型三圖比較之 Correct Rejection 盒狀圖

由圖 27 混合模型三圖比較之 Correct Rejection 盒狀圖發現 RF、model1、model3、model4 相較於另外 4 組模型呈現數值分佈集中，且 100%抽樣 correct rejection 數值接近平均值之分佈情況。反之，model2、model5、model6、model7 則呈現數值分佈較為分散，且 correct rejection 數值隨著抽樣比例增加而遞增之情形，並以 100%抽樣之數值為最大。

由圖 28 混合模型三圖比較之 Hit 盒狀圖發現混合模型的 Hit 盒狀圖分佈與品質因素之盒狀圖分佈相似。各模型之數值皆隨著抽樣比例增加而遞增，並 100%抽樣之數值為最佳。

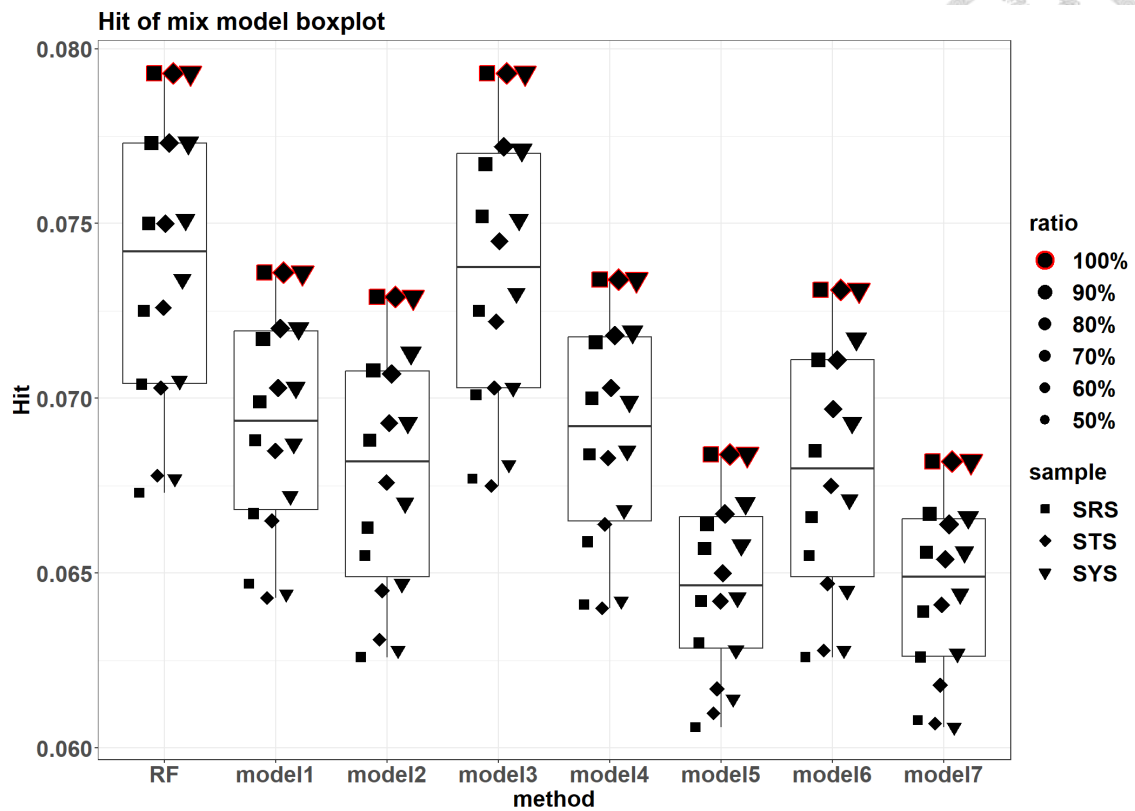


圖 28 混合模型三圖比較之 Hit 盒狀圖

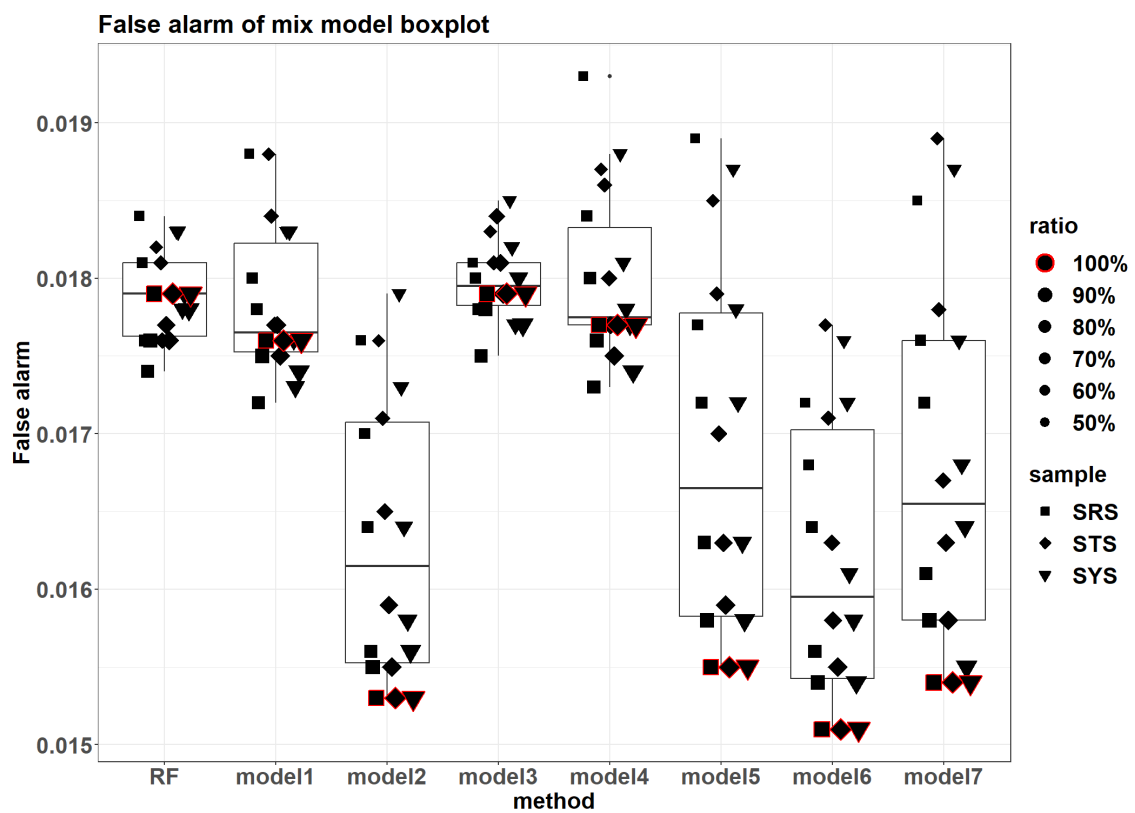


圖 29 混合模型三圖比較之 False Alarm 盒狀圖

由圖 29 混合模型三圖比較之 False Alarm 盒狀圖看出，混合模型之 False Alarm 盒狀圖分佈明顯呈現兩種不同分佈情形；RF、model1、model3、model4 呈現 False Alarm 數值分佈較為集中，且 100%抽樣之數值接近平均值之分佈；model2、model5、model6、model7 則呈現數值分佈較為分散，且 False Alarm 數值隨著抽樣比例增加而遞減之情形，並以 100%抽樣之 False Alarm 值為最小。

