

數值土地利用資料準確性之研究

Accuracy Analysis of Digital Land-use Data

賴進貴*

Jinn-Guey Lay*

Abstract

The quality of data is critical to the development of a GIS. All spatial data are of limited accuracy. Thus, the investigation and analysis of digital data is necessary in order to assure the reliability of results generated by GIS. For efficient control of land use change, GIS technology has been adopted for land use management in Taiwan. From the nature and production method per se, land use data poses some uniqueness and deserves a close investigation. This research investigates the quality of land use data using an existing database of western coastal Taiwan as an example. It identifies the amount and magnitude of errors associated with the tested database, and then traces the source of data errors. A series of tests were also conducted using two different approaches for error-elimination. This research addresses the multi-faceted problems associated with data quality and is aimed to enhance our understanding to data quality. Three directions for future research are suggested in the end of this paper.

Keywords: GIS, data quality, land use

摘 要

資料品質優劣是地理資訊系統發展成敗的關鍵。所有地理資料的準確性皆屬有限，為了確保資料的適用性，資料品質的檢核與分析是地理資訊系統研究的重要課題。為了有效地掌握台灣地區急遽的土地利用變遷資訊，地理資訊系統的技術已經被引用至土地利用資料的管理與分析。就資料的生產方式和本質而言，土地資料都具有其特性與意義，值得吾人加以探討。本研究以台灣西部海岸地帶的土地利用資料為例，針對土地利用資料的品質問題加以探討，利用疊圖和邏輯分析來偵測誤差的數量、大小，進而探究誤差的來源，最後並選擇兩種方法進行一連串的誤差消除測試。本研究凸顯資料管理上的各相關層面問題，增進吾人對於誤差量度、來源、消除等各個面向的知識。針對本研究的結論，文末提出後續研究的三個方向。

關鍵字：地理資訊系統、資料品質、土地利用

* 台灣大學地理系副教授

Associate Professor, Department of Geography, National Taiwan University, Taipei, Taiwan, Republic of China

前言

一個完整正確的資料庫是地理資訊系統的核心。就以往的經驗顯示，建立資料庫的費用往往佔據整個系統建構總預算的百分之八十以上（Aronoff, 1989）。以台北市政府的地理資訊系統推動工作為例，其資料庫建立的經費達數億元之多，建立資料庫所需經費之昂貴可見一斑（林峰田等，1995）。由於其費用之昂貴，大型資料庫的建立也就有「只許成功，不許失敗」的壓力，而成功與否的關鍵之一即在於資料的品質是否合乎所需。從使用者的角度而言，由地理資訊系統所推導出來的結論都是建立在資料庫之上，若是資料本身有問題的話，則其所推演出來的結論勢將難以立足，連帶地也可能造成錯誤的決策。是以，就系統的建立和使用而言，資料品質的管理都是一個必須嚴肅以待的課題。

由於具有模糊（fuzzy）及碎形（fractal）的特質，許多地理現象是無法明確地展現（Lowell, 1991；Altman, 1994）。這種本質上的特性再加上測量及數化工作所可能引進的誤差，所有地理資料都一定含有誤差，然而在地理資訊系統的推廣和應用中，有關資料的品質經常被有意無意地迴避，例如資料建製單位往往很不喜歡聽到他們的所生產的資料被批評。這種心態對於GIS的長期發展是不健康的。唯有透過嚴謹的檢測與分析，才能確保資料的品質，也唯有正確適用的資料，才能確保地理資訊系統的應用價值。基於這種的認知，學術界對於資料品質的研究一直非常關心，將誤差及其相關議題視為GIS研究上的重要工作。以美國國家地理資訊及分析中心（NCGIA, National Center of Geographic Information and Analysis）的任務規劃為例，誤差研究即是其所選定的第一個研究議題，而該中心所發表的第一本研究成果專書也是以資料品質為主題（Goodchild *et al.*, 1989）。

地理資料品質所牽涉的面向雖然極其廣泛，但研究的重心不外乎是誤差的量度和處理上所衍生的相關事宜。針對誤差研究的内容和階層順序，Veregin (1989)提出了五個層級的工作重點，如圖1。



圖1 誤差研究工作的五個層級

這個架構標示了誤差研究的五個階層，由誤差的偵測、量度、到誤差傳導的模擬、管理策略的提出，以期達到消除誤差之最高層目標。這個架構所指出的僅是誤差處理的過程，就以往的研究題目來分析，可以發現研究題目的範疇非常廣泛，涵蓋了上述架構中底層到高層問題探討，並不見得是依序解決一層層的問題。由於地理資料的内容和格式繁多，各種資料的品質問題不盡相同，研究者所需探討的議題也就出現許多不同的面向。可以從個案的探討著手，歸納出誤差的理論，也可以透過演繹的方法，推論出誤差的模式或理論。

