

國立臺灣大學生物資源暨農學院森林環境暨資源學系



碩士論文

School of Forestry and Resource Conservation

College of Bioresources and Agriculture

National Taiwan University

Master Thesis

林業巨量資料應用策略之探討

The Application Strategy for Big Data of Forestry

沈玉婷

Yu-ting Shen

指導教授：邱祈榮博士

Advisor: Chyi-Rong Chiou, Ph.D.

中華民國 104 年 7 月

July 2015

國立臺灣大學碩士學位論文  
口試委員會審定書

林業巨量資料應用策略之探討

The Application Strategy for Big Data of Forestry

本論文係沈玉婷君（學號 R99625037）在國立臺灣大學森林環境暨資源學系研究所完成之碩士學位論文，於民國 104 年 7 月 28 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

邱新宇 (簽名)

賴嘉任 (指導教授)         

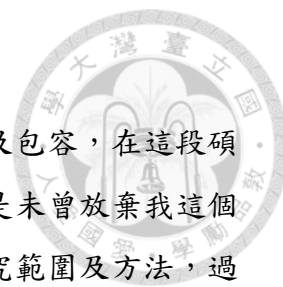
\_\_\_\_\_

\_\_\_\_\_

         (簽名)

系主任、所長

## 謝誌



這篇論文得以完成，由衷感謝恩師邱祈榮教授的教導、協助及包容，在這段碩士班的修習過程當中，因為工作關係常感到力不從心，但老師總是未曾放棄我這個學生，不厭其煩地指導我如何選定一個適宜的研究題目，調整研究範圍及方法，過程也給我許多的打氣及鼓勵，著實讓我受益良多，在此致上最深的謝意。同時感謝口試委員余坤勇教授及賴彥任學長對於論文的詳細審閱，給予我許多專業的建議及論文撰寫上的指導，讓論文架構及內容朝向更為明確、完整，雖然最終修改結果不甚完美，在我有限的時間能力下無法將所有的建議一一納入，但每項建議我都謹記在心。

感謝與我一起修習碩士班課程而認識的好同學們，詩涵、惠如及書勤，能認識妳們真的是件非常開心的事，我們都是在工作後重拾學生身份打拼的人，當中是我最後結束這段歷程，即將再次離開學生身份這個角色，希望我們的未來都能很好。感謝研究室的惇淳學姊，過程中如果沒有妳不斷在背後的幫忙和鼓勵，我真的很難想像怎麼完成這份人生的功課。感謝舉行口試時幫忙我的舒婷、海之及研究室的小學弟妹們，慌亂之中有你們無微不至的幫忙真的讓我安心不少。感謝工作上的同事及長官們，林務局的沈簡任技正、峻銘科長、恬恬、怡平、舒婷、小恬、農糧署北區分署的陳課長…，太多在我修課過程中幫忙我的人，無法一一列舉出來，真的謝謝你們無私的幫忙。

最後，感謝這一路關心我、陪我走過來的家人們，爸爸、媽媽、大姊、大姊夫、二姊、二姊夫及我的老公阿肯，你們是我永遠最深刻的精神支柱。

沈玉婷撰 2015年8月

## 摘要

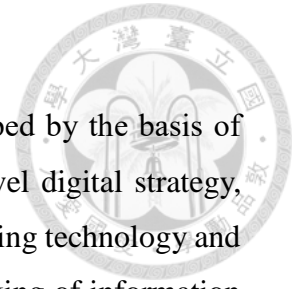
巨量資料 (Big Data) 議題本質上並非一個全新的議題，而是基礎過去近十年間資訊技術的發展，在高度數位化的結果長期累積了大量資料，因為資料量的龐大，使得過去的資料處理技術及應用方式受到挑戰，也使得資訊、知識的擷取有了新的思維。

本研究首先透過文獻回顧，針對巨量資料定義、技術發展及應用等面向進行闡述：一般對於巨量資料的討論，其特性主要可從資料數量、資料速度及資料多樣化等幾個面向來看，其實務上的定義則依組織而異，因應使用者需求不同而有所差異。在技術面，其包含了從資料的蒐集、整合、操作分析及視覺化的呈現等，同時也涉及到多元資料來源的異質性資訊整合串聯；而對應技術上的議題，又可依資料特性可分為動態及靜態巨量資料兩種類別，希望能分別針對資訊效率及資訊擷取能力做提升。而另外在應用面，目前台灣仍處於發展初期，未來需能面對使用者的需求，以決策發展為主要方向；應用發展過程，需考慮以下幾個面向，包含是：資料取得、技術應用、資料政策與組織改變及能力等。

本研究探討國外林業巨量資料應用案例「全球森林觀測網 (GFW) 專案」的執行成果，該專案結合了分布世界各地有關組織之資料及研究人力，提供了森林覆蓋、森林變遷、火災監測及林業資源管理分析等各種決策資訊，進行全球森林資源狀態監測，推動資源保育；台灣林業雖長期持續進行森林資源調查工作，但在即時性變遷資訊的掌握能力相較明顯不足。本文嘗試提出國內林業巨量資料應用發展策略，並以衛星影像變遷偵測為個案分析，討論如何透過跨年期時序資料，觀察地區的植生特性及變遷情形。最後再依巨量資料分析過程所需面對的各項議題，提出國內後續應用發展的建議。

關鍵字：巨量資料、林業應用、全球森林觀察、森林監測、變遷偵測

## Abstract



“Big Data” issue is essentially not an “all new” issue. It developed by the basis of information technology over the past decade. Because of the high-level digital strategy, large amounts of data has been collected. It makes the past data processing technology and application mode to be challenged, and also makes a new way of thinking of information and knowledge capture.

Through this study, we review the property of Big Data: its main characteristics are “data volume”, “data velocity” and “data variety” (3V). Actually the definition varies depending on the needs of users. In the technical part, which contains the process of data collection, integration, analysis and visualization of operational presentation, etc. It also related to the heterogeneity of the sources of information integration. Corresponding the issue of technology, the data characteristics can be classified into dynamic and static two categories, and be expected to enhance the efficiency and information are for information capture capabilities. The other side, in application, it is still in the early stages of development of Big Data currently in Taiwan. There are some main development direction to be considered, including that: data acquisition, technical skill, data policy, and also organizational change and capability.

This study describes Big Data case of forestry applications: “Global Forest Watch (GFW).” The project combines lots of research organizations around the world. It provides information of forest cover, forest change, fires monitoring and decision making tool of forest resource management, in order to monitoring global forest resources and promote resource conservation. In Taiwan, although long-term survey of forest resources, but the ability of get near-real time change information is still insufficient. In this article, using the satellite images try to identify the temporal-changes, discuss how to observe vegetation growth trends from the process of analysis. Finally, try to make suggestions of the future development of Big Data application in Taiwan.

Keywords: Big Data, Forestry Application, Global Forest Watch, Forest Monitoring, Change Detection

# 目錄



中文摘要.....	I
英文摘要.....	II
目錄.....	III
圖目錄.....	IV
表目錄.....	IV
第一章 緒論.....	1
第一節 背景說明.....	1
第二節 研究動機.....	2
第三節 研究方法.....	2
第二章 文獻回顧.....	3
第一節 巨量資料定義.....	3
第二節 巨量資料技術發展.....	4
第三節 巨量資料應用.....	7
第三章 國外林業巨量資料應用案例介紹：全球森林觀察.....	11
第一節 案例介紹.....	11
第二節 案例分析.....	17
第四章 國內應用發展.....	20
第一節 國內林業巨量資料現況.....	20
第二節 林業巨量資料應用之策略架構.....	23
第三節 植生變遷監測之個案分析.....	29
第五章 結論與建議.....	45
第六章 參考文獻.....	49

## 圖目錄

圖 1	YARN 運作流程示意圖.....	7
圖 2	大武山自然保留區 2006 年至 2015 年間 FORMA 發布警訊區位.....	13
圖 3	大武山自然保留區 2009 年至 2010 年森林變遷之損失分析共 1,038 公頃.....	13
圖 4	國家檔案：以日本為例.....	14
圖 5	台灣火災監測點位及風向、空氣品質情形（日期：2015/4/14-21）.....	15
圖 6	林業巨量資料應用策略發展架構.....	28
圖 7	大肚山地區研究範圍示意（背景衛星影像來源：Google Satellite，2015）...30	
圖 8	光譜分析樣點分布位置示意圖.....	31
圖 9	2004/4/5 大肚山地區 SPOT 衛星影像（左）及植生指標（右）計算結果.....	33
圖 10	2004/3/10（左）與 2004/4/5（右）NDVI 值對照.....	34
圖 11	2004/3/10 與 2004/4/5 NDVI 差值頻度分布圖.....	35
圖 12	2004/3/10 與 2004/4/5 植生指標差值依標準化計算結果.....	36
圖 13	依網格大小比較各年期 NDVI 平均值及標準差情形.....	37
圖 14	2004/3/10 與 2004/4/5 NDVI 差值依 50m（左）、100m（右）網格計算.....	38
圖 15	依表 2 圖示各年期 NDVI 標準化差值佔全域百分比之變化情形.....	39
圖 16	2004/3/10 與 2004/4/5 各樣點前後期 NDVI 數值對.....	41
圖 17	清水火燒樣點各時期 NDVI 值依時間序變化情形.....	43
圖 18	橫山火燒樣點各時期 NDVI 值依時間序變化情形.....	43

## 表目錄

表 1	本研究使用之 SPOT 衛星影像資訊.....	30
表 2	依網格大小比較各年期 NDVI 變遷區域識別度（NDVI 變遷差值>1）.....	38
表 3	各類型地景光譜樣點數.....	40



## 第一章 緒論

### 第一節 背景說明

巨量資料 (Big Data) 議題，基礎過去近十年間資訊技術的發展，在高速光纖網路的興起及普及，允許大量的資料透過網路快速的傳輸 (林芳邦, 2013)，而電腦的計算處理速度及存儲性能也不斷提升，高度數位化的結果，全世界每天累積、建置的資料量已達到了百萬兆位元組 (Exabyte)，使得過去的資料處理技術及應用方式受到挑戰，也使得資訊、知識的擷取有了新的思維。

正如資訊化帶動了許多研究及學術領域的新紀元，巨量資料議題的影響範圍也遍及各種領域。在空間資訊領域部分，可從我國政府推動國家級地理資訊系統的建置與整合應用計畫「國土資訊系統 (National Geo-spatial Information System)」來看。國土資訊系統建置工作中，歷年討論的重心隨著資訊領域的新議題不斷變化：計畫一開始著重於空間資料建置，將全國性的地理資訊來源配合行政權責劃分為九大主題分責推動；隨著資料量逐漸累積，為了利於資料管理及流通分享，引入資料倉儲 (Data Warehouse) 概念及技術，建立圖資詮釋資料 (Metadata) 及資訊檢索機制，但因各單位自有倉儲系統及建置資料格式的差異，限制了跨單位間資料即時更新及整合應用；為了加速跨單位間資料流通應用，順應國際間由國際標準組織 (ISO, International Organization for Standardization) TC211 委員會及開放空間資訊協會 (OGC, Open Geo-spatial Consortium) 所提出的各項空間資訊標準逐步被實作，圖資的供應及運用開始朝網路服務發展，並以資訊領域在應用發展上提出之服務導向架構 (SOA, Service Oriented Architecture) 模式為基礎，其方向為：使用端應用定義出所需圖資或功能分析之服務元件，供應端進行服務元件開發，並於仲介端進行服務註冊供使用者整合應用，此架構下，當資料量龐大、需求者眾多時，供應端的服務效能成為相當大的挑戰，如何能將巨量資料 (Big Data) 結合雲端運算 (Cloud Computing) 技術，便成為目前各單位致力推展的目標。

而林業地理資訊屬於國土資訊系統「生態資源資料庫」的一部分，除了持續擴充建置更完整的林業資料庫內容，亦致力於發展出更為完善的資料倉儲及應用分析服務，以支援森林資源經營管理的決策工作。我國對於森林資源現況的掌握，目前主要作法係透過人工航攝影像判釋及現場樣區調查，資料內容固然較為精確，但更新頻率難免



受限；隨著遙測分析技術提升及感測網應用日益廣泛，提供更多元的資料取得管道，有助於更即時性資訊的蒐集及掌握，但此類型的資料量相對龐大且異質，要如何從這些「巨量資料」內獲取真正有用的資訊，讓相關決策人員更有效的管理及利用這些資訊來源，是個值得探討的議題。



## 第二節 研究動機

「巨量資料」議題伴隨著全球性資訊技術的不斷提升及突破，近年受到了廣泛的注目，並已在電子商務及網路社群應用上大放異彩。巨量資料會如此受到重視，在於這些資料分析者相信，看似混亂、龐雜的資料當中，潛藏著一些從表面或少量資料處理過程無法獲取的訊息等待挖掘，而這些潛藏的資料價值將有機會突破領域內現有的思維及視野。

而對應到林業領域而言，「巨量資料」這樣一個課題，與林業發展之間的相關性為何？怎樣類型的資料堪稱為「巨量資料」？目前國內的資料現況，是否已具備跨足「巨量資料」研究領域的可能性？而一旦進入「巨量資料」領域的範疇，後續應用發展的方向又會是什麼？本文即嘗試就上述這些問題進行探討，希望由目前被廣泛討論的「巨量資料」課題，探討其在林業上的應用性，延伸討論在應用面可能存在的貢獻及所面臨的挑戰，期對於國內後續林業巨量資料應用策略提出建議。

## 第三節 研究方法

本研究首先透過文獻回顧，將逐一就國內外文獻就巨量資料的定義、涉及之相關技術及應用課題進行探討，整理出巨量資料的內涵及在應用發展過程可能涉及的各個層面；進一步，經由國外林業巨量資料應用之案例介紹及分析，觀察此案例在應用發展上的特點；再回頭檢視國內林業資料建置現況，參照國外經驗，提出國內未來發展相關的監測資訊平台，在巨量資料導入應用可嘗試的策略方向；同時配合個案分析，嘗試提出一項在森林變遷偵測流程可運用的資訊分析方法，並據以提出最後的結論與建議。



## 第二章 文獻回顧

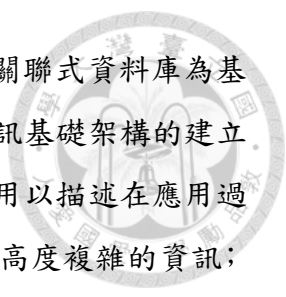
### 第一節 巨量資料定義

Big Data (巨量資料), 亦被稱為在中文譯名常被翻譯為「巨(鉅)量資料」、「海量資料」、「大資料」或「大數據」等(本研究中取用「巨量資料」譯名)。在討論巨量資料時, 普遍均採以 META 集團 (META Group; 現已被資訊科技研究及顧問機構 Gartner 公司收購) 研究員 Doug Laney 於 2001 年提出之 3 種資料管理維度的未來挑戰做為討論架構 (Laney, 2001) (最初發表之文章中簡稱為 3D, 後續延續此架構之研究者多改以 3V 稱之), 包含:

- 一、資料數量 (Data Volume): 為了對訊息充分掌握, 大量的資料不斷被蒐集、累積, 資料儲存媒介及技術是第一步需要克服的部分, 而更進一步, 則在於如何從龐大的資料中找出資料間的關聯性, 篩選擷取獲得真正有用的訊息。
- 二、資料速度 (Data Velocity): 對於資料速度的要求, 包含資料處理存取及後續分析管理, 在實作面的解決策略, 諸如定期將原始資料萃取、重組及整合, 建立快取提升即時處理效能, 並依決策流程設計適合之結構供近即時 (Near Real Time) 的決策分析等。
- 三、資料多樣化 (Data Variety): 資料來源多元且異質, 需能有效率地處理各種不相容的資料格式, 納入非結構化的資料及不一致的資料語義, 讓資料間彼此可以建立關聯性、共同檢索, 進行語義標準化及整合應用。

在 2001 年的研究報告中, Laney 於文中並未直接提出「巨量資料」一詞; 而至 2012 年, Gartner 基礎於 3V 的架構, 針對「巨量資料」提出更明確的定義, 指巨量資料為大量 (High Volume)、高速 (High Velocity) 及多樣化 (High Variety) 的資訊資產, 需要採用新的處理形式, 以發揮更強的決策能力、洞察力及處理過程的最佳化 (Beyer et al., 2012)。

許多國際資訊公司及組織, 也都分別提出其對於巨量資料的看法及定義, 列舉如下 (Ward et al., 2013): 國際商業機器公司 (IBM, International Business Machine Corporation) 在定義巨量資料時, 採用延伸自 3V 特性的 4V 架構, 其第 4 個 V (Veracity) 指的即是資料真實性, 指出巨量資料的龐雜資訊中存在不確定性, 需要能過濾這些偏誤的資料以維持資訊的正確性; 以發展大型資料庫軟體為營運核心的甲骨文公司