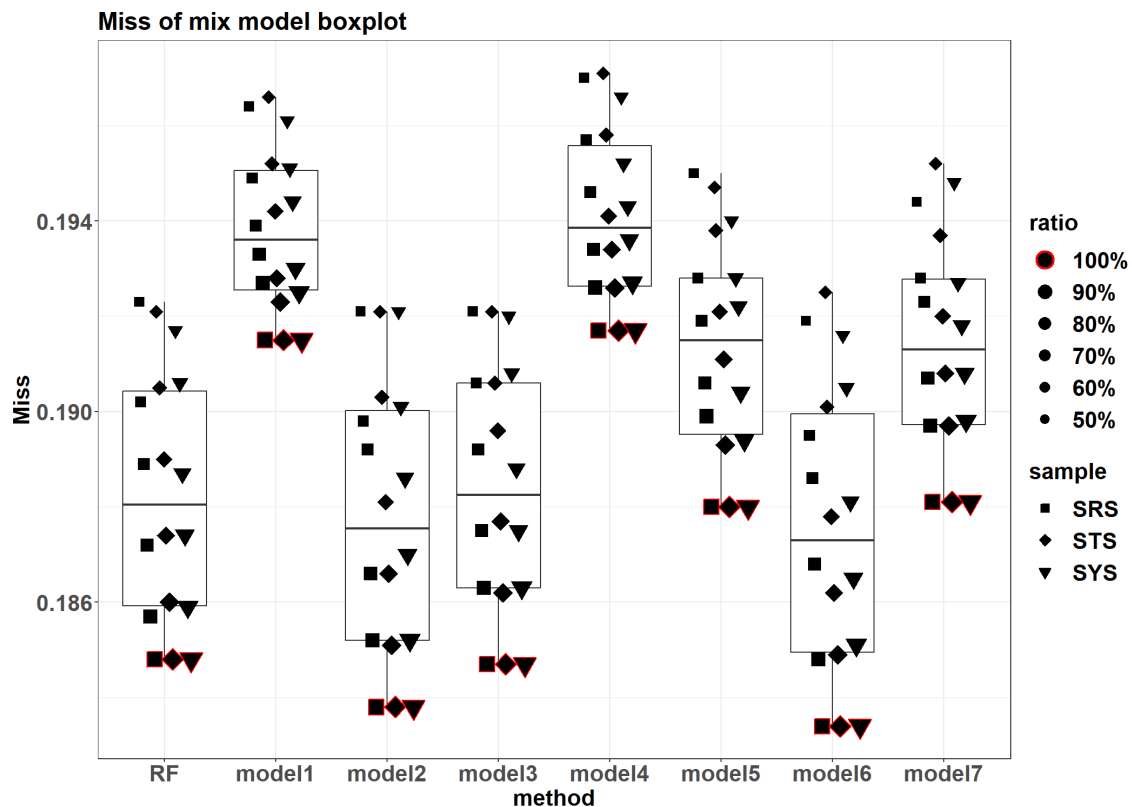


圖 30 混合模型三圖比較之 Miss 盒狀圖

由圖 30 混合模型三圖比較之 Miss 盒狀圖可知，混合模型之 Miss 數值皆呈現隨著抽樣比例增加而遞減之分佈，並以 model6 之 100%抽樣時之 Miss 數值為最小

由圖 31 混合模型三圖比較之 Wrong Hit 盒狀圖可知，混合模型之 Wrong Hit 的盒形圖分佈呈現兩種不同分佈情形；RF、model1、model3、model4 的 Wrong Hit 數值分佈較為分散，且數值隨著抽樣比例增加而遞減，並以 100%抽樣之數值為最小；model2、model5、model6、model7 則呈現數值分佈較為集中，且 Wrong Hit 數值隨著抽樣比例增加而遞減之情形較不明顯，且 100%抽樣數值較接近平均值。

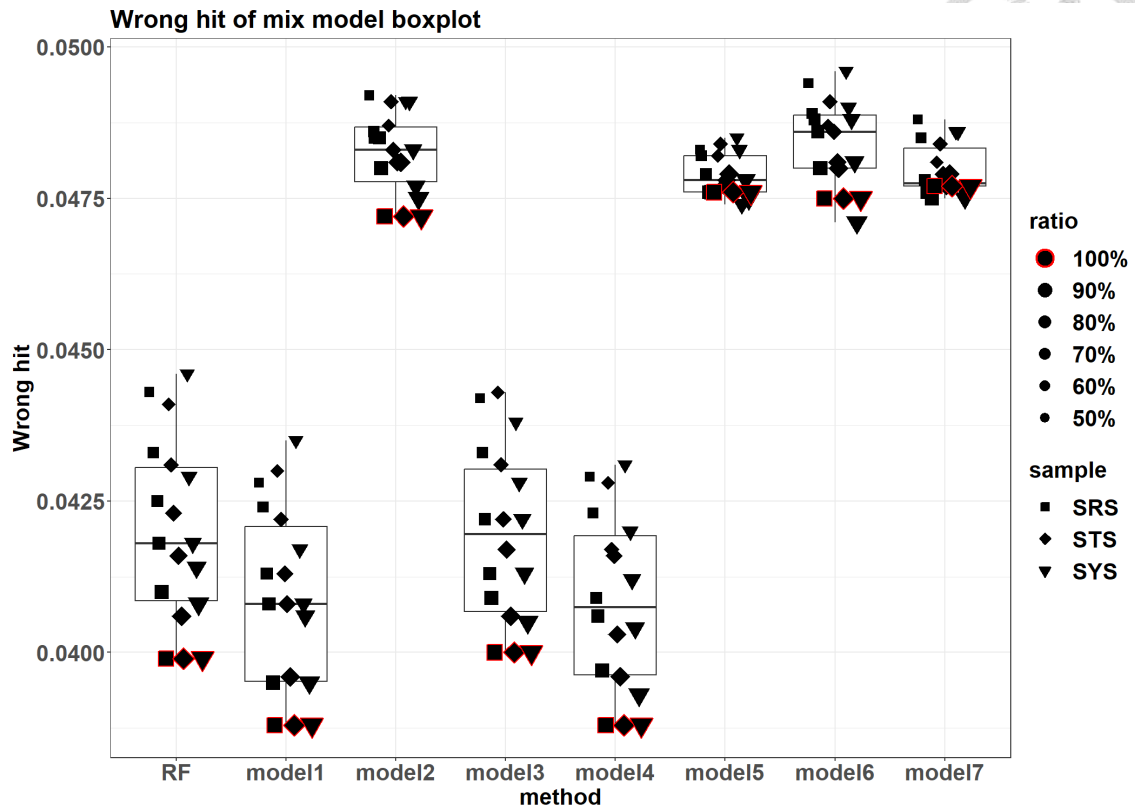


圖 31 混合模型三圖比較之 Wrong Hit 盒狀圖

表格 25 混合模型三圖比較與品質因素之 100%抽樣模型數值結果

排 序	模型 類別	抽樣 比例	Correct Rejection	False Alarm	Hit	Miss	Wrong Hit	品質 因素 FoM
1	RF	100%	0.6781	0.0179	0.0793	0.1848	0.0399	0.2463
1	model3	100%	0.6781	0.0179	0.0793	0.1847	0.04	0.2463
15	model6	100%	0.6809	0.0151	0.0731	0.1834	0.0475	0.2291
16	model1	100%	0.6785	0.0176	0.0736	0.1915	0.0388	0.2289
17	model2	100%	0.6808	0.0153	0.0729	0.1838	0.0472	0.2284
19	model4	100%	0.6784	0.0177	0.0734	0.1917	0.0388	0.2282
55	model5	100%	0.6805	0.0155	0.0684	0.188	0.0476	0.2141
58	model7	100%	0.6806	0.0154	0.0682	0.1881	0.0477	0.2135

各混合模型與隨機森林之三圖比較與品質因素結果如表格 25 所示。該表格為依照品質因素由大到小排序而成。其中，在所有混合模型中以隨機森林 100%抽樣及 model3 100%抽樣為最佳，完整替換之混合模型 model7 為最差。