土地利用型態是一項重要的國土資訊，上自國土計畫的重點，下至各種區域計畫、環境規畫、工程開發設計、環境影響評估等工作都必須利用到它。有鑑於土地利用資料在自然資源及農林資源經營管理的重要性，行政院農業委員會從民國 70 年起開始生產五千分之一的土地利用圖，其後並委託學術單位進行數化和轉檔工作，先後完成民國 65 年和 71 年等兩個不同年度的數值化土地利用檔案（賴進貴、孫志鴻，1994）。張長義等（1993）也分別在不同的研究計畫建立了大量的海岸地區土地利用資料。

本研究屬於個案研究，以一個現成的土地利用資料庫為例，探討土地利用資料的品質。希望透過個案的探討，可以增進吾人對於資料誤差的瞭解，進而提出一般性的誤差處理策略，提昇資料品質。

研究問題及目標

本研究針對台灣西部海岸地帶的土地利用資料加以探討。該資料庫包括四個不同年度、四個縣市濱海鄉鎮的土地利用資料。本研究僅抽取其中的部份資料來加以分析。這些數值資料是利用現有地圖數化而來，原始地圖是農林航空測量所生產的五千分之一相片基本圖，由人工判讀地圖上的資料，標示出一個個土地利用單元。所有的土地利用型態被劃分成十四類。每一個單元依照其土地利用的型態而被賦予一個屬性值，以代表其土地利用的型態。然後再利用點圖板 (digitizer) 將每一個土地利用單元的邊界和屬性輸入到電腦中，產生 ARC/INFO 格式的圖檔。每一個年度的土地利用資料都依據這個程序分別產生。由於原始圖檔所涵蓋的面積廣大，故整個資料庫建立的工作是由數名助理合力完成，每一名助理分別負責部份地區的判讀和數化，再經由接圖的方式將一張張圖幅的檔案拼湊成一個完整的圖檔。

這些土地利用資料在建構過程中，具有幾個特點。第一點，整個資料庫是分別由一群助理判讀和數化。第二點，四個不同年度資料分別重複建構，可以預期有大量的邊界資料會重複數化。第三點，整個資料庫是由原始地圖數化而得，歷經原始地圖的測量、繪製、印刷，乃至人為的判讀、點圖等，整個資料建構過程中所牽涉的層面繁複。本研究的目的是在瞭解這類土地利用資料的品質問題，希望透過這個研究，不僅可以提高這批資料的應用價值，並且增進吾人對於其他具有相同特性資料品質之瞭解。本研究的主要工作在於分析這些資料的品質，分析其誤差的類型及尺度，進而探討造成誤差的原因，最後並就所研擬的誤差消除方法加以比較、分析。

原始資料共包括四個年度的土地利用類型，限於篇幅，本報告僅就其中兩個年度的資料加以測試和分析。研究的進行兼顧實測和分析，由於原有圖檔是以 ARC/INFO 系統儲存，所以測試工作也是在 ARC/INFO 系統進行。

資料處理與分析

透過 ARC/INFO 系統所提供的基本查詢功能，本研究整理出這兩個年度的土地利用的統計資料。茲將這兩個年度的土地利用資料統計列舉於表 1。

本研究區僅涵蓋沿海鄉鎮，呈南北狹長之形狀，區內包括大面積之潮汐灘地。研究區的兩個不同年度資料，分別是利用該年度的五千分之一相片基本圖進行判釋、數化所產生。因海岸地形的變遷以及航照拍攝時間的不同，造成基本圖上陸地範圍的差異，資料所顯示的總面積也就略有差異，兩者之間相差 2174 公頃，約為總面積的百分之八。由於本文之研究方法是透過圖層套疊偵測土地利用變遷，進而探討資料之品質問題，研究區面積的些微差異無礙於研究分析之進行，所以本研究未就總面積的變化加以深究。

表1 1976和1992的土地利用統計表

年 代	1976			1992		
土地利用類型	多邊形個數	平均面積(m ²)	總面積(公頃)	多邊形個數	平均面積(m ²)	總面積(公頃)
1.農地	396	511605	20259.5	419	357938	14997.6
2.果園	23	32995	75.9	1	12745	1.3
3.林地	92	66932	615.7	76	99371	755.2
4.草地	5	27758	13.9	34	23119	78.6
5.建築地	1866	7391	1379.1	2519	7321	1844.1
6.墓地	57	12661	72.2	93	10902	102.5
7.工業用地	27	6849	18.5	20	95804	1916.1
8.運輸用地	42	122877	516.1	29	228517	662.7
9.養殖魚池	1758	17234	3029.8	1487	56350	8379.3
10.養雞場	1	10295	1.0	81	5846	47.4
11.灌溉設施	111	21339	236.9	136	22873	311.1
12.荒地	142	633723	2661.6	44	35620	156.7
13.水域	84	61178	513.9	45	35525	159.9
14.其他	107	52499	561.7	23	53361	93.9
總 計	4611	64966	29955.9	5008	55475	27781.8

