

研究報告

評估兩種取樣設計對於香桂適生育地模式 預測能力之影響

羅南璋¹ 張偉顛² 黃凱易³

【摘要】香桂 (*Cinnamomum subavenium* Miq.) 為常綠闊葉樹種，廣泛分布於台灣中、南部山區。本研究矩形試區位於台灣中部，涵蓋惠蓀林場，香桂為此試區的優勢樹種之一，故選為研究的對象。研究目標係藉GIS疊合GPS定位之香桂樣株圖層與海拔、坡度、坡向、坡面位置及SPOT-5影像導出植生指標圖層，協同多變量統計模擬試區香桂之空間分布型態。研究建立邏輯思複迴歸 (LMR)、抉擇樹 (DT) 及區別分析 (DA) 三種模式，預測並繪製全區的香桂適生育地。建模與驗模採兩種取樣設計，兩者建模樣本相同，取自東峰溪流域，惟驗模樣本分別取自東峰溪與關刀溪兩流域。準確度評估顯示，DT優於LMR，而前二者又遠優於DA。三者於建模、驗模與繪圖之執行效率相當。重要的是DT和LMR於首次模擬，大幅縮小實地調查面積，節省可觀的經費及人力，故兩者更適用於香桂適生育地之模擬。SPOT-5影像導出植生指標改善模式預測能力效用很小，乃因其光譜及空間解析度皆不足，無法分辨散生香桂。三種統計法建立「東峰模式」雖都通過東峰驗模組檢測，但皆未通過關刀驗模組檢測，凸顯此模式無法僅透過地形變數跨越空間精確外推無建模樣本區域。未來研究將嘗試從高空間、高光譜解析度遙測資料萃取物種光譜資訊作建模用變數，期能跨越空間精確外推無建模樣本區域。

【關鍵詞】香桂、地理資訊系統、遙感探測、全球定位系統、數值高程模型、抉擇樹、邏輯思複迴歸、區別分析、適生育地。

Research paper

The Effects of Different Sampling Designs on the Ability of Model for Predicting the Suitable Habitat of *Cinnamomum subavenium*

Nan-Jang Lo¹ Wei-I Chang² Kai-Yi Huang³

【Abstract】Randaishan cinnamons (*Cinnamomum subavenium* Miq.), one of the evergreen broad-leaved tree species, are generally distributed in central and southern Taiwan. The species was chosen as target

1. 國立中興大學實驗林管處

Experimental Forest Management Office, Chung-Hsing University.

2. 農委會林務局森林企劃組組長

Chief, Forest Planning Section, Forest Bureau, Council of Agriculture.

3. 國立中興大學森林學系 教授，通訊作者

Professor, Department of Forestry, National Chung-Hsing University, Corresponding Author, e-mail: kyhuang@dragon.nchu.edu.tw

for this study because it is one of the dominant species in the Huisun study area in central Taiwan. GIS technique was applied to overlay the tree samples positioned by GPS on the layers of elevation, slope, aspect, terrain position, and vegetation indices derived from SPOT-5 images for modeling the tree's suitable habitat. Decision tree (DT), discriminant analysis (DA), and logistic multiple regression (LMR) models were developed to predict and map the tree's suitable sites in the study area, and to determine the optimum one in terms of accuracy and efficiency. Two sampling designs were created for model development and validation. They used the same set of training samples from Tong-Feng watershed for model development but different sets of test samples for model validation, one from Tong-Feng and the other from Guan-Dau watershed. Accuracy assessment showed that the accuracy of DT was slightly better than that of LMR, accuracies of the two models were much better than that of DA; and the three models were highly efficient in implementation of model development and validation. More importantly, DT and LMR can be applied to predict the tree's suitable habitat because they greatly reduced the area of field survey to 4-7 % of the entire study area at the first stage. Vegetation indices derived from SPOT-5 images could not improve the predicting ability of models for the widely distributed species because of SPOT imagery lacking fine spectral resolution and spatial resolution. The "Tong-Feng models" developed from three methods failed to pass validation by Guan-Dau test samples despite passing validation by Tong-Feng test samples. The outcome emphasized that the "Tong-Feng models" only based on topographic variables could not perform spatial extrapolation accurately from a smaller area with training data to a larger area without any training data. Follow-up studies will attempt to extract spectral information associated with the species from high spatial, spectral resolution remotely sensed data and use it as variable for model development so that the ability of spatial extrapolation with a model can be improved.

【Key words】 Randaishan cinnamons, Geographic Information System (GIS), Remote Sensing, Global Positioning System (GPS), Digital Elevation Model (DEM), Decision Tree (DT), Logistic Multiple Regression (LMR), Discriminant Analysis (DA), Suitable Habitat.

一、前言

香桂 (*Cinnamomum subavenium* Miq., Randaishan cinnamon) 又名巒大桂為樟科常綠大喬木，因其樹皮及葉均具肉桂香味，因而得名 (劉業經等, 1994)。香桂分布華南地區及台灣，台灣多分布於中、南部海拔500–2,000 m之森林中，數量甚多，耐陰性強，在大型孔隙或崩塌地區，極少發現其小苗；但在林下其小苗數量眾多，且在擾動較少及土壤層較厚的森林中，常成為大喬木 (樹高可超過20 m)，佔據森林的上層。香桂在惠蓀林場為優勢樹種之一，於海拔650 m (松風山區) 即可發現它的蹤跡，但它主要出現在海拔900–2,000 m環境較

為潮濕的森林中。根據羅南璋 (1992) 於東峰溪流域設置49個樣區進行植群分析調查，分布於東峰溪流域、北東眼山區的香桂，多為胸高直徑60 cm以上，樹高20 m以上之大樹，然分布於關刀流域者胸高直徑多不及30 cm，且樹高低於20 m以下，在形態上因兩個流域的生長環境不同，而有明顯的差異存在。香桂樹形通直圓滿，木材可供建材、器具，樹皮提煉的精油可供化粧品等之香精原料；從葉提煉的葉油亦可作食品，香煙的香料，或供作為殺菌劑材料 (羅南璋, 2010)。因香桂為廣泛型的分布樹種，而國內、外多數研究相對上較少將GIS結合多變量統計應用在廣泛分布型植物的生育地

模擬上，所以本研究選擇香桂為探討對象，以瞭解此方法在廣泛分布型植物的適用性。

Felicísimo *et al.* (2002) 於GIS上疊合現存及非現存之林型與海拔、坡度、日輻射潛能 (potential insolation)、至海距離等環境變數，並取樣匯入邏輯思複迴歸 (Logistic Multiple Regression, LMR)，模擬西班牙北部六種林型之潛在空間分布，可有效繪製出潛在植群圖，以提供森林管理規劃之用，惟該研究未驗證潛在植群圖的推測準確度。Gross *et al.* (2002) 使用LMR模式推測位於美國科羅拉多州的山羊適生育地 (山羊活動區與非活動區)，模式自變數有坡度、海拔、坡向、至庇護地形的距離。研究結論指出此模式提供簡易不貴的方法，可迅速精確地從大區域篩出山羊適生育地，主因山羊與顯著地理特徵 (環境變數) 有極強關聯，其中以至庇護地形的距離的影響力最強，而模式在納入影響力越強的環境變數後，有效提高模式推測準確度。O'Brien *et al.* (2005) 以GIS整合五個解釋變數坡向、植物群落、到乾河谷距離、坡度、土壤，連結LMR建立模式，推測美國西南部及墨西哥地區珍稀瀕危叉角羚 (Sonoran pronghorn) 潛在生育地，以作為新族群移居地。該研究使用LMR係因其廣泛用於生育地模擬且能建立連續機率而非二位元之潛在生育地統計面。Wang *et al.* (2007) 建立浙江天目山自然保護區GIS，並藉由LMR找出坡向、土壤、年降雨量等重要自變數，從而建立最佳柳杉生育地模式，劃設核心保護區、緩衝區及實驗區，有效監測與管理柳杉珍稀樹種。Gude *et al.* (2009) 應用LMR模式進行野生動物管理和保護機制評估，並在研究中成功篩選出洛磯山脈北部一帶之灰狼 (*Canis lupus*) 適生育地，作為灰狼群繁殖復育之用。

De'ath and Fabricius (2000) 利用分類迴歸樹 (Classification and Regression Tree, CART) 亦稱抉擇樹或決策樹 (Decision Tree, DT) 分析澳洲大堡礁珊瑚棲地調查資料，並指出CART相當適合分析生態資料。Felicísimo and

Gómez-Muñoz (2004) 在GIS平台上以LMR、DT、MARS (Multivariate Adaptive Regression Splines) 等多變量統計有效模擬與測繪西班牙伊比利半島櫟樹群的潛在分布圖。O'Brien *et al.* (2005) 以GIS結合CART及LMR建立模式，推測美國西南部及墨西哥地區珍稀瀕危叉角羚 (Sonoran pronghorn) 潛在生育地，以作為新族群移居地。該研究使用CART係因其不需受限於統計假設且已成功運用於生育地模擬，研究結論指出CART模式無論在建模或驗模的準確度皆優於LMR模式。Bourg *et al.* (2005) 以GIS疊合海拔、坡度、林型及火災頻率等環境變數圖層，並取樣匯入CART預測位於美國維吉尼亞州西北部喬治華盛頓國家森林中之稀有草本植物火雞芒 (turkeybeard) 的適生育地，其結果顯示CART能有效且精確地預測火雞芒的適生育地。Landenburger *et al.* (2008) 以改良式DT結合Landsat ETM+ 影像及大量白皮松 (whitebark pine) 林木參考資料，有效模擬與測繪大黃石生態系白皮松的分布圖。Ke *et al.* (2010) 利用QuickBird高空間解像力多譜影像和光達資料以影像切割方法和DT模式進行紐約Heiberg人工林針葉樹種與闊葉樹之分類，分類成果符合要求，堪稱成功。

