

台灣生育地因子群集資料庫 查詢及應用於WWW

◎高堅泰／國立中興大學森林學研究所研究生

◎馮豐隆／國立中興大學森林學系教授

一、動機

若「生育地分類」之目的在於區分生育地之試驗設計，而非單純生態系統之分類，則宜採用多因子之設計，以這些因子作為取樣設計之分層，因子選擇越多，則分層越顯複雜，若以平常試驗設計之方式，將因子分等級，行交叉分層，選取樣區，則有以下多種設計問題：

1. 生育地因子間之交互作用難以分離出來，2. 試驗過於龐大，3. 難以顧及眾多因子。近年來，雖倡導生態系經營理念，然其驗證工作，皆仍以小面積實驗，固定單因子為主效果，鮮少提及生育地因子之交互作用。

為解決此問題，乃採取「群集分析」法，俾能一併解決交互作用及多因子試驗規模太大之問題，因此，建立生育地因子群集資料庫，為當務之急。

群集分析結果，多仍有主觀因素存在。本群集資料庫雖生育地因子分佈狀況良好，仍不能確認與生態之相關性，因此，需收集多種已有資料，及新設試驗結果之驗證，才能使此資料庫邁進實用階段，但驗證工作之範圍廣大，

且驗證實驗亦僅能以一般生態試驗之一部份進行，作為取樣分層方法的改進，亦即所謂適應性取樣（adaptive sampling），因此，群集之起始值（seed）的開放查詢及應用，為生態系經營理念試驗、生育地因子群集資料庫實際應用之始。

本文針對本資料庫之查詢及應用進行一瀏覽，以期各界先進參考指正，並提供更新資料，使生態系經營理念試驗能更精進。

二、材料

生育地因子群集資料庫由高堅泰、馮豐隆（1998）進行生育地群集研究，收集影響生育地之因子資料，共引用海拔、坡度、坡向、年均溫、最高溫月份（七月）高均溫、最低溫月份（一月）低均溫、年雨量、冬半年雨量（10-3月）、夏半年雨量（4-9月）、土壤有效深度、土壤型、土壤級等共12因子，各因子資料經最適空間推估，完成全省1km*1km之面資料共35990筆，再以群集分析將其分類成948生育地群集。並分別以ACCESS 97及ARC /

VIEW 3.0a 建立資料庫。現利用 ESRI 公司之 IMS (Internet Mapping Server) 1.0 套裝軟體、IDC (Internet Database Connector) 程式, 透過 Microsoft 公司之開放式資料庫介面 (open database connectivity, ODBC) 將資料庫連上 Microsoft 公司之 IIS (Internet Information Server) Web 伺服器。架構如圖 1, 本資料庫所採用之軟體除 IMS 外, 多為 Microsoft 公司之 Windows NT 所內建之功能, 軟體資源易於取得, 更新容易, 配套多樣化, 如 IDC 亦可用 Java、Active X 等撰寫, 而 ODBC 資料庫介面更僅需適當的驅動程式, 就能如 IDC 程式存取各種現存資料庫一般, 可存取 dBase、MS Access、MS Excel、Oracle、SAS、SQL Server、Sybase、Visual FoxPro... 等資料庫, 不致為資料提供上網而另覓新的資料庫或重新建構。而圖上三部分可以分別在不同機器硬體上, 以減少系統負荷。



圖 1 生育地因子群集資料庫軟體架構

三、查詢瀏覽

由中興大學森林資源調查測計研究室網站 (<http://140.120.93.10>) 之生育地資料庫超連結按鍵進入生育地因子群集資料庫查詢系統, 首先為說明首頁 (圖 2), 其右側之按鍵「查詢一」至「查詢四」說明如下。



圖 2 生育地因子群集資料庫首頁

1. 利用圖上查詢因子 (圖 3)