由表1所列的資料顯示，研究區中的多邊形數量增加、平均面積減小，顯示土地利用單元有分割化、複雜化的趨勢。就各類土地利用總面積的變化而言，其中最明顯的變化是，農地的面積由近 20,259 公頃降為 14,997 公頃，養殖魚池的面積則從 3,029 公頃左右增加至 8,379 公頃。這種面積變化的探討在傳統的土地利用研究常見到。利用地理資訊系統的技術，若將不同年代的土地利用資料予以套疊，研究者可以歸納出土地利用變遷的型態，或找出土地利用變遷的機制。Muller et al. (1994) 甚至嘗試由土地利用資料發展蒙特卡羅模式 (Monte Carlo Model) 來預測土地利用變遷的趨勢。這些研究都利用地理資訊系統所提供的分析和模擬功能，在研究方法上頗具新意，然而，其分析結果的可信度受制於資料的品質而含有不確定性 (uncertainty)，必須視資料的品質高低而定。針對這些因資料品質而引起的不確定性，本研究利用 GIS 中的疊圖功能來進行資料品質的分析。

疊圖 (overlay) 不僅是 GIS 中常用的一種運算，也是分析資料品質的主要方法之一，可結和不同圖層上的資料以產生一張新的圖層，並利用此一新的圖層中進行分析。疊圖功能可以是數學上的加、減、乘、除運算，也可以是邏輯上的 Boolean 運算，如聯集、交集等。在 ARC/INFO 系統中所經常被使用的 intersect 指令，即是一種 Boolean 式的運算。它是利用幾何位置來構出新的區域，這個區域承續了原有圖層上的所有屬性 (Chrisman, 1989)。本研究的第一項測試即採取疊圖來檢測資料品質。首先以 ARC/INFO 的 intersect 指令將兩個不同年度的圖層套疊。在第一次的套疊作業中，並未設定額外的參數，但是由於原始圖檔設有 0.57m 的 fuzzy_tolerance 值，所以套疊結果自動承續這個設定。這兩張圖層套疊的結果，可以顯示各類型土地利用變遷的面積和空間分佈情形。表2列舉這兩個年度間的土地利用變遷之統計數字。

表2 1976至1992年的土地利用變遷

1976 1992	農地	果園	林地	草地	建築地	墓地	工業地	運輸地	養殖池	養雞場	灌溉設施	荒地	水域	其他
農地	824 17409 1434	34 20562 70	178 1478 26	×	2004 574 115	110 1426 16	30 1186 4	762 853 65	859 1021 88	1 415 0	333 2248 7	12 1679 2	95 2240 21	30 47593 143
果園	×	×	×	×	×	×	×	×	×	×	×	×	×	1 12745 12745
林地	188 2098 39	2 3021 1	2 12900 26	×	81 818 7	×	1 1200 0	26 2732 7	10 7281 7	×	8 2667 2	8 215477 172	1 2 0	7 734 1
草地	31 16718 52	×	4 284 0	×	14 1122 2	2 859 0	×	10 275 0	26 9310 24	×	4 475 0	×	3 832 0	×
建築地	3633 1538 589	6 653 0	161 1441 23	×	1936 5415 1048	10 1495 2	21 2901 6	1121 459 52	1545 816 126	1 9854 1	108 430 5	36 2835 10	80 577 5	38 1088 4
墓地	215 1783 38	×	3 11808 4	×	26 2567 7	51 10105 52	×	8 374 0	28 362 1	×	10 433 0	×	6 508 0	1 3569 0
工業地	502 1412 71	×	8 3483 3	×	384 592 23	3 388 0	8 8927 7	95 6985 66	124 443 6	1 30 0	70 751 5	12 4133 5	14 533 1	5 2438 1
運輸地	886 2642 234	7 958 1	64 1336 9	1 866 0	847 787 67	14 395 1	19 313 1	312 5688 178	965 1072 104	×	102 1453 16	9681 36 35	58 1575 9	46 1690 8
養殖池	89 23485 209	19 3291 6	89 3138 28	6 22988 14	1612 608 98	50 391 2	10 345 0	900 986 89	1825 13907 2538	×	166 2740 46	80 53162 425	136 8541 116	165 19928 329
養雞場	89 3956 35	×	2 508 0	×	36 1390 5	×	×	13 337 0	33 1739 6	×	2 974 0	×	4 662 0	1 2094 0
灌溉設施	401 2553 102	×	21 1317 4	×	79 366 3	5 258 0	2 1159 0	150 3136 47	163 3132 51	×	117 58196 800	12 10511 13	38 2374 9	37 1054 4
荒地	40 27982 112	×	16 948 2	×	41 601 3	1 405	×	11 2667 3	35 6411 22	×	2 1877 1	×	6 3166 2	9 14591 13
水域	126 4667 59	×	5 142 0	×	11 817 1	4 1226 1	×	23 396 1	39 12764 50	×	7 4368 3	7 2109 2	32 11634 37	37 3922 15
其他	18 5500 10	1 1156 0	1 22708 2	×	7 2805 2	×	×	2 60 0	17 1236 2	×	5 2642 1	11 49617 55	3 11298 3	7 19443 0

註記 1. × 表示該欄並沒有資料。

2. 每一格的最上層數值表示該種變化的多邊形數目；中間的數字顯示平均面積（單位：m²）；最下層數字顯示總面積（單位：公頃）。

疊圖結果分析

兩張圖層套疊的結果可以顯示土地利用變遷的類型，透過地理資訊系統中的空間分析功能也可顯示土地利用變遷的空間分佈形式。然而，這些所偵測到的變遷是實際發生的變遷或是資料誤差所造成的「變遷」，則有待進一步分析。本研究分別從空間幾何部份和屬性部份來探討疊圖結果的可信度。