(Oracle) 另從實作角度，提出巨量資料是針對業務決策由傳統以關聯式資料庫為基礎之應用，進一步將非結構化的資料來源納入分析，將重點放在資訊基礎架構的建立及應用；微軟公司 (Microsoft) 對於「巨量資料」一詞的定義，是用以描述在應用過程需要繁重的運算能力(像是機器學習及人工智慧等)之大量且多半高度複雜的資訊；另由開放源社群維繫的整合性知識環境方法論 (The Method for an Integrated Knowledge Environment) Mike2.0 專案中，對於巨量資料提出了另一種反向思考的觀點，「巨量資料的資料量可以很小，且並非所有大型資料集都具有巨量特性 (Big Data can be small and not all large datasets are big)」，將「巨量」的概念定義在資訊處理上的複雜程度，而非單純以資料量衡量；美國國家標準與技術研究院 NIST(National Institute of Standards and Technology) 提出與 Mike 專案類似的觀點，認為巨量資料指的超出現有方法及系統運算能力所能處理的資料，將「巨量」視為一種相對的概念，並非新的觀點或技術，隨著技術不斷演進、資料量不斷增加，永遠都存在巨量資料的挑戰。

儘管上述在巨量資料的定義，不同組織間其討論範圍及觀點均有差異，其核心的論點均不出：資料量、資料複雜度及分析技術能力等 3 個面向。當組織現有的資料其資料數量或複雜度，超出既有的分析處理能力，或需突破既有的分析模式，以獲得更有效率及潛在決策資訊，即可視為一項巨量資料的處理議題。

## 第二節 巨量資料技術發展

巨量資料使用到的技術層面相當廣泛，而新的技術亦持續在發展及精進當中，包含在巨量資料的整合、操作分析及視覺化 (Visualize)，結合統計學及電腦科學及各專業領域，屬於跨領域的應用 (Manyika et al., 2011)。

巨量資料屬於資料科學 (Data Science) 的一部分，資料科學早期又稱為資料學 (Datalogy)，該名詞於 1996 年在國際會議 (國際分類協會聯合會 (IFCS, International Federation of Classification Societies) 東京雙年會議) 正式提列為討論議題，定義為研究自資料中萃取知識的科學 (the Extraction of Knowledge from Data)。

巨量資料興起，包含了遙測探測技術、無線感測網路、相機及監視器及無線射頻辨識等的發展，同時配合儲存技術的發展，讓資訊的蒐集可以更快速、更即時，使得各資訊的來源越廣，所可以提供的資訊也更多。




而從空間資訊科學來看，無論是在自動製圖、遙感探測及或其他各種地理資訊的呈現及處理，巨量資料的本質已幾乎成為通則，其面對的課題像是：需發布、整合應用不同感測器（遙測衛星）來源取得的遙測影像，對於資料取得方法之描述、以至處理校正過程資訊，都成為資料正確應用的必要條件，對應的作法像是依循圖書館學提出的都柏林核心集（Dublin Core）及 ISO 系列等詮釋資料標準建置、註冊及分享作業等；而此時地理資料處理（Geo-processing）由資料分析進入資料探勘（Data Mining），透過機器學習（Machine-learning）的方法，以利擷取龐大資料內隱藏的關聯性；又地理資訊是對真實世界建模量測的結果，當資訊來源於各種不同的方法、尺度及觀點，透過資料品質指標（DQI, Data Quality Indicators）的建立，可提供資料使用的準則及限制資訊，亦需各種演算法或工具將資料重整，以針對這些龐大且具不確定性的資料取得一致性的描述，轉化為知識；透過知識本體（Ontology）的建立，將專業領域的知識架建立起來，將各個來源的資料、表面看來異質性的資料、甚至決策上面對到的真實問題，都可透過知識類別對應、串聯起來，以促成知識層面的整合利用（Jeansoulin, 2015）。

引用 IDC 的定義（Olofson, 2012），其將巨量資料技術（Big Data Technologies）定位為新一代的技術及架構，設計用來有效率地萃取數量龐大且廣泛多樣化的資料，並允許操作過程能高速的存取、探索及分析，包含了硬體、軟體及服務層面以整合、組織、管理、分析及呈現（Visualization）資料。基礎於巨量資料基本特性的 3V（Volume, Velocity and Variety）架構外，另增加 1 個 V（Value），指的是資料價值，其價值的評斷包含了技術投入成本及資料使用效益兩個層面，其中技術成本是相當關鍵的因子，需考量投入軟硬體等基礎資訊架構耗費的資金成本、提升資料操作效率所減省的人力成本及強化組織內部流程的經營成本。

而在討論巨量資料的技術時，亦可將其區分為 2 種類別（Olofson, 2012）：

- 一、動態巨量資料（Big Data in Motion）：指的是大量、快速串流的資料，需要於資料接收的當下被立即過濾及正規化，並進一步反應、回饋事件的發生，如複合事件處理（CEP, Complex Event Processing）的系統概念。
- 二、靜態巨量資料（Big Data at Rest）：多數對於巨量資料的討論則是在靜態性的資料處理技術，能在最有效率的狀態下，就蒐集資料進行重整、轉化分析，讓資



料維持一定的狀態，提供有意義的搜尋、探勘及問題分析。資料處理包含了結構化（如關聯式資料庫）及非結構化資料（如 XML 或一般使用的自然語言），關聯式資料庫一般通用的查詢語言為 SQL 指令，因此在處理非關聯式資料庫的查詢語法則統稱為 NoSQL，目前被提出來討論的技術包含：Key-Value、Grid-based key-value、List-processing、Graph 等，而現行被廣為使用的 Hadoop 便是其中一種解決方案。

Hadoop 專案是由 Apache 軟體基金會發展的一套開放源(Open-source)軟體框架，以 JAVA 為開發語言，可在跨平台作業系統環境運作，支援電腦叢集(Computer Clusters)之間大型資料集的分散式運算，並允許偵測容錯。Hadoop 顯然成為近年巨量資料處理的主流技術解決方案，目前包含 Yahoo、Google、Amazon 及 Facebook 等知名網站均採用 Hadoop 為基礎進行開發。其名稱來源並非是任何英文縮寫的簡稱，而是以原始作者小孩的一隻黃色絨毛小象玩具為名，黃色小象目前也成為 Hadoop 的吉祥物。其程式框架包含許多子模組及關聯的開發方案，其中 HDFS(Hadoop 分散式檔案系統，Hadoop Distributed File System) 為處理資料儲存議題的核心，而 MapReduce 及 YARN（亦稱為 MRv2，即 MapReduce 2.0）部分則主要在於資料分散式運算議題。

HDFS 的分散式檔案儲存方式，係將檔案分割成許多特定大小的區塊 (block)，其實體檔案分割之區塊同時以複本方式分散存放於不同的資料節點 (Data Node)，而透過名稱節點 (Name Node) 處理檔案系統的詮釋資料，即管理檔案與各區塊間、以及各區塊與各資料節點間的對應關係，控制跨節點間的資料讀寫流程；其複本分散存放的作法，亦提升其資料之安全性，可因應硬體故障情形下搜尋複本自動回復。而 MapReduce 則是由 Google 提出的演算法，可用來協調電腦間資源配置，達到程式分散式的運算處理，其中 Map 及 Reduce 均為一種演算函數，Map 函數的運算模式，是將輸入的鍵值 (Key-value) 序列透過相同的運算輸出為另一組的中繼 (Intermediate) 鍵值序列，Reduce 函數則是負責將相同的中繼鍵合併其所有相關聯的中繼值；過程中由使用者進行函數的定義，而 MapReduce 框架則可進行後端電腦叢集間的的平行運算，加速處理資料處理效能。YARN 則是基礎於 MapReduce 之上，主要在進行整個分析流程的工作計劃 (Job Scheduling) 及電腦叢集資源的配置管理。其處理流程可參考下圖所示，由資源管理者 (Resource Manager) 進行整體工作計劃，而主要應用

(ApplicationMaster)係與節點管理者(NodeManager)共同運作，負責每個 Map-Reduce 工作的實際運作，其運算分配並依檔案實際存放架構進行配置 (圖 1)：

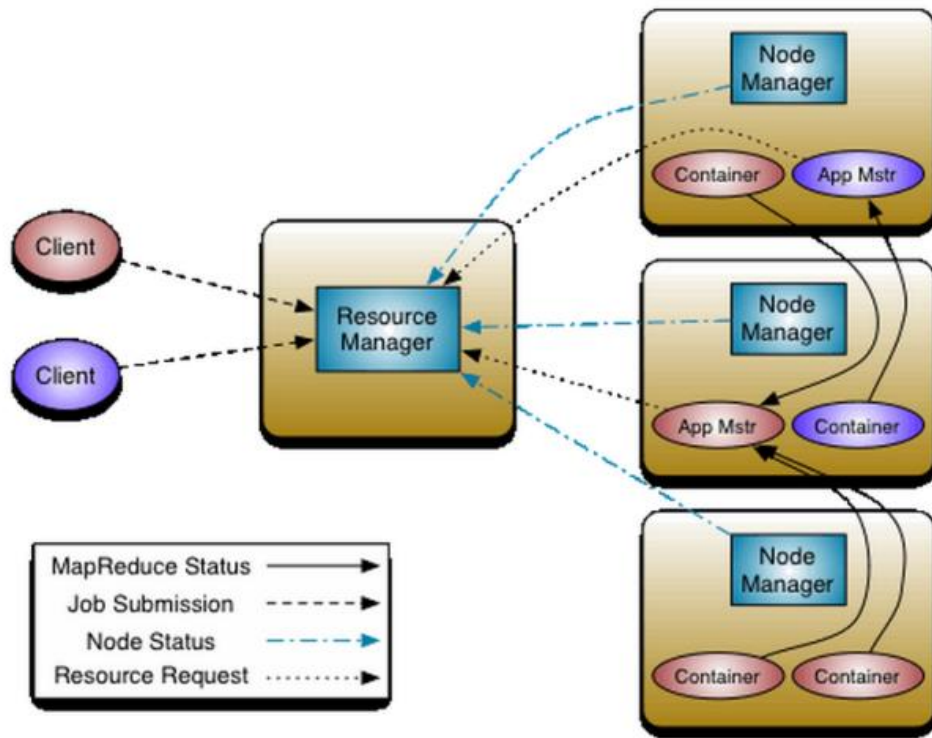


圖 1 YARN 運作流程示意圖

(資料來源：Hadoop 官方文件網站)

### 第三節 巨量資料應用

「物聯網」及「機器對機器 (M2M, Machine to Machine)」是創造巨量資料的重要來源，透過諸如感應裝置、行動裝置、穿戴式裝置及嵌入式裝置等，資料被大量的創造及取得；要從資料轉化為資訊再晉升為實際有用的「情報」，包含了從資料創造、資料蒐集 (儲存、萃取)、資訊分析及情報決策等 4 個階段。從產業的觀點，2013 年全球巨量資料產業價值鏈市場規模排名前 3 名之次產業，分別為「決策」專業服務、「蒐集」運算力及「儲存」設備，而台灣產業目前在使用者端、行動裝置或伺服器軟硬體部分的整合創造能力尚可，但在巨量資料庫的建置、演算法理論 (如資料探勘及機器學習等) 及由應用需求產出分析模型及專業分析等部分則較弱，重硬輕軟的產業

價值觀，限縮了相關產業的創新服務思維，並巨量資料應用重點之一是在處理大量的非結構化資料、從中獲取人工無法輕易判讀的資訊，而此部分目前在台灣也未被廣泛蒐集及分析（童啟晟，2014）。由趨勢可知，台灣目前仍處於巨量資料導入初期的發展階段，未來勢必也需將重點逐步落在決策服務部分，讓巨量資料應用價值有效發揮。

一般常見對於巨量資料的誤解（Olofson, 2012），像是認為巨量資料分析是一項全新的任務，而事實上巨量資料概念已經討論多年，改變的是近年網際網路高速發展允許跨電腦間相互存取支援，讓更多資料來源能相互連結應用，以及人們開始察覺到巨量資料應用能帶來的新的競爭優勢；而另一方面的誤解，則是認為巨量資料主要討論的議題就在「資料」本身，是對於資料探勘新的流行語，或只要透過架構起足以承載巨量資料運作之軟硬體環境，就成功的解決巨量資料議題，而正如前述所言，巨量資料同時涉及到資料蒐集、管理、組織及分析等環節，不僅在於軟硬體的架構及部署，而技術方案的實作也只是巨量資料分析前哨，更重要的在於在使用者需求及未來趨勢資訊分析的掌握。

對於擁有巨量資料的政府部門、企機或機構組織，資料收集多數基礎於特定目的，由因、果關係蒐集而來的資料，過去往往侷限於特定用途，聚焦在表象的功用，未進一步加值及延伸利用，當過去單純的數字分析，要走向圖片、影像、聲音及文字等各種形式，從有限的因果分析擴大成不同變因項間的關聯分析，實需要有嶄新的思維面對此一領域的發展；而要成功應用巨量資料，成功的關鍵因素至少包含了：(1)資料的可獲得性、取得成本及使用技能、(2)以巨量資料達成設定目標所需時間、(3)能克服障礙而快速有效的處理巨量資料、(4)及能提出達成使用者需求的有效解決方案與結果（楊純明，2014）。

農業試驗所楊純明博士在其「鉅量資料及其農業生產應用潛力」一文中，提出巨量資料在農業應用上的可行方向，像是：透過生物基因資訊技術（基因體(組)學，Genomics），配合栽培管理提升農作物對於生物或非生物逆境之耐受力；整合氣象資料、遙測與地理資訊、資通訊技術及環境控制技術，達到農業生產環境監控穩定生產，進一步併聯農作物遺傳及栽培管理資訊，建構精準的農業生產系統；結合農產品生產體系及各項食品機能、營養資料，健全安全農業等。（楊純明，2014）

而另一方面，資通訊科技帶來科學探索工具及方法的變革，最直接影響在「觀測」

層面，數位化的觀測連帶迅速大量的資料產出，像是衛星遙測及環境監測等，也帶動了以資料為導向的資訊架構發展。

以全球湖泊生態觀測網 (GLEON) 為例，其最初目的在以全球性均質化的湖泊觀測浮標載具，透過雙向網路連結的環境生態觀測監測器系統平台，以達到自動化資料蒐集、遠端操作監測器及監測資料集中管理與開放性資料分享，我國國家高速網路中心亦以鴛鴦湖監測參與其中，近 10 年發展下來，全球湖泊生態觀測網已達到全球 25,000 個湖泊資料連結，並由 10 個以上的國際監測站同步提供即時監測資料服務，並建構 CONDOR 平台提供全球性計算分析之設施服務，並將資料整合至單一入口湖泊庫 (Lakebase)。資料導向架構之資訊架構，很重要的一個層面則在於知識的萃取，國家高速網路中心自 2004 年底與墾丁核三廠、中研院及海生館合作，以核三廠入水口之海底珊瑚礁生態長期監測，以每小時 20GB 累積的海底觀測影像，英義荷等地研究單位合作，以這些大量的影像資料為基礎，發展自動化海底觀測影像海洋魚類萃取 (包含魚類種類及在高濁度水質、藻類及植物等環境內辨識分布情形)、海洋魚類行為分析及智能介面發展等技術，建立知識本體，將欲進行分析相關生物科學議題之查詢描述，進一步轉譯系統語彙，並轉化為影像處理序列及資料存取流程。而台灣由國家高速網路中心建置地球科學觀測知識庫 (ESOK, Earth Science Observation Knowledge)，整合國家研究院大型觀測資料並提供高階查詢服務，包含海洋中心之海洋環境資料、太空中心之福衛二號遙測影像及三號全球大氣層及電離層觀測資料，颱洪中心大氣水文觀測資料及地震中心之強震測站場址工程地質資料庫等，支援全球地球觀測系統 (GEOSS, Global Earth Observation System of System)，其建置架構從前端與各觀測資料蒐集單位之資料交換做為整合性資料典藏保存，進行屬性分類、品質管控及審查回饋等，資料基礎設施如智慧化之資料處理工作流程及資料管理詮釋，並提供資料服務像是展示、供應及應用介面，提供各項研究應用 (林芳邦, 2013)。

應用巨量資料時，需思考下列幾個面向的議題 (Manyika et al., 2011)：

- 一、資料取得 (Access to Data)：巨量資料往往需要整合更多的資料來源，而資料如何取得、以及資料取得後得否將相關資訊進一步分享他人應用，都是應用過程需面對的議題。
- 二、科技及技術 (Technology and Techniques)：巨量資料的處理分析需要運用到新



的科技及分析技術，組織資料建置的成熟度會影響相關科技及技術的適用性，同時也需透過創意讓相關的科技及技術在應用領域內能充分發揮。

三、資料政策 (Data Policy)：如巨量資料涉及個人資料，像探討公眾健康議題，如何在應用時同時保護資料安全性及個人隱私便相當重要；而也因資料來源多元且廣泛，資料應用的合法性、智慧財產權的保障及誰應負起資料(訊)正確性的責任等都應被釐清，以發揮資料應用的潛在價值。

四、組織的改變及能力 (Organizational Change and Talent) 及產業結構 (Industry Structure)：在面對巨量資料的洪流，組織必需要有新的思考及處理能力，才能獲得巨量資料帶來的價值，提升組織的決策能力；其中，組織的領導者及決策者扮演相當重要的角色，當其需要能去思考如何將組織的運作模式帶入巨量資料的世界，才能讓組織改變並嘗試從巨量資料獲得其潛在的價值。