表格 26 混合模型與隨機森林之三圖比較及品質因素差異比較



模型類別	替換土地利用	Correct Rejection (比例)	False Alarm (比例)	Hit (比例)	Miss (比例)	Wrong Hit (比例)	品質因素 FoM (比例)
model3	LU5	0 (0.00%)	0 (0.00%)	0 (0.00%)	-0.0001 (-0.05%)	0.0001 (0.25%)	0 (0.00%)
model6	LU5 +LU6	0.0028 (0.41%)	-0.0028 (-15.64%)	-0.0062 (-7.82%)	-0.0014 (-0.76%)	0.0076 (19.05%)	-0.0172 (-6.98%)
model1	LU2	0.0004 (0.06%)	-0.0003 (-1.68%)	-0.0057 (-7.19%)	0.0067 (3.63%)	-0.0011 (-2.76%)	-0.0174 (-7.06%)
model2	LU6	0.0027 (0.40%)	-0.0026 (-14.53%)	-0.0064 (-8.07%)	-0.001 (-0.54%)	0.0073 (18.30%)	-0.0179 (-7.27%)
model4	LU2 +LU5	0.0003 (0.04%)	-0.0002 (-1.12%)	-0.0059 (-7.44%)	0.0069 (3.73%)	-0.0011 (-2.76%)	-0.0181 (-7.35%)
model5	LU2 +LU6	0.0024 (0.35%)	-0.0024 (-13.41%)	-0.0109 (-13.75%)	0.0032 (1.73%)	0.0077 (19.30%)	-0.0322 (-13.07%)
model7	LU2 +LU5 +LU6	0.0025 (0.37%)	-0.0025 (-13.97%)	-0.0111 (-14.00%)	0.0033 (1.79%)	0.0078 (19.55%)	-0.0328 (-13.32%)

各混合模型與隨機森林之土地利用分歧則如表格 26 所示。依照數值差異由小到大排列依序為 model3、model6、model1、model2、model4、model5、model7，且從替換之土地利用類別中可以看出對於品質因素之影響由小到大依序為 Landuse 5:埤塘< Landuse 2:草地< Landuse 6:荒地。此結果與土地利用分歧所得之結果稍有出入，替換 Landuse 2:草地、Landuse 6:荒地的分類演算法對於土地利用變遷模式之量化數據之影響將於下一小節討論。

第二節、討論



本節為討論本研究之研究結果與目前已知相關文獻之關聯比較。以下將分為五小節分析本研究之抽樣方式對土地利用變遷模式的影響、土地利用資料穩定性及代表性、兩時期土地利用地圖驗證模式比較、混合模型之過度擬合及對土地利用結果之影響。

(一)抽樣方式對於土地利用變遷模式的影響

土地利用資料因具有時間及空間分布上的自相關性，因此過往研究為避免資料間的自相關性所造成的內生性誤差影響變數的檢定力及預測準確性，傾向透過空間抽樣的手段控制自相關性，其中又以空間系統抽樣最為普遍。空間系統抽樣為固定距離抽取樣本點之抽樣方式，普遍認為空間自相關性會隨距離增加而遞減，因此在相隔一定距離之樣本點可視為互相獨立，並符合一般線性回歸模型之假設。(Overmars et al., 2003)

本研究的研究結果，可依照與土地利用變遷模式分析階段大致分為三階段；變遷模式分析前、變遷模式分析以及變遷模式分析後。在變遷模式分析前，藉由比較抽樣方法差異與 Moran's I 數值變化趨勢，發現不同抽樣方式下的 Moran's I 數值並無顯著差異，可以推論抽樣方式對土地利用模型的影響較為有限。而在變遷模式分析階段，比較重複抽樣 AUC 數值的結果發現，抽樣方式對 AUC 平均值的影響依照分類演算法不同及土地利用類別不同而有差異，然而抽樣比例則普遍呈現抽樣比例較高的組別 AUC 平均值較高的趨勢。最後在變遷分析後階段中，透過比較土地利用分歧數值發現抽樣比例因子在三種分類演算法下皆為顯著，此情形在隨機森林的三圖比較數值中更為明顯，散佈圖呈現數值隨著抽樣比例增加而上升的趨勢，並以 100%(全樣本)的數值為最佳。



表格 27 土地利用空間資料代表性之土地利用分歧數值排行

邏輯斯迴歸(LR)				
排序	抽樣比例	正確配置(H)	量分歧(QD)	配置分歧(AD)
1	100%	0.7024	0.0217	0.2759
2	90%	0.7020	0.0227	0.2753
3	60%	0.7020	0.0223	0.2757
4	50%	0.7019	0.0226	0.2755
5	80%	0.7019	0.0226	0.2754
6	70%	0.7018	0.0226	0.2756
廣義可加性模式(GAM)				
排序	抽樣比例	正確配置(H)	量分歧(QD)	配置分歧(AD)
1	100%	0.7096	0.0225	0.2678
2	60%	0.7095	0.0227	0.2678
3	80%	0.7095	0.0226	0.2679
4	90%	0.7095	0.0226	0.2679
5	70%	0.7094	0.0226	0.2679
6	50%	0.7092	0.0226	0.2682
隨機森林(RF)				
排序	抽樣比例	正確配置(H)	量分歧(QD)	配置分歧(AD)
1	100%	0.7575	0.0216	0.2209
2	90%	0.7557	0.0217	0.2226
3	80%	0.7534	0.0216	0.2250
4	70%	0.7512	0.0215	0.2272
5	60%	0.7483	0.0216	0.2301
6	50%	0.7453	0.0215	0.2332

表格 28 土地利用空間資料代表性之三圖比較與品質因素數值排行

邏輯斯迴歸(LR)							
排序	抽樣比例	Correct Rejection	Hit	False Alarm	Miss	Wrong Hit	品質因素 FoM
1	100%	0.6626	0.0398	0.0334	0.2152	0.0489	0.1180
2	90%	0.6622	0.0398	0.0338	0.2149	0.0492	0.1179
3	50%	0.6621	0.0398	0.0339	0.2150	0.0491	0.1179
4	60%	0.6622	0.0398	0.0339	0.2152	0.0489	0.1178
5	70%	0.6621	0.0398	0.0340	0.2151	0.0490	0.1178
6	80%	0.6621	0.0398	0.0339	0.2150	0.0492	0.1177
廣義可加性模式(GAM)							
排序	抽樣比例	Correct Rejection	Hit	False Alarm	Miss	Wrong Hit	品質因素 FoM
1	100%	0.6659	0.0436	0.0301	0.2086	0.0517	0.1306
2	60%	0.6659	0.0436	0.0301	0.2086	0.0517	0.1305
3	90%	0.6660	0.0435	0.0301	0.2087	0.0517	0.1303
4	80%	0.6661	0.0435	0.0300	0.2087	0.0517	0.1303
5	70%	0.6659	0.0435	0.0301	0.2087	0.0518	0.1302
6	50%	0.6658	0.0434	0.0303	0.2090	0.0515	0.1299
隨機森林(RF)							
排序	抽樣比例	Correct Rejection	Hit	False Alarm	Miss	Wrong Hit	品質因素 FoM
1	100%	0.6781	0.0793	0.0179	0.1848	0.0399	0.2463
2	90%	0.6783	0.0773	0.0177	0.1859	0.0408	0.2403
3	80%	0.6784	0.0750	0.0177	0.1873	0.0416	0.2333
4	70%	0.6784	0.0728	0.0177	0.1889	0.0422	0.2265
5	60%	0.6779	0.0704	0.0182	0.1904	0.0431	0.2185
6	50%	0.6777	0.0676	0.0183	0.1920	0.0443	0.2098

抽樣方式及抽樣比例差異對土地利用變遷模式的影響依照分析階段與使用的分類演算法而有差異。以桃園地區的研究結果為例，土地利用分歧以及三圖比較的數值排序可整理為表格 27 及表格 28。由上述 2 表可得知抽樣方式及抽樣比例對邏輯斯迴歸模型與廣義可加性模式模型的土地利用變遷模式較無肉眼可見的影響。然而從隨機森林的數值可以知道，抽樣比例越高，品質因素則越高。