在空間幾何方面，Veregin (1989) 曾引述 McAlpine 和 Cook (1971) 所提出的一個有關疊圖誤差的檢驗式。這個檢驗式是利用圖層套疊前後各相關圖檔中的多邊形數目，來分析圖層中的可能誤差。這個由實驗推導出來的經驗公式如下：

$$m_c = \left[\sum_{i=1}^n m_i^2 \right] \quad (1)$$

其中， m_c 為疊圖後所產生圖層中的多邊形數目， m_i 為疊圖前各個圖層中的多邊形數目， n 為圖層數。此一檢驗式所代表的意義是，在常態狀況下，疊圖後所產生圖層的多邊形數量應該接近於各原始圖層的多邊形數量開根號之和的平方，亦即數學式 (1) 的關係。若是疊圖結果所得到的實際多邊形數量，遠大於利用式 (1) 所預測的數值，則顯示原始圖檔中可能含有大量的誤差。這個檢驗式雖然只是一個經驗公式，而非絕對的指標，但是仍可作為評估資料品質的參考之一。

利用這個檢驗公式來驗證本研究的結果，兩個原始圖層分別有 4,611 及 5,008 個多邊形，套疊結果預期有 19,230 個多邊形，實際結果則有 28,085 個多邊形，遠超過預期的數字，這個差距顯示套疊的結果中可能存有許多實際上並不存在的偽多邊形 (pseudo-polygons)。若就本研究所處理的土地利用資料的本質來探討，則前述的套疊結果所得之多邊形數應該要比等式 (1) 所預測的數字來得少。因為，就土地利用變遷的型態而言，一般的土地利用變遷可大致分成三類：第一類，土地利用單元的邊界並沒有改變，但是單元中的土地利用形式改變，例如：一塊農田變成魚塢。在這種情形下，不同圖層間套疊所產生的多邊形數量並沒有改變。第二類，舊有的土地利用單元被切個成數個新單元，其中部份單元上之土地利用並未改變，另一部份則有改變，例如：一大塊空地上蓋了一間工廠，原有的一個空地單元變成兩個單元，一個單元依然是空地，另一個單元則是由空地變成工廠。在這種情形下，新的土地利用單元之邊界包含了舊有的邊界，疊圖結果的多邊形數量就是新圖層的多邊形數量。第三類情形是，原有的土地利用單元有擴充或合併的現象，例如：工廠的擴建、或數個單元被合併成一間學校的用地。在這種情形下，原有的土地利用邊界消失，疊圖結果的多邊形數量比原圖中的多邊形數量還多。這三種類型中，只有第三類的土地利用變遷型態在疊圖之後會增加大量的多邊形，第一、二類則不會。綜合這三種可能的變遷而言，不同年度的土地利用圖在疊圖後所產生的多邊形數量應該要遠小於等式 (1) 所預估的數量。目前所得到的疊圖結果與這個推論相反，這種現象顯示上述的套疊過程中產生大量實際上並不存在的偽多邊形，這些多餘的多邊形暗示著誤差的存在。

針對點圖誤差，Veregin (1989) 也引述了 Goodchild (1978) 所提出一個誤差模式，該模式認為：將同一條線用點圖板數化兩次，不管是由不同人各點一次或同一人點兩次，兩條數值線的坐標點會有所差異。若以 V_1 和 V_2 來代表兩次點圖所分別選取的點數，在隨機狀況下這兩條數值線套疊後所產生的偽多邊形數量 S ，和二次點圖所選取的點數 V_1 、 V_2 存有下列的關係式：

$$s = \frac{2v_1v_2}{v_1 + v_2} - 3 \quad (2)$$

這個公式是由數學推導而來。運用在點圖檢核上，這個公式能用來估計兩個不同點圖者可能引入的 *slivers*。這個公式所預測的 *slivers* 是針對隨機誤差所引起的偽多邊形。利用這個公式來檢證本研究的兩個不同年度資料，第一年圖層中的點數為 91,192 點，第二年圖層中的點數為 60,960 點。由於這兩個圖層中的邊界不完全相符，無法直接套用公式（2）。然而若僅欲求得大概的數值，則可以把點數較少的圖層為基準代入公式（2），如此所估計產生的 *slivers* 數將有數萬個之多。本研究所產生的實際 *slivers* 數目遠小於這個估計值。顯示表二中所顯示的可能誤差，不全然是點圖時的隨機性誤差所引起。

就屬性資料的檢核，Huang *et al.* (1992) 曾將誤差分為邏輯不一致、非邏輯的不一致、和分裂誤差。其中，若有邏輯不一致的現象出現，則可以用來偵測誤差。所謂的「邏輯不一致」是指資料出現了邏輯上不合理的現象，這種誤差可以由邏輯推理來檢核。非邏輯上的誤差通常是指資料的屬性部份和真實世界的現象不相符，但是這種不符卻又不構成邏輯上的矛盾，也就無法直接由邏輯上去推論資料是否準確，只能透過實際地面的檢核來確認其準確性。分裂誤差指的是在資料更新或疊圖的時候，產生許多細小的多邊形 (*slivers*)，這些多邊形大都是由數化過程所引進。本研究採用這一套架構中的「邏輯不一致」概念，來驗證土地利用資料的可能誤差。