近年巨量資料的價值逐漸受到全球性的關注，而台灣政府方面，由行政院科技會報辦公室，自 2012 起正式將「公開資料 (Open Data)」列為焦點工作，相關政策的推動及重視，提供了台灣社會巨量資料發展的機會；以空間資訊而言，近年智慧生活及 SoLoMo (Social, Local, Mobile, 結合社交及適地性的行動服務) 等創新服務應用的興起，讓空間資訊的商機無限，而在台灣的空間資訊產業環境，除民間業者與關聯產業在空間資訊軟硬體及應用增值服務的開發，由政府端推動國土資訊系統建置計畫在公開資料政策下，亦開始重視民間增值，如何運用現有空間資訊發展創新服務模式成為關鍵 (林蔚君，2013)。

### 第三章 國外林業巨量資料應用案例介紹：全球森林觀察

#### 第一節 案例介紹

「全球森林觀察 (GFW, Global Forest Watch)」專案由世界資源組織 (World Resources Institute) 自 1997 年起開始推動，最初即以建置全球森林監測網絡目標，希望透過呈現全世界的森林資源空間分布狀況，讓世界各地的森林經營者的決策更為透明，促使森林資源被更妥善的利用及保育，此專案之推動，亦讓許多大面積、完整的原始森林區域得以受到保存。

近年結合衛星科技及雲端運算技術發展，由國際間超過 60 個參與者共同合作，其間包含 Google、ESRI、DigitalGlobe、NASA 及各國進行土地利用與森林保育組織之技術合作及資料提供者，全球森林觀察專案於 2014 年 2 月間發表了新的資訊平台，建立了一個互動式的線上森林監測及警示平台，提供歷年時間序列之全世界森林資源空間分布狀況及變化資訊，並對於疑似森林覆蓋消失之地點發出近即時的警訊。

平台主要內涵依其功能設計架構及提供之資訊，可分成以下幾個區塊來看：

##### 一、全球森林觀察互動地圖 (GFW Interactive Map)：

在此地圖平台功能，提供使用者瀏覽全球各地森林資源分布及近即時之變遷警示資訊，供使用者依關注的地點、議題進行分析查詢，並展示全球森林資源概況及利用管理等分布資訊。

目前系統內涵蓋的主題資訊，在森林覆蓋 (Forest Cover) 部分，包含了全球性森林分布區域、現存尚未受人為破壞的完整原始森林區域、熱帶雨林碳匯儲量及紅樹林分布等；森林利用 (Forest Use) 部分，呈現了像是伐採、採礦、棕櫚油及木材纖維造林等人為活動分布區域；保育 (Conservation) 部分，呈現世界各地保育區劃設、生物多樣性熱點及瀕危物種現存分布區域；同時亦將世界人口 (People) 分布密度、部分國家現有建立之少數民族或原住民特設之土地資源權利等範圍共同呈現，提供對應分析森林資源管理情形。

而在森林變遷 (Forest Change) 部分，經由 Landsat 多光譜衛星影像分析產生的林木覆蓋變遷 (Tree Cover Change) 資訊，其採用監督式學習演算法，以光譜特徵建立模式，偵測識別各影像網格 (grid, 地面解析度約 30 公尺) 森林覆蓋百分比，配合時序性分析，以年度為頻率呈現林木覆蓋範圍及變化情形；

為利於使用者分析，因應使用者的觀察尺度(依地圖縮放比例)，透過顏色深淺，呈現冠層密度變化在地理分布的集中程度，讓使用者容易識別潛在大規模森林減少的區位，並可由使用者依時間序列比較不同起迄年度間的森林變遷資訊，及調整欲呈現為森林覆蓋增加、減少的門檻值(森林覆蓋變化超過多少百分比方視為「變遷」資訊)，讓分析功能更具彈性。目前系統所提供 2000 年至 2013 年的森林變遷資訊，所處理之原始資料總計已超過近 100 萬張的 Landsat 影像(依時間點不同，包含了 Landsat 系列 5、7 及 8 號衛星等)。

進一步，系統亦串聯呈現數個研究組織所發布之林木覆蓋減少警訊(Tree Cover Loss Alerts)資訊，其多針對特定的觀測區域如拉丁美洲、熱帶雨林、巴西亞馬遜流域等，以較密集的時間頻率，像是每季、每月甚至每週發布特定區域內監測到林木覆蓋面積減少的訊息，每項資訊發布均獨立聲明其採用之原始資料來源及分析演算方法。以 FORMA 專案而言，其原始資料來源為每日取得 NASA 之 Terra 及 Aqua 衛星搭載 MODIS 感測器獲取的光譜資訊，以月為頻率，由各網格 NDVI 值變化分析其林木覆蓋情形，透過此監測方法可能造成判斷為林木覆蓋減少的情形包含：森林皆伐、火災及落葉等。此項資料解析度為 500 公尺網格，並於林木覆蓋減少比例超過 50% 時做為發布警訊的門檻值。

不同的主題圖層之間，除了以地圖圖層套疊瀏覽方式供使用者觀察森林資源概況，亦可進行套疊分析。下圖 2 以台灣大武山自然保留區為例(目前該系統所建置之台灣保護區域體系資料尚不完全，涵蓋部分國家公園、自然保留區及國家風景區範圍等)，透過選定關注的區域，可進行區域內森林變遷警訊的分析，可知在 2006 年至 2015 年間 FORMA 專案共偵測到 226 點疑似森林消失的警訊區位；而圖 3 則就同樣區域依時序就森林損失面積進行觀察，可發現於 2009 年至 2010 年間大武山自然保留區就有近 1,038 公頃面積的森林覆蓋損失(設定以林冠密度減少超過 50% 為門檻值)，而期間 2009 年正是莫拉克颱風(八八風災)發生的時間點。

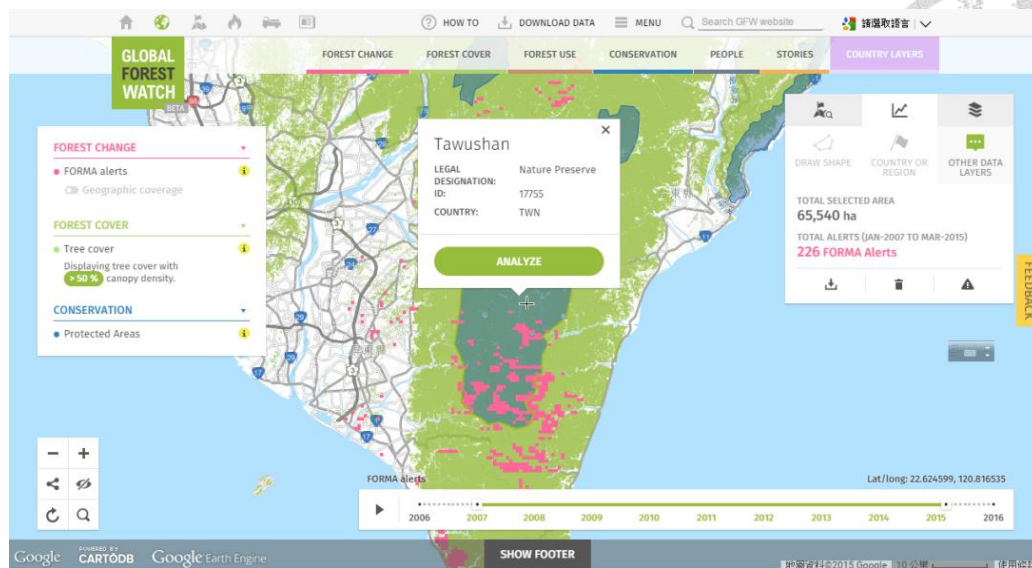


圖 2 大武山自然保留區 2006 年至 2015 年間 FORMA 發布警訊區位  
(資料來源：全球森林觀察網)

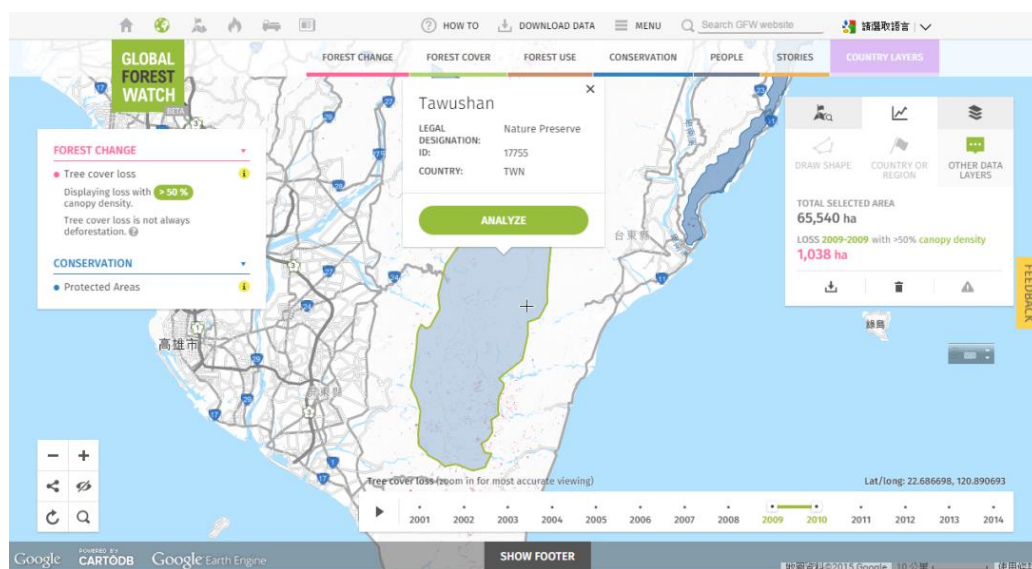


圖 3 太武山自然保留區 2009 年至 2010 年森林變遷之損失分析共 1,038 公頃  
(資料來源：全球森林觀察網)

## 二、國家檔案及評比 (Country Profile and Ranking)：

此功能係以國家為角度(可再區分針對各行政區域進行分析)，結合前述互動地圖之各項主題資訊，以空間統計計算出各國森林覆蓋及面積變遷資料，同時整合呈現聯合國糧農組織 (FAO, Food and Agriculture Organization of the

United Nations) 各國林業相關之基礎統計資料，包含了森林類型（原生林、再生林及人工林）、林業經濟產值、林業就業人口、各國森林資源權利分配情形及土地變遷與林業經營佔該國溫室氣體排放量之比例等，供作了解各國家林業政策及森林資源利用情形。目前並非每個國家都已建立完整檔案資料，而台灣非聯合國之一員，並未建立國家檔案。以鄰近國家日本呈現如下：



圖 4 國家檔案：以日本為例

(資料來源：全球森林觀察網)

此網站中並就各國林木覆蓋及其增減情形進行全球的評比排名；同樣從日本來看，依據此平台計算結果，2000 年至 2013 年間日本國境內林木覆蓋損失比例排名為 151 名計 1.69%（44 萬 5,960 公頃；以全球 203 個國家進行排序），增加面積為第 33 名（2 萬 1,430 公頃）。

### 三、森林火災（Forest Fire）

此功能中有關森林火災的動態資訊主要由美國國家太空總署（NASA）提供，透過以 MODIS 衛星影像分析（解析度為 1km），呈現近 7 日內監測的火災區位，並以每日為更新頻率，目前資料建置範圍主要在東南亞範圍印尼（Indonesia）一帶；而在資料聲明中，也特別點出受限於衛星影像尺度、拍攝時間及雲遮等情形，並非所有的火災發生均能完整掌握。在觀測地圖中，同時整合呈現前述的森林利用、保育區域、土地覆蓋等資訊，並提供空氣品質包含每日風向資訊，可供使用者觀察火災發生、風向及空氣品質間的關聯性。使用者可以針對特定區域（於圖上描繪或上傳 shapefile 檔案）註冊，以電子郵件或簡訊方式獲取火災發生的警訊通知，即時獲知火災發生。

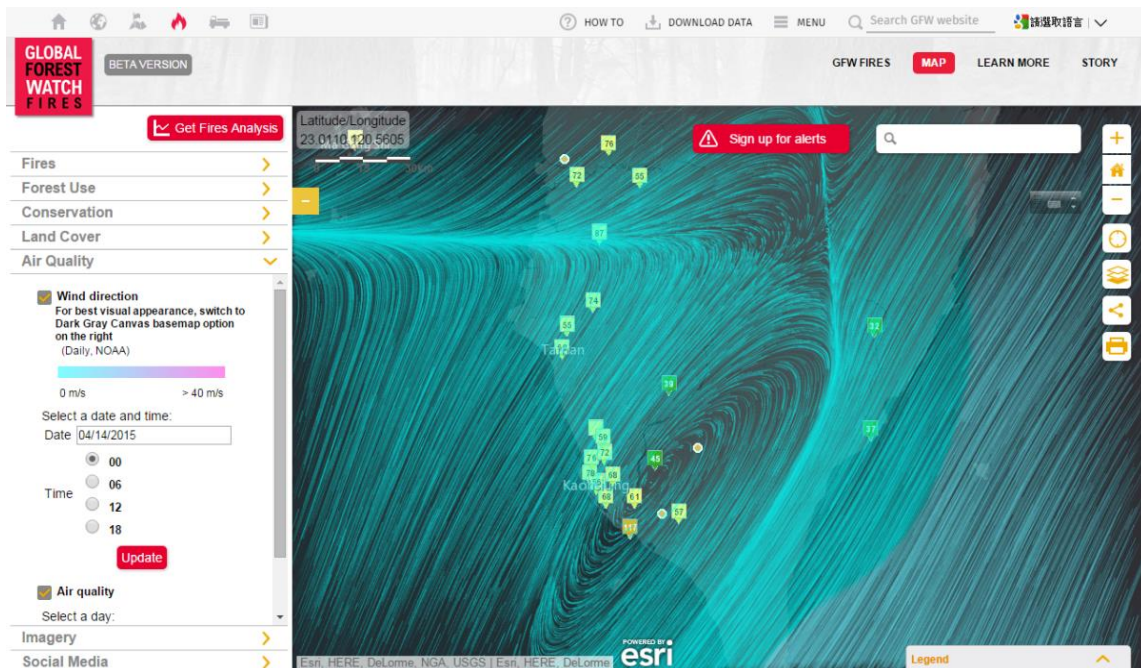


圖 5 台灣火災監測點位及風向、空氣品質情形（日期：2015/4/14-21）

（資料來源：全球森林觀察網）



#### 四、商品 (Commodity)

GFW 商品提供的是較具主題性的客製化分析工具，包含森林變遷分析 (Forest Change Analysis)、適宜性製圖 (Suitability Mapper) 及 RSPO 工具 (Round-table on Sustainable Palm Oil tools)。在森林變遷分析中，提供空間及時間軸的分析角度，除了在互動地圖以全球的觀察角度提供之基礎資訊外，增加由使用者依行自行區劃範圍及上傳 shapefile 檔進行分析之工具，並提供更多區域性資訊；而在適宜性製圖及 RSPO 工具應用上，目前以棕櫚油之生產區位分析為主，配合土壤特性、海拔、坡度、雨量及保育區域劃設情形等，繪製出潛在適宜區位，做為資源合理應用區劃之參考。

此外，本平台並設計了許多使用者參與的介面，可供使用者反映及利用其發展之電腦及行動載具應用系統 APP (此功能尚在發展當中)，供使用者透過此 GFW 平台專案，從地面觀測結果給予資訊的回饋修正；使用者並可運用平台上所展示的森林覆蓋、變遷資訊，結合地圖空間區位的描繪標註，透過故事描述介紹實際發生在各地的林業故事，透過發布即時展現於 GFW 互動地圖中。

GFW 同時發展了公開的應用程式介面 (API, Application Programming Interface) 及開放程式碼 (Open Source Code)，程式碼透過 MIT License 授權條款，被授權人可自由散布、複製、合併、出版發行、和販售軟體與軟體副本，但需包含以上版權聲明和本許可聲明，以相同方式授權；而其程式碼透過 GitHub (版本控管伺服器) 開放原始碼。其網站上同時對於平台所提供的資料發表其資料政策聲明，包含：針對 GFW 所產製的資料成果，包含主題地圖及統計圖表等，其開放資料政策採行創用 CC 授權 4.0 (Creative Commons BY 4.0)，允許使用者下載、分享及重製相關資料，並可用於營利及非營利的使用用途；而針對由其他組織所發展出來的資料項目，則另以連結方式供使用者原始的資料下載途徑，資料使用上則遵循原始提供者的授權模式。使用者在相關資料運用上需依循其所訂定之使用者合約，確保資料合法、合理運用，並應就其資料使用負責。



## 第二節 案例分析

承上一章節的討論內容，針對「全球森林觀察網」在林業巨量資料應用的具體案例，可就其發展特點歸納如下：

### 一、資料面

#### (一)大範圍、多方來源之資料整合

在巨量資料應用的實作上，本平台成功彙集了全球森林覆蓋及時序變遷資訊，應用遙測資訊為主要技術，展示以衛星影像估算而得之全球森林覆蓋面積及火災警示區位，有效透過大尺度、全球性的視野，達到森林經營的決策輔助及國際監督效果。平台現有資訊類別包含從資源的現況、時序監測及利用保育情形等，其中又以目前最受到人類活動發展威脅的熱帶雨林區域為發展重點；部分資訊的蒐集建置集中於特定區域，像是油棕櫚的栽植區域，針對拉丁美洲區域發布的每月森林覆蓋減少警訊等，對於特定應用需求的使用者可獲取更進一步的資訊。