Lowell (1991) 在GIS空間環境下，透過區別分析 (Discriminant Analysis, DA) 模擬位於美國密蘇里州Boone郡的BWREC野生生物研教中心之長期森林演替狀況，其結論指出DA統計法在空間環境下對檢視長期時間現象甚為有用。Marnell (1998) 利用區別分析判別陸地和水體的環境因素，對在愛爾蘭的鱗螈以及常見的蛙類的棲地加以區分。陳永寬等 (1998) 利用海拔、坡向、坡度、土壤和地質等五項環境因子透過區別分析、因素分析和群聚分析進行林地分級，求導出臺灣杉的適生育地。李培芬等 (2005) 利用GIS整合約50幅生態與環境因子圖層，包括至海拔3,000 m及其以上區域之最近距離、坡度、數值地形模型 (Digital Terrain Model, DTM)、溫度因子、森林

分布及規整差植生指標 (Normalized Difference Vegetation Index, NDVI) 等，並以五種方法，包括DA、LMR及其他三種方法，來預測全台灣八色鳥 (fairy pitta) 之棲地。Riordan and Rundel (2009) 以主成份分析與區別分析預測美國加州山艾低灌木叢植物群落 (sage scrub) 空間分布並繪製適生育地圖。

上列引用文獻都是以GIS疊合不同環境因子 (預測變數) 圖層，再透過一至多種不同多變量統計法如LMR、DT、DA或其他方法，建立預測模式，推測動植物適生育地及非適生育地。如上所述那般的資料處理方式，本質上仍與多頻譜遙測影像分類相同，只是後者通常專門處理電磁能波段一致率定資料，而前者處理資料類型則更為多樣化，甚至也包含遙測影像及其衍生資料。儘管如此，兩者仍都屬於圖樣辨識及分類 (pattern recognition and classification) 範疇。這些引用文獻多數採用兩種或以上的圖樣辨識方法來建立模式，其目標不外是比較各模式於物種適生育地的預測能力，以評比方法的優劣差異所在。Jensen (2005) 指出沒有任何一種圖樣分類法在先天上優於任何其他方法；分類問題的本質、試區的生物物理特性、遙測資料分布的形態及既有的知識共同決定哪一個分類演算法將產生有用的結果。Duda *et al.* (2001) 提出可信賴的忠告：「對那些研究旨在證明某一特定學習或辨識演算法的全面優越性，我們應該抱持健康而合理的懷疑」。因此影響物種適生育地模式預測能力的因素，不惟只有圖樣辨識法 (即指前述引用文獻所採用多變量統計法如DT、LMR、DA等等)，也需考量地面實測資料品質 (尤指精度)、數量與分布 (做為建模用訓練樣本及驗模用檢測樣本)；遙測資料品質 (尤指空間、光譜、時間解析度)；預測 (自) 變數本身的描述與解釋能力等等。

本研究的核心目標係利用GIS平台，將惠蓀林場試區的東峰河流域及關刀河流域GPS定位香桂樣株圖層與海拔、坡度、坡向、坡面位

置及衛星影像導出植生指標等圖層加以疊合並抽取樣本，再匯入抉擇樹、邏輯思複迴歸、區別分析等三種多變量統計方法，建立預測模式，模擬及測繪整個試區的香桂適生育地。除了從推測準確度與執行效率來評估這三種統計方法建立預測模式之優劣所在，更重要的是分析香桂樣株數量與分布對於模式預測能力的影響。因此，本研究將東峰與關刀兩流域內GPS所定位香桂樣株加以分割與組合，形成兩種取樣設計，從而評估此兩者於模式預測能力的影響。另外，本研究也將評估前述五個預測變數對模式預測能力之影響及它們的相對重要性。

二、研究區域

惠蓀林場在行政區域上隸屬南投縣仁愛鄉，為中興大學所有四個實驗林場之一，其位置約在南投縣東北角，如圖1紅點所示，座標範圍東經121°0′–121°8′，北緯24°2′–24°7′，略呈東北至西南走向，海拔由454 m至2,419 m。其境內有五個主要集水區，從西到東分別為蘭島溪 (629 ha)、關刀溪 (2,216 ha)、東峰溪 (2,061 ha)、椿谷溪 (976 ha) 及合水溪 (715 ha)，流向均由南向北，匯流入北港溪，加上數個小集水區 (面積有計入總數但未列出)，總面積約7,477 ha。惠蓀林場全區氣候溫潤，植物種類繁多，約有1,100種之多，為台灣中部地區極具代表性之森林。本研究為配合使用衛星影像，故採矩形區域，涵蓋整個不規則形狀之惠蓀林場，總面積約為17,136.0 ha，分析模擬係根據該矩形區域實施。目標樹種香桂的地面定位調查區域是在東峰河流域及關刀河流域，兩個主要集水區之相關位置如圖1所示。

三、材料與方法

本研究共分為六項主要步驟，茲分項條列並搭配圖2之研究流程詳述如後。

(一) 資料蒐集

本研究蒐集的資料項目，包括數值高程模型 (Digital Elevation Model, DEM) 資料檔，像

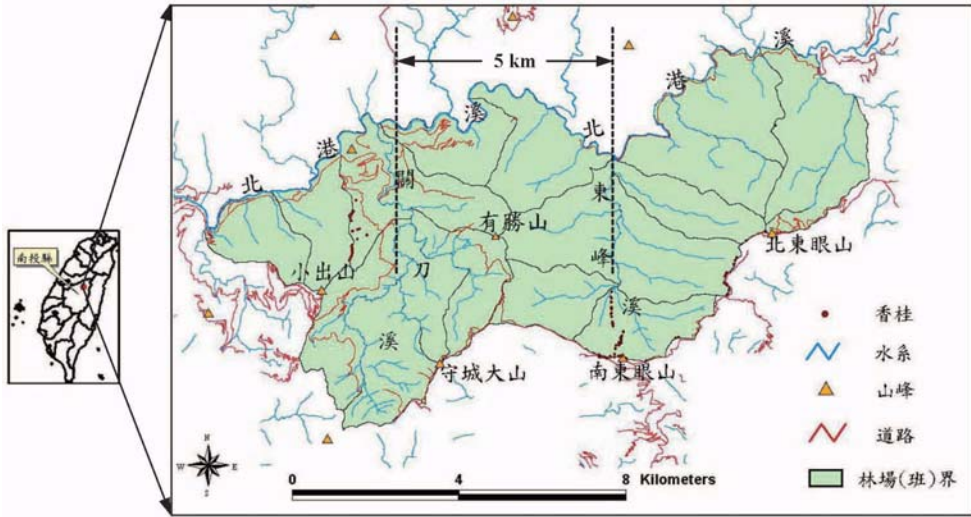


圖1. 惠蓀林場研究區域位置圖
Fig. 1. Location map of the Huisun study area

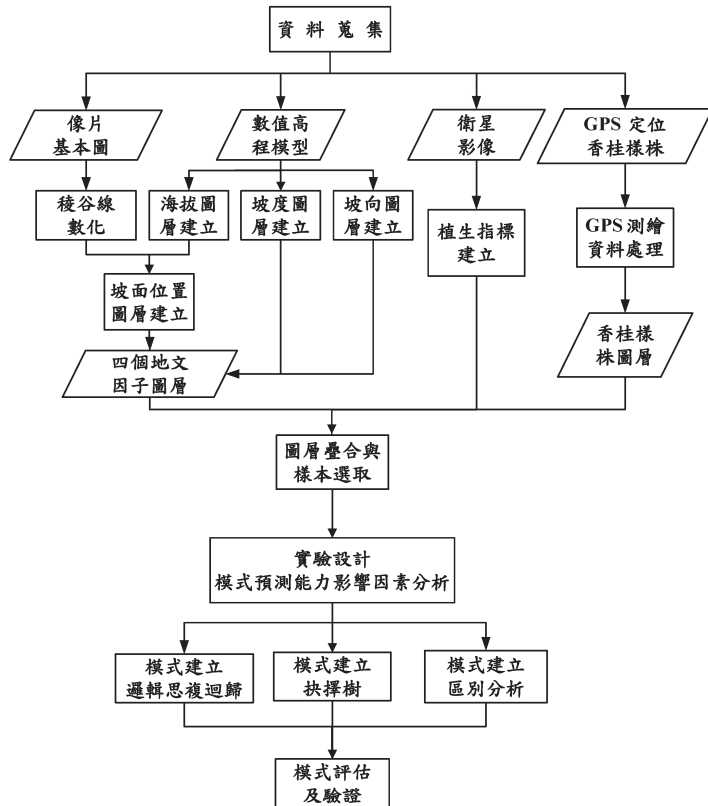


圖2. 本研究的流程圖
Fig. 2. The flowchart of this study.