就土地利用變遷的型態而言，大多數的土地利用變遷都是由低密度的利用演變成高密度的利用，例如：由農地轉換成工業用地、建築、交通用地等等，而不是由建築物轉換成農地。本研究的兩張圖層分別包括 14 個類別的土地利用，兩個年度的土地利用型態可以有 196 種組合，但是其中的某些組合含有邏輯不一致的謬誤，發生的可能性應該很小，例如由建築物變為農田的情形在理論上是不可能發生的，所以這一類的多邊形個數也就理應非常少。然而，在本研究的疊圖結果中，筆者卻發現共有 2,004 個多邊形是由建築地變成農地，這些多邊形是一種可能的誤差，值得進一步檢測。

本研究將具有這種變換的多邊形列舉、統計，發現其中的面積都很小，遠小於原來房舍的面積，顯見這些都不是實際的土地利用轉換，而是因資料建立過程所產生的 *slivers*。

表 3 農地之個數及面積統計

年 代	農地多邊形數 (A)	面積小於 100 m ² 農地多邊形數 (B)	耕地之平均面積 (m ²)	面積小於 100 m ² 所佔百分比 (B/A)
民國 65 年	396	2	511605	0.5
民國 81 年	419	10	357938	2
建築地->農地	2004	791	574	39.47

誤差消除測試

疊圖過程所產生的 *slivers*，是普遍存在的資料品質問題。這個問題吸引多位學者的注意 (Chrisman, 1987)，在 ARC/INFO 的軟體中也設想到 *slivers* 的存在，並且提供因應的辦法。第一個方法是執行 *intersect* 指令進行疊圖時，使用者可以設定一個 *fuzzy_tolerance* 值，將距離小於這個容許值的相鄰點合併在一起，如此可以將點圖時所產生的鄰近點合併，進而減少產生 *slivers* 的可能；另一個方法是直接利用 *eliminate* 指令將小於一定面積的多邊形去除，由於大多數的 *slivers* 面積都非常小，因此這個指令也可以消除 *slivers*。

本研究使用這兩種方法來去除 *slivers*，以測試這兩種方法的效果。測試過程中，分別以不同的 *fuzzy_tolerance* 值來執行 *intersect* 指令，也分別以不同面積為 *eliminate* 的參數，試圖消除

slivers。由於建築地變成農地是一種邏輯上不一致的土地變遷型態，所以本研究以建築地的變遷情形做為資料誤差的指標，分析各種不同的誤差消除方法之效果，若是有大量的多邊形是由建地變為農田和魚塭，則顯示套疊結果仍有誤差。表四所列的是在不同的處理方式之下，原有建築地轉換成各類型土地利用的多邊形之數量。

表4 不同誤差消除方法結果之比較
(建築地轉為其它土地利用的多邊形數)

土地利用類別 測試項目	農地	果園	林地	草地	建築地	基地	工業用地	運輸用地	養殖魚池	養雞場	灌溉設施	荒地	水域	其他
採 5 m 容許值之多邊形數	1428 771 1100420	×	65 1008 65514	10 1601 16010	1801 5943 10702899	27 2515 67916	255 785 200043	688 903 621559	1227 775 950330	31 1686 52277	51 468 23889	29 836 24231	9 984 8536	6 3342 20052
採 10 m 容許值之多邊形數	940 1074 1009338	×	48 1293 62052	10 1683 16827	1784 6451 11507608	23 3063 70445	139 875 121581	436 863 376337	888 1022 907739	28 1971 55179	20 787 15749	24 911 21870	6 1237 7424	3 6865 20595
採 20 m 容許值之多邊形數	445 1884 838509	×	28 2369 66326	6 2369 14216	1434 8435 12095491	16 4531 72487	33 1578 52062	116 1339 155308	453 1568 710384	15 2813 42189	9 752 6771	12 1497 17964	4 1790 7162	2 9501 19001
採 'eliminate' (area < 10m ²) 之多邊形數	1784 694 1237925	×	77 881 67860	10 1571 15709	1872 22914 42895773	25 2736 68401	344 1127 387643	802 1250 1002236	1455 683 993679	35 1430 50037	72 814 58581	37 718 26560	11 817 8982	6 3271 19626
採 'eliminate' (area < 100m ²) 之多邊形數	1213 1254 1520613	×	47 1412 66378	9 10824 97413	1726 23071 39821144	22 3115 68533	239 2544 608003	588 2549 1498649	1061 1027 1090122	31 1618 50158	50 1762 88083	28 973 27235	8 1126 9005	6 3322 19931
採 'eliminate' (area < 500m ²) 之多邊形數	531 3693 1960978	×	26 3213 83539	6 16247 97484	1393 28355 39499119	17 4221 71751	116 10853 1258900	280 7236 2026075	503 2483 1249213	21 2587 54317	14 5839 81748	16 1783 28533	7 1425 9972	2 9414 18828

