

疊圖誤差模擬之研究*

A Simulation Study for Errors of Maps Overlay

賴進貴** 王慧勳***

Jinn-Guey Lay Hui-Shiung Wang

摘要

一個完整正確的資料庫是地理資訊系統的核心。基於地理資料的本質和測量及數化工作的限制，所有地理資料的準確度皆屬有限，亦即一定含有誤差。為了確保地理資訊系統分析結果的可信度，資料品質是相關研究的重要課題。疊圖是地理資訊系統中非常普遍的一個功能，這個功能對於資料品質的優劣極為敏感，可以清楚地顯示資料品質的瑕疵。所以它除了被廣泛應用在區位選址、變遷偵測等分析工作之外，也是資料品質檢核的一項利器。本研究利用電腦模擬疊圖操作以探討資料品質問題。研究工作包括兩個部份：第一部份，利用 S-Plus 軟體產生數種不同類型的誤差，再將這些誤差加到原始圖檔多邊形的坐標中，以模擬點圖過程的誤差，然後再將誤差圖層和原圖套疊；第二部份，針對疊圖後的結果，分析多邊形的數量和幾何特性，以試圖找出 slivers 並加以消除。本研究分別以多邊形之面積和形狀指數進行分析，結果顯示 slivers 的形狀不見得如傳統所認定的細長。此一發現有待進一步研究證實。

關鍵字 Key words

地理資訊系統

GIS

誤差模擬

error simulation

形狀指數

shape index

* 本文曾於第十四屆測量學術及應用研討會中發表（1995）

An earlier version of this paper has been presented at the 14th Symposium of Science & Technology of Surveying and Mapping(1995).

** 台灣大學地理系副教授

Associate Professor, Dept. of Geography, National Taiwan University.

*** 台灣大學地理系研究生

Graduate student, Dept. of Geography, National Taiwan University.



ABSTRACT

A complete and correct database is fundamental to reliable applications of GIS. Due to the nature of geographic data and limitation of measurement, all geographic data are of limited accuracy. In order to assure the credibility of GIS analysis, the quality of data is an important issue within GIS research. Overlay is a common operation of GIS, which is very sensitive to the quality of data involved. As a result, besides of its widely use in site selection and change detection, overlay is often adopted for error detection of GIS data. This research uses overlay as a tool to investigate errors in two steps. First, a number of erroneous map layers are created using the S-Plus software and ARC/INFO.

These maps contain errors of various types and magnitude. Through overlay, this research investigates the relation between data quality and the results of overlay. Second, this research uses spatial factors, such as acreage and shape index, to identify and eliminate sliver polygons. Results show that slivers are not thin and long as we used to think. This controversial finding is subject to further research.

一、研究介紹

一個完整正確的資料庫是地理資訊系統的核心。基於地理資料的本質和測量、數化工作的特性，所有地理資料都是有限度的準確度（of limited accuracy）（Goodchild et al., 1989），亦即資料一定含有誤差。疊圖（overlay）是地理資訊系統中很普遍的一個分析功能，被廣泛地使用在如區位選址、變遷偵測等各種工作上。這個運算功能所產生結果的可信度，和原始資料的準確度有極大的關連。本文第一作者曾嘗試使用疊圖指令來偵測兩個不同年度之間的土地利用變遷（賴進貴 1995；Lay, 1995）。然而，在初步的測試中發現，由地理資訊系統中所顯示的土地利用變遷，有極大部份是資料品質的瑕疵所引起的假變遷，而非地面上實際發生的變遷。從現有文獻的閱覽中顯示，疊圖過程中所隱含的誤差大小和原始資料品質之間的關連性，是目前所知較少的一環，值得吾人深入探討。本研究的主旨在於探討誤差和疊圖結果準確性之間的關連，透過誤差模擬和疊圖分析，以瞭解資料誤差的傳遞，並且進一步測試消除誤差的方法。

二、誤差的來源與類型

數值地理資料庫的資料來源雖然日趨多元，然而現有的紙張地圖仍然是一個最重要的來源，這些既有地圖的數化工作不外乎是由數位板（digitizer）數化和掃描儀（scanner）數化，這兩者之間目前仍以數位板數化較為普遍，本研究的模擬和分析乃以數位板點圖的誤差為重點。由於目前所使用的電腦都是以數位（digital）方式而來儲存資料，一般地圖上的線性地理特徵（linear geographic features）在數化之後，都被轉換成一連串離散點（discrete points）的

組合。在點圖過程中，如何選取一組點來代表一條線，牽涉到線的概括化 (line generalization) 作業。同一條線由數個人重複數化，每個人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，也不可能會選取完全相同的一組點 (Warner et al., 1991)。當然，也有一些例外的情形，以地籍圖的數化工作為例 (台北市政府地政處測量大隊，1991)，每一塊地籍宗地是由一個個界址點所界定，每一個界址點的位置都明確標示，點與點之間則是由直線相連，在數化時也就不會出現線概括化的差異。一般的地圖數化時，是難以避免選點差異所造成的誤差，而當所數化的線性特徵是複雜曲折時，產生誤差的機率和誤差的大小隨之增加。