片基本圖 (1/10,000)、赴現場以GPS搭配5 m可伸縮加長天線桿及雷射測距系統測繪香桂樣株資料。另外，本研究向中央大學太空及遙測研究中心選購九幅SPOT-5衛星影像，採第二級正射糾正處理，座標系統設定為TWD67基準及

橫麥卡托投影經差二度分帶 (TM 2°)，並以最近鄰法重取樣為5 m × 5 m。本研究再由其中篩選品質最佳兩幅做為本次實驗之用，影像之雲量皆要求在15 % 以下，入射角或視角皆在 ± 15°以內，其影像基本資料如表1所示。

表1. 惠蓀林場研究區 SPOT-5 衛星影像基本資料

Table 1. The basic information of SPOT-5 satellite images over the study area

影像編號	季節	接收日期	視角 (°)	平均雲量 (%)	方位角 (°)	高度角 (°)
I0009406	夏	2004/07/10	-7.8	11	90.7	70.8
I0009409	秋	2005/11/11	-13.2	3	155.9	45.4

(二) 資料處理

1. 稜谷線數化

本研究使用Calcomp Drawing Board III (A0尺寸) 的高精度數化板將研究區內所涵蓋1/10,000像片基本圖上之稜谷線加以數化及建檔，而其數化結果將用於後續求導與建立坡面位置圖層。

2. 地文因子圖層建立

應用ERDAS Imagine影像處理及分析軟體模組從DEM資料產生海拔圖層，並求導與建立坡度、坡向圖層。方位角採360° 為圓周式循環數列，而非直線數列，不利於比較，故在納入資料分析前，應予以轉換處理 (Chang, 2010)。本研究實施行前預備測試，分別採用360° 方位角、坡向分為八級 (每級45°)、坡向度數轉換為餘弦等三種計量方式，所得結果實質上均無明顯差異，故決定坡向採用360° 方位角的計量方式，以維持坡向與海拔、坡度皆為比率或間距式數值變數之本質 (比率或間距變數的資料含量大於級序與類別變數)。

3. 坡面位置圖層建立

一般而言，生育地在山坡上之相對位置，對於土壤性質及林木生長深具關係 (劉榮瑞等, 1983)。求導及建立坡面位置圖層，首先需將數化的稜脊線與山谷線分別轉換為網格

檔，並依據Skidmore (1990) 所提相對位置計算式子，計算每一網元至最鄰近稜線與谷線的歐幾里得距離 (Euclidean distance)，並求出該網格點之相對位置比例 P_{ij} ，其公式如下所示：

$$P_{ij} = PV / (PV + PR) \dots\dots\dots (1)$$

PV = P點到與最鄰近山谷線的歐幾里得距離；

PR = P點到與最鄰近稜脊線的歐幾里得距離；

上式中的P為某一檢定點 (網元)， P_{ij} 為在 i 列， j 行之相對位置比例。本研究分別以稜脊線及山谷為最高及最低坡面位置，到山谷線之 $P_{ij} = 0.0$ ，而到稜脊線之 $P_{ij} = 1.0$ ，再將其間分為六級，連同稜脊線、山谷線共分為八級 (謝立忻等, 2005)。

4. 植生指標圖層建立

Hoffer (1978) 指出植物在夏、秋兩季，近、中紅外光反射因葉子內部結構與葉層數及含水量而有明顯之差異，因此可作為分辨不同種類植物的重要依據。植生指標原始構想除了源自於植物、土壤於近、中紅外光譜反射原理及特性，再加上植物物候之特性，在夏、秋季節葉片細胞壁結構及含水量有甚大差異，使這兩季的近、中紅外波段差值比可以作為樹種之間區別的依據。本研究由夏 (2004/07/10) 和秋

(2005/11/11) 兩期SPOT-5衛星影像求導出植生指標，其求導的公式如下所示：

$$[(NIR - MIR)_{秋} / (NIR - MIR)_{夏}] \dots\dots\dots (2)$$

5. GPS測繪資料處理

本研究赴試區現場獲取香桂樣株資料，係使用Trimble Path finder Pro XR移動站搭配5 m可伸縮加長天線予以定位，另視衛星訊號實際接收狀況，輔以雷射測距系統協助定位。於調查結束返回學校後，再以設置於系館建築物頂層Trimble基站資料融合移動站資料，做後續差分改正，經改正後樣株位置之二維平均定位誤差在1 m以內。最後，將其轉換為ArcGIS所能讀取及處理的shapefile圖層資料，且座標系統設定為TWD67及TM 2° 分帶，東峰溪及關刀溪流域香桂樣株的分布如圖1內暗紅色斑點所示。

(三) 圖層疊合及樣本選取

本研究將香桂樣株及上述五項因子圖層之座標系統皆設定為TWD67與TM 2° 分帶，並以ERDAS Imagine軟體中Image Interpreter\ Utilities\ Layer Stack疊合形成GIS資料庫，供後續取樣分析與模擬之用。接著再以ERDAS Imagine軟體中Image Interpreter\ Utilities\ Subset分別切取目標樹種區塊(即主體)與該樹種之互補區塊(即背景)。本研究之目標樹種為惠蔴林場內以GPS調查之香桂樣株。依據Sperduto and Congalton (1996) 之研究，將背景(非香桂生育地)樣本對主體(香桂生育地)樣本以至少三倍的比例來選取背景樣本。同時依據Pereira and Itami (1991) 從東峰溪與關刀溪流域採亂數隨機分散抽取背景樣本，以降低樣本空間自相關(spatial autocorrelation)至最小，然後分別計算主體區塊和背景區塊在這五項因子的相關統計值。本研究調查香桂樣株共有149株，其中108株來自東峰溪流域，而其餘的41株來自關刀溪流域。本研究採用分割樣本建模與驗模之取樣設計方法(Chang, 2010)，選取全數樣本中佔某一比例的一部份(例如2/3或1/2)作為模式建立

用之訓練樣本，而剩餘的另一部份(例如1/3或1/2)則作為後續模式驗證用之檢測樣本，詳見下面第(四)項的說明。

(四) 模式預測能力影響因素分析

模式預測能力一般除了考量建模用的多變量統計方法之外，也需考量其他因素如樹種地面實測樣株數量及分布、預測(解釋或自)變數本身解釋能力及其組合的效用、遙測資料品質(空間、光譜解析度及精確度等)、地面調查資料品質(例如水平位置及垂直高程之精度)等等。本研究針對多變量統計方法、目標樹種的樣株數量及分布、預測變數等因素採不同取樣實驗設計，以評估模式預測能力之優劣，而其他因素受限於時間關係及若干極重要遙測資料(試區之高光譜影像及光達資料)無法及時取得，故現階段不納入本研究之考量，但無可否認，這些因素為將來後續研究之重點所在。由於上述三項因素組合變化繁複，計算量極為龐大，為化繁為簡，本研究以最早完成調查東峰溪流域之香桂樣本為建模與驗模起始點，再評比不同預測變數組合，決定評比所用之基準模式，以作為後續不同樣本配置設計、不同統計方法比較的基準。

1. 取樣設計

本研究香桂的樣株主要來自先完成調查的東峰溪流域及後完成調查的關刀溪流域，兩相鄰但分隔約5 km之流域，藉由GPS地面定位調查所獲得。在模式建立使用的訓練樣本及模式驗證使用的檢核樣本採取不同數量及分布的配置，以評估與瞭解「香桂樣株數量及分布」對模式推測能力的作用所在。本研究採用兩種取樣設計：

- (1) 取樣設計-A：「東峰模式」建模用之「主體」訓練樣本與驗模用之「主體」檢核樣本皆來自東峰溪流域(但都不含關刀溪流域樣本)，且建模用之「背景」訓練樣本與驗模用之「背景」檢核樣本皆從整個惠蔴林場試區(包含東峰溪與關刀溪兩流域)以亂數隨機抽取出來。全數

樣本(含主體與背景)中之三分之二(2/3)作為模式建立之用,而其餘之三分之一(1/3)則作為模式驗證之用。因東峰香桂樣株最先完成調查,將此樣本配置用於基準模式之建模與驗模,以作為其他取樣設計相互比較的基準。

- (2) 取樣設計-B:此一設計仍然使用與取樣設計-A相同的「東峰模式」,即建模使用與取樣設計-A相同的一套主體與背景樣本(建模不納入任何關刀河流域香桂樣本),「背景」驗模樣本也相同,而與取樣設計-A唯一的差別在「主體」驗模樣本。此設計的驗模樣本是將全數關刀河流域之「主體」樣本及來自兩流域的「背景」樣本(佔全數「背景」樣本的三分之一)用於驗證「東峰模式」,以評估推測錯誤究竟發生在「關刀驗模樣本」,或者是發生在「東峰驗模樣本」。此一設計也在瞭解模式(預測變數僅使用地文因子)在空間外推之能力,能否跨越空間上兩相分隔的流域(東峰流域及關刀河流域相隔約5 km上下)。

2 預測變數組合

本研究使用之模式預測變數,皆由航遙測資料求導出的面型資料,包括海拔、坡度、坡向、坡面位置、SPOT-5衛星影像導出植生指標。實驗設計上將採多種配置組合,以評比某一預測變數在模式有此一變數與無此一變數下,對模式推測能力影響之差別。本研究也可從不同預測變數組合下,瞭解模式推測準確度的差異,從而推斷是否還有遺漏而未考量之重要環境變數。

3. 三種統計法建立模式

本研究使用SPSS統計分析軟體進行三種多變量統計模式之運算,茲分述如下:

(1) 邏輯思複迴歸

邏輯思複迴歸(LMR)模式的應變數為類別型態(存在或不存在),而自(解釋)變數可為類別型或數值型。LMR模式之應變數是採二

元資料(0或1)記錄的,其應變數對自變數的反應結果是呈S形曲線而非直線關係,比較接近生物對環境的反應,故較適合建立模式(李培芬,2000)。因為線性迴歸在考慮解釋變數的選取時,必須要注意所選出來的解釋變數和應變數是否存在著因果關係,透過資料中的某些已知的訊息,便可對未知的變數做預測。除此之外,如果解釋變數之間的關係非常密切,則彼此之間或許存在有共線性的關係,顯然不適合放在同一個模型中。LMR最大的優點是不需符合線性迴歸及區別分析的兩個假設:應變數必須為常態分布、線性關係(王保進,2004;吳明隆,2008)。