Chakir and Le Gallo (2021)認為透過空間抽樣方式控制空間自相關性(*autocorrelation*)為無效手段，且無法量化空間自相關性的影響。此一觀點與本研究之結果吻合；由 Moran's I 的結果可知，空間抽樣可有效降低空間自相關性，但抽樣方式差異影響不顯著；再由重複抽樣 AUC 之結果，最佳抽樣方式及抽樣比例因土地利用分布型態及分類演算法而有所差異，因此不存在可廣泛適用之抽樣架構；同時由土地利用配置分歧及品質因素的結果可知，抽樣方式的影響不顯著，且在三種演算法下 100%抽樣時的土地利用變遷數值最佳，則可推論若透過抽樣方式降低未空間自相關性可能造成的內生性誤差，反而更容易損失其他未被抽樣的網格特徵，進而使得土地利用變遷模型擬合程度下降。

另外，由土地利用變遷模式結果如表格 27 土地利用空間資料代表性之土地利用分歧數值排行、表格 28 土地利用空間資料代表性之三圖比較與品質因素數值排行，可知無論演算法為何皆呈現當 100%抽樣時，為該演算法下土地利用配置正確配置最高，分歧最小的結果。由此可推論在本研究桃園地區的土地利用變遷分析，100%抽樣為最能代表該研究區整體土地利用資訊的最小代表性樣本大小。

Dörfler and Stierand (2019)中定義資料代表性為包含所有母體顯著特徵及統計頻率。由抽樣方式對於土地利用變遷模式之討論可知，空間抽樣會使得樣本失去一些空間特徵，並使得模型分數下降。由於土地利用資料之空間抽樣牽涉時間及空間因素，缺少可廣泛適用之不偏估計量(Strand, 2017)，故無法與一般抽樣使用樣本平均及變異數作為檢視代表性之標準。因此，本研究就重複抽樣 AUC 及土地利用變遷研究結果廣泛討論，在邏輯斯迴歸、廣義可加性模式與隨機森林等三種分類演算法中，100%抽樣之 AUC 平均值、土地利用分歧及品質因素數值皆為最佳，同時誤差最小，因此我們推論在桃園地區的土地利用變遷研究中 100%抽樣為最具有代表性之樣本。



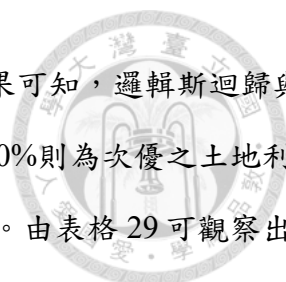
(二)兩時期土地利用地圖驗證模式比較

部分土地利用研究礙於土地利用圖資較難取得及土地利用變遷緩慢等原因，研究中僅使用 1、2 個時期之土地利用圖資建立變遷模型並預測未來土地利用，缺乏穩定的模型驗證程序可能使得預測結果與實際土地利用出入較大 Ahmed et al. (2014)

藉由本研究的成果，可透過空間抽樣切割時期一之土地利用圖資建立訓練資料集(training dataset)與校正資料集(calibration dataset)再以時期二之土地利用圖資作為驗證資料集(validation dataset)便可在僅有的土地利用圖資下完成模型驗證，以確保預測結果與實際土地利用相符。

表格 29 1995 至 2007 模型驗證數值比較

土地利用分歧(Disagreement)								
土地利用時期	演算法	抽樣比例	正確配置(H)		量分歧(QD)		配置分歧(AD)	
2007	GAM	60%	0.6979		0.0750		0.2272	
	LR	90%	0.6930		0.0750		0.2321	
	RF	90%	0.6898		0.0749		0.2352	
	GAM	100%	0.6986		0.0750		0.2265	
	LR	100%	0.6940		0.0750		0.2310	
	RF	100%	0.6891		0.0750		0.2360	
品質因素與三圖比較(FoM and Three map)								
驗證時期	演算法	抽樣比例	Correct Rejection	Hit	False Alarm	Miss	Wrong Hit	品質因素 FoM
1995 與 2007	GAM	60%	0.6587	0.0392	0.0288	0.2411	0.0322	0.1149
	LR	90%	0.6558	0.0372	0.0317	0.244	0.0314	0.1081
	RF	90%	0.6814	0.0083	0.006	0.2945	0.0097	0.0261
	GAM	100%	0.6623	0.0363	0.0251	0.2457	0.0305	0.1204
	LR	100%	0.6594	0.0346	0.0280	0.2484	0.0296	0.1016
	RF	100%	0.6828	0.0063	0.0046	0.2973	0.0090	0.0197
真實土地利用 變化比例			0.6875	0.3125				



以本研究之桃園地區為例，由表格 27 土地利用空間資料代表性的結果可知，邏輯斯迴歸與隨機森林中 90%之抽樣比例為次優之模型，廣義線性可加式模式中 60%則為次優之土地利用變遷模型。以上三個模型的 2007 年之驗證結果則可概括為表格 29。由表格 29 可觀察出透過空間抽樣所建立之土地利用變遷模式再以時期二的土地利用驗證後的數值與 100%抽樣的正確配置誤差約為-0.11%~0.15%，量分歧數值誤差為-0.01~0.02，配置分歧誤差為-0.45%~0.31%。整體而言抽樣模型與 100%模型的土地利用分歧數據差異甚小，然而從三圖比較中可以發現，空間抽樣模型與 100%模型 Hit 數值差異為-24.1%~-6.99%，False Alarm 數值差異則介於-23.33%~-11.67%，相較於土地利用分歧數值，三圖比較數值顯示出空間抽樣模型較能正確反映出土地利用的轉變行為。

由此可知以空間抽樣分割時期一(1995)土地利用資料為訓練組與校正組的土地利用模式可以在有限的資料下較穩定輸出土地利用預測結果。但其中以隨機森林為分類演算法之土地利用變遷模型的數值則明顯差於時期二(2007)至時期三(2014)的驗證結果(見表格 28 土地利用空間資料代表性之三圖比較與品質因素數值排行)。綜合上述結果可知，如僅以兩時期之土地利用圖資建立之土地利用變遷模式時以線性模型作為分類演算法較能完整描述土地利用的變化情形，也不易產生過大之量分歧與 Miss 誤差。



(三)混合模型之過度擬合

由混合模型 Mix Model 的研究結果可以知道完整替換土地利用模式之混合模型 model7 在各項數據中皆為表現最差的模型。由此可以討論混合演算法之土地利用變遷模式的過度擬合情形。過度擬合(overfitting)可被定義為複雜度高且變異性高的模型(Ghojogh & Crowley, 2019)。由於 CLUE-s 模式為結合土地利用事宜度(suitability map)、土地利用轉換彈性矩陣及空間限制進行土地利用配置之變遷模式，無法衡量變異性，因此本研究僅比較混合模型與單一分類演算法之 AUC 數值與品質因素數值差異。

表格 30 土地利用變遷模式過度擬合程度

AUC				FoM			
土地利用	RF 100%	model 7 100%	比例	土地利用	RF 100%	model 7 100%	比例
0:農地	0.8641	0.8643	0.03%	0:農地	0.2864	0.2647	-7.6%
1:林地	0.9720	0.9720	0.00%	1:林地	0.1081	0.0434	-59.9%
2:草地	0.7904	0.7976	0.91%	2:草地	0.4527	0.3815	-15.7%
3:建地	0.8623	0.8623	0.01%	3:建地	0.0000	0.0000	0.0%
4:水體	0.9968	0.9969	0.01%	4:水體	NA	NA	NA
5:埤塘	0.7095	0.7238	2.02%	5:埤塘	0.2382	0.1399	-41.2%
6:荒地	0.6756	0.7158	5.95%	6:荒地	0.6093	0.5576	-8.5%