註記 1. × 表示該欄並沒有資料。

2. 每一格的最上層數值表示該種變化的多邊形數目；中間的數字顯示平均面積（單位：m²）；最下層數字顯示總面積（單位：m²）。

由上述的測試結果顯示，資料產生後的後續誤差處理工作，無法完全消除各種類型的誤差。而這些處理方法也存在一個兩難的問題：若是 fuzzy_tolerance 的值或 eliminate 指令所設定的面積太小，勢必無法將所有的誤差消除；然而，若是一味地增加 fuzzy_tolerance 的值或 eliminate 指令的面積，則可能會將一些真正改變的多邊形給刪除。就面積大小和點之間的距離而言，錯誤的資料和真實資料之間並沒有明確的分野，所以這些方法也就無法明確地消除所應該去除的多邊形。

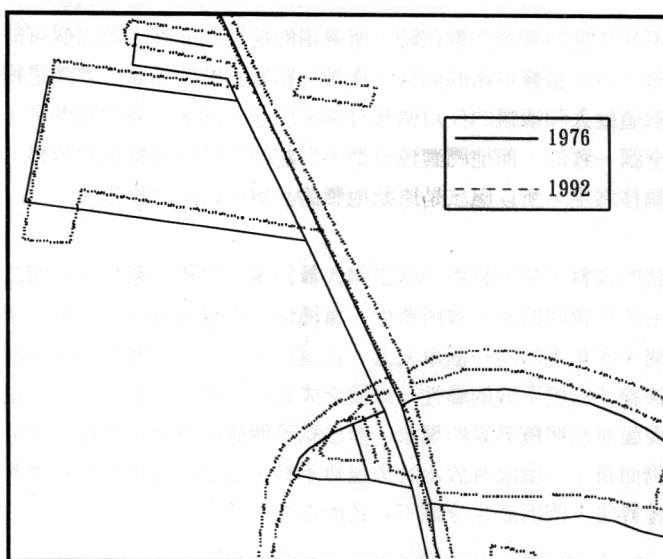
本節的測試結果顯示，單純地使用 fuzzy_tolerance 或 eliminate 指令並無法完全消除誤差。為了提昇誤差處理過程中的分辨能力，本研究認為應該從誤差的本質和發生原因來加以探討。在找出誤差的類型之後，才能進一步研擬消除誤差的方法。

誤差的探討

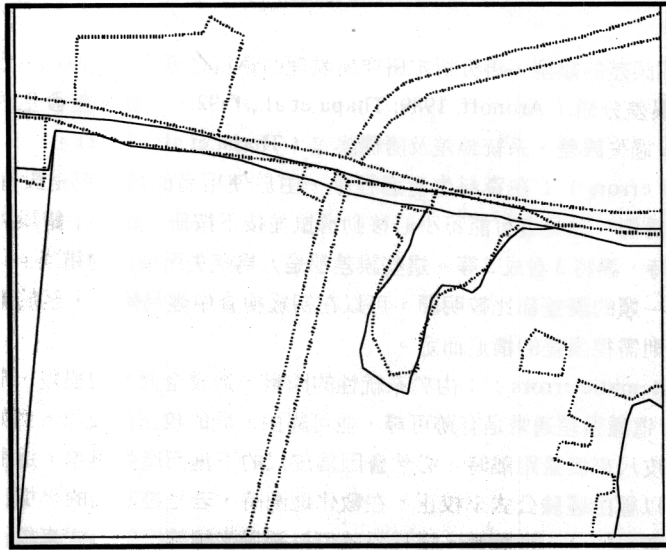
本節將首先介紹誤差的類型，再分析本研究所處理的資料之可能誤差原因。現有文獻中，已經提出多種不同的誤差分類（Aronoff, 1989; Thapa et al., 1992）。就誤差發生的本質而言，誤差可以分成三大類：過失誤差、系統誤差及隨機誤差（Thapa et al., 1992）。

- (1)過失誤差（gross errors）：在資料生產過程中，由於使用者的疏忽或是設備使用不當所造成。以點圖工作為例，點圖者可能不小心移動滑鼠並按下按鈕，輸入了錯誤的坐標值，或是在輸入屬性數值時，誤將3看成5等。這些誤差都是人為疏失所造成的錯誤。以資料的空間幾何部份而言，這一類的誤差都比較明顯，可以在初級檢查中輕易除去，至於屬性部份的誤差能否輕易去除，則需視誤差的情形而定。
- (2)系統性誤差（systematic errors）：由於系統性的因素，造成全面性的誤差，形成統計上的偏差（bias）現象。這種誤差通常是有跡可尋，並可經由系統的校正而去除。例如傳統測量工作中，若懸空使用皮尺來度量距離時，必然會因為皮尺的下垂而產生誤差，這種誤差即是一種系統性誤差，可以經由經驗公式來校正。在數位化地圖時，若是控制點的坐標位置有誤的話，也會造成所數位化資料的全面性誤差，這種誤差可以透過坐標轉換的方式來校正。
- (3)隨機誤差（random errors）：這種誤差產生的主要原因，是儀器設備的精度不足或人類感官能力上的限制，以致無法取得精確的結果。例如，在讀取刻度時，只能準確到刻度的單位，再下一位數字就必須用估計的，而這個估計值可能高估也可能低估，估計的誤差即是一種隨機誤差。隨機誤差應該是有正有負，可以互相抵消，理論上也可以經由重複的量度來取得平均值，以得到較準確的坐標值。就實務工作而言，基於經費和工作量的考量，卻不太可能重複數位化一筆資料以求取其平均坐標值。因此，這種隨機誤差無法完全去除，其大小和原始地圖的品質、測量儀器的精密度、點圖員的小心程度有關。