LMR模式屬於廣義的線性模型使用邏輯思鏈結函數(logistic link function),主要用於應變數為二元性資料。LMR與傳統的迴歸分析性質相似,差別在於它是用來處理類別性(categorical)資料的問題,由於類別性資料是屬於離散型的資料,所以必須先將連續型資料轉為介於0與1之間的離散資料型態,才可以對轉換過後的連續型資料作迴歸。LMR主要目的是為了找出類別型態的應變數和一連串解釋變數之間的關係,因此LMR和迴歸分析中最大的差別在於應變數型態的不同。然而,對於離散型變數有很多分析方法可供採用,而選用LMR是根據兩個主要的理由:第一個理由是基於數學觀點而言,它是一個極賦彈性且容易使用的函數;第二個理由則是因為它適用於解釋生物學上的意義(鄭順林、康惠雯,2006)。

邏輯思複迴歸公式使用如下,假設某一樹種(例如,本研究探討的目標樹種為香桂)在樣區中之分布情形,若存在者記為1,不存在者記為0,有 h 個樣本像元,假設 Y_h 為應變數, x_1, \dots, x_n 為自變數, q_h 為第 h 個樣本像元為生育地的機率,因此

$$P(Y_h = 1) = q_h = \frac{\exp(B_0 + B_1 x_1 + \dots + B_n x_n)}{1 + \exp(B_0 + B_1 x_1 + \dots + B_n x_n)}$$

.....(4)

為邏輯思複迴歸公式，其中 B_0, B_1, \dots, B_n 為未知參數通常以最大概似法 (maximum likelihood) 加以估計，並以newton-raephson疊代法求解 (Freeman, 1987)。本研究採用統計軟體SPSS之Analyze\ Regression\ Binary Logistic運算資料。 $(B_0 + B_1 x_1 + \dots + B_n x_n)$ 項乃由統計軟體SPSS自動疊代產生迴歸式。

自變數是否納入迴歸式視其影響力之大小並經由軟體統計檢測來決定。各自變數 (環境因子) 迴歸係數的大小可顯示其對生育地的相對重要性 (Narumalani *et al.*, 1997)。將迴歸式代入式子 (4) 可求得研究區域中任一位置出現某一樹種的機率，再與門檻值比較確認該位置是否為適生育地。本研究是以香桂出現於某一位置 (像元) 與否當作LMR模式應變數，而海拔、坡度、坡向及坡面位置等因子作為自變數。某一位置出現目標樹種之機率在0.0–1.0之間，通常以0.5為出現與否之門檻，出現之機率要大於或等於0.5，不出現之機率小於0.5。惟可經測試後酌量調整 (例如0.4、0.45或0.55)，找出整體推測準確度達最高之反轉點 (cut-off point)。如上所述，LMR迴歸式可藉由SPSS軟體運算獲得，並將迴歸式代入ERDAS Imagine軟體中之Spatial Modeler，從而推測產生目標樹種適生育地之全區分布圖。

(2) 抉擇樹

分類迴歸樹 (Classification and Regression Trees, CART) 演算法為 (Breiman *et al.*, 1984) 所開發，Verbyla (1987) 首次運用於生態研究上，其為抉擇 (決策) 樹的演算法之一。此方法因在統計上無常態分布的假設，故屬於無母數統計方法之一 (鄧家駒, 2004)。抉擇樹同時也是一種資料探勘 (data mining) 領域中兼具分類與預測功能的技術，藉由分析給定的已知原始資料，來建立適當的分類規則，再根據此規則對未知類別的資料進行預測，以一個單一解釋 (預測) 變數函數，來建構一個二分式抉擇樹。樹狀圖形頂端的節點稱為根節點 (root node)，最下方的節點則稱為葉節點 (leaf node)，每個

葉部節點擁有一個類別標記 (class label)，如圖3所示。圖中X, Y, Z為分支，即測試之結果，中節點代表測試條件，即輸入之解釋變數 (如海拔、坡度等環境因子)，葉節點分為「0」與「1」，在本研究即分別為香桂的「非生育地」與「生育地」。當給定的原始資料從根部節點進入抉擇樹後，可應用各種不同的演算法，來決定每筆資料該進入下一層的哪一個子節點 (child node)，此過程不斷地反覆，直到所有的資料到達葉部節點，而停止分割的情況有六種，詳見林震岩 (2008)。

從根部到每一個葉部節點都有一條獨特的路徑，每一條路徑皆表示一個用來分類資料的規則。當一筆未知的資料進入抉擇樹時，該筆資料將會依其符合的分類規則，而得到預測的結果 (O'Brien *et al.*, 2005)。統計軟體SPSS 13.0版之後才具有CART分析功能，放置於Analyze\ Classify\ Tree功能之下的演算法。CART的建樹過程，必須藉由分析訓練資料集中之資料特性，反覆地找出最適當的分類屬性及其分岔點 (splitting point)，將資料集予以分類。分散度 (diversity) 或稱混雜度 (impurity)，則是用來評估一個分隔屬性變數的衡量標準，對於一組資料的分散度指標以Gini (1912) 開發吉尼指標 (Gini index) 計算其公式如下所示：

$$Gini\ index = 1 - \sum c^2 \dots \dots \dots (3)$$

上式之中的c表示每一組資料的反應比例，而最好的分隔變數是能夠降低一個資料組最多的分散度 (De'ath and Fabricius, 2000)。模式準確度評估 (accuracy assessment) 是將樣本分割為建模組與驗模組資料，而以驗模組樣本進行評估 (Breiman *et al.*, 1984)。因此，在建構CART模式之前，必須先選定已分類好的原始資料來源，而資料來源通常包含了許多的紀錄，每筆紀錄擁有相同數目的屬性及一個目標屬性。依照資料的特性和資料探勘的目標，必須將原始資料分割成建模組與驗模組資料，提供CART使用。此外，為了避免過度學習

(over-fitting) 而使CART樹狀結構出現過於冗長複雜的情形，可利用CART驗模組模式修正其結構。其原理係利用建模資料組建立預測模式後，再將驗模資料組代入模式進行驗證，可經由總資料組之中不斷選取部分資料作為建模組

與驗模組，並觀察不同樹狀結構大小分類錯誤率之變化，來反覆修正的過程，此稱為折疊刀法 (jack-knifing procedure) 或交叉驗證法 (cross validation) (Breiman *et al.*, 1984)。

(3) 區別分析

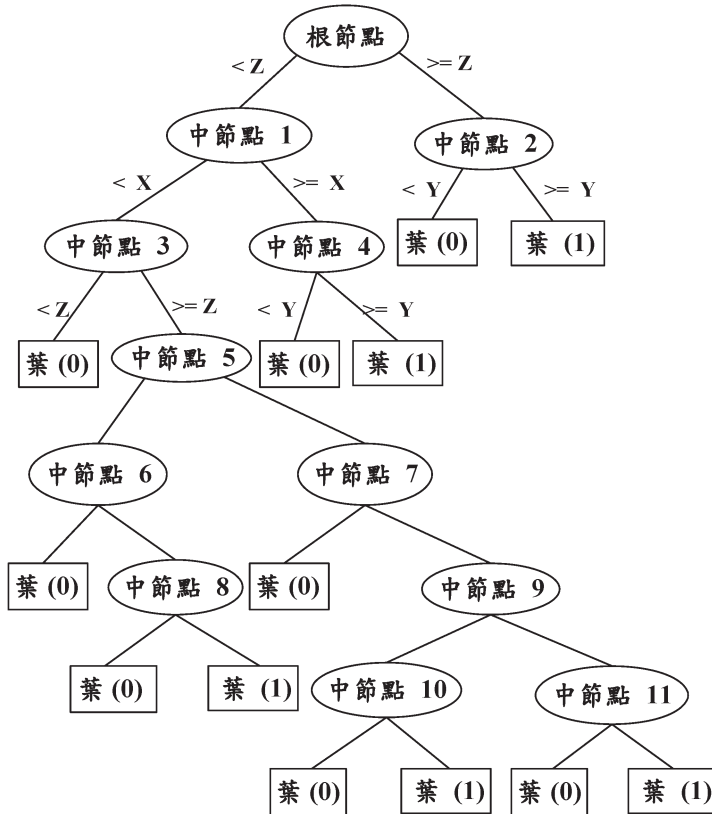


圖3. 抉擇樹之樹狀結構示意圖
Fig. 3. The tree structure of a decision tree

區別分析，或稱判別分析、鑑別分析 (Discriminant Analysis, DA) 為 Fisher (1936) 所創，最早應用於貸款信用評分模型以區分好壞之貸款者；Williams (1983) 首次將此演算法應用在生態上，便對於多個植物族群的分類獲得很好的分類效果。區別分析是一種劃分群體之技術，事先將已知類別的觀察值，選取具有分類效果的樣本，利用類別變數 (grouping

variable, g類) 當作反應變數 (如下式之 y)，而多個計量的區別變數 (discriminant variable) 當作解釋變數 (如下式之 x_i)，建立區別函數 (discriminant function)，公式如下所示：

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_ix_i + \dots + b_nx_n \dots\dots\dots (5)$$

y為區別函數值 (或稱區別分數 discriminant score)， x_i 為區別變數， b_i 為區別係數

(discriminant coefficient or weight)，利用區別函數將新觀察值進行適當分類(林震岩，2008)。如圖4所示，假設有兩個變數 X_1 和 X_2 ；兩橢圓代表兩個族群I和II有部份相互重疊影響。族群I可能為香桂分布區域、族群II可能為鄰近香桂分布區域之其他樹種，其生態環境是交疊重複難以區分開的。區別分析的原理便是找出能劃分I、II兩族群的交點所連成之直線DF，與另一條引自原點而與直線DF垂直的直線Y交於C點。直線Y即是區別函數圖形，線上任何一點觀測值可由 X_1 和 X_2 兩個變數轉換而得；直線Y上的C點，稱為分界點，將Y值分為兩部份，作為區分I、II兩族群之依據。兩個族群的 X_1 和 X_2 變異數不同，要使組間變異數對組內變異數比值達最大時方能找到分界點；就可以得到每個區別變數的權重係數，其意義與迴歸分析類似。標準化後的區別係數值愈大，代表此區別變數對群體的區別能力愈高(黃俊英，2007)。本研究採用統計軟體SPSS之Analyze\ Classify\ Discriminant運算資料。 $(b_0 + b_1x_1 + \dots + b_nx_n)$ 項乃由統計軟體SPSS自動疊代產生之線性判別式。