為 model 7 與隨機森林(RF)模型的各土地利用類別之 AUC 數值及 FoM 數值比較表。其中，因為本研究將水體設為限制土地利用轉變區域，並無透過土地利用變遷模式配置土地利用網格，因此 FoM 欄位數值皆為 NA。由於 model 7 為結合各土地利用類別最佳的演算法的模型，因此可明顯看出其 AUC 數值皆較 RF 模型高。然而，結合各土地利用最佳演算法的 model 7 之品質因素(FoM)在所有土地利用類別中皆明顯較 RF 模型差。其中 FoM 數值下降比例由大到小依序為，1:林地 59.9%>5:埤塘 41.2%>2:草地 15.7%>6:荒地 8.5%>0:農地 7.6%。由此結果可知，結合各土地利用類別之最佳演算法的混合模型並未能提升該土地利用類別的配置準確度，反而使得其他土地利用類別的配置結果下降，因此可證明混合模型存在過度擬合的風險。

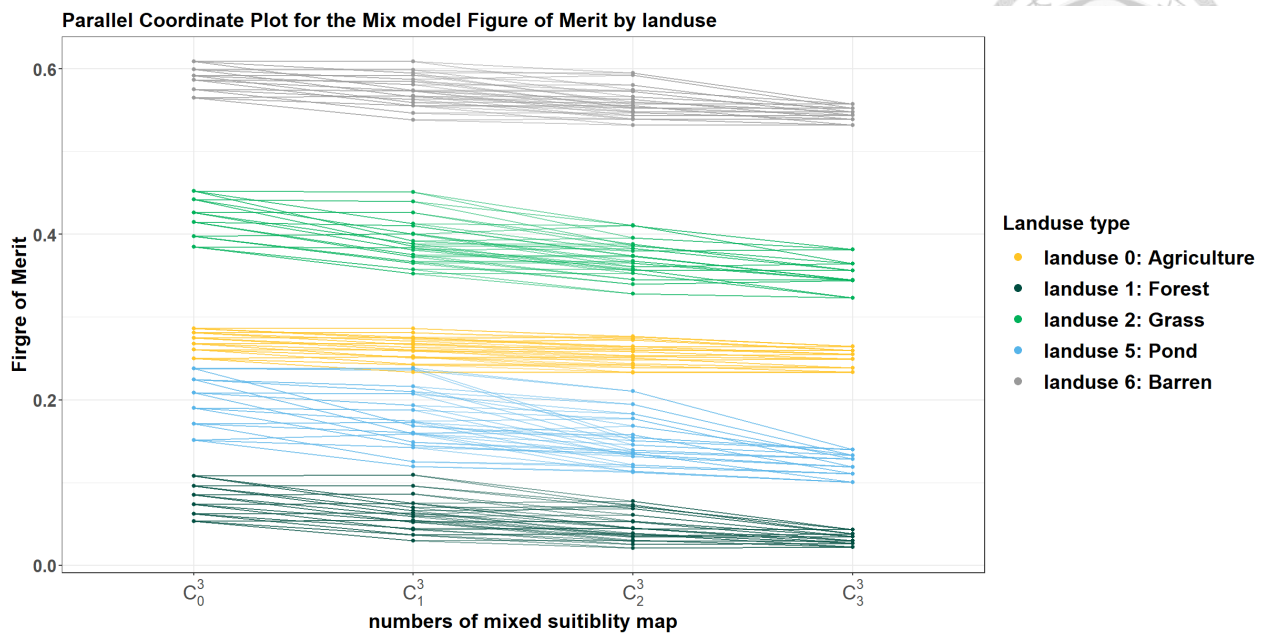


圖 32 各土地利用品質因素變化平行座標圖

圖 32 為各土地利用類別之品質因素在不同混合模型下的變化圖。縱軸為品質因素，橫軸為置換演算法之各數， C_0^3 為 landuse 2, landuse 5, landuse 6 皆未被置換的情境，在本研究中為隨機森林模型； C_1^3 為 landuse 2, landuse 5, landuse 6 其一被置換的情境，於本文中為 model1, model2, model3； C_2^3 為其二被置換的情境，於本文中為 model4, model5, model6； C_3^3 則為全部皆被置換之混合模型，在本研究中為 model7。由圖 32 可知所有土地利用的 FoM 數值皆明顯隨著替換的演算法增加而下降，其中下降幅度以林地、草地、埤塘較為顯著。針對土地利用適用其最佳演算法的混合模型有過度擬合之風險，且過度擬合程度隨著混合演算法個數增加而上升。下一節則將討論以本研究桃園地區土地利用變遷為例，替換演算法對於土地利用配置結果的影響。



(四)替換演算法之土地利用配置結果之影響

相較於其他土地利用變遷模式，CLUE-s 模式為透過土地利用間的競爭同時模擬多種土地利用類別配置之變遷模式(Verburg et al., 2002)。如此的特徵在混合模型中更能明顯看出，表格 31 為土地利用水體、埤塘、荒地的土地利用適宜度之描述統計量，由此表可知相較於隨機森林(RF)的轉變機率，邏輯斯迴歸(LR)、廣義可加性模式(GAM)下的轉變機率較低，如此的差異可能造成土地利用配置階段中，採用邏輯斯迴歸、廣義可加性模式的土地利用類別被分配到的網格較以隨機森林的網格少。以下將分別討論演算法差異對於土地利用分歧與三圖比較及品質因素數值之影響。

表格 31 水體、埤塘、荒地土地利用適宜度描述統計量

landuse2: 水體						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
LR	0.0002	0.0082	0.0148	0.0213	0.0253	0.6415
GAM	0.0000	0.0058	0.0116	0.0199	0.0225	0.4151
RF	0.0000	0.0020	0.0096	0.0324	0.0247	0.9793
landuse5: 埤塘						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
LR	0.0000	0.0000	0.0025	0.0095	0.0120	0.1918
GAM	0.0000	0.0001	0.0031	0.0099	0.0074	0.3143
RF	0.0000	0.0020	0.0055	0.0189	0.0131	0.8380
landuse6: 荒地						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
LR	0.0004	0.0179	0.0464	0.0463	0.0656	0.4322
GAM	0.0004	0.0133	0.0477	0.0453	0.0665	0.6076
RF	0.0000	0.0074	0.0350	0.0694	0.0752	0.9805



變更演算法之土地利用類別對土地利用變遷量化數值之影響，評估方法為，將 144 組混合模型，在固定相同抽樣方式及抽樣比例之條件下，依照模型組合差異，將兩模型之土地利用分歧數值、三圖比較及品質因素數值相減，相減之後的數值即為變更某一土地利用演算法所致之差異，模型配對如表格 32 所示。

表格 32 混合模型差異驗證評估組合

土地 利用	Landuse 2 草地	Landuse 5 埤塘	Landuse 6 荒地
配對 組合	model1-RF	model3-RF	model2-RF
	model4-model3	model4-model1	model5-model1
	model5-model2	model6-model2	model6-model3
	model7-model6	model7-model5	model7-model4



1. 替換演算法之土地利用類別對於土地利用分歧的影響

圖 33 為變更土地利用分類演算法後，影響土地利用分歧的差異值之盒狀圖。橫軸為土地利用分歧類別，縱軸為差異數值；其中，綠色為 Landuse 2:草地、藍色為 Landuse 5:埤塘、灰色為 Landuse 6:荒地。由圖 33 及表格 33 得知，影響土地利用分歧數值的土地利用類別由小到大依序為 Landuse 5:埤塘< Landuse 6:荒地< Landuse 2:草地。平均而言，將 Landuse 2 的演算法由 RF 置換 GAM，正確配置數值下降 0.0042，配置分歧數值增加 0.0039；將 Landuse 6 的演算法由 RF 置換 LR，正確配置數值下降 0.0036，配置分歧數值增加 0.0031；將 Landuse 5 的演算法由 RF 置換 GAM，正確配置數值下降 0.0001，配置分歧數值下降 0.0001。此外，對於量分歧數值的影響三者差異較不明顯，置換 Landuse 2 後增加量分歧平均增加 0.0003；置換 Landuse 5 後增加量分歧平均增加 0.0002；置換 Landuse 5 後增加量分歧則平均增加 0.0005。由此結果可知，結合各土地利用類別之最佳分類演算法的混合模型會使得正確配置數值下降，配置分歧及量分歧上升。整體而言，混合模型無法增加土地利用變遷模式之預測準確度，為過度擬和之模型組合。