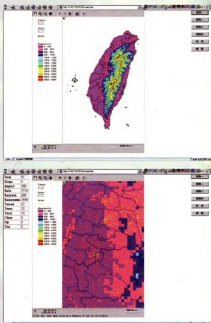


圖 3 查詢一版面及結果

以研究區域與資料庫套疊, 藉以得到因子資料及分類群集代號, WWW 上僅提供因子

資料供生態研究對未測定因子之查詢，資料內容為12項生育地因子，資料精密密度為 1km^2 。

2.利用座標查詢因子(圖4)



圖4 查詢二版面及結果

在知曉樣區(點、線)概略之TM2⁺座標值時，可藉由座標查詢因子資料，所得之資料及精密密度亦同於查詢一，且速度較快(因省略圖形界面之所有步驟)，對小區域之因子查詢甚為方便。

3.利用因子查詢群集代號(圖5)

適合針對細部個別的研究作分類依據，在各領域研究中，依個別所需之精密密度，列出特別觀測之因子值，較不重視(影響較小或無)

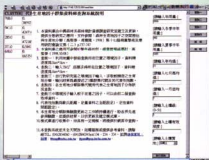
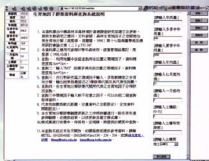


圖5 查詢三版面及結果(接續查詢二結果)

之因子值，則可由前查詢一、二取得，輸入資料不限精密密度，輸出為5種最近之分類群集代號，為輸入資料之特定資訊，無精密密度，並列出各分類之代表性指數(越小越佳，但不擔憂土壤因子者，不在此限)，代表性指數由歐氏(Euclid)距離計量，定性資料間距為1，定量資料全距為1。

4.利用群集代號查詢因子分佈狀態(圖6)

承接查詢三之結果，研究人員可以就研究對象之容許範圍，對於前述產生之5種分類，查詢因子分佈狀態，訂定分類群集合併與否之規範。

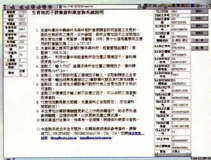
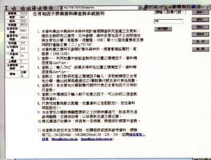


圖 6 查詢四版面及結果 (與查詢二結果比對)

四、應用瀏覽

1. 永久樣區設立之調查設計

- 材 料：僅有研究區域圖層。方法：套疊分析。
- 適 用：對大面積、不瞭解之試驗作預調查設計。
- 限 制：未曾到實地取得研究所重視因子(主效果)而採用資料庫1km*1km之資料，精密度較差，不適合小

面積或移動區域小之生物。

預期結果：得到過於精細之分類，須於預調查後合併分類群集。

例 子：杉木造林林分之永久樣區設立。

2. 永久樣區之檢討

材料：對研究區域已有預選樣區或舊有樣區。方法：交叉分析。

適用：對預選樣區檢討及舊有樣區整合。

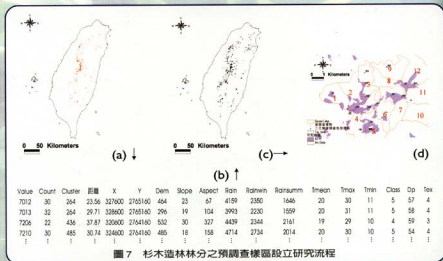


圖 7 杉木造林林分之預調查樣區設立研究流程

(a) 杉木造林林分區域圖 (b) 套疊所得生育地因子資料表 1668 筆 (c) 所得 406 種群集取樣之樣點 (d) 樣點位置圖共 80 幅 (圖示其一慈惠林場西半部)

限制：未有生育地因子值時，會有上例之限制。自訂精密度之生育地因子值時，則僅能挑出過剩樣本，資料庫無法建議應加選樣區資訊及位置。

預期結果：得到一交叉分析表，可藉以挑出過剩樣本。欲得加選樣區資訊，須有全區生育地資料或位置（位置僅能得到 $1\text{km} \times 1\text{km}$ 精密度資料）。