#### (二)開放的資料政策

本平台提供了各項資料集、程式碼及應用程式介面（Application Program Interface, API）的下載及引用，同時針對平台上發布的各項資訊，均建立了完整的詮釋資料資訊，包含從原始分析的資料來源、資訊發布的權責單位、資料解析度、建置時間、更新頻率、資料使用限制及引用授權資訊等，利於資料分享及再加值利用。

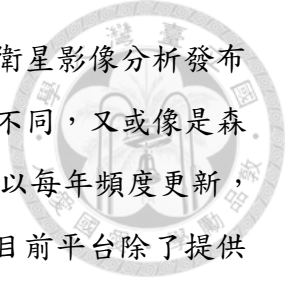
### 二、系統面

#### (一)網際網路地理資訊系統（WebGIS）技術

平台提供之資訊，主要均透過 WebGIS 技術進行展示及發布。WebGIS 的發展，突破了過去使用者端需購置昂貴的 GIS 軟體及運算效能強大的硬體設備，僅需具備輕量型的網際網路瀏覽器，即可進行 GIS 資訊的套疊查看及分析，讓資訊的應用更普及；而 GIS 視覺化的優勢，讓決策資訊更為直觀，符合全球森林觀察專案的「觀察」本質。同時，基礎於地理空間分布的資訊整合，可掌握森林資源與生育環境或利用管理空間關聯性之分析。

其中一點值得思考的是，平台中對於各項資料的整合工作，尚止於同一





系統平台之「共同呈現」，像是由不同機構運用各自取得的衛星影像分析發布的森林面積損失警訊，其空間解析度不同、解析的結果也不同，又或像是森林火災監測（以每日為頻度更新）與森林面積損失監測（以每年頻度更新，部分區域更新頻度可達到每月）間可能存在的生態意義，目前平台除了提供空間分布共同套疊外，並未就分析結果之關聯性進行處理。

## (二)線上即時互動分析及使用者參與

平台除了以線上地圖方式展示已完成處理分析的全球森林變遷及火災監測資訊，並可由使用者依感興趣的地理區位，進行基礎的空間統計分析，其分析流程的參數設定相對較為彈性，包含可選定不同時間點的時序比較，以及選定不同的森林覆蓋變遷之門檻值（視使用者需求定義，同一網格單元內林冠密度變遷比例達多少的百分比視為發生變遷）；而同時，對於感興趣的區位，經由資訊註冊可取得後續的監測資訊發布服務，提升使用者重複造訪率。而此平台在地圖展示效能的表現上也相當優異，即使是在全球性的森林覆蓋瀏覽過程中，也都可達到立即性的回應，這在使用者系統接受度有相當程度的影響。


## 三、組織面

### (一)跨產官學組織之合作分工

平台無論從技術發展及資料提供，參與者均相當多元，包含如民間產業界 Google、ESRI、Digital-Globe、各國非營利組織如英國熱帶雨林基金會（Rainforest Foundation UK）、國際級組織如國際自然保護聯盟 IUCN、聯合國糧食及農業組織 FAO、國家級機構如美國國家航空暨太空總署 NASA、學研單位如美國哥倫比亞大學等，超過 60 個以上的合作者共同參與，也因此具有整合展示全球森林資源及經營資訊的條件及能力。

### (二)目標任務設定

「全球森林觀察」專案的目標任務在於建立全球森林監測網絡，協助世界各地的人去「看（Watch）」這片土地的森林狀態，包含選定以 WebGIS 技術為發展核心，整合遍及全球的龐大資訊，透過視覺化直觀的呈現，使用者不需自行投入高額的研究成本、也不需具備高深的技術門檻，透過此平台可



取得透明化的公開資訊，讓使用者易於參與共同投入森林資源的監測，以促進森林資源在更廣泛的監督下妥善利用。此專案不僅在技術的發展及平台的開發投入，並透過使用者需求的結合，策略性達成組織運作目標，對於未來巨量資料應用發展上實具有參考性。

#### 四、應用面

本平台的應用發展上，從巨量資料性質來看，包含了靜態的歷史時序森林分布資料及動態的即時林火監測警訊，對於不同的應用需求，所需的原始資料、對應資訊更新頻度及分析過程設計上均有不同，同時也反映在訊息的呈現展示方式。延續前述在資料、系統及組織面各項課題，如何設定明確的專案目標，結合跨組織的合作，達成巨量資料整合、設計具有高可用度的系統將資訊妥適呈現及推動後續之資訊公開，可做為台灣發展類似應用時的參考作法。



## 第四章 國內應用發展

### 第一節 國內林業巨量資料現況

由第二章文獻討論中，綜合各界對於巨量資料的討論，可知巨量資料之定義是隨組織而異的，需視組織本身的應用目的、處理的資料特性以及領域內現有的技術能力不同，去定義在導入巨量資料應用時所需處理的議題。


而從林業的觀點來看，森林資源本身就是一個動態、長期且複合性的生態系統，台灣整體面積雖然不大，但以單位面積而言，生物多樣性相當高，要能完整建立國內森林資源狀況及監測變化情形，在資料蒐集及分析處理上仍具相當的複雜度。以下從巨量資料的 3V (Volume、Velocity、Variety) 特性，針對國內林業資料數量、處理速度或多樣化資料型態等議題，分別討論如下：

#### 一、資料數量：

林業資料數量，相當程度反映在資源調查作業上，需要長時間尺度的資料觀測及蒐集，以確實觀察掌握生態系的動態變化情形，甚而在相關經營管理過程做出合理的決策。而在空間尺度上，從全島性的航遙測影像資訊，地域性如氣象觀測站之溫度、雨量資訊，細至各項研究調查工作設置的地面樣區等，都是林業不可或缺的資料來源。

以我國林業主管機關林務局所進行各項森林資源調查建置業務為例，目前依據不同的觀測尺度及經營目的，包含全國性林區航攝影像蒐集、土地利用型(林相)判釋，依現行事業區、林班、小班之森林經營管理區劃辦理之事業區檢訂調查，長期監測體系之地面樣區調查，從生態保育面向辦理之各項生態研究調查計畫，全國植群圖繪製，野生動物紅外線自動相機監測，溪流魚類調查作業，以及個別研究區域、物種等調查成果等，多年累積下來具有「巨量資料」規模的潛力。

雖則國內林業在資料數量上，應能跨足巨量資料應用領域的門檻，但實際在資料取得應用上，則面臨到一個根本的問題，即待建立、運作長久倉儲管理及統合取用各類研究調查資料的流程及機制。早期許多資料建置後，除了當次調查應用報告外，因原始資料留存完整性不足，無法進一步加值運用，而因為意識到資料累積建置的重要性，現行多數國家型計畫的調查成果均同時要求調




查者原始資料的提供，希望將資料妥善保存，供未來利用，像是航攝影像蒐集及全國森林資料調查等重要工作，也都透過持續性的經費編列，進行資料更新建置，逐步達到跨時間尺度資料的長期累積，做為歷史資料分析及未來資料預測的基礎。

現行國內所建立的林業資料倉儲，包含目前在林務局內部開發的「森林地理資訊整合供應系統」，主要整合了林業經營管理過程所需及產製的各項 GIS 資料，以 WebGIS 為核心技術建立地圖平台做為林業資訊展示的基礎，惟在已進行空間數化之 GIS 資料外，林業尚涵蓋了許多原始調查之資料表或資料庫型式資料，甚至影像及物種聲紋資料等，目前尚未建立出較適宜的串聯整合模式；而林務局農林航空測量所近年著眼在航遙測影像資料庫建置工作，被國土資訊系統推動小組視為重點計畫項目，除建立歷年期航攝正射影像的倉儲管理機制，同時將影像服務雲端化，提供線上即時正射影像查詢調閱，近年並與國家太空中心合作，將福衛二號影像同時納入雲端服務當中，惟目前影像服務仍屬於瀏覽性質的 WMS 或 WMTS 圖磚服務，尚未能提供線上即時運算分析；此外，針對個別性的研究計畫調查資料，由美國生物複雜性知識網 KNB (Knowledge Network for Biocomplexity) 發展、林業試驗所推動導入及中文化的詮釋資料管理系統 (Metacat)，在國內也已嘗試於用於林業原始調查資料的蒐集管理，惟目前實際推動上，仍受限研究人員對於調查資料公開意願及完整詮釋資料建置能力，成效有待觀察。

綜上，國內林業資料雖然朝向巨量資料規模已逐步成形，但在統合的倉儲管理及取得分析的便利性上，都仍有許多待努力之空間，此部分也是台灣林業要走向巨量資料應用發展的關鍵限制之一。

## 二、資料速度：

對於資料速度的討論，可以從資料取得效率及決策時效性兩方面來討論。在資料取得效率部分，目前森林資源調查資料的取得來源基本上可區分為現地調查及遙測技術，遙測部分以衛星影像而言，國內福衛二號影像取像可達到以日為單位的頻率（正常拍攝流程將台灣分為 8 個航帶輪流拍攝，惟常受限於氣候雲遮，真正可用之資料頻度不高），航攝影像部分，全國普及性的資料蒐集現



行以平地每年兩次、山區兩年一次的頻率辦理，近年國內研究單位及災害防救單位則主力在無人航空載具的發展，讓遙測資訊的取得更為機動性，惟所能蒐集調查的範圍自然也較局限，視當次飛行目的而定。目前國內林業上已廣泛運用衛星影像及航攝影像進行林型及崩塌地等災害區位判釋工作，此亦為巨量資料來源的重要項目之一。

現地調查一般受限人力及交通可及性，資料蒐集更新頻度較低，但可搭配遙測技術，以取樣概念推估獲取全域性的資訊，此部分仍見於個案研究應用為主；近年國內林業試驗所嘗試以無線感測器應用於植物物候監測及動物行為進行研究，林務局亦採紅外線自動相機監測野生動物，大幅提升了調查資料量及取得速度，正如同「全球感測網」概念的提出，廣布的感測器及感測器網絡建立，將成為未來在即時性生態監測的利器，其對應的是大量建置成本的投入及後續資料處理能力，目前國內林業經營管理過程尚未普遍納入應用。

而在決策資訊的時效性需求，主要針對災害防救面向的森林保護工作，一般所稱之森林災害包含：森林火災、濫墾濫伐及極端氣候下各種颶風災害造成的森林資源損失，對於天然災害發生時，需要於災害發生時，迅速地掌握災害影響範圍及資源損失情形，而對於人為或潛在型的災害（近年如堰塞湖的形成），則需能監控、迅速獲悉災害發生，扼止不當的人為破壞及避免影響擴大。目前國內針對森林災害資訊之掌握，包含每日更新發布林火危險度預警系統，或在天然災害發生時，由人工進行的森林災害統計查報工作，以及經由衛星影像自動分析及航攝影像結合人工判釋以掌握災害發生影響之區位範圍，近年並搭配衛星影像自動判釋建立歷年崩塌地範圍資訊等，以監控歷年崩塌地植生復育情形等。多數的資料分析作業，都是以歷史性資料分析及災害發生時被動監測影響範圍為主，如何在廣大複雜的森林資源組成當中，以自動化流程主動偵測到資源變化，達到近即時性的災害監控，目前國內尚處於研究發展階段。

### 三、資料多樣性：

以資料格式來看，現有林業資料的建立，從常見的文數字記錄、資料庫及具空間屬性的數化向量資料與網格影像資料為主，而近年在野生動物調查上，多了紅外線相機或錄影之自動化監測、聲紋錄製等，需整合各種不同的資料格

式及資料語義，共同分析應用。

目前國內林業資料整合應用上，主要以地理資訊系統(GIS)建置出發，從空間分布觀點整合運用各項森林資源調查及經營管理資訊，擷取各項林業資料之地理資訊，透過地圖平台的直觀呈現，協助經營管理上的判斷及理解，並可透過空間分析方式獲得更多訊息。整合過程中，藉由地理分布可快速建立資料在空間上的關聯性，但仍有待處理資料間語義關聯性的問題，像是：特定的經營區位如劃設之「保安林」範圍與森林資源調查結果，其林木分布是否完整、有無遭受破壞，其設立目的下之受益及保護對象是否仍維持完善等；其間的關聯性可透過空間分析方法如疊合分析方式處理，將屬性透過空間建立關聯性，再經由統計方式計算資源情形據以推論，而當資料量持續累積，希望透過跨年期的資料進行時間區間的觀察，以人工方式需要相當多的時間，如果能將關聯性透過語義(意)方式串聯，可結合以自動化流程分析，定期獲得需要的決策資訊。針對國內在此類語義網概念的相關發展，主要仍以資訊技術上的搜尋引擎、自然語言或詞典、目錄學為主，針對各專業領域的知識本體發展，目前亦尚處於學術研究階段。

綜上，對於國內林業資料的「巨量特性」，可具體表現在時序性、空間複雜度、即時監控、資料語義複雜度等幾個面向，說明在林業推動上確實有發展巨量資料應用模式的潛力，但現階段國內尚未發展出巨量資料應用上成功的典型案例。

## 第二節 林業巨量資料應用之策略架構

延續第三章對於國外「全球森林觀察」專案推動成果的分析，檢視台灣現行林業資料建置及發展現況，目前台灣林業發展上尚缺乏類似「全球森林觀察網」這樣一個整合性、透明公開的森林資訊平台，完整地對外界展示台灣森林資源分布及經營管理現況；總體來看，台灣現階段林業發展上多半仍是在組織內部或專業領域內，配合管理過程及個別研究需求進行資料蒐集及平台建置，而「全球森林觀察」專案無疑提供了一個值得台灣去學習的開放式、參與式的應用架構。

在台灣，森林面積超過五成以上為國有林，又因地狹人稠、環境開發壓力大，同時地勢變化劇烈，極端氣候常造成人民生命財產上的威脅，森林向來負有相當高的國

土保安、水源涵養及生態保育價值重任，具有十足鮮明的公共財特性；為此，林業相關經營管理資訊的確應朝向更透明、公開的發展方向，讓相關決策受到全民監督。以下參考「全球森林觀察網」專案的執行成果，嘗試以建立台灣在地森林監測網絡的開放資訊平台為目標，探討如何運用巨量資料輔助森林監測工作，以及發展過程需面對的各種課題。

#### 一、應用需求定義：


為能達到森林監測目的，在應用上可依資訊需求之頻度定義出以下兩種監測需求：一是偏向靜態性的時序森林資源調查資料，頻度可採年或更長區間，一般在正常生育狀態下，短時間內森林資源狀態相對穩定，透過較長的時間尺度可觀察其變化情形。另一部分，則是偏向動態性的(近)即時森林災害監測，透過變遷事件的識別，進行災害因應及決策。

對於時序森林資源調查資料，其觀察的變化資訊可以區分為不同的面向及空間尺度，包含從：森林覆蓋區位(如同「全球森林觀察網」的森林覆蓋及變遷資訊)、土地利用型態及森林組成(植群、林相及物種)等，據以掌握台灣森林生態長期動態變化。而在森林災害監測部分，需觀察的資訊可從災害面向區分，針對森林火災、崩塌地、堰塞湖及濫墾濫伐地等不同災害類型，監測事件發生，並掌握其災害影響範圍及後續復育。

#### 二、資料取得及處理：

從前一節對於國內資料現況的討論，航遙測影像為目前國內發展較為成熟巨量資料來源，從資料蒐集到倉儲供應均已建立例行、穩定的資料來源，透過各種影像分類及演算法的設計，將歷次取得的航遙測影像萃取、轉化為上揭森林監測所需資訊，配合時序性資料的呈現及變遷分析，可建立由巨量資料分析達到自動化監測的流程。而在影像處理過程，包含從影像取得頻度及解析度、影像前處理的空間對位、輻射校正(雲遮及陰影干擾)、地物光譜特徵的解析、以至於演算法的應用，均相當重要。

目前國內已運用衛星影像作為長期、例行監測對象者，像是內政部營建署執行之國土利用監測計畫(2001年起即開始執行)，透過衛星影像自動化監測土地利用變遷及變異點查報，進行土地違規使用通報及取締；而從林業應用的



角度而言，在資料取得上，可參考國土利用監測計畫與國立中央大學太空及遙測中心的合作方式，隨時就最新取得衛星影像資料進行森林覆蓋及災害資訊的演算萃取，並應針對植生之判釋及分類發展一套自動化、標準化的演算流程，做為森林監測資訊來源。