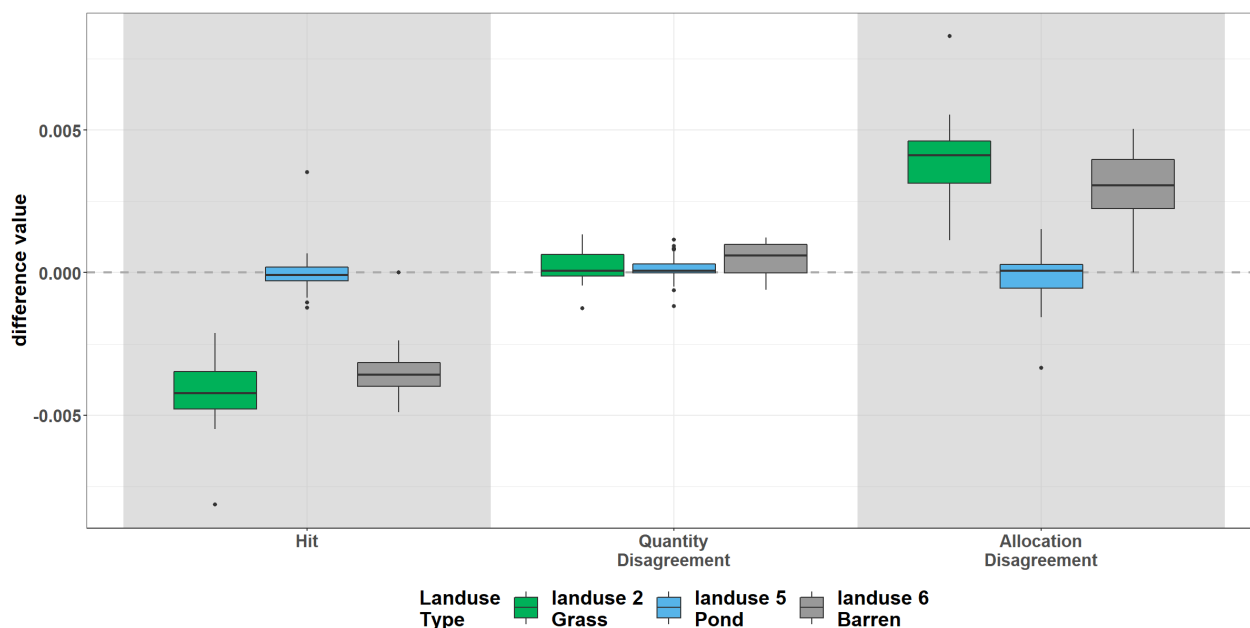


圖 33 變更土地利用之土地利用分歧差異盒狀圖

表格 33 變更土地利用之土地利用分歧差異



土地利用	統計量	正確配置 Hit	量分歧 QD	配置分歧 AD
Landuse 2 草地	最小值	-0.0081	-0.0013	0.0011
	最大值	-0.0021	0.0013	0.0083
	平均值	-0.0042	0.0003	0.0039
	標準差	0.0009	0.0005	0.0012
Landuse 5 埤塘	最小值	-0.0012	-0.0012	-0.0033
	最大值	0.0035	0.0012	0.0015
	平均值	-0.0001	0.0002	-0.0001
	標準差	0.0006	0.0004	0.0008
Landuse 6 荒地	最小值	-0.0049	-0.0006	0.0000
	最大值	0.0000	0.0012	0.0050
	平均值	-0.0036	0.0005	0.0031
	標準差	0.0007	0.0005	0.0010



2. 替換演算法之土地利用類別對於三圖比較的影響

圖 34 為變更土地利用分類演算法後，影響三圖比較的差異值之盒狀圖。橫軸為土地利用分歧類別，縱軸為差異數值；其中，綠色為 Landuse 2:草地、藍色為 Landuse 5:埤塘、灰色為 Landuse 6:荒地。

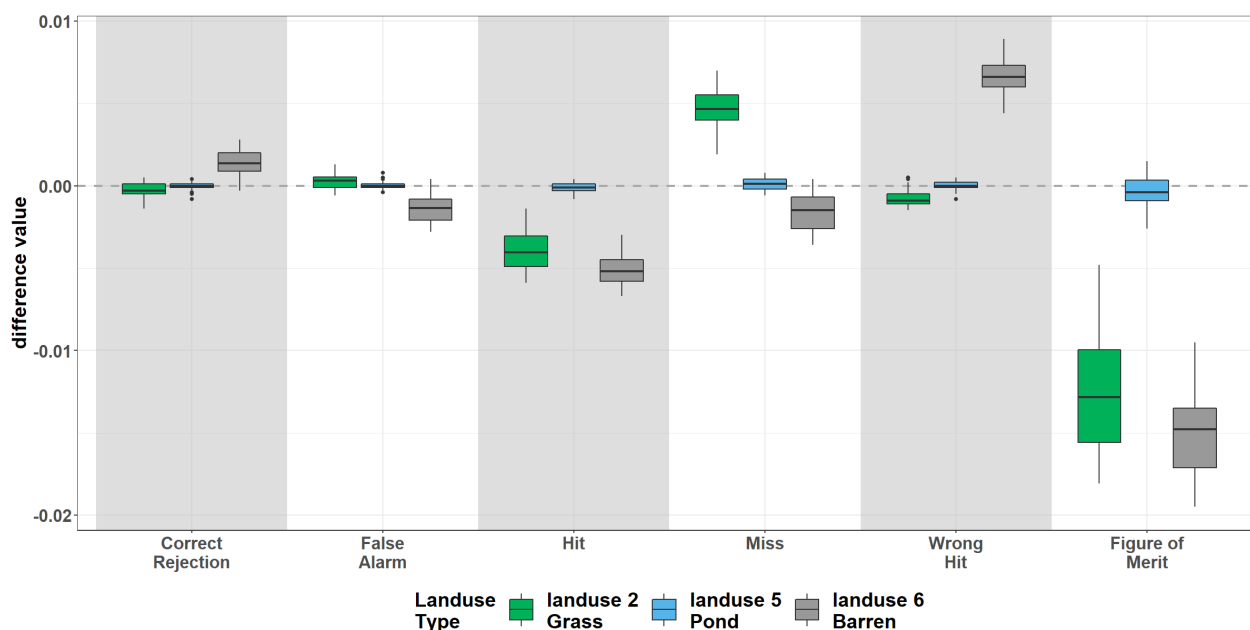
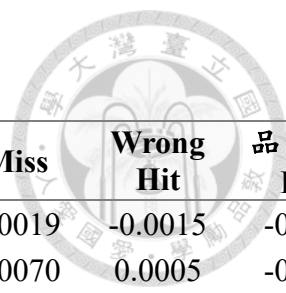


圖 34 變更土地利用之三圖比較差異盒狀圖

由圖 34 及表格 34 得知，影響品質因素的土地利用類別由小到大依序為 Landuse 5:埤塘< Landuse 2:草地< Landuse 6:荒地。平均而言，將 Landuse6 的演算法由 RF 置換為 LR，品質因素下降 0.0149；將 Landuse2 的演算法由 RF 置換為 GAM，品質因素下降 0.0125；將 Landuse5 的演算法由 RF 置換為 GAM，品質因素下降 0.0003。

另從圖 34 可知，變換 landuse 2 的分類演算法後，使得 Miss 數值增加幅度明顯大於其他數值。由此可以觀察出，變換分類方法後，使得混合土地利用變遷模式較無法辨別出實際有改變土地利用的網格；此外變換 landuse 6 的分類演算法後，使得 Wrong Hit 數值增加幅度明顯大於其他數值。由此可以觀察出，變換分類方法後，使得混合土地利用變遷模式雖已掌握實際有改變土地利用的網格，但無法配置為正確的土地利用類別。