例子：八仙山事業區永久樣區預選樣區檢討。（圖 8）

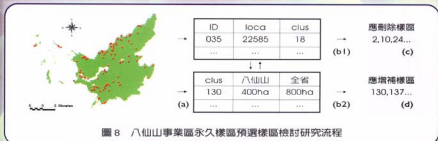


圖 8 八仙山事業區永久樣區預選樣區檢討研究流程

(a)八仙山事業區永久樣區預選樣區 105 筆(3 筆無法確定位址) (b) 預選樣區及八仙山事業區範圍套疊資料庫得 180 性群集分類 (c) 預選樣區經交叉分析(land use * cluster)選出刪除樣區 16 筆 (d) b 表經比對後，得最少應增補樣區 12 筆之群集代號 9 種。

3. 細部分類

材料：研究區域全面資料。方法：K-means 群集分析。

適用：小區域研究且結果欲外推至全省。

限制：小區域之分類結果良好，但外推區域之精密度較差（ $1\text{km} \times 1\text{km}$ ）。

預期結果：得到精密度較高之分類，而用於其他研究。

例子：關刀溪長期生態研究區引用全省生育地群集。（圖 9）



圖 9 關刀溪長期生態研究區引用全省生育地群集研究流程

原 $1\text{km} \times 1\text{km}$ 網格經加郵資料增補後成為 $40\text{m} \times 40\text{m}$ 網格 1668 筆資料，經 K-means 群集分析後得可推測至全省之生育地群集 32 個。

4. 全省生育地分類

材 料：K-means 之 seed。方法：群集分析。

適 用：應用一不同因子、結果之群集，將其集合分類。

限 制：集合分類內含其他因子之集合，集合分類間不能明確指示界線，若是單純生育地區分，僅以該因子界定，結果較明顯。

預期結果：得到層次分明、分類較粗獷之群集集合。

例 子：全省氣候區分類。（圖 10）

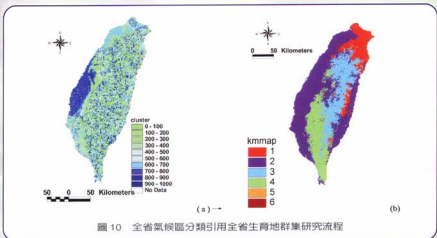


圖 10 全省氣候區分類引用全省生育地群集研究流程

原 948 群集經氣候因子選擇、規整、均連法群集分析後區分 6 種氣候分類

五、討論

生態系經營理念試驗選取因子為多方面的，這些因子亦成為樣區設立之分層，因子選擇越多，則分層越顯複雜，若以平常將因子分等級，行交叉分層，選取樣區，則有以下多種設計問題：樣區屬性資料於多度空間中，呈現如圖 11 (a) 方形之切割狀，與生物於生態環境中之分佈-生態幅度現象不符；取樣設計若以拉丁方格設計，則生物於生態因子上之眾多交感作用，易於分析中流失；若要對交感作用做詳細研究，則樣區數量勢必過於龐大，假設

以地形、土壤等共 5 因子，各因子分 5 級，3 重複，則樣區數已接近一萬 (55*3)，實非任何一實驗計畫所能負擔，故近年來，雖倡導生態系經營理念，然其驗證工作，皆仍以小面積實驗，固定單因子為主效果，鮮少提及生育地因子之交感作用。而群集分析明顯劃分同質區域，其結果應用於分層，可以減少分層之層數，顧及實驗計畫之負擔；因子交感作用之影響，亦計入於群集間，可於分析時提出探討；且群集為最原始的理由聚集相似之樣本資料，其中以相關、相似性、距離來產生如圖 11 (b)