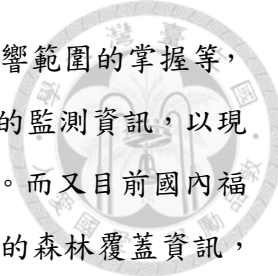
較大尺度的森林監測，透過遙測技術配合林冠層的分析可達到的監測效果，如針對森林內物種組成層級，則無法直接透過遙測技術處理，仍需配合地面調查資料，像是目前國內已設置的永久樣區及動態樣區等，以及各種研究專案進行的森林調查工作，結合長期累積的調查資料進行森林組成變化的分析，其巨量資料的特性將主要反映在森林組成動態分析的複雜度及語義整合工作。目前國內針對物種層級之調查資料整合，已可見於台灣生物多樣性資訊入口網 TaiBIF 整合之物種觀測資料地理分布資訊，以達爾文核心集 (Darwin Core) 為欄位串聯各界調查資料庫，整合物種出現紀錄為生物多樣性監測之基礎。而對於森林樣區調查資料而言，也應朝向建立統合的資料倉儲及流通及資訊語義之整合機制，做為監測資訊來源的基礎。在語義整合部分，可分成幾個面向的課題，一是針對調查工作蒐集的資料項目要針對調查方法及欄位屬性清楚定義，以利未來資料整合運用；另一方面，可就樣區調查結果轉換為林分統計值及監測指標，甚結合衛星影像進行全台性生物量或碳匯量的演算推估等，以提供更多面向的監測資訊。

而在建立森林監測平台時，也需就分析輔助資訊共同蒐集整合，包含台灣行政區界、地籍資料及森林管理區劃等資訊，為監測資訊對應通報權責單位之基礎，另包含參考性的資訊如道路、水系、DTM，生態分析時所需的氣象測站長期記錄資料、土壤調查資料等，災害管理時所需之地質調查、災害潛勢區域圖等，都應共同納入森林監測平台資訊當中。

### 三、資料限制考量：

航遙測影像在台灣地區應用，很大區塊的限制在於雲遮影響，尤其在林業應用上，森林覆蓋多分布於山區，台灣四面環海對流旺盛下，讓資料實際可取得的頻度大幅降低，在災害事件發生前後雲遮發生尤其明顯，可用之資料頻度降到以月甚至季為單位。台灣因面積小，雲遮發生所造成資訊偏誤的比例也就





相對較高，如是在於靜態形式資訊的需求，或是於災後進行影響範圍的掌握等，尚可透過多期的影像篩選排除掉限制因素，但如在於即時性的監測資訊，以現有航遙測影像取得情形，先天環境因素的影響很難完全克服。而又目前國內福衛影像解析度為 2 公尺，在森林監測資訊適合較大空間尺度的森林覆蓋資訊，對於林相組成和小區塊或立木形態的盜伐等則相對難掌握。從長遠來看，有待於新的資料來源及技術導入，像是結合光達資料建立之地面三維資訊，其具高精度、高準確度且天候限制少等特性，有機會真正突破現階段不可抗力因素造成的影像應用限制，並可進一步透過三維資訊時序變化，估算森林生長情形變化，及獲得更細節的變遷資訊，惟目前此技術在台灣還在發展階段，且相關資料係透過各次計畫專案進行建置，暫未朝向高頻度、例行性蒐集機制發展。

#### 四、系統展示及分析：

對於巨量資料的應用上，視覺化處理呈現相當重要，尤其如以公開資訊平台為發展目標，需重視資訊的直觀及可讀性，而地理資訊系統技術的應用是相當好的途徑。針對以航遙測資訊為基礎的森林監測資訊，可透過地圖平台展示歷年在森林覆蓋、土地利用型或植群、林相分布區位，可結合變化時序模擬動畫展示變化趨勢，同時配合時序變遷區位突顯及前後期資訊比對，並就變遷區域與其他參考氣象、地物資訊等進行空間關聯分析，讓使用者由資訊接收者成為參與分析的資訊產出者，將可做為森林監測工作的輔助，在地理資訊系統的展示上國內已經有相當成熟的技術基礎。

而在森林調查資訊部分，如欲進行視覺化的展現，可以包含以下幾種做法：首先最基礎的在於樣區（樣木）設置區位，可由樣區結合後端調查資料庫調閱歷年原始資料，提供分析統計圖表，並可搭配時序性統計呈現樣區調查變化；而要讓資料更具可讀性，須將單點轉化為面資訊，包含前述所提，將樣區資料計算出各種林分統計值、監測指標及全台推估資訊，即可透過屬性值結合林分區位呈現，或透過網格值呈現數據的空間分布特徵，讓資訊更豐富多樣。而未來進一步，資訊呈現可評估三維模型建立的可行性，結合空達資料，透過林分的三維結構及立木虛擬建模，從景觀角度去呈現森林資源的變化動態細節，讓觀察資訊更具真實度。

## 五、組織合作關係：

組織間的合作，除了資料面的整合，需求整合也相當重要，惟有在能滿足多數組織的共同需求及發展模式下，才有利於建立長久的合作關係。不同組織之目標任務不同，對於應用需求的定義也會不同，而目前台灣在林業推動上，尚缺乏一個可凌駕於各機關本位需求之上的合作組織，能包容不同面向的參與者，像是經營管理單位（如林務局、各縣市政府）、試驗研究單位（如林業試驗所、特有生物研究保育中心、大專院校研究系所）及環境保育團體等，確保以整體利益為考量，提出台灣林業發展上所必須共同建立發布之資訊。

而目前台灣林業資源資料，散見於行政機關林務局、地方縣市政府及公所、研究機構林業試驗所、特有生物研究保育中心及各大專院校之森林系所及相關試驗研究單位，從更廣義的角度來看，涉及林業應用相關資料尚可包含像是國家太空中心之遙測衛星影像、中央氣象局之氣象測站資料、經濟部水利署的水文資料及內政部地政司之地籍管理資料等。目前各單位個別間多以資料申購方式取得，要能有效率地整合各方資料來源共同分析應用，及定義後續資料長期合作及流通應用的資料政策，將是監測平台推動過程需重視的組織合作課題。

綜合上述討論，提出國內未來林業巨量資料應用策略架構，如圖 6 所示。在發展林業巨量資料應用時，無論是在資料面、系統面、應用面及組織面等，各面向都有待發展的課題和未來需進一步努力的地方，而在應用策略的規劃上，在初期發展階段，需基礎於現有的資料及技術成熟度進行發展，而未來更應持續導入新的資料、演算法及系統展示技術，克服現有的資料限制及應用障礙，據以達到更好的應用成效。

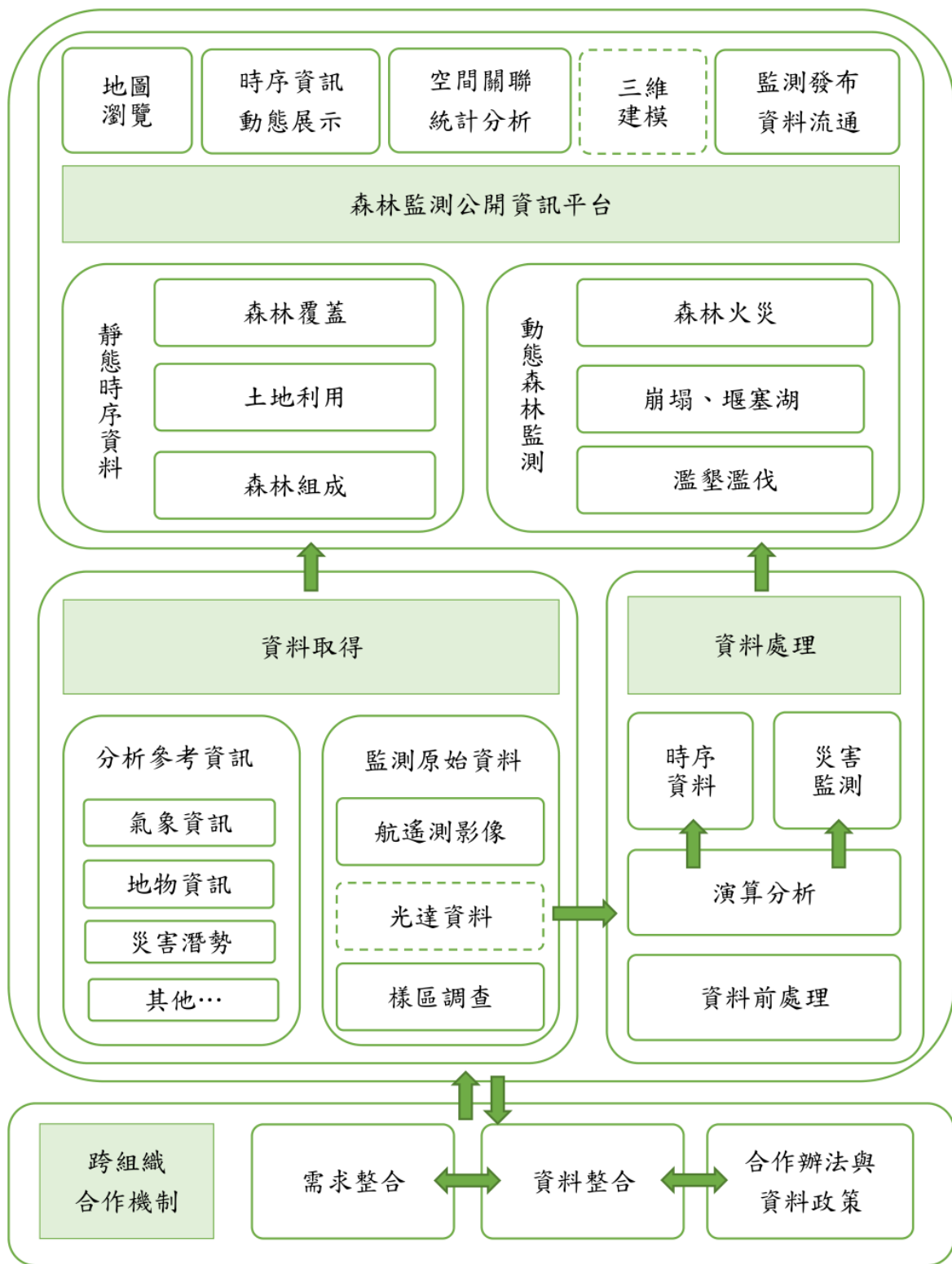


圖 6 林業巨量資料應用策略發展架構

### 第三節 植生變遷監測之個案分析

在資訊平台發展過程，資料面為相當核心的環節，如何從取得的原始資料轉化為森林監測過程所需資訊相當重要，決定了平台內可呈現的資訊內容。由前述討論，遙測影像的取得及分析技術相對較為完整，而考量資料成本及可取得性，延續過去在大肚山地區植被光譜研究蒐集之影像資料，嘗試以個案案例分析進行討論，透過時序性影像處理應用流程進行探討。

#### 一、案例情境：

##### (一) 研究範圍

本研究範圍設定在台中市大肚山台地範圍，大肚山地區位處於台中盆地以西、清水海岸平原以東，南北鄰接著大甲溪及大肚溪，行政區域則涵跨台中市之清水、沙鹿、龍井、大肚、神岡、大雅、西屯、南屯及烏日等區之部分範圍。整體地勢是由北北東往南南西延伸，因受東南地殼板塊運動影響，台地中央向西北方突出，西坡較高較陡、東坡則較低較緩。排除掉北區部分軍事機場區域，針對本次研究地理區位標示如下圖 7。

##### (二) 區域植被特性

大肚山地區植物物種以相思樹 (*Acacia confusa*) 及大黍草 (*Panicum maximum* Jacq.) 為優勢族群，相思樹早年因薪材需要而成為當地主要的造林樹種，大黍草則於 1900 年代引入種植做為馬匹糧草，因生長勢強成為草本植群中的優勢物種。依據先前研究計畫引用林務局林火資料統計結果，因此地區植被特性屬易燃的輕質燃料堆積，加以鄰近都會區及墓地，人為活動頻繁、接觸火源機率高，此區域平均每年約發生 10 次火燒事件，火燒頻度相當高 (曾仁鍵, 2004)。而也正因鄰近都會區，此區域地景較破碎且複雜，由相思林、大黍草及農地、人工建物 (墓地) 等不同土地覆蓋型態混雜分布。

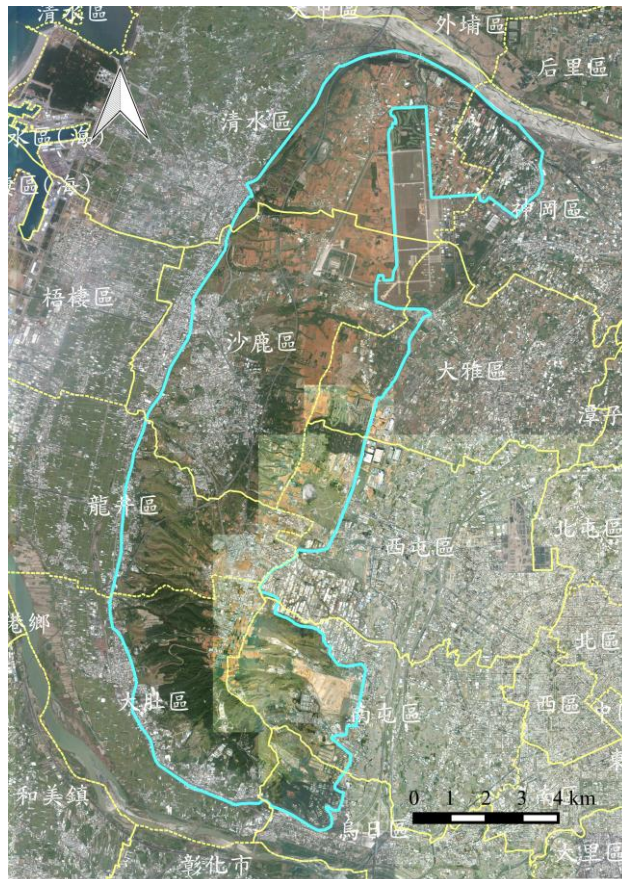


圖 7 大肚山地區研究範圍示意 (背景衛星影像來源：Google Satellite，2015)

### (三) 研究資料

1. 本研究取用自 2003 年 1 月至 2004 年 4 月期間共計 12 幅經輻射校正之 SPOT 衛星影像，其具體影像日期、攝像衛星、原始影像解析度及各波段輻射校正值詳下表 1 所示。

表 1 本研究使用之 SPOT 衛星影像資訊

影像日期	攝像衛星	影像解析度	G	R	NIR	MIR
2003/1/13	SPOT 2	12.5m	1.30509	1.2725	1.53995	
2003/2/27	SPOT 4	20m	2.03416	2.63587	1.2735	8.19106
2003/3/11	SPOT 2	12.5m	1.43774	1.2662	1.15374	
2003/7/25	SPOT 2	20m	0.77097	0.97893	1.17835	
2003/8/24	SPOT 4	20m	0.608	0.786	0.852	5.47

影像日期	攝像衛星	影像解析度	G	R	NIR	MIR
2003/10/19	SPOT 4	20m	0.91261	1.17979	1.278	8.19406
2003/11/14	SPOT 4	20m	1.368	1.753	1.278	8.19406
2003/12/22	SPOT 2	20m	0.85047	0.97747	1.15374	
2004/1/14	SPOT 5	10m	0.9	1.054	1.172	6.411
2004/2/26	SPOT 4	20m	1.368	1.763	1.278	8.19406
2004/3/10	SPOT 2	20m	1.43774	1.2662	1.50824	
2004/4/5	SPOT 2	20m	1.43774	1.2662	1.50824	

2. 本研究樣點選定，主要依據「大肚山望高寮地區林火生態學研究計畫」針對研究範圍內之各植被類型與對照地物樣點，並結合鄰近期間國內航攝正射影像補充部分人工建物及農地之設置點，共計擇定 161 個光譜分析樣點，其分布位置示意如下圖 8；樣點選定以自然植被為主、人工植被為輔。



圖 8 光譜分析樣點分布位置示意圖



#### (四) 研究方法及限制

本研究以前後期衛星影像光譜資訊之變化，監測區域植被變遷情形、識別變遷區位，討論由累積的時序資料觀察地區之植被特性及偵測變遷發生。本研究將以變遷偵測流程之設計為討論重點，而分析工具及後續自動化模式建構未列入討論。

### 二、案例分析流程：

#### (一) 植生指標計算

本研究監測重點設定於植被監測及生長變化，為能從衛星影像快速掌握森林資源概況，可透過影像光譜波段進一步計算植生指標 (Vegetation Index)，運用綠色植被吸收紅光、反射紅外光之特性，用以呈現植生分布情形，做為分析植被變遷之基礎。

植生指標的種類相當多，其中又以常態化差異植生指標 NDVI (Normalized Difference Vegetation Index) 被廣為應用。NDVI 的計算需要透過紅光及近紅外光波段數值進行處理，因植被分布地點近紅外光較高、紅光反射較低 (被植物吸收)，透過紅光與近紅外光波段數值差額除以兩者相加 (即名稱所謂差異之「常態化」)，數值會落在 1 至 -1 之間，如果為水域、雲層或下雪會以紅光反射為主，故會產生負值，土壤、岩石數值近於 0，植被分布密集區域則傾向紅外光之反射，數值主要在 0.1 至 0.6 之間，其公式為：

$$NDVI = \frac{NIR - R}{NIR + R}$$

植生指標的應用為對於植生資訊的影像增揚過程，突顯植生的光譜特性，可有效降低許多幅射、地形所帶來分析上的限制，但同時其也有應用上的限制，即陰影區和裸露區難以區分，因此分析過程當中，亦有許多研究係採用混合性指標共同計算。本研究中僅採用 NDVI 為案例進行討論。