表格 34 變更土地利用之三圖比較及品質因素差異



土地利用	統計量	Correct Rejection	Hit	False Alarm	Miss	Wrong Hit	品質因素 FoM
Landuse2 草地	Min	-0.0014	-0.0059	-0.0006	0.0019	-0.0015	-0.0181
	Max	0.0005	-0.0014	0.0013	0.0070	0.0005	-0.0048
	Mean	-0.0003	-0.0039	0.0003	0.0047	-0.0007	-0.0125
	SD	0.0005	0.0012	0.0005	0.0013	0.0005	0.0034
Landues5 埤塘	Min	-0.0008	-0.0008	-0.0004	-0.0006	-0.0008	-0.0026
	Max	0.0004	0.0004	0.0008	0.0008	0.0005	0.0015
	Mean	-2.0833E-05	-0.0001	4.1667E-06	0.0001	0.0000	-0.0003
	SD	0.0002	0.0003	0.0002	0.0004	0.0003	0.0009
Landuse6 荒地	Min	-0.0003	-0.0067	-0.0028	-0.0036	0.0044	-0.0195
	Max	0.0028	-0.0030	0.0004	0.0004	0.0089	-0.0095
	Mean	0.0014	-0.0051	-0.0014	-0.0016	0.0067	-0.0149
	SD	0.0008	0.0009	0.0008	0.0011	0.0011	0.0024

伍、結論與建議

土地利用變遷為長久以來人類行為與自然環境及生態系統間互動而留下的軌跡，其變遷過程為一連串複雜且不可逆的程序。土地利用變遷研究致力於尋找變遷過程中的驅動力及可能轉變模式，並預測未來的變動趨勢。隨著資料科學的演進，土地利用變遷研究的涵蓋範圍更加廣泛且細緻。本研究成功以桃園地區為研究區域討論抽樣模式及抽樣比例差異對於土地利用變遷模式之影響；並進一步討論結合各土地利用類別最佳之二元分類演算法之混合模型是否過度擬合。

第一節、結論

1.抽樣方式降低空間自相關性誤差

本研究使用桃園地區 1995 年及 2007 年之歷史土地變化利用作為土地利用變遷模式校準區間，再以 2007 年及 2014 年之歷史土地利用變化做為模型驗證區間，發現雖然土地利用資料間有高度自相關性，但抽樣方式並不能有效降低因空間自相關性而產生的內生性誤差，反而容易因為損失其餘空間特徵而使得模型驗證分數下降。Chakir and Le Gallo (2021)建議以空間計量模型取代一般計量模型，透過增加空間自相關變量的誤差項，降低因空間自相關所產生的內生性誤差。Ay et al. (2017)則提出空間自相關性反而可以提升模型準確度，可透過增加空間權重矩陣(spatial weight matrix)考慮空間自相關性的影響，提升總和模型(aggregate model)的精確度。

2.抽樣模式與抽樣比例對於 CLUE-s 模式之影響

長期以來土地利用研究皆認為土地利用資料具有空間自相關性，傾向透過抽樣方法等手段以避免資料間的自相關性造成分析偏誤。本研究以桃園地區為研究區，應用重複抽樣取後放回的手法針對土地利用資料重覆抽樣並建立 TOC 曲線下面積 AUC 信賴區間、多因子變異數分析與事後檢定，發現最佳抽樣方式及抽樣比例因土地利用分布型態及分類演算法而有所差異，不存在廣泛適用之抽樣架構。

抽樣比例之差異則顯著影響相同二元分類演算法下之土地利用變遷模式表現，且不論邏輯斯迴歸(LR)、廣義可加性模式(GAM)及隨機森林(RF)皆呈現當土地利用資料為 100%抽樣時，其土地利用變遷結果與真實土地利用配置誤差最小，並顯示出在桃園地區的土地變遷研究中 100%抽樣最具有代表性之空間樣本大小。然而，若是考慮模型演算效能或是其他原因無法取 100%樣本的話，則可參考本研究之結果適用次佳的抽樣比例，以桃園地區為例，邏輯斯迴歸建議以 90%抽樣比例，廣義可加性模式為 60%抽樣比例，隨機森林則以 90%抽樣比例為有效率之網格抽樣比例。

3.兩時期土地利用地圖變遷模式驗證

有鑑於部分土地利用變遷研究中僅使用 1~2 時期之圖資作為原始資料，本研究提出藉由空間抽樣分割時期一土地利用資料為訓練組與校正組以建立土地利用變遷模型，再以時期二之土地利用圖資進行土地利用變遷模式驗證，以確保預測結果符合實際土地利用變遷趨勢。以桃園地區為例，1995 到 2007 的三圖比較數值中，邏輯斯迴歸(LR)、廣義可加性模式(GAM)及隨機森林(RF)等三種演算法下抽樣模型的 Hit 數值接明顯較 100%的數值高，差異幅度為 24.1%~6.99%，顯示在僅有 2 時期土地利用地圖情況下，使用空間抽樣方式建立土地利用變遷模式較能預測土地利用轉變行為。

4.混合模型的過度擬合程度

在本研究中亦嘗試結合各類土地利用之最佳分類演算法之土地利用適宜度進行土地利用配置。由研究結果可知，混合模式會降低土地利用變遷模式的配置正確度，且隨著混合的替換程度上升而下降。CLUE-s 模式相較於其他土地利用變遷模式，透過土地利用間相互競爭的關係進行空間配置。在本研究中，替換演算法之後，各類土地利用之 AUC 數值增加 0.00%~5.95%，但品質因素(FoM)下降 8.5%~59.9%。清楚呈現混合模型並無法提升對應土地利用之正確配置比例，亦連帶影響其他未變換演算法之土地利用的配置結果。



第二節、後續應用與研究建議

本研究在抽樣方式對土地利用變遷模式的影響為初步探討，對於爾後抽樣方式、抽樣比例於土地利用變遷模式之研究，以下提出後續研究建議以供參考：

1.其他土地利用變遷模式

從本研究結果可以知道土地利用變遷模式 CLUE-s 模式下抽樣方式及抽樣比例對模型的影響依照土地利用類別及演算法而有差異。後續研究上或許可以考慮應用本研究之研究架構於其他土地利用變遷模式，可更加深入討論土地利用變遷模式與抽樣方式間的關係及可能的影響程度範圍。

2.其他機器學習演算法

本研究中僅比較邏輯斯迴歸(LR)、廣義可加性模式(GAM)及隨機森林(RF)等三種二元演算法之結果，後續研究可嘗試比較其他機器學習演算法，如人工類神經網路(ANN)、支撐向量機(SVM)等，或許可以增加土地利用變遷模型的配置結果及準確性。且本研究混合模式為結合邏輯斯迴歸、廣義可加性模式及隨機森林，由於邏輯斯迴歸、廣義可加性模式所計算出之土地利用適宜度的機率較小，故與隨機森林所計算出的土地利用適宜度結合為混合模式時易產生較大之過度擬合情形，故若結合同為機器學習之二元分類演算法的土地利用適宜度，再進行土地利用配置，或許可降低過度擬合之程度。

3.空間計量模型

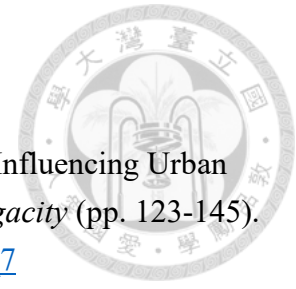
空間自相關性為空間資料研究中無法避免之潛在誤差來源。本研究發現抽樣手段無法有效降低因空間自相關性所產生的內生性誤差，反而存在損失未被抽樣空間網格特徵之風險。因此後續研究上可以使用空間計量模型代替一般線性模型及邏輯斯迴歸模型，較能有效控制因空間自相關性而產生的內生性誤差。