之近乎圓形之分類，此乃與生物於生態環境的分佈實況相似。故採用生育地群集為生態系經營理念試驗設計之分層是為可行之道。

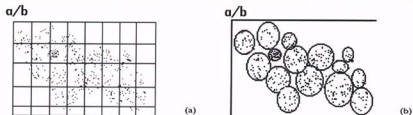


圖 11 雙因子二種取樣分層示意圖

假設雙因子試驗設計因子 a, b 之組成分佈如圖中黑點所示，圖(a)所示為將 a, b 因子分級，其形狀如方形，而可以為取樣單位者共29種因子分級組合，圖(b)則為以 a, b 因子自然組成之群集，其形狀如圓形，將其視為取樣單位，可得15種。群集取樣之樣區數將比各因子分級分層取樣為少，主因子效果及交互作用的效果皆存在於群集間，群集內之變異除少部份選取因子效果，少量機差外，可以藉以探討其他因子。

現有的研究，大部份為小型、某單一領域方面之研究，並未加以整合。而在生態系理念下經營，則樣區重新設計，將相當耗時耗力，且以往花費大量經費所獲得之各類調查資料，束之高閣不用，則未免太可惜且浪費，因此，積極恢復過去研究、經營之樣區、整合過去已有之資料，檢討已有之樣區，實為生態系經營理念下試驗之第一步。

過去已有的資料，因年代不同、目的不同，因此，資料型態、內涵亦可能不同，而其生育地因子之記錄不全，更為應用推廣之大傷，因此建立生育地因子地理資訊資料庫，以供已有資料之生育地因子查詢，且方便生態系經營理念試驗之整合。

而欲查詢過去已有資料之生育地因子，除現場重新調查外，亦可以在生育地因子地理資訊資料庫中查詢，一般可能有四種情況：

1. 從地圖上直接輸出研究區域，處理方式以套疊、轉化出資料給查詢者較為方便。（如查詢一。）
2. 樣區點座標查詢生育地因子及群集代號，處理方式則不經圖形資料庫轉化，直接以座標核對所屬位置代號，讀出生育地因子及生育地群集代號。（如查詢二。）
3. 生育地因子查詢生育地群集代號及位置，因為資料不全或定位座標、尺度不同，而以僅知生育地因子輸入，查詢所屬生育地群集代號，或基繪出位置圖，以利研究者研究。（www上尚未完成，如查詢三，但不需全部的因子。）
4. 混合查詢，包含上述任2種以上一併應用之查詢。

生態系經營應該在多種尺度下探討，所以探討之因子才會如此複雜。而且本研究所探究之因子，包含了氣候、地形、土壤等因子，依

蘇鴻傑(1987)所述陸域生態系中環境影響因子之層次中,可知氣候為大層次生育地之影響因子,地形為中層次生育地之影響因子,而土壤、微氣候等因子則為小層次生育地之影響因子,此亦為本研究所包含之不同尺度的生育地因子,故可應用於不同尺度之資料,且先前採用 K-means 法進行群集分析,可將大尺度(全省)所產生之群集結果製成 seed 檔,以供更新之資料及較小尺度之資料檔群集分析之用,而對於點尺度因子資料不全,亦能以 seed 檔判斷所屬可能群集集合,以供後續合併、分類研究之用。

Seed 檔可謂本資料庫應用之根本,除藉以分類、推測全省性應用,亦能用以針對氣候等因子,再行區分、分類,其結果不若一般以地區分類,呈現外觀完整之地景單元,而在邊界形成破碎之不連續面(如圖 10),不能明顯指示其邊界,因子界限,是其缺點,但在另一面,卻也告知了自然界的模糊特性,分類時之不確定性。

前述多種採用此法之優劣點及原因,但最終仍是資料庫之建立與共享,才能使得生態系經營理念、試驗得以整合,若是各家儘以自行分類系統試驗,在整合上則多有困難,甚至造成剩餘資料,形成浪費。本文不敢以本資料庫為標準,僅希望在本文提供之實例應用下,喚起各界生態研究之先進,一起訂定可行之生育地資料庫,將資料庫上網,並進而建立模式庫。試將本資料庫建立所涉及不定因素列舉於下,以為有關單位生育地資料庫建立之討論。