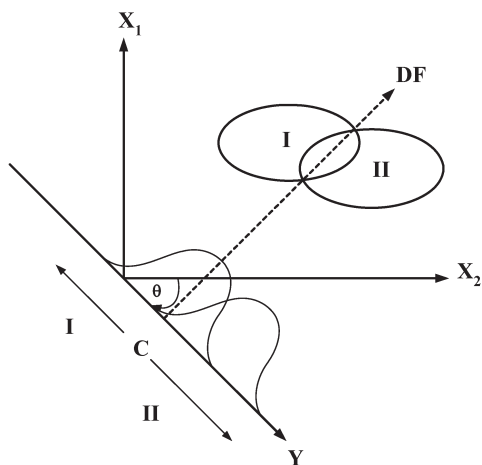


圖4. 區別分析的幾何圖形觀念

Fig. 4. The concept of geometry for discriminant analysis

區別分析在應用上需遵守三個前提假設，詳見林震岩(2008)所列。由於赴現場調查相當耗工耗時，在進行統計分析時先暫時略過前提假設檢定，改以事後檢定的方式，採多組獨立樣本回推驗證模式的可靠性。此外區別分析不只停留在描述各類組間變異上的關係，還能夠對未知所屬之類組加以區別分類，故本研究較重視後者預測香桂空間分布的情形，與前提假設違反並無直接相關。但為避免複共線性會有不利分類的影響，採逐步估算法(step-wise estimation)，逐漸加入新的解釋變數，將不適合的剔除直到模式合理為止。

(五) 模式評估及驗證

模式準確度評估是將推測結果與實際情況相互比較，並計算兩相符合之正確百分比率，其中應包含整體準確度、kappa相符係數(coefficient of agreement)、主體(目標)推測準確度、背景推測準確度。惟因採全區逐格做像元對像元之比對殊為不可能，故採抽樣方式兩兩比對。由全區抽取樣本除須符合統計取樣原則外，所選總樣本通常分為建模用之訓練樣本與驗模用之檢核樣本兩組，再分別實施模式準確度評估，此方法普遍應用在多光譜遙測影像分類之準確度評估，旨在避免建模樣本自身評估所隱含之偏頗與降低空間外推(spatial extrapolation)之可能風險(Jensen, 2005; Lillesand et al., 2008)。

SPSS統計軟體LMR、DT與DA模組具準確度評估功能，「建模組」訓練樣本於輸入樣本建立模式之時一併完成準確度評估。DT與DA的「驗模組」與「建模組」的準確度評估是於SPSS軟體DT與DA模組執行時一併完成；LMR則需藉由ERDAS Imagine軟體Spatial Modeler完成驗模。此三模組於繪製香桂潛在生育地全區預測圖皆需藉由ERDAS Imagine軟體Spatial Modeler的協助，來完成繪圖工作。

四、結果與討論

(一) 香桂空間分布特性之統計分析

本研究之試區為矩形，其左上角之TM 2°

分帶座標為 (248,845.500, 2,668,192.500)，右下角為 (264,590.500, 2,657,317.500)，涵蓋面積約為17,136.0 ha (包含惠蓀林場不規則區域面積7,477 ha)。本研究所用之DEM網格資料，是根據前列矩形區域的座標，以ERDAS Imagine的Interpreter\ Utilities\ Subset功能切取出來，其範圍大小為3,150行 (columns) × 2,176列 (rows)，總共有6,854,400個像元，像元尺寸為5m × 5 m。

本研究以GPS搭配伸縮加長天線及雷射測距系統來定位香桂樣株，總共調查149株 (東峰溪、北東眼山108株、關刀溪41株)，並以shapefile向量檔儲存。惟在轉換為網格檔資料後，有若干株因間距小於一個網元 (亦稱像元) 尺寸5 m而遭合併，故東峰溪之樣本實際上為102個網元，而關刀溪則為35個網元。全區及香桂於五個因子之基本統計值如表2所示。香桂樣本分布海拔範圍1,073–2,097 m，較一般天然分布範圍500–2,000 m稍高，在惠蓀林場

海拔650 m處 (松風山區) 即可發現香桂蹤跡，但因其胸高直徑多未達10 cm，所以未將其定位作為樣木。表2顯示香桂分布所在平均坡度23°，明顯較全區平均坡度34°為平緩，這表示香桂喜好生長在水份不易流失而較為潮濕的緩坡上。就北半球而言，西南坡向最乾燥，東北向最陰濕來看，香桂喜好較潮濕的東北向，此與表2香桂坡向之眾數落在北東北向16°大致相符。羅南璋 (1992) 於東峰河流域以傳統調查方式對香桂主要分布樣區環境描述為：「多位於山坡中、上側，及稜脊較平坦，土層深厚處。」，及現場觀察得知，香桂喜生長在寬平稜線帶上或近稜線兩側下方較平緩處及中段段潮濕且日照充足處，與表2香桂坡面位置平均值為「6」且眾數為「7」大致符合。在研究區中常可見許多香桂之小苗，表示它們在此一環境中具有自行更新之能力，可穩定存在一段時間。

表2 香桂及全區於五項預測變數的相關統計值

Table 2. The statistic of five predictive variables for the study area and Randaishan cinnamons sites

統計值	全區					香桂				
	海拔 (m)	坡度 (°)	坡向 (°)	坡面位置	植生指標	海拔 (m)	坡度 (°)	坡向 (°)	坡面位置	植生指標
平均值	1314	34	—	5	25	1639	23	—	6	24
眾數	1239	37	127	6	22	1165	29	16	7	21
最大值	2418	89	361	8	119	2097	46	359	8	51
最小值	445	1	1	1	5	1073	1	1	4	20

註：在ERDAS Imagine軟體內，坡度361°是指定為平坦地。

(二) 預測變數組合對模式準確度之影響

因東峰香桂樣株最先完成調查，本研究將前面第三節材料及方法所敘述的「取樣設計-A」用來建立與檢驗基準模式，並以此基準模式評估預測變數組合將如何影響模式的推測準確度。海拔、坡度、坡向、坡面位置及植生指標為本研究建立模式的五個預測變數。這五個變數有30種組合，剔除五取一的單一變數模

式、五取二的雙變數模式之後，仍有五取三、五取四、五取五等之組合，總共還有15種組合，計算工作量仍然相當龐大。所以，本研究使用之SPSS套裝軟體的LMR、DT及DA模組，可讓我們從輸入的預測變數自動篩選出最有效用之變數，大幅減少計算工作量，從而節省工作時間。

表3. SPSS套裝軟體多變量統計模組評比五預測變數之相對重要性

Table 3. The relative importance of the five predictive variables assessed by multivariate statistical modules in the SPSS software packages

預測變數		LMR		DT		DA	
		Wald 統計量	排序	標準化 重要百分比	排序	標準化 判別係數	排序
海拔	B1	16.327	3	100.0%	1	.430	3
坡度	B2	16.426	2	87.1%	2	-.503	2
坡向	B3	1.674	4	54.4%	3	-.126	4
坡面位置	B4	33.726	1	45.7%	4	.616	1
植生指標	B5	.270	5	23.3%	5	-.007	5

在透過該軟體三種模組執行之後，表3顯示SPSS套裝軟體內三種多變量統計模組評比五項預測變數重要性之參數值，見林震岩 (2008) 說明。LMR、DA均篩選出坡面位置、坡度、海拔分佔前三名最有效用的變數，而坡向及植生指標則殿後，分居第四、第五。DT排序稍有出入，海拔、坡度為頭兩名最有效用之變數，坡向反而超越坡面位置，排名第三，坡面位置則落居第四，惟兩者之重要百分比相差有限。本研究認為多變量統計法在變數重要性排序不同，有兩項主要原因，敘述如後。第一，DT與DA、LMR明顯不同。DA、LMR為「參數式統計」，有常態分布等不同假設之限制，而DT為「非參數式統計」，無常態分布等假設之限制。第二，坡面位置為分八級的級序變數，而海拔、坡度、坡向為比率或間距式之數值變數，數值變數在資訊含量與差異分辨能力的本質上大於級序與類別變數。儘管如此，本研究為了利於後續的評估比較，統一採用DA、LMR篩選結果為依據，以海拔、坡度、坡面位置等三個變數來分別建立這三種模式，且以它們所得之推測準確度作為比較基準。其次，再分別將「植生指標」、「坡向」強制納入由這三個變數形成之基準模式，所得推測準確度與基準模式之結果相互比較，可評估「植生指標」、「坡向」是否具有改善模式推測能