本報告從測試資料中抽取部份樣本，以顯示圖檔中所隱藏的誤差類型。圖二展示了不同類型的誤差，圖中的實線和虛線分別表示同一個地區在不同年度下的土地利用單元界線。



(a)系統性誤差



(b) 隨機性誤差

圖2 誤差範例

這些誤差顯示了在原始的兩張圖檔之中，分別包括隨機性的誤差、系統性的偏移、和過失性的誤差。這些誤差的產生可以追溯到資料的生產過程。資料生產過程中的每一個步驟都可能引進誤差，而這些誤差累積的結果造成疊圖後所顯示的種種問題。本研究所測試的資料，其誤差的來源可以從三個主要層面來探討。

1. 原始地圖的品質

在分屬兩個不同年度的圖層中，出現整條道路偏移的誤差。這種誤差的發生的可能原因，第一個是由於兩張不同年度的資料在數化時，所選用的控制點偏移。第二個可能是，地圖印製時的清繪、套色不準確，造成整條道路的偏移，連帶地影響點圖的品質。當然這種偏移也可能是點圖時的控制點坐標數值輸入和滑鼠的使用偏移有誤所引起。兩者之間的差別是，控制點坐標問題所造成的偏移將是全圖一致的，而地圖套色不準所引起的則是局部特徵的偏移。本研究所觀測到的是以局部地區的偏移為主，所以應該是原始地圖的品質所引起的誤差。

2. 判讀誤差

本研究所測試的資料，分別是由一組工作人員判讀、點圖而產生。從圖二的範例中顯示，有些誤差的發生是由於判讀時的不一致所產生。這種情形出現在同一個土地利用單元跨越兩張相鄰的地圖上、或兩個不同年度的同一個單元上。在這些情況下，分屬於兩張地圖的土地利用是相同的，但是判讀者卻給予它們不同的屬性，以致造成資料判讀不一致的誤差。這種誤差頗為普遍，判讀者可能常會面臨無法明確界定的現象，無法給予明確的分類，也就造成分類上的不一致，因而導致誤差。舉例而言，一個沒有放水的魚塢算不算魚塢？這種模糊的特性常見於地理現象上，很難一一加以明確界定；因而產生分類不一致所造成的誤差。

3. 點圖誤差

利用點圖機數化是一個方便卻也非常容易產生誤差。由於目前所使用的電腦都是以數位

(digital) 方式來儲存資料，一般線性的地理資料 (linear geographic features) 在數化後大都被轉換成一連串離散點 (discrete points) 的組合。如何選取一組點來代表一條線，牽涉到線概括化 (line generalization) 的處理，這個處理又往往具有高度的主觀判斷。不同的人數化同一條線時，所選的點不可能完全相同，即使由同一個人數化同一條線二次，也不太可能會產生兩組坐標完全相同的點。這種誤差也正是 Goodchild (1978) 所提到的偽多邊形，目前還沒有徹底的解決辦法。

除了人為選點的差異之外，即便是一組標示清楚的點，在數化時也無法取得一致的結果，點圖板的品質和人為操作上，都會造成資料誤差。就儀器方面，點圖板的穩定性、精密度、線圈分佈的均勻與否，都會影響所讀取坐標的準確性。點圖人員的視覺極限是點圖誤差的最重要因子。目前一般的點圖機大都可以精密到千分之一吋的解析度，而人眼感官上的精度和地圖上線條的寬度，都遠比於機器的精密度來得粗略，所以相對地造成所讀取坐標不一致的誤差。以台北市圖解地籍圖數化工作的經驗為例，地籍圖上的每一個點都被重複點兩次，兩組坐標之間的差距還會有超過 0.2mm 的誤差 (台北市政府測量大隊，1991)，如果碰到地圖的品質不良，點圖人員管理不善的話，則點圖的誤差當更甚於此。點圖之不確定性可見一斑。