針對監測區域範圍內取得的多光譜 SPOT 衛星影像，取其紅光 (R) 及近紅外光波段 (NIR)，透過 NDVI 計算公式進行轉換。而因本研究中取得之 SPOT 衛星影像，影像解析度共計包含 10 公尺、12.5 公尺及 20 公尺不等，為能在相同基礎下比較各期間變化情形，透過網格重新取樣，將影像光譜資

訊統一以 20 公尺網格大小呈現，即設定 20 公尺網格（400 平方公尺）為監測單元，做為變遷識別基礎，總計本研究區共涵蓋 27 萬 4,344 個網格數。針對 SPOT 衛星影像計算 NDVI 植生指標圖情形，就 2004 年 4 月 5 日之衛星影像計算情形如下圖 9，其中衛星影像的呈色，因 SPOT 衛星影像第 1-3 個波段依序為綠光、紅光及紅外光，透過各波段與 RGB 值對應，採 Red 對應紅外光、Green 對應紅光及 Blue 對應綠光方式，將植被覆蓋區域以鮮紅色顯示以利於觀察，對火災跡地的探測也較為容易（曾仁鍵，2004）。

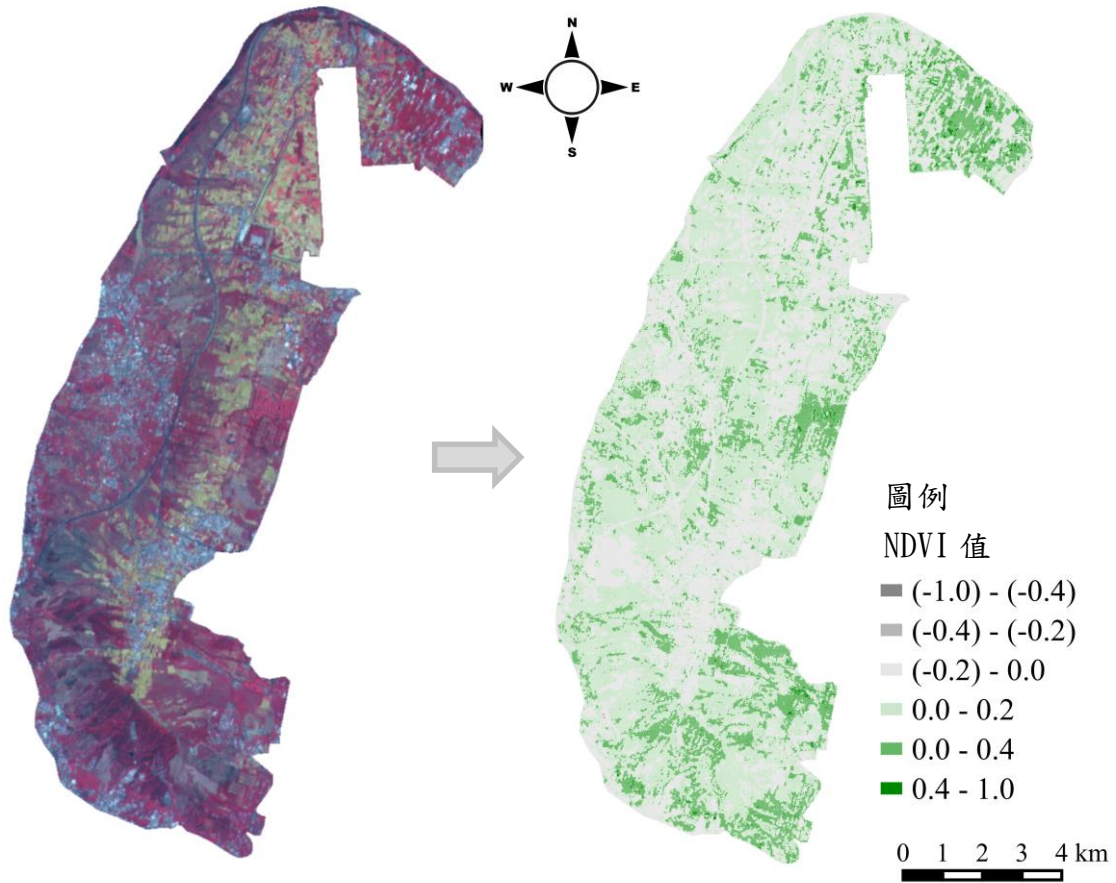


圖 9 2004/4/5 大肚山地區 SPOT 衛星影像（左）及植生指標（右）計算結果



## (二) 變遷區域偵測及識別

為分析研究區域植生變化，並能監控變遷事件、地點的發生，可透過影像相減法，計算前後期網格 NDVI 差異，呈現前後期植生指標變化的空間分布。計算方式係以同網格內後期影像 NDVI 值減去前期影像 NDVI 值，因此如計算結果差值為負值，表示隨著時間變化，前期至後期植生有減少之趨勢；差值為正值則表示植生增加。

以 2004 年 3 月 10 日及 4 月 5 日影像 NDVI 值為例，由圖 10 初步比對觀察可見，2004 年 3 至 4 月期間，全域均為植生增加趨勢，再透過前後期 NDVI 差值計算，在研究區域全域之 NDVI 差值平均值為 0.09 (植生增加)，標準差 0.08，透過直方圖可見差值的頻度(圖 11)，其數值分布偏態值為 0.51、峰度為 1.70。

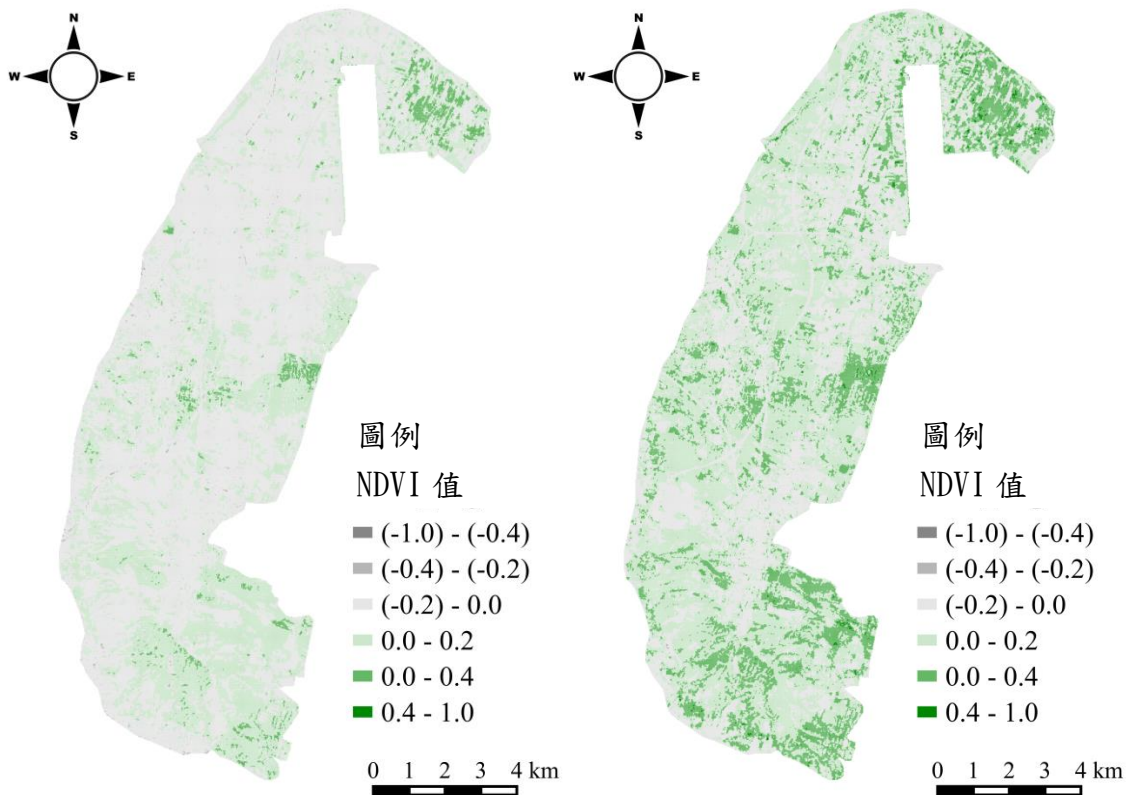


圖 10 2004/3/10 (左) 與 2004/4/5 (右) NDVI 值對照

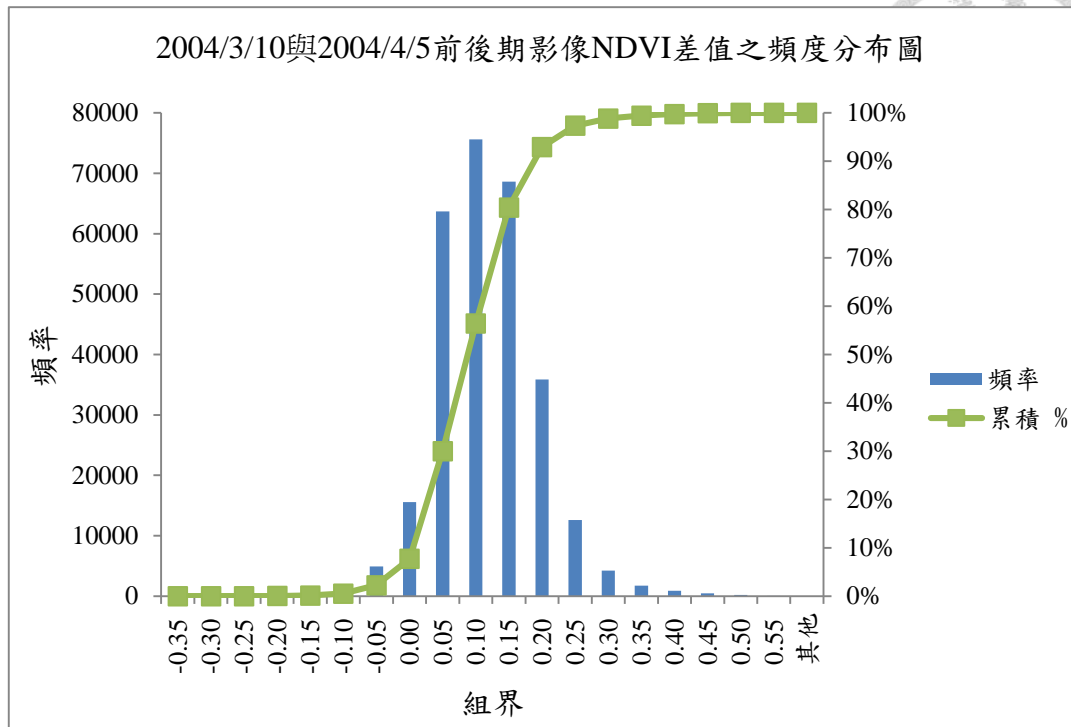


圖 11 2004/3/10 與 2004/4/5 NDVI 差值頻度分布圖

而一年當中不同時間點的植生狀態變化，受到季節性氣候因素影響，植生會有自然榮枯之週期，如果單純以 NDVI 值的差異來看時，當全域植生普遍均為減少狀態時，其他可能影響植生變遷之因子，如林火發生，就不容易在分析過程被識別出來。同時，受限於影像可及性，所能取得的各期衛星影像距離時間間隔不一，單從影像前後期 NDVI 差值來評斷，亦較容易誤判其間變化程度，難以識別特殊的變遷事件是否發生。

為能減低自然季節性變化對於變遷事件偵測的影響，本研究假定不同的地景類別在全域 NDVI 值之數值分配內均有其特定分布，將各網格 NDVI 值 ( $X$ ) 透過統計學上標準化的定義，依全域 NDVI 值之平均值 ( $\mu$ ) 及標準差 ( $\sigma$ ) 計算標準分數 ( $Z$ )，再進行前後期差異的比較計算：

$$Z = \frac{X - \mu}{\sigma}$$

其標準化結果如圖 12 所示，可見透過標準化可篩濾掉部分全域性植生指標之變化趨勢，有助於突顯出局部區域不同於整體趨勢的變遷現象。

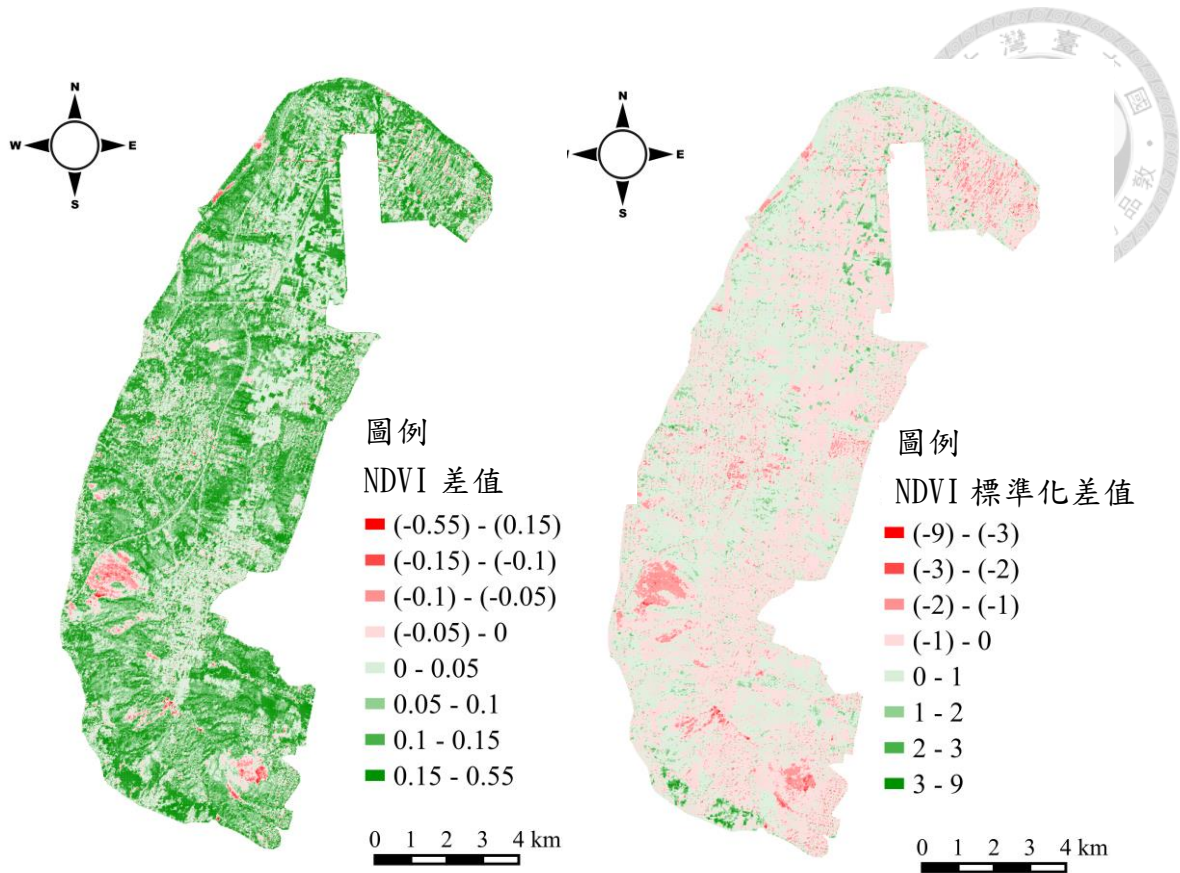


圖 12 2004/3/10 與 2004/4/5 植生指標差值依標準化計算結果

而在圖 12 標準化計算前後的差異，亦可觀察到一個現象，即原 NDVI 差值為正者，經過標準化計算之後轉變為負值。而這可以從原始影像 NDVI 值來看，由 2004 年 3 月 10 日影像計算之 NDVI 平均值為 -0.01、標準差為 0.09，2004 年 4 月 5 日 NDVI 值平均值為 0.08、標準差為 0.12，即在標準化過程一減、一增的調整下，相對性的變化可能造成部分維持植生狀態的地景誤判為植生減少。另一部分應思考的是，NDVI 值原本即會受到季節植生生長變化影響，無法完全對應到地景變遷事件的發生，仍需透過進一步的資料處理及判斷，方能做為變遷監測的資訊基礎。

### (三) 不同網格大小之變遷識別差異

在變遷偵測過程中，監測單元大小的選擇連帶影響到變遷識別的能力及結果。一般而言，監測單元面積越大，單元內混合像元情形會越普遍，對於

面積較細小的地物或較小範圍的地景變遷事件，在監測過程易被稀釋、忽視掉；而監測單元面積越小，對於變遷識別精度越高，惟可能造成識別到的變遷資訊較於龐雜、難以聚焦，或是識別結果在空間分布較為破碎化，造成後續管理決策上的困擾。

為討論不同監測單元大小對於變遷分析的差異性，本研究將分析區域以網格劃分為 20 公尺（原影像解析度）、50 公尺及 100 公尺之網格大小，依前述流程，計算各時期 NDVI 值及前後期 NDVI 標準化差值之變化情形；其中針對 50 公尺及 100 公尺網格 NDVI 值，係取其網格範圍內原有影像 NDVI 值之平均值。