1. 確立欲收集之生育地因子

本資料庫以王子定(1962)所立之生育地因子分類,及蘇鴻傑(1987)所述環境因子影響層次加以規劃出 12 因子。而各生態領域之研究所專注之生育地因子亦會有些許不同,如何規劃收集因子,使其能有生態環境之代表性,亦有全省面資料調查之可能(如水域深度即不能),為確立欲收集生育地因子之重要考量。

2. 各因子之測定及推估

本資料庫因子資料引用水資會(1989)及林務局(1995)第三次全省森林資源調查之資料,自行推估。然資料來源、準確度、推估方法、及推估參數,對資料的可信度影響頗大,且對每一因子皆需加以測驗,方能找出較正確之推估面,加上往後變遷之更新,實為一繁複之工作。

3. 群集分析之規整化方式

未經規整化之群集分析,會造成對較大變域之因子具較大影響之結果,如許秀英(1995),如此將失去收集眾多因子之原意,因此規整化是為必要,自無疑義,但方式卻也會造成結果差異,本資料庫採百分位法,旨在消除推估時產生之不合理極大(小)值,並增加平均值附近之分類種數,亦即在多變的中低海拔山區,進行較為細緻之分類。

4. 群集個數

群集分析多有提供尋找群集個數之指標(如本資料庫採用 CCC 值),然其計算及反覆群集分析,如同一非線性求解,所得最佳解(群集個數)只是區域最佳解,但這是必然

的，因其全區最佳解（群集個數），必為資料總數，即原母體，如此則失去群集分類之意義。同樣分類個數過多，分割過細，群集分類之意義不大，預調查亦太費時費力；分類個數過少，分類粗略，群集內部變異太大，與同類中均質之期望，相差太大，亦非良好分類。個數反應因子多元及群集內部均質之假設，應加以考量，而林務行政機關方便管理、林業研究機構試驗執行能力、全面調查之可行性，均是應加以考量之外在因素。

5. 驗證結果輸入之檢查

本資料庫雖尚未進行此步驟，仍有一些問題發現，將其提出加以討論：1.查詢資料庫不設限，但對驗證結果輸入應加以設限於相關研究單位，以避免過多無關資料、不正確資料，甚至為害資料庫本體之病毒資料進入；2.驗證結果輸入之檢查，需專業，也需客觀；3.驗證結果輸入之檢查及資料庫維護為長期之工作，而利益為共有，似應納入常態業務，而非以計畫執行，以免計畫完成後，因經費無著，或其他因素而停擺，殊為可惜。

六、結論

在複雜多樣化的生態環境中，欲加以整合生態系經營理念之適應試驗，利用K-means法群集分析進行分類，是為可行且不錯的方法，

礙於生育地因子之多樣化，而所學無法兼顧所有生態系統研究，故僅以所識，建立生育地因子群集資料庫，盼各界先進加以驗證，並提供結果以供更新及他人查詢。

七、參考文獻

1. 王子定 1962 育林學原理 國立臺灣大學叢書 p.324。
2. 林務局 1995 第三次台灣森林資源調查及土地利用調查 台灣省農林廳林務局
3. 高堅泰、馮豐隆 1998 森林生育地因子之空間推估與生育地分類 第二屆兩岸測繪學術研討會暨第十七屆測量學術及應用研討會論文集（二）p.737-747
4. 許秀英 1995 應用群落分析法於林地分類之研究 - 以八仙山事業區為例 宜蘭農工學報 10:19-28。
5. 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) p.47-70
6. 陳永寬、鄭祈全、賴晃宇 1994 數值地形模型應用於潛在崩塌地之預測 第四屆ESRI&ERDAS User Conference 論文集 p.221
7. 經濟部水資源統一規畫委員會 1989 台灣地區雨量記錄
8. 蘇鴻傑 1987 森林生育地因子及其定量評估 中華林學季刊 20(1):1-14