關於點圖的研究吸引了大量的研究者。曾有專門論文加以實驗探討，提出經驗數字，這些數字可以用來推估點圖作業的準確性 (Warner and Carson, 1991)。Dunn 等人 (1990) 進一步指出點圖品質是和數化模式、原圖線的寬度、操作者的誤差有關。細緻的線條比較能明確的數化，減少操作者主觀的判定或疏忽所造成的誤差，當然數化線段的複雜度也和誤差大小極其相關。Maffini et al. (1989) 依據資料的平滑特性，分別測試點圖工作的誤差，以探討影響數值化誤差的因子。他將資料分兩方面來討論，一是連續性的圓滑資料，一是非連續性資料。他發現連續性圓滑資料的點圖會受其點圖時間及比例尺所影響，不連續資料則只受比例尺影響，與時間無關。這兩種資料的誤差等級皆會因比例尺的變小而增加。點圖的過程所引發的種種誤差，還是一個有待研究解決的領域。

以本研究所測試過的誤差消除方法，fuzzy_tolerance 參數的設定僅能消除點圖時的隨機誤差，卻無法消除其他類型的誤差。而其他類型的誤差，也不見得都是面積狹小的多邊形，因此 eliminate 指令也無法完全消除它們。ARC/INFO 系統所提供的兩個誤差消除方法之限制，也就顯露無遺。由此可見目前的技術水準尚無法完全去除點圖的誤差。

未來研究方向

從現有的資料建立來源到後續的誤差消除，吾人發現數值地理資料一定含有誤差，並且是無法完全消除的。就地理資料的本質，即使將來的科技更加純熟，誤差的存在依然是無可避免的。針對這些資料和數化的特性和本質而言，本研究認為土地利用資料的分析尚有三個方向可以努力。

1. 技術層面：目前的 fuzzy_tolerance 作法是以點之間的距離遠近作為合併的依據，此種作法無法有校地整合重複數化的同一條線。未來研究應該嘗試發展以線為基礎的 fuzzy_tolerance 操作，透過線和線之間重合程度的分析來決定兩條線是否可能為同一條線的重複數化。這種作法在目前一般軟體中尚未出現。如何將這個作法落實到實際的系統運作，是未來可以努力的一個研究方向。
2. 觀念層面：所有的測量數據，都有其準確度的極限，一定都包含誤差。因此，考慮資料品質，不應該單方向的追求誤差的完全消除，而應該更務實地以資料的適用性 (fitness of use) 來思

考。資料一定有誤差，不是好不好的問題，而是適用性的問題。基於這種考量，研究者應該反過來思考，在現有的資料品質之下，土地利用變遷資料的解析度以及所處理的土地利用單元大小兩者之間如何取捨。這個觀念亦即在地理資料處理分析上，要加入「信賴區間」的觀念，小於一定尺度的分析並無意義。至於這個信賴區間如何量度，則是一個有待解決的研究問題。

3. 資料處理上：在誤差的偵測上，未來的研究應該嘗試利用空間分析、空間統計，來區分偽多邊形和正常多邊形，然後加以去除。本研究認為利用形狀指數（shape index）即屬於一個值得進一步測試的方法，因為通常 slivers 是比較狹長的，所以利用形狀指數應可以增加區辨的效果。

本研究透過測試、分析，對於誤差的處理提出一份報告，未來的研究將就三個研究方向繼續努力，以期對於誤差的解決提出更進一步的方法。

謝 辭

本研究所使用的測試資料是由台大地理系張長義教授所提供，研究過程中的部份電腦操作工作是由王慧勳同學協助進行，謹在此一併致謝。

引用文獻

- 林峰田、朱子豪、賴進貴（1995）台北市地理資訊系統整體規畫之研究（第三期），中華民國都市計劃學會研究報告，台北市政府委託研究。
- 張長義、劉英毓、蔡博文（1993）彰雲海岸敏感地區土地利用變遷之研究，國立台灣大學地理研究所。
- 賴進貴、孫志鴻（1994）台灣地區數值土地利用資料庫建立之研究，國立台灣大學地理研究所研究報告，農委會委託研究。
- Altman, D. (1994) Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *International Journal of Geographical Information Systems*, 8(3): 271-289.
- Aronoff, S. (1989) *Geographic Information Systems: A Management Perspective*, Ottawa, Canada: WDL Publications, 133-149.
- Chrisman, N.R. (1987) The accuracy of map overlay: a reassessment, *Landscape and Urban Planning*, 14: 427-439.
- Chrisman, N.R. (1989) Modeling error in overlaid categorical maps. In: Goodchild, M.F. and Gopal, S. (eds.) *Accuracy of Spatial Database*, New York: Taylor & Francis, 21-34.
- Dunn, R., Harrison, H.R. and White, J.C. (1990), Positional accuracy and measurement error in digital databases of land use: an empirical study, *International Journal of Geographical Information Systems*, 4(4): 385-397.
- Goodchild, M.F. (1978) Statistical Aspects of the Polygon Overlay Problem, *Harvard Papers on Geographic Information Systems*, Vol. 6, Reading, MA: Addison-Wesley.
- Goodchild, M.F. and Gopal, S. (1989), *Accuracy of Spatial Databases*, New York: Taylor & Francis, 290p.
- Huang, S., Douglas, D. and Zhang, A. (1992) A logic-based approach for GIS error handling, *GIS/LIS Proceedings*, 343-359.
- Lowell, K.E., (1991) On the incorporation of uncertainty into spatial data systems, *GIS/LIS Proceedings*, 484-493.