在 NDVI 值部分，針對不同網格大小，在全域 NDVI 平均值呈現差異不大，而標準差則隨網格單元面積增大降低，說明了較大的網格單元將去除掉極端值的呈現，全域各年期的 NDVI 值依不同網格大小其平均值及標準差變化情形如圖 13 所示。

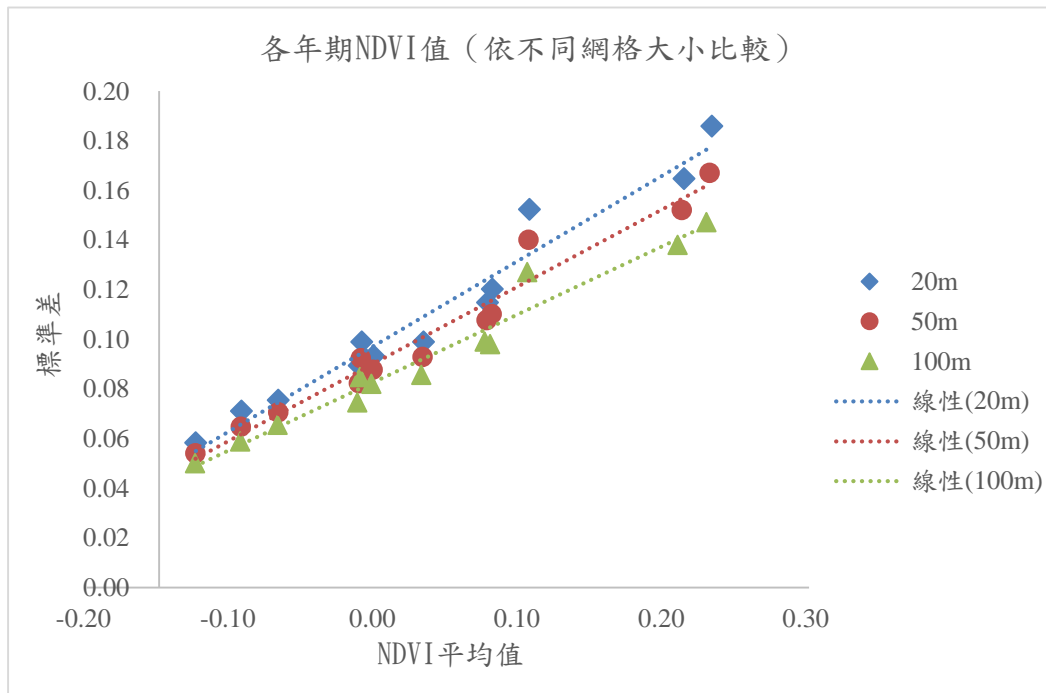


圖 13 依網格大小比較各年期 NDVI 平均值及標準差情形

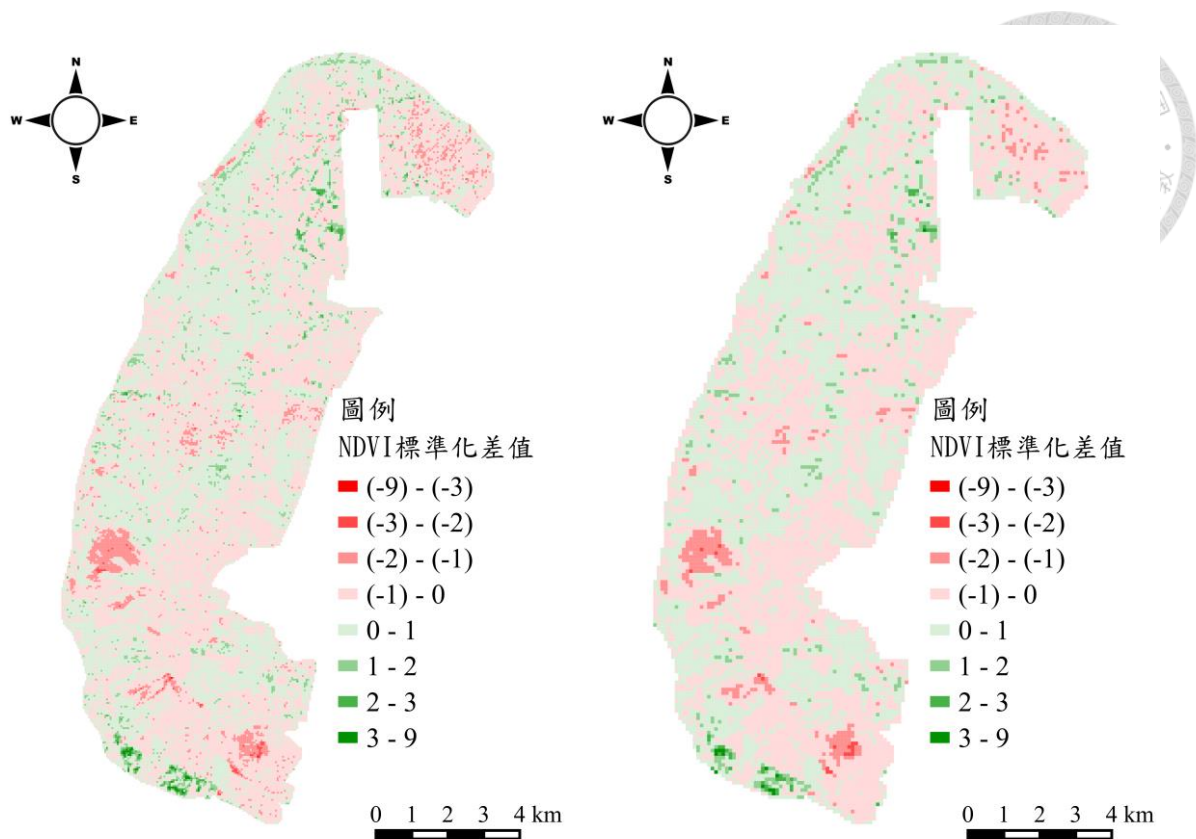


圖 14 2004/3/10 與 2004/4/5 NDVI 差值依 50m (左)、100m (右) 網格計算

針對就不同網格大小進行 NDVI 標準化差值計算，所得各年期變遷分析結果如圖 14 所示，由圖中可見，隨著網格面積變大確實減低了變遷區域破碎化情形，有助於識別出較集中、大區域的變遷事件。而針對變遷區位範圍中，設定 NDVI 標準化差值大於 1 者作為植生顯著減少之區域，可將有顯著變遷之網格數可統計如表 2 及圖 15 所示，可見在變遷識別的區域面積，隨網格大小的增加而降低，此結果亦符合前段討論到較大的監測網格排除掉細微面積極端值的呈現，有助於減低變遷偵測結果之破碎化程度。

表 2 依網格大小比較各年期 NDVI 變遷區域識別度 (NDVI 變遷差值>1)

變遷區間	前期影像	後期影像	20m 網格數	佔全域 百分比	50m 網格數	佔全域 百分比	100m 網格數	佔全域 百分比
1	2003/1/13	2003/2/27	16089	5.63%	1665	3.60%	362	3.08%
2	2003/2/27	2003/3/11	9571	3.35%	697	1.51%	136	1.16%

變遷區間	前期影像	後期影像	20m 網格數	佔全域 百分比	50m 網格數	佔全域 百分比	100m 網格數	佔全域 百分比
3	2003/3/11	2003/7/25	20389	7.13%	2412	5.22%	485	4.13%
4	2003/7/25	2003/8/24	20974	7.34%	2537	5.49%	587	4.99%
5	2003/8/24	2003/10/19	26799	9.38%	3896	8.43%	978	8.32%
6	2003/10/19	2003/11/14	22312	7.81%	3009	6.51%	698	5.94%
7	2003/11/14	2003/12/22	18065	6.32%	1968	4.26%	353	3.00%
8	2003/12/22	2004/1/14	11495	4.02%	1267	2.74%	284	2.42%
9	2004/1/14	2004/2/26	12856	4.50%	1025	2.22%	234	1.99%
10	2004/2/26	2004/3/10	10497	3.67%	742	1.60%	146	1.24%
11	2004/3/10	2004/4/5	16553	5.79%	1923	4.16%	394	3.35%

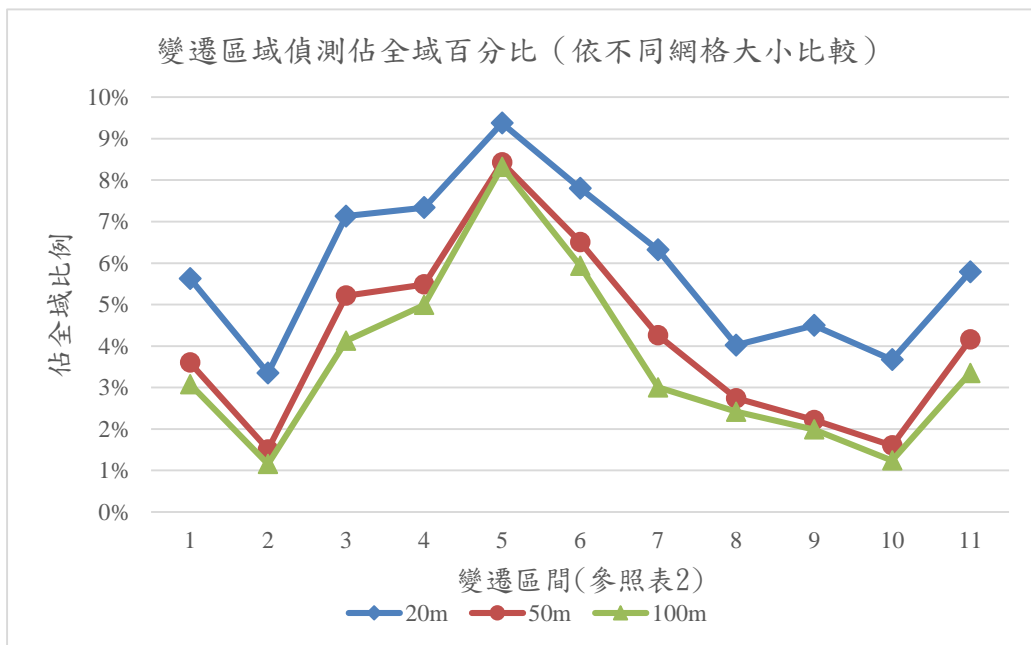


圖 15 依表 2 圖示各年期 NDVI 標準化差值佔全域百分比之變化情形

綜上所述，可知不同的監測單元大小，對於研究範圍全域性植生變遷監測結果，在整體趨勢相近，而差異主要發生在局部區變遷識別的區位及能力，較大的網格將排除掉局部零星極端值的呈現；也因此較大的監測單元下，適合用於偵測影響範圍接鄰且較廣泛的變遷事件；

較小的監測單元則能獨立識別小範圍的變遷事件，惟可獲得資訊在空間上相對較為破碎，不易聚焦。

監測單元面積的選定，一方面直接受限於原始影像解析度，另一部分則需考量變遷事件的特性。如以大肚山地區林火變遷的監測而言，依據「大肚山望高寮地區林火生態學研究計畫」中針對 1991 年至 2003 年間植群火燒歷史紀錄的統計，以主要優勢植被物種大黍草及相思樹林而言，草生地火燒次數統計 49 次、面積計 108 公頃，相思樹林火燒次數 47 次、面積計 225 公頃，平均每次火燒影響範圍近 2-4 公頃，本研究擇定網格大小由 0.04 公頃、0.25 公頃至 1 公頃，均符合大肚山地區林火變遷事件之特性。

#### (四) 光譜樣點 NDVI 值之多期時序變遷

延續前段「(二)變遷區域偵測及識別」的討論，單純由各網格 NDVI 值前後期的差異，並無法完全對應到地景變遷事件的發生，以下嘗試配合地物樣點，由不同地景類型在光譜上之時序變化表現來探討。本研究中針對已知地物之參考樣點光譜資訊進行分析，各類型地景與樣點數如表 3 所示：

表 3 各類型地景光譜樣點數

地被物	樣點數
大黍草	85
相思樹林	46
農地	19
人工建物	8
清水火燒	1
橫山火燒	2
合計	161

為從各光譜樣點歷年 NDVI 值變化情形，觀察不同地物之光譜特性，依各樣點座標萃取歷年影像 NDVI 值計算結果，以前後期 NDVI 標準化數值分別為座標之橫縱軸，透過 XY 散布圖繪製，圖 16 為運用 2004 年 3 月 10 日與 2004 年 4 月 5 日的 NDVI 值數值對之繪製結果：

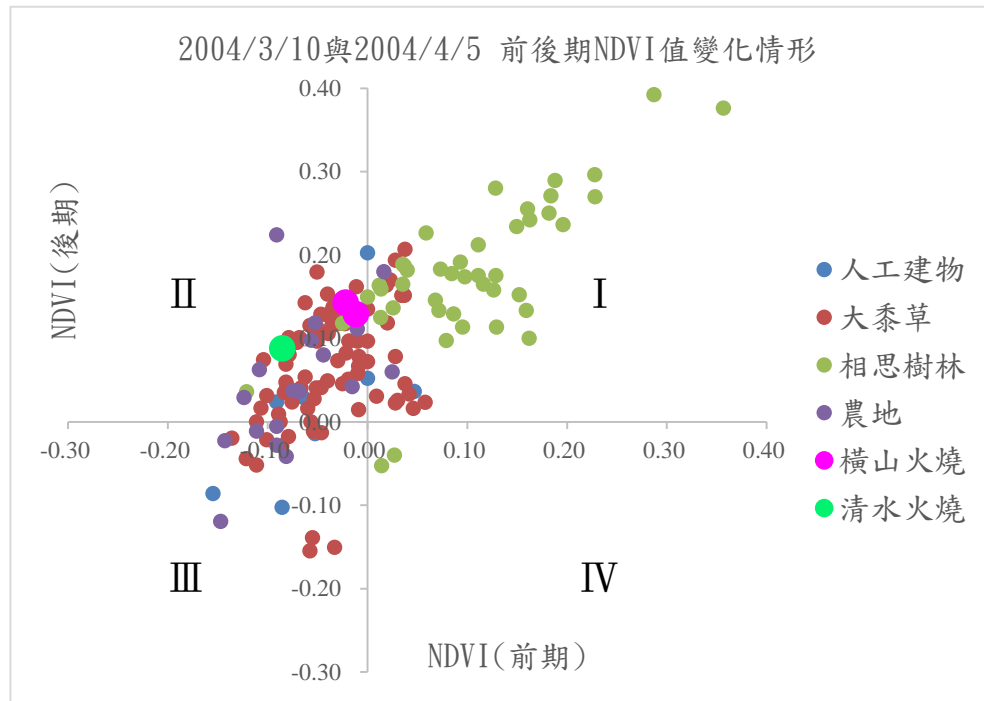
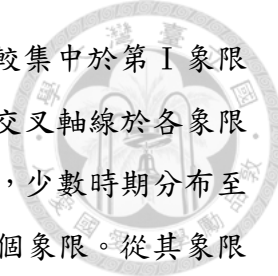


圖 16 2004/3/10 與 2004/4/5 各樣點前後期 NDVI 數值對

透過各期間各地景類別前、後期影像 NDVI 值比對，其各年期變動情形，依  $X=0$  及  $Y=0$  之等值線直角座標區分為四個象限：I (+, +)、II (-, +)、III (-, -) 及 IV (+, -)。由圖 16 可見，樣點落於第 I 象限表示前、後期 NDVI 值維持為正值，第 III 象限則表示前、後期 NDVI 值維持為負值，如樣點落於此兩個象限當中，可推斷該樣點仍維持在相同的地景類型；相對的，樣點如落於第 II 象限，表示前期為負值、後期為正值，即樣點從非植生轉變回植生覆蓋的狀態，第 IV 象限表示前期為正期、後期為負值，即樣點從植生變遷為非植生的狀態發生，在監測資訊來看可推斷該區域期間發生了特定的變遷事件。





依地景類別來看，「相思樹林」樣點在各年期較集中於第 I 象限右上部；「建物」樣點則分布於第 III 象限，或於接近交叉軸線於各象限間變動；「農地」樣點主要分布在第 II、III、IV 象限，少數時期分布至第 I 象限；「大黍草（草生地）」則相對平均散布四個象限。從其象限分布狀態可知，「相思樹林」及「建物」都是相對穩定的地景，「農地」受到作物種植及收成影響，常有短期 NDVI 值增減變化情形，而「大黍草」除常受到火燒干擾，生長也相當快速，NDVI 亦處於短時間內不斷增加循環之情形。其中草生地及農地的差異較不明顯。