力之效用。

表4顯示LMR、DT及DA於三種預測變數組合下所建立模式之推測準確度。三種預測變數組合分別是：海拔、坡度、坡面位置 (B1)；海拔、坡度、坡面位置、植生指標 (B2)；海拔、坡度、坡向、坡面位置 (B3)。本研究探討重點之一是SPOT-5衛星影像導出植生指標對模式預測能力之改善，而B2與B1的差別在植生指標。先就「植生指標」來看，不論在建模與驗模部份，分別比較三種方法在B1與B2的整體、生育地及非生育地的推測準確度，三者大多數在B1與B2都相等，而僅有少數在加入植生指標後，推測準確度有小量上升 (大約1-3%，LMR生育地的推測準確度例外上升3%)，但亦有極少數推測準確度小量 (1%) 下降。所以SPOT-5衛星影像導出植生指標整體上並無法有效改善這三種模式的推測準確度。本研究認為，SPOT-5影像近紅外波段空間解析度為10 m，而中紅外波段則為20 m。香桂樹冠直徑通常約5-10 m，樹冠展開面積不足一個像元。因此，本研究認為SPOT-5影像無法完整解析單株香桂，且無法與周遭其他樹種分離隔開。同等重要的是，上列兩個波段皆為寬頻波段，光譜解析度亦不夠細緻，難以區別香桂與周遭其他樹種之光譜反射的差別。

表4. 三種不同變數組合建立三種模式於預測香桂適生育地之推測準確度

Table 4. The accuracies of predicting the suitable habitat of Randaishan cinnamons by the three models based on three combinations of different variables

組別	LMR			DT			DA		
	準確度 (%)			準確度 (%)			準確度 (%)		
建模組	B1	B2	B3	B1	B2	B3	B1	B2	B3
非生育地	90	90	92	98	98	98	80	80	80
生育地	72	72	71	78	78	78	88	88	85
整體 (%)	88	88	89	96	95	96	81	81	81
驗模組	準確度 (%)			準確度 (%)			準確度 (%)		
	B1	B2	B3	B1	B2	B3	B1	B2	B3
非生育地	95	95	96	95	95	95	83	83	83
生育地	58	61	65	62	62	62	85	85	88
整體 (%)	80	81	81	91	91	91	84	84	84

註：B1 = 海拔、坡度、坡面位置；B2 = 海拔、坡度、坡面位置、植生指標；B3 = 海拔、坡度、坡向、坡面位置。

同樣地，第三種組合 (B3) 添加了坡向變數，分別比較這三種方法在B1與B3的整體、生育地及非生育地的推測準確度，三者大多數在B1與B3都相等，而僅有少數在加入坡向後，推測準確度有小量上升 (約1-3%)，而LMR生育地的推測準確度例外上升7%)，但亦有極少數推測準確度小量下降。因此，整體上這三種模式在納入坡向變數並無法有效改善這三者的推測準確度。

(三) 三種模式預測能力之比較

根據前面評比之結果，本研究採用海拔、坡度及坡面位置三項變數，來建立基準模式，進一步來評比LMR、DT及DA方法之優劣。表5顯示以三項變數建立這三種模式所獲得推測準確度之比較。先從整體精度來看，DT「建模組」整體推測準確度96%，高於LMR的88%，而前二者皆遠高於DA的81%；DT「驗模組」整體推測準確度91%，與LMR者相等，而前二者也皆高於DA的84%；DT建模及驗模整體推測準確度的平均值94%，高於LMR的88%，而前二者又皆高於DA的83%。而更重要的是DT在建模與驗模之kappa平均值0.68高於

LMR的0.57，而前二者又遠高於DA的0.40。從整體推測準確度及kappa值而言，DT在香桂適生育地的預測能力上優於LMR，而前二者的預測能力又遠優於DA。

其次，再就表5內三種模式的生產者精度與使用者精度來評比三者之優劣差別所在。為了能更清楚瞭解表中數字的意義，本研究以建模與驗模平均值來簡化說明。就驗模組而言，DT在「非生育地」的建驗模生產者精度平均值97%，略高於LMR的93%，而遠高於DA的81%約12-16%，但DT在「生育地」的建模與驗模生產者精度平均值70%，略高於LMR的68%，然而前二者反低於DA的87%約17-19%。再深入檢視DA數據，DA「生育地」推測之漏授 (100% - 生產者精度平均值 = 100% - 87%) 僅13%，但DA「生育地」推測之誤授 (100% - 使用者精度平均值 = 100% - 41%) 卻高達59%，與DT的27% (100% - 73%) 相較，幾達2.2倍；其與LMR的42% 相較，也達1.4倍。由此可研判DA推測誤差主要是發生在將「非生育地」(背景) 誤授為「生育地」(主體或目標樹種—香桂)，大幅犧牲背景之準確度，來

表5. 海拔、坡度及坡面位置建立三種模式預測試區香桂適生育地之誤差矩陣表

Table 5. The error matrix for the two models based on elevation, slope, and terrain position predicting the suitable habitat of Randaishan cinnamons in the study area

模式	整體精度 (%)	<i>kappa</i>	類別	驗證	生產者精度 (%)	使用者精度 (%)		
LMR	建模	88	生育地	建模	72	平均	51	平均
	驗模	91		0.60	驗模	64	68	64
	平均	90	非生育地	建模	90	平均	96	平均
	0.57	驗模		95	93	95	95	
DT	建模	96	生育地	建模	78	平均	84	平均
	驗模	91		0.57	驗模	62	70	62
	平均	94	非生育地	建模	98	平均	97	平均
	0.68	驗模		95	97	95	96	
DA	建模	81	生育地	建模	90	平均	41	平均
	驗模	84		0.33	驗模	85	87	41
	平均	83	非生育地	建模	80	平均	98	平均
	0.40	驗模		83	81	98	98	

平均：分別是指建模及驗模的整體精度、*kappa*、生產者精度、使用者精度的平均值。

換取提高目標樹種「生育地」之準確度。這裡要特別強調的，是背景數量在整個試區所佔的比例通常遠大於主體數量所佔的比例，因此模式在預測「非生育地」時，應儘可能優先維持背景的高推測準確度，最好能在90%以上。惟DA統計法產生的線性判別式，有將「非生育地」與「生育地」兩群體均等分割的傾向，從而放寬「生育地」判定之標準，導致不少的「非生育地」被誤判為「生育地」。這正是DA方法的缺點所在，而此缺點更不利於找出模式尚未考量到之預測變數。因此，從推測準確度的角度來看，DT模式確實較LMR為佳，而前二者又遠較DA模式來的好。

表6顯示三種模式推測香桂適生育地面積與百分比之統計，DT推測所得面積佔試區總面積的4.3%，LMR為6.4%，而DA則為11.7%。這三種模式首次推測香桂適生育地即可從全區面積約17,136.0 ha篩選出約4–12%的高潛力區，大幅減少後續地面調查之面積，從而降低所需投入經費、時間與人力。根據上面表5結果顯示DT及LMR模式的推測準確度遠高於DA模式，而在表6之內DT及LMR推測香桂適生育地之面積明顯小於DA者 (DA面積分別約為DT、LMR的2.72、1.85倍)，表示DT、LMR模式在推測香桂適生育地均遠較DA更為嚴謹。

表6. 海拔、坡度、坡面位置建立三種模式預測試區香桂潛在生育地之分布統計

Table 6. The distribution statistics of the potential habitat of Randaishan cinnamons in the study area predicted by the three models based on elevation, slope, and terrain position

預測類別	LMR		DT		DA	
	面積 (ha)	百分比 (%)	面積 (ha)	百分比 (%)	面積 (ha)	百分比 (%)
生育地	1088.88	6.4	739.45	4.3	2013.38	11.7
非生育地	13547.12	93.6	16396.55	95.7	15122.62	88.3
合計	17136.00	100.0	17136.00	100.0	17136.00	100.0

為能更深入探討這三種模式在香桂適生育地判定標準鬆緊之別，本研究疊合三種模式預測香桂適生育地圖，並以像元為基準分別將「三種模式皆推測為背景—非香桂適生育地」、「僅有一種模式推測為香桂適生育地，

另二則否」、「有二種模式推測為香桂適生育地，另一則否」、「三種模式皆推測為香桂適生育地」，之情況予以分類編碼，並各指定一顏色顯示於圖5。同時以像元為基準就上面不同組合情況加以統計呈現於表7。

表 7. 以像元基準比較三種模式預測結果之統計

Table 7. Statistics acquired from comparison based on the pixel among the results predicted by the three models

推測類別	推測結果相符模式	代碼	顏色	像元數目	百分比 (%)	
非生育地	LMR, DT, DA	0	白	□	5976544	87.20
生育地	DT	1	亮綠	■	70019	1.02
生育地	LMR	3	灰	■	1742	0.02
生育地	DA	5	紅	■	340596	4.97
生育地	DT, LMR	4	水藍	■	744	0.01
生育地	DT, DA	6	橘	■	31689	0.46
生育地	LMR, DA	8	藍	■	239738	3.50
生育地	DT, LMR, DA	9	墨綠	■	193328	2.82
總計	—	—	—	—	6854400	100.00

圖5 (a) 為全區三種模式推測香桂適生育地疊合圖，而圖5 (b) 顯示圖5 (a) 中亮綠圓圈線圍繞區的放大圖。圖5 (b) 最明顯的是僅有DA推測為香桂適生育地的紅色區 (表7中，像元數居第二) 圍繞在三種模式皆推測為香桂適生育地的墨綠色區 (表7中，像元數居第四) 之外。這表示且印證，如前所述，DA判定標準最寬鬆，而DT、LMR相對上遠較DA嚴謹，所以三者共同交集之墨綠色區落在核心帶，最外圍是紅色區。其次，有兩種模式推測為香桂適生育地，而另一則否的情況 (如 [LMR, DA]、[DT, DA]、[DT, LMR]) 分別為藍色 (表7中，像元數居第三)、橘色 (表7中，像元數居第六)、水藍

色表示，大多數都落在紅色外圍區與墨綠核心區之間，或偶有交錯穿插出現。這三種模式推測結果若以小比例尺圖分開展示，總會予人三者推測結果非常相似之錯覺。這種錯覺可藉由圖5 (b) 內另一非常明顯之現象來消除。圖5 (b) 內僅有DT推測為香桂適生育地的亮綠色 (表7中，像元數居第五) 近矩形區塊出現在分岔樹枝狀的尾端，或有出現於樹枝與主幹接合處，可作為這三種模式推測結果明顯有差別之最有力明證。