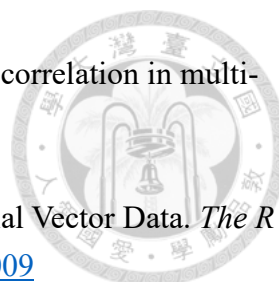
4.其他地區土地利用資料

本研究為以桃園地區 1995 年、2007 年、2014 年之土地利用情形為研究資料，而得出 100% 抽樣之樣本大小為該區域最具代表性之土地利用樣本。但目前仍不清楚空間自相關性對於土地利用變遷模式之確切影響模式及程度，無法斷言是否在所有區域中 100%抽樣皆為最具代表性之抽樣大小。後續研究可應用本研究之方法於其他地區的土地利用資料，或許可以更加深入討論空間自相關性對於土地利用變遷模式之影響模式及程度。

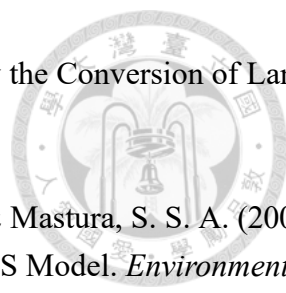
參考文獻



- Ahmed, S. J., Bramley, G., & Verburg, P. H. (2014). Key Driving Factors Influencing Urban Growth: Spatial-Statistical Modelling with CLUE-s. In *Dhaka Megacity* (pp. 123-145). Springer Netherlands. https://doi.org/10.1007/978-94-007-6735-5_7
- Ay, J.-S., Chakir, R., & Gallo, J. L. (2017). Aggregated Versus Individual Land-Use Models: Modeling Spatial Autocorrelation to Increase Predictive Accuracy. *Environmental Modeling & Assessment*, 22(2), 129-145. <https://doi.org/10.1007/s10666-016-9523-5>
- Chakir, R., & Le Gallo, J. (2021). Spatial Autocorrelation in Econometric Land Use Models: An Overview. In A. Daouia & A. Ruiz-Gazen (Eds.), *Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan* (pp. 339-362). Springer International Publishing. https://doi.org/10.1007/978-3-030-73249-3_18
- Cheng, J., & Masser, I. (2003). Modelling Urban Growth Patterns: A Multiscale Perspective. *Environment and Planning A: Economy and Space*, 35(4), 679-704. <https://doi.org/10.1068/a35118>
- Dörfler, V., & Stierand, M. (2019). Extraordinary: Reflections on Sample Representativeness. In I. Lebuda & V. P. Glăveanu (Eds.), *The Palgrave Handbook of Social Creativity Research* (pp. 569-584). Springer International Publishing. https://doi.org/10.1007/978-3-319-95498-1_36
- Eastman, J. R. (2020). TerrSet 2020 Geospatial Monitoring and Modeling Software Manual. In C. U. Clark Labs (Ed.). Clark Labs, Clark University.
- Fitzpatrick-Lins, K. (1981). Comparison of Sampling Procedures and Data Analysis for a Land-Use and Land-Cover Map. *Photogrammetric Engineering and Remote Sensing*, 47(3), 343-351.
- Ghojogh, B., & Crowley, M. (2019). The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial.
- Li, Y., & Huang, S. (2015). Landscape Ecological Risk Responses to Land Use Change in the Luanhe River Basin, China. *Sustainability*, 7(12), 16631-16652. <https://doi.org/10.3390/su71215835>
- Lin, Y.-P., Chu, H.-J., Wu, C.-F., & Verburg, P. H. (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling – a case study. *International Journal of Geographical Information Science*, 25(1), 65-87. <https://doi.org/10.1080/13658811003752332>
- Mazloum, B., Pourmanafi, S., Soffianian, A., Salmanmahiny, A., & Prishchepov, A. V. (2021). The fate of rangelands: Revealing past and predicting future land-cover transitions from 1985 to 2036 in the drylands of Central Iran [<https://doi.org/10.1002/ldr.3865>]. *Land Degradation & Development*, 32(14), 4004-4017. <https://doi.org/https://doi.org/10.1002/ldr.3865>



- Overmars, K. P., De Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164(2-3), 257-270.
[https://doi.org/10.1016/s0304-3800\(03\)00070-x](https://doi.org/10.1016/s0304-3800(03)00070-x)
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10, 439-446, Article 1. <https://doi.org/10.32614/RJ-2018-009>
- Peng, K., Jiang, W., & Deng, Y. (2021, 3-5 Nov. 2021). Simulating urban land-use changes by incorporating logistic regression and CLUE-S model: a case study of Wuhan city. 2021 28th International Conference on Geoinformatics,
- Peng, K., Jiang, W., Deng, Y., Liu, Y., Wu, Z., & Chen, Z. (2020). Simulating wetland changes under different scenarios based on integrating the random forest and CLUE-S models: A case study of Wuhan Urban Agglomeration. *Ecological Indicators*, 117, 106671.
<https://doi.org/10.1016/j.ecolind.2020.106671>
- Pontius Jr, R. G., & Si, K. (2014). The total operating characteristic to measure diagnostic ability for multiple thresholds. *International Journal of Geographical Information Science*, 28(3), 570-583. <https://doi.org/10.1080/13658816.2013.862623>
- Pontius, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407-4429. <https://doi.org/10.1080/01431161.2011.552923>
- Pontius, R. G., Peethambaram, S., & Castella, J.-C. (2011). Comparison of Three Maps at Multiple Resolutions: A Case Study of Land Change Simulation in Cho Don District, Vietnam. *Annals of the Association of American Geographers*, 101(1), 45-62.
<https://doi.org/10.1080/00045608.2010.517742>
- Puertas, O. L., Henríquez, C., & Meza, F. J. (2014). Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010–2045. *Land Use Policy*, 38, 415-425.
<https://doi.org/https://doi.org/10.1016/j.landusepol.2013.11.024>
- Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., & Taleai, M. (2017). Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64, 297-308.
<https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2017.04.002>
- Strand, G.-H. (2017). A study of variance estimation methods for systematic spatial sampling. *Spatial Statistics*, 21, 226-240. <https://doi.org/10.1016/j.spasta.2017.06.008>
- Turner, R., & Baddeley, A. (2005). SPATSTAT: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12. <https://doi.org/10.18637/jss.v012.i06>
- Varga, O. G., Pontius, R. G., Singh, S. K., & Szabó, S. (2019). Intensity Analysis and the Figure of Merit's components for assessment of a Cellular Automata – Markov simulation model. *Ecological Indicators*, 101, 933-942. <https://doi.org/10.1016/j.ecolind.2019.01.057>

- 
- Veldkamp, A., & Fresco, L. O. (1996). CLUE: a conceptual model to study the Conversion of Land Use and its Effects. *Ecological Modelling*, 85(2), 253-270.
[https://doi.org/10.1016/0304-3800\(94\)00151-0](https://doi.org/10.1016/0304-3800(94)00151-0)
- Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002). Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. *Environmental Management*, 30(3), 391-405. <https://doi.org/10.1007/s00267-002-2630-x>
- Wagner, P. D., & Waske, B. (2016). Importance of spatially distributed hydrologic variables for land use change modeling. *Environmental Modelling & Software*, 83, 245-254.
<https://doi.org/10.1016/j.envsoft.2016.06.005>
- Waiyasusri, K., Yumuang, S., & Chotpantarat, S. (2016). Monitoring and predicting land use changes in the Huai Thap Salao Watershed area, Uthaitani Province, Thailand, using the CLUE-s model. *Environmental Earth Sciences*, 75(6), 533. <https://doi.org/10.1007/s12665-016-5322-1>
- Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1 - 17.
<https://doi.org/10.18637/jss.v077.i01>
- 內政部國土測繪中心. (2000-2020). 國土利用現況調查.
<https://www.nlsc.gov.tw/cl.aspx?n=13705>