針對其中「清水火燒」及「橫山火燒」兩個已知發生火燒事件的地點來看：將樣點依時間序列（箭頭方向）之前、後期 NDVI 數值對繪製連線，觀察其時序變化，「清水火燒」記錄之火燒時間為 2003 年 9 月 7 日，對照圖 17（2003 年 8 月 24 日與 10 月 19 日前後期之 NDVI 數值對）亦確實落於第 IV 象限，並於這一年間於 4 個象限當中呈現循環狀態：由維持植生狀態、植生轉變為非植生、維持非植生狀態，再由非植生區域恢復為植生等。而「橫山火燒」記錄之火燒時間為 2003 年 7 月 7 日（時間落於 2003 年 3 月 11 日至 7 月 25 日間），但對照圖 18，該區間的數值並未落於第 IV 象限，反而另外在 2003 年 10 月 19 日與 11 月 25 日間落入第 IV 象限。從 2003 年 3 月 11 日至 7 月 25 日橫山火燒樣點的具體數值可見，3 月 11 日植生指標為 0.10、而 7 月 25 日為 0.05，在 3 月至 7 月間在季節上屬於植生生長階段、但該樣點 NDVI 值卻不升反降，可推測因火燒發生之時間點前後期影像取得頻度較低（相隔將近 4 個月），加上火燒發生日期 7 月 7 日至影像拍攝日期 7 月 25 日相隔近 3 個禮拜，此時正值植生生長快速期間，部分區域植生在火燒過後已開始恢復生長。另由時間序列變化情形，可進一步推測除已有記錄在 2003 年 7 月 7 日發生火燒外，橫山區域極可能於 2003 年 10 月至 11 月間有再次火燒事件發生（樣點數值落入第 IV 象限）。

比較以前、後期影像 NDVI 差值去監測變遷區域，透過 NDVI 值

時間序列上變化趨勢，結合 NDVI 值以 0 為門檻值區分植生與非植生地景類別的特性，有助於掌握會造成地景上「植生變化為非植生」一類災害事件的發生。

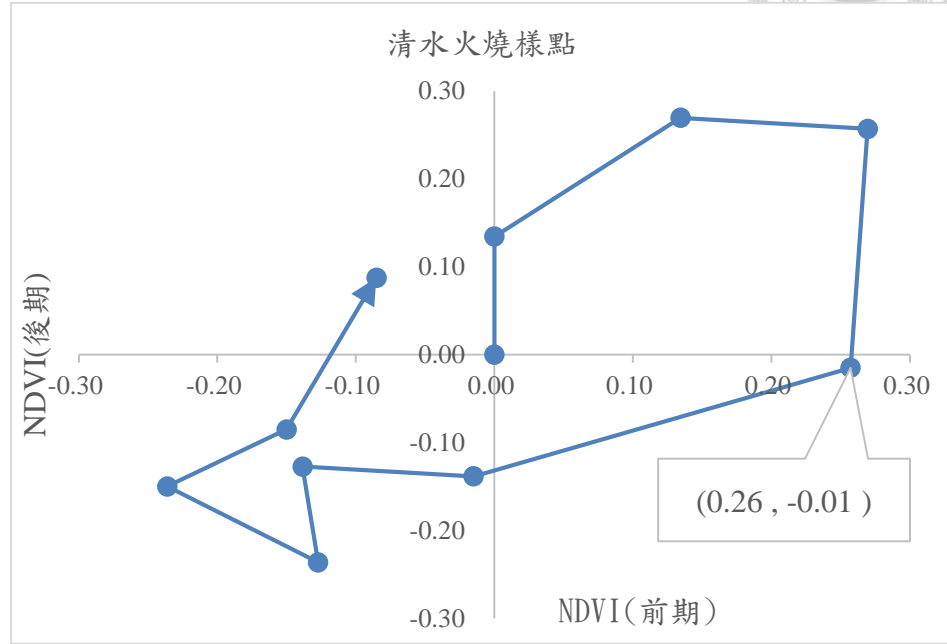


圖 17 清水火燒樣點各時期 NDVI 值依時間序變化情形

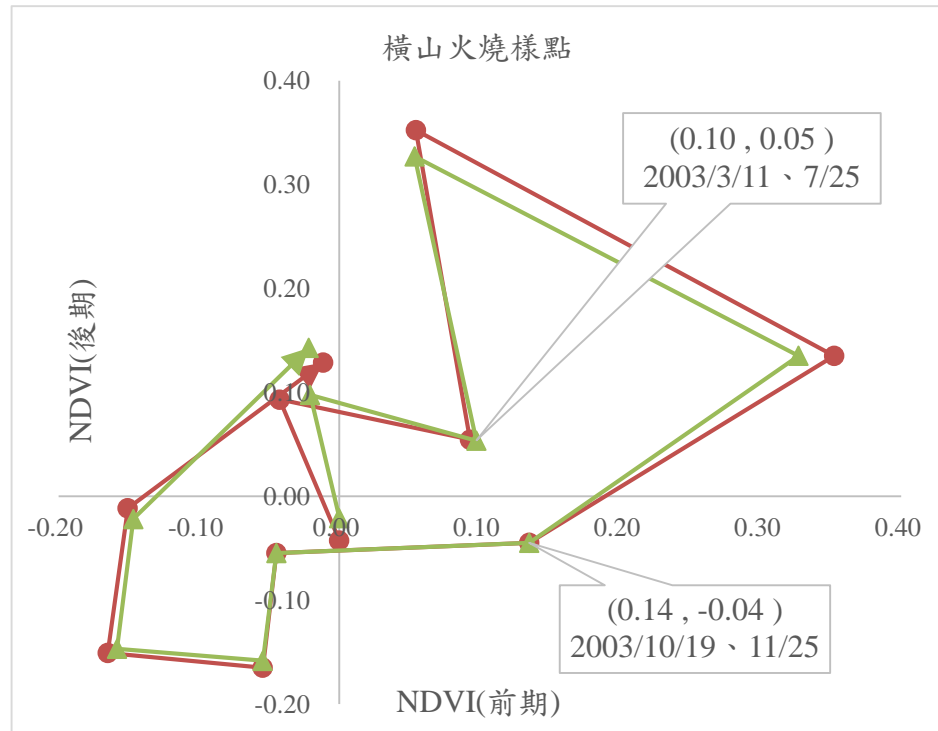


圖 18 橫山火燒樣點各時期 NDVI 值依時間序變化情形

## (五) 小結

本研究嘗試對於變遷偵測方式進行討論，由上述討論中可見，當就以月為資料頻度之衛星影像進行變遷分析時，如透過傳統的影像相減法的演算規則觀察前後期的 NDVI 值變遷，其易受到季節性植生生育變化增減的干擾，先將研究區域的 NDVI 值進行標準化處理後，可篩濾掉全域性植生生長差異的訊息，保留具顯著差異變遷資訊；另一方面，變遷偵測網格大小的選擇需視變遷事件特性而調整，較大的網格排除掉極端值的呈現，適合用於偵測影響範圍接鄰且較廣泛的變遷事件，較小的監測單元則能獨立識別小範圍的變遷事件，惟可獲得資訊在空間上相對較為破碎。此外，如進一步以時序性累積多期的衛星影像觀察各監測網格 NDVI 值的變化趨勢，可看出植生地區在歷經火燒干擾時 NDVI 值的增減循環變化，並具體呈現火燒之變遷時點；惟影像資料頻度將是影響能否確實偵測出火燒變遷事件發生的關鍵。

## 第五章 結論與建議

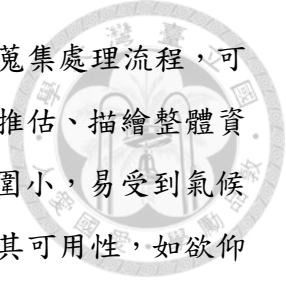
巨量資料除了帶來分析技術上的挑戰，更重要在應用思維的建立。要能夠發揮巨量資料價值，需要透過業務創新想法帶動，而開放資料更是要實踐巨量資料應用環境不可或缺的環節，要如何達到從資料到應用間的無縫接軌，這部分台灣的整體氛圍仍有待努力。我國對於森林資源現況的掌握，目前主要作法係透過人工航攝影像判釋及現場樣區調查，資料內容固然較為精確，但更新頻率難免受限；隨著遙測分析技術提升及感測網應用日益廣泛，提供更多元的資料取得管道，有助於更即時性資訊的蒐集及掌握，但此類型的資料量相對龐大且異質，要如何從這些「巨量資料」內獲取真正有用的資訊，讓相關決策人員更有效的管理及利用這些資訊來源，為有待研究發展的議題。

本文檢視了國外林業巨量資料應用案例：全球森林觀測網（GFW）專案，該專案結合了分布世界各地有關組織之資料及研究人力，提供了森林覆蓋、森林變遷、火災監測及林業資源管理分析等各種決策資訊，進行全球森林資源狀態監測，推動資源保育；而 GFW 專案推動係以全球性角度發展，落實到在地的林業管理，檢視現行台灣林業管理資訊建置現況，雖則長期持續進行森林資源調查工作，但在即時性變遷資訊的掌握能力明顯仍有不足，此部分仍有賴未來相關調查技術及自動化分析流程的導入方能有效突破。本文同時嘗試透過衛星影像進行時序分析，討論變遷偵測的處理過程及議題，從植生指標轉換、前後期影像比較及時間序列上觀察地景變遷特性，提出變遷偵測應用流程的基本架構，並針對國內未來發展巨量資料應用應關注的幾項議題進行討論，希望提供未來研議林業巨量資料發展推動策略之參考。

本研究雖未實際操作巨量資料、未能就實際之技術及分析方法進行探討，透過衛星影像時序分析，由影像資料取得、監測單元選擇、變遷區域分析及光譜變遷特性分析等過程，亦初步提出了時序變遷監測應用方法。而針對未來實際發展林業巨量資料應用，嘗試從以下方向提出推動建議：

### 一、基礎資料之建立及取得

討論「巨量資料」的議題，最核心的關鍵還是在資料面，要先具備或取得足以堪稱為巨量的資料本身，才能進行下一步的資料處理及訊息擷取。資料是需要被創建的，林業資料在現地調查蒐集相當重要，但資料取得速度受限，而



將現地資料透過結合航遙測技術，以自動或半自動化的資料蒐集處理流程，可提升資料建置的速度及更新頻率，並有助於短時間內大範圍推估、描繪整體資源狀態。惟以台灣而言，因四面環海對流旺盛，加以面積範圍小，易受到氣候因素影響，拍攝之航攝或衛星影像常受到大片雲遮影響降低其可用性，如欲仰賴巨量影像資料進行近即時變遷偵測及分析，由現行所取得可用之資料頻度來看，仍有一段距離；除了於適宜氣候儘量取得空照資訊外，亦可結合其他資料來源共同分析，如國內發展中的空載光達 (LiDAR)，即具備有高精度及低天候限制之特性，或結合地面感測網的建置，都將有助於長期、密集的資料蒐集，以導入後續之應用分析。

## 二、組織間的合作及資料流通

除為了特定研究或經營目的自行蒐集資料外，巨量資料的精神更在於能有效整合多元的資料來源共同運用。以林業而言，除了森林資源調查本身，包含航遙測影像、氣象、地形、水系及土壤等資料，都是林業經營過程需參考的基礎。而每項資料的蒐集均有相對應需投入的時間、金錢及人力成本。以福衛二號影像購置費用來看，政府或學術機關購置單幅 2 米彩色全台融合影像需 33 萬 7,500 元，如希望取得以月為頻度並累積跨時序之資料進行監測分析，每年需成本相當高昂，並非每個組織都能力負擔，在資源有限狀態下，組織間資料的合作、共享便格外重要。近年台灣政府積極推動開放資料政策，對於資料公開的作法日趨開放，相信對於巨量資料應用的推動環境也將更為友善。

## 三、長期穩定的資料來源

巨量資料的累積並非一朝一夕，而資料蒐集本身也必須是長期性、持續性的工作，如希望能在林業上充分發揮巨量資料能帶來的效益，自然也需建立一套林業所需基礎資料長期蒐集、倉儲及存取應用的機制。以本文的研究案例來看，植生除了持續性生長，每年度並有隨季節性的變化，單以一年的影像資料，實則無法真正觀測到當地植被生態之全貌。對於如要真正推動巨量資料應用，在資料蒐集、建置之初，應同步規劃後續的更新機制，讓資料得以長期累積，提供日後分析。



#### 四、設計自動化流程

要能確保巨量資料的處理效率，首要是建立自動化流程，無論在資料的取得階段、前處理或是實際分析階段，適時運用資訊技術，對於資料流程進行系統化的定義及串聯，結合模式的建構，將取得之龐大資料自動轉化為可用的資訊，讓管理者及決策者能在最短時間內獲取資訊進行判讀及後續因應。台灣林業近年頻面臨極端氣候的考驗及震撼，災害應變需要在最短時間內掌握發生位置及影響範圍，自動化之處理機制無疑是達成近即時的決策的一項途徑；而巨量資料最讓人期待仍是在潛在價值部分，要能從龐大的資料當中找到隱藏的關聯性，需適時運用資料探勘工具，搭配演算法及模式分析去找尋資料的特徵，當資料量龐大難以倚賴傳統人工分析方式進行，自動化流程的建立可對於累積的資料庫反復分析，協助我們在資料茫茫中尋找有用的資訊。

#### 五、儲存環境及雲端運算架構

如在自有的儲存環境下存放歷年的影像資料做為分析之用，以全島福衛二號鑲嵌正射影像產品約 150GB 來看，如約 10 天取得一次涵蓋全島影像估算，每年約有 5.5TB 以上的資料量，加上加值過後的 NDVI 影像、計算後的變遷監測結果，歷年累積勢必需要相當龐大的儲存空間。在發展巨量資料應用架構時，儲存及運算環境架構是相當重要的成本考量，如將所有資料集中儲放，無疑是另建一個資料中心，除了建置成本龐大，未來長久維護成本也是關鍵。台灣國內整體資源有限，後續在發展上，應盡量朝向跨組織間之合作及雲端資源整合方式進行，避免各組織持續重複投置經費於軟硬體設施，而能儘可能專責投注在自身領域之資料分析及應用上。


#### 六、組織的態度及能力

除了資料分析工作本身外，巨量資料應用能否發展成功，除資料及技術能力的配合，其中很大一部分的關鍵決定在組織。舉例而言，巨量資料蒐集及分析需要投入相當多的成本，組織的支持度影響是否得以跨足巨量資料領域；管理階層是否能清楚定義組織任務及發展目標，影響到一開始資料蒐集整合分析方向，不明確的問題可能讓資料處理花了大半時間、投入大量資源，分析結果卻對組織起不了任何幫助；組織並需願意接受改變，因巨量資料應用導入意謂

著資訊可能來得更快速、更密集，連帶挑戰著各個組織既有的管理流程及作業方式，組織需能將巨量資料分析的結果反映在實際執行作為上，才能真正發揮巨量資料應用的價值。



## 第六章 參考文獻

- 
- 王順生、王淑卿、嚴國慶、王信傑 (2013) 在物聯網建構以服務為導向的架構和作業流程--以防災監測服務為例。2013 年資訊科技國際研討會 (朝陽科技大學)。
- 吳岸明、張桂祥 (2014) 福衛二號影像處理系統研發之回顧與展望。航測及遙測學刊，18(1):1-12。
- 林芳邦 (2013) 空間巨量資料分析之前瞻資料基礎設施發展。國土資訊通訊，88:2-11。
- 林蔚君、詹雅慧 (2013) 巨量資料分析與 GIS 應用價值創造。國土資訊通訊，88:12-20。
- 紀麗美 (2005) 林務地理資訊系統推動概況。國土資訊通訊，50:35-48。
- 曾仁鍵 (2004) 衛星影像於大肚山地區植被光譜變遷之監測。國立台灣大學森林環境暨資源學研究所碩士論文。
- 陳銘憲、林與絜 (2013) 雲報專欄：巨量資料 (Big Data) 的技術發展與挑戰。www.twcloud.org.tw。
- 楊純明 (2014) 鉅量資料及其在農業生產上之應用潛力。作物、環境與生物資訊，11:51-56。
- 童啟晟 (2014) 台灣發展巨量資料產業之機會及挑戰。證券櫃檯 103 年 12 月號，174:45-49。
- Beyer, Mark and Doug Laney (2012) The Importance of Big Data: A Definition. Stanford CT: Gartner.
- Jeansoulin, Robert (2015). Big data: How Geo-information Helped Shape the Future of Data Engineering. AutoCarto Six Retrospective. arXiv preprint arXiv:1501.04832.
- Kumar, Rakesh, Neha Gupta, Shilpi Charu and Kumar Jangi (2014) Architectural Paradigms of Big Data. National Conference on Innovation in Wireless Communication and Networking Technology, 2014.
- Laney, Doug (2001) 3D Data Management: Control Data Volume, Velocity, and Variety. Application Delivery Strategies.



Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh Angela Hung Byers & McKinsey Global Institute (2011) Big data: The next frontier for innovation, competition, and productivity.

Olofson, Carl W. and Dan Vesset (2012) Big Data: Trends, Strategies, and SAP Technology.

Ward, Jonathan Stuart and Adam Barker (2013) Undefined By Data: A Survey of Big Data Definitions. arXiv:1309.5821v1.

參考網站：

<http://hadoop.apache.org/> Hadoop 官方文件網站

<http://ngis.afasi.gov.tw/> 台灣航遙測圖資供應平台

<http://www.globalforestwatch.org/> 全球森林觀察網