最後是這三種模式在執行效率上之比較。三者是在SPSS軟體裏有各自的模組，DT與DA在點選模式各輸入參數後，皆可高度自動化執

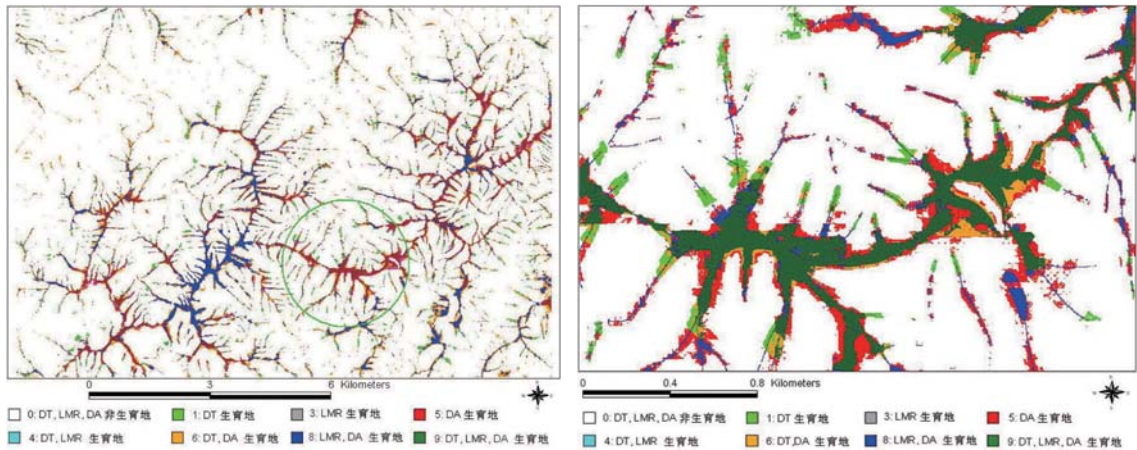


圖5. 以像元基準比較三種模式預測結果：(左) 全區；(右) 放大左圖之亮綠圓圈區

Fig. 5. Comparison based on the pixel among the results predicted by the three models: (left) entire study area, (right) enlarged area with light-green circle in the left diagram.

行模式之建立與驗證，產生各式的統計參數報表及誤差矩陣表，供模式效能評比之用，兩者可謂旗鼓相當。惟LMR只有建模可在SPSS軟體上高度自動化執行，而驗模部份必須轉換平台至ERDAS Imagine軟體的Spatial Modeler上執行。其次，三者是最後繪製香桂全區的適生育地分布圖時，都須跨軟體平台作業，轉移到ERDAS Imagine軟體的Spatial Modeler模組上來執行。DT模式在Spatial Modeler模組下建立與執行繪圖程序較LMR、DA模式略顯耗工費時。

(四) 兩種取樣設計影響模式預測能力之評估

在探討統計模式預測能力的影響因素時，往往非僅評估統計方法與執行效率，另也需考量資料的品質、預測變數、目標樹種樣株的數量和分布等因素，而其中目標樹種樣株的數量及分布為本研究探討的重點之一。

本研究採用之取樣設計有兩種，已敘述於第三節材料及方法。此處再強調的是這兩種取樣設計都是採用東峰香桂「主體」樣本之中的三分之二來建立模式（兩種取樣設計的「背景」建模與驗模樣本皆相同，係從東峰與關刀

兩個流域以亂數隨機抽取而出），而唯一差別是驗證模式的樣本。「取樣設計-A」是以東峰樣本中剩餘的三分之一來驗證模式，而「取樣設計-B」是以關刀全部的樣本來驗證模式。為何要採用如此設計？這兩個流域的主流中心線相距約5 km，本研究擬瞭解建立「東峰模式」只納入東峰香桂樣本而不納入關刀香桂樣本，在空間阻隔狀況下，所謂「東峰模式」究竟是否可以跨越空間阻隔，精確地預測於關刀河流域之香桂適生育地？

表8顯示的結果為兩種取樣設計採用東峰與關刀驗模樣本對於三種模式預測準確度之影響評估。首先，從表8分別就三種模式各自比較兩取樣設計之整體精度。LMR、DT分別由取樣設計-A的整體精度92%、91%，明顯下降至取樣設計-B之83%、83%，降幅分別達9%、8%，而相對上DA由取樣設計-A的整體精度84% 小幅下降至取樣設計-B的81%，降幅僅3%。這個結果表示兩者對於LMR、DT推測準確度的影響超過對於DA的影響，似乎也意味DA在跨越空間分隔之推測能力優於LMR、DT，但是否果真如此？

表8. 兩種取樣設計對於三種香桂適生育地模式預測準確度影響之評估

Table 8. Assessment of the effects of two sampling designs on the accuracies of the three models for predicting the suitable habitat of Randaishan cinnamons

模式	整體精度 (%)	類別	取樣設計	生產者精度 (%)	使用者精度 (%)
LMR	取樣設計-A 東峰驗模	生育地	A 東峰	64	50
			B 關刀	0	0
	取樣設計-B 關刀驗模	非生育地	A 東峰	95	95
			B 關刀	95	87
DT	取樣設計-A 東峰驗模	生育地	A 東峰	62	62
			B 關刀	0	0
	取樣設計-B 關刀驗模	非生育地	A 東峰	95	95
			B 關刀	95	87
DA	取樣設計-A 東峰驗模	生育地	A 東峰	85	41
			B 關刀	63	34
	取樣設計-B 關刀驗模	非生育地	A 東峰	83	98
			B 關刀	83	94

註：漏授 = 100% - 生產者精度；誤授 = 100% - 使用者精度。

由表8先分別就三種香桂模式各自比較兩種取樣設計「生育地」的生產者精度，再分別就三種香桂模式各自比較兩種取樣設計「非生育地」的生產者精度。先就「生育地」的生產者精度來看，LMR及DT分別從取樣設計-A於「生育地」之生產者精度為64%及62%，劇跌至取樣設計-B之0%及0%；相對地，DA從取樣設計-A的85%明顯下降至取樣設計-B之63%，降幅達22%。三種香桂模式「生育地」的生產者精度都受到取樣設計-B建模不含關刀溪香桂樣本的影響，而都有明顯下跌，惟LMR及DT所受影響更大，劇降至0%，將關刀溪香桂驗模樣本全部推測錯誤。再就「非生育地」的生產者精度來看，LMR及DT兩種取樣設計於「非生育地」之生產者精度相等，LMR是95%、95%，而DT也是95%、95%；DA亦是如此，是83%、83%。值得注意的是，三種模式「非生育地」生產者精度雖未受兩種取樣設計差別之影響，但正如前面表5所指出的，DA模式的缺點是放寬「生育地」判定之標準，導

致不少的「非生育地」被誤判為「生育地」，所以DA「非生育地」之生產者精度在83%，明顯低於LMR的95%及DT的95%，高達12%。這就是DA犧牲「非生育地」之生產者精度，來換取提高「生育地」的生產者精度的代價。

接著，由表8先分別就三種香桂模式各自比較兩種取樣設計「非生育地」的使用者精度，再分別就三種香桂模式各自比較兩種取樣設計「生育地」的使用者精度。先就「非生育地」使用者精度來看，LMR及DT都從取樣設計-A於「非生育地」之使用者精度95%下跌至取樣設計-B之87%，下跌8%；相對地，DA從取樣設計-A的98%下降至取樣設計-B之94%，降幅4%。三種模式「非生育地」的使用者精度都受到取樣設計差別的影響而小有升降，雖DA變化幅度略小一些，但三種模式仍維持在高準確度的狀態。再就「生育地」使用者精度來看，LMR及DT分別從取樣設計-A於「生育地」之使用者精度50%及62%，劇跌至取樣設計-B之0%（陰影區）；相對地，DA從取樣設

計-A的41% 下降至取樣設計-B之34%，降幅7% (陰影區)。三種模式「生育地」的使用者精度都受到取樣設計之差別影響而有大幅之升降，尤其是LMR及DT都受到取樣設計-B無關刀溪流域香桂建模樣本的影響而遠大於DA，使LMR及DT於取樣設計-B「生育地」的使用者精度低於DA，惟LMR及DT在取樣設計-A的使用者精度仍明顯皆高於DA。

綜合上面所做之討論來看，在取樣設計-A，LMR及DT模式推測準確度明顯優於DA模式，確實毋庸置疑，然而在取樣設計-B下，DA卻優於LMR及DT。惟實際上是DA統計模式對「生育地」判定之標準遠較LMR及DT模式來的寬鬆，導致不在少數的「非生育地」被誤判為「生育地」，大幅犧牲「非生育地」的推測準確度，以換取提昇「生育地」的推測準確度所造成，這就是所謂「歪打正著」。因此，在大多數狀況下，LMR及DT模式確實優於DA模式。現在將三種模式優劣差異，即主要力道之一卸除之後，剩下的是另一主要力道，即兩種取樣設計下香桂與蘇鐵蕨樣本分布與數量差異之影響。這就表示「東峰模式」僅含海拔、坡度、坡面位置等地形變數，並藉由東峰香桂樣本所在位置三地形變數特性代入這三種統計法，雖然可以正確推測東峰溪流域之香桂適生育地，但是卻無法跨越空間阻隔正確地外推到關刀溪流域。

五、結論與建議

根據不同準確度指標評估，DT模式的表現優於LMR模式，而前二者又明顯優於DA模式。DA的缺點是將背景「非生育地」誤授為主體「生育地」，導致DA預測能力明顯不及DT及LMR。就節省地面調查所需的經費及人力而言，DT及LMR的成效明顯優於DA。就執行效率來看，DA略優於LMR、DT，惟差距有限。整體而言，DT及LMR在香桂適生育地預測上均較DA更加適用。

本研究採用衛星影像是為了跨越空間上的

阻隔，提昇模式空間外推的能力，然而此次所採用之SPOT-5影像導出植生指標幾無提昇模式預測能力之效果，乃因SPOT-5的近、中紅外波段為寬頻，其光譜解析度不足以分辨樹種間細微光譜反射差異，而且空間解析度亦不足以分辨台灣複雜、散生之林相，再加上大氣效應、光照與感測儀觀測幾何效應及地形效應所導致光譜變異之影響所致。

比較兩種取樣設計驗證三種統計法建立「東峰模式」，結果顯示LMR、DT模式均將取樣設計-B的關刀香桂樣本完全推測錯誤，而DA模式因以犧牲背景來換取主體之缺陷得以倖免全軍覆沒。因此，僅藉由東峰香桂樣本所建立的「東峰模式」，雖然可以精確推測同區的香桂適生育地，但卻無法跨越空間阻隔精確外推至關刀溪流域，亦即模式的適用性仍有大幅改善空間。

本研究從前面討論可推論出四個要點。第一，惠蓀林場內這兩個流域雖毗鄰相連(緯度相同)，但它們香桂樣株所在之地文特性卻無法互通共用，表示兩者微地形仍有顯著差異存在。第二，因香桂所在兩個流域微地形環境之差異，造成這三種統計模式無法僅藉由地文因子跨越空間阻隔，從一個流域正確地推測另一個流域的香桂適生育地，此即空間外推 (spatial extrapolation) 今後尚待突破之重點。第三，在目標物的空間推測上應嘗試從它找出具備跨越空間阻隔之特性作為模式的自變數，例如目標物與背景之間光譜反射的細微差異，是未來後續研究探討的重點之一。第四，這三種模式在香桂「生育地」的使用者精度在40–65% 範圍之間，這表示模式還有遺漏某些重要的生育地因子未納入考量中，諸如微地形、微氣候、土壤及周遭生物等因子。

影響模式推測能力並非僅有使用的統計方法，也包含資料的品質、預測變數和目標樹種樣株的數量和分布等，其中遙測資料的品質、所萃取的光譜預測變數及目標樹種樣株數量及分布均甚為重要，都是後續研究的重點。在未

來之後續研究將考慮使用高空間解析度影像、高光譜影像及光達資料，期能改善模式之預測能力。

六、參考文獻

- 王保進 (2004) 多變量分析：套裝程式與資料分析。高等教育出版社。第438頁。
- 李培芬 (2000) 以遙測和GIS探討臺灣地區之脊椎動物多樣性 (I)。行政院農委會。
- 李培芬 白梅玲 林瑞興 (2005) 利用遙測與GIS探討瀕危物種八色鳥之棲地喜好與分布。農委會94年度遙測應用計畫成果發表會專刊。第1-10頁。
- 林震岩 (2008) 多變量分析：SPSS的操作與應用。智勝文化。第489-530頁。
- 吳明隆 (2008) SPSS操作與應用多變量分析實務。五南圖書出版公司。第134-151頁。
- 陳永寬、鄭祈全、廖錦偉 (1998) 應用地理資訊系統篩選台灣杉造林地。航測及遙測學刊 3 (2): 47-70。
- 黃俊英 (2007) 多變量分析，第七版。華泰文化。第137-163頁。
- 劉棠瑞、蘇鴻傑 (1983) 森林植物生態學。臺灣商務印書館，第462頁
- 劉業經、呂福原、歐辰雄 (1994) 台灣樹木誌。國立中興大學農學院。第440頁。
- 鄭順林、康惠雯 (2006) 資料採礦中的統計預測與分類方法 (一) <http://web.thu.edu.tw/sljeng/www/Datamining/predict.htm>
- 鄧家駒 (2004) 多變量分析。華泰文化。第369-394頁。
- 謝立忻、羅南璋、黃凱易 (2005) 應用3S地球空間技術於植群空間分布型態之探討。林業研究季刊27 (4): 37-46。
- 羅南璋 (1992) 惠蓀實驗林場東峰溪集水區植群分析 國立中興大學森林研究所碩士論文。
- 羅南璋 (2010) 3S技術結合多變量統計模擬測繪林木之潛在生育地。國立中興大學森林研

究所博士論文。

- Bourg, N. A., W. J. Mcshea, and D. E. Gill (2005) Putting a CART before search: successful habitat prediction for a rare forest herb. *Ecology*, 86 (10): 2793-2804.
- Breiman, L., J. H. Friedman, R. A. Olsen, and C. G. Stone (1984) *Classification and Regression Trees*. Chapman and Hall, New York, USA. 357 pp.
- Chang, K. T. (2010) *Introduction to Geographic Information Systems*, 5th ed., The McGraw-Hill Companies, New York.
- De'ath, G. and K. E. Fabricius (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis, *Ecology*, 81 (11): 3178-3192.
- Duda, R. O., P. E. Hart, and D. G. Stork (2001) *Pattern Classification*, 645 p., John Wiley & Sons, New York.
- Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7: 179-188.
- Felicísimo, A. M., E. Frances, J. M. Fernandez, A. Gonzalez-Diez, and J. Varas (2002) Modeling the potential distribution of forests with a GIS, *Photogrammetric Engineering & Remote Sensing*, 68 (5): 455-461.
- Felicísimo, A. M. and A. Gómez-Muñoz (2004) GIS and predictive modeling: a comparison of methods applied to forestal management and decision-making, *Proceedings of GIS Research UK 2004*, University of East Anglia, Norwich, Norfolk, UK, pp. 143-144.
- Freeman, D. H. Jr. (1987) *Applied Categorical Data Analysis*, Marcel Dekker, Inc, pp. 237-274.
- Gini C. (1912) *Variabilità e mutabilità*, Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della Regia

- Università di Cagliari, Anno III, parte 2, also reproduced in C. Gini (1939) op. cit.
- Gross, J. M., M. C. Kneeland, D. F. Reed, and R. M. Reich (2002) GIS-based habitat models for mountain goats. *Journal of Mammalogy*, 83 (1): 218–228.
- Gude, J. A. Gude, M. S. Mitchell, D. E. Ausband, C. A. Sime, and E. E. Bangs. (2009) Internal validation of predictive logistic regression models for decision-making in wildlife management. *Wildlife Biology*, 15 (4): 352–369.
- Hoffer, R. M. (1978) Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data. In: Swain, P. H. and S. M. Davis (Eds.), *Remote Sensing: The Quantitative Approach*, McGraw-Hill, Inc., New York, pp.227–289.
- Jensen, J. R. (2005) *Introductory Digital Image Processing—A Remote Sensing Perspective*, 3rd ed., Pearson Education, Inc, New Jersey.
- Ke, Y., L. J. Qackenbush, and J. Im. (2010) Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification, *Remote Sensing of Environment*, (114): 1141–1154.
- Landenburger, L., R. L. Lawrence, S Podruzny, and C. C. Schwartz (2008) Mapping regional distribution of a single tree species: whitebark pine in the Greater Yellowstone Ecosystem, *Sensors*, 8: 4983–4994.
- Lillesand, T. M., R. W. Kiefer, and J. W. Chipman (2008) *Remote Sensing and Image Interpretation*, 5th Edition. John Wiley & Sons. Inc., New York.
- Lowell, K. (1991) Utilizing discriminant analysis with a geographical information system to model ecological succession spatially. *Int. J. Geographical Systems*, 5 (2): 17–191.
- Marnell, F. (1998) Discriminant analysis of the terrestrial and aquatic habitat determinants of the smooth newt (*Triturus vulgaris*) and the common frog (*Rana temporaria*) in Ireland. *Journal of Zoology London* 244: 1–6.
- Narumalani, S., J. R. Jensen, J. D. Althausen, S. Burkhalter, and H. E. Mackey, Jr. (1997) Aquatic macrophyte modeling using GIS and logistic multiple regression. *Photogrammetric Engineering & Remote Sensing*, 63 (1): 41–49.
- O'Brien, C. S., S. S. Rosenstock, J. J. Hervert, J. L. Bright, and S. R. Boe (2005) Landscape-level models of potential habitat for Sonoran pronghorn. *Wildlife Society Bulletin*, 33 (1): 24–34.
- Pereira, J. M. C. and R. M. Itami (1991) GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham red squirrel. *Photogrammetric Engineering & Remote Sensing*, 57 (11): 1475–1486.
- Riordan, E.C. and P. W. Rundel (2009) Modeling the distribution of a threatened habitat: the California sage scrub. *Journal of Biogeography*, 36 (11): 2176–2188.
- Skidmore, A. K. (1990) Terrain position as mapped from a gridded digital elevation model. *International Journal of Geographical Information Science*, 4: 33–49.
- Sperduto, M.B. and R. G. Congalton (1996) Predicting rare orchid (small whorled Pogonia) habitat using GIS. *Photogrammetric Engineering & Remote Sensing*, 62 (11): 1269–1279.
- Verbyla, D. L. (1987) Classification trees: a new discrimination tool. *Canadian Journal of Forest Research*, 17: 1150–1152.
- Wang, T., M. Li, and N. Wang (2007) Protecting

China Cedar (*Crypyomeria fortunei*) habitat using GIS-based simulation, modeling of existence probability, and function zoning. Mountain Research and Development, 27(4):

352–358.

Williams, B. K. (1983) Some observations of the use of discriminant analysis in ecology. Ecology, 64: 1283–1291.