即使是地圖上所欲數化的節點都已標示清楚，同一組點被數化兩次所得到的坐標值將無法完全相同，因為點圖作業本身會引入許多額外的誤差。影響點圖數化準確性的因子包括儀器及人為兩項。就儀器方面，數位板的電子線圈若分佈不平均，會影響坐標的連續性，且其誤差範圍將呈波動的型態。不同機種的數位板的品質有所不同，造成輸出結果不一致的差異。此外，數位板的解析度高低，也直接影響數化作業的誤差大小，理論上解析度為 1000 dpi 的數位板，其輸出結果之準確度就比 250 dpi 的數位板之準確度要來得高。除了儀器的誤差之外，點圖人員的在操作上的極限和地圖的品質，都會引入點圖過程中的主要誤差。目前，一般中上水準的數位板都可以精確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，卻都遠遜於數位板的解析度，所以造成坐標讀取上的不一致。以台北市圖解地籍圖數化的工作為例，所欲數化的界址點大都可以明確辨認，所採用的儀器在操作上限制了點圖的速度，間接提昇點圖過程的審慎性，在任務派遣上也以輪班的方式來維持點圖人員的體力和注意力。然而，即使是在這種多重防範的制度之下，所產生的數值資料還是有超過 0.2 mm 的誤差存在 (台北市政府地政處測量大隊，1991)。一般的數化作業通常沒有採行如地籍圖數化作業般有系統的品管作業，其所生產資料之品質當遠遜於地籍圖數化的品質，點圖作業所引進的誤差尺度之大可見一般。

有關於點圖作業的誤差之探討，曾吸引了大量研究者的興趣，並有許多專文加以探討。Dunn 等人 (1990) 指出點圖品質和數化模式、原始地圖上線條的寬度、點圖者的操作誤差有關，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，所欲數化的線段之複雜度和誤差大小也極其相關。Maffini 等 (1989) 依據資料的平滑特性，分別測試點圖工作的誤差，以探討影響數化作業誤差的因子。他將資料分成兩種類型來加以討論，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點圖作業的速度及地圖比例尺所影響，不連續資料則只受比例尺大小影響，與點圖速度無關。這兩種類型的資料誤差尺度會隨比例尺的變小而增加。

在一般測量作業上，誤差依其本質不同而被分成三大類：過失誤差、系統誤差及隨機誤差 (Thapa et al., 1992)。這三大類的誤差，同樣地也出現在點圖數化的工作上。

(1) 過失誤差 (gross errors)：在數化過程中，由於使用者的疏忽或是設備使用不當所造成。

就空間幾何的資料而言，錯誤發生的情形可能包括如：記錄或讀取數字時的疏忽，誤將 5 6 視為 36，或是點圖時不小心移動滑鼠並按下按鈕，以致輸入了錯誤的坐標值。就屬性方面，過失誤差可能包括輸入錯誤的屬性數值等。一般而言，這一類的誤差都比較明顯且出現的頻率較小，可以利用目視檢查來加以去除。

(2) 系統誤差 (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

離 (bias) 現象。這種誤差通常是有跡可尋，可以經由系統的校正而去除。例如：測量時，量尺的熱漲冷縮即是一種系統誤差，可以經由經驗公式來校正；點圖時，若是控制點的坐標位置有誤差或彼此不一致時，也會造成全面性的坐標誤差。理論上系統性誤差有可能利用坐標轉換的方式來校正，然而實際作業時並不一定能取得轉換所需之參數。

(3) 隨機誤差 (random errors)：這種誤差產生的原因主要是因為儀器設備的精度不足或操作人員感官能力上的極限，無法取得非常精確的結果。這種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測量來取得平均值，以得到較準確的估計值。地圖數化的作業是不太可能重複數化一筆資料數次以求取平均坐標值，此類誤差也就無法完全去除，受點圖者的細心程度和點圖板的品質所影響，其大小可以視為資料精度的一項指標。

上述三種類型的誤差，系統性誤差和隨機性誤差是數化過程中比較常見的類別。系統性誤差可能發生在地圖紙張的伸縮、控制點坐標誤差等。隨機誤差和點圖板的精確度、點圖者的細心程度有關。本研究乃針對這兩種不同類型的誤差加以模擬和分析。

三、研究問題及方法

在地理資訊系統的作業中，疊圖本身純粹是一種坐標幾何的運算，利用電腦的快速計算能力求取兩個圖層套疊之後，線段和線段的交叉點，並利用位相關係 (topological relations) 的索引，將一條條相連的線弧 (arcs) 連接成多邊形，並賦予新圖層中各個多邊形屬性。疊圖運算過程的精密度是受所使用電腦系統的有效位數影響，所以疊圖運算的誤差大抵是來自於有效位數的多寡。由於大部份的地理資訊系統都是以倍準 (double precision) 的格式來儲存和處理坐標值，所以由系統運算過程中所引入的誤差其實非常有限。疊圖過程的誤差主要來自於原始圖層的品質。原始圖層中的邊界是否明確，會影響到資料的品質，並連帶影響疊圖的準確性。也正是基於同樣的因素，疊圖作業常被作為檢定資料品質的一種方法。

本研究所探討的資料品質問題是以向量式資料為主，利用電腦模擬的方式來檢驗資料品質對於疊圖結果的影響。研究問題包括兩個：第一、各種不同類型和不同大小誤差對於疊圖的作業，會造成什麼樣的影響；第二、疊圖結果中所顯現的誤差多邊形，具有什麼空間特性？如何利用這些空間因子來偵測誤差，進而加以消除？

針對這兩個問題，本研究利用電腦系統的運算功能，在人工數化的數值圖層中，分別加上不同尺度和類型的模擬誤差，再進一步分析各種不同的資料誤差對於疊圖作業結果的影響。其次，本研究針對所產生誤差的空間特性加以分析，以試圖區分正常多邊形和偽多邊形，進而偵測出圖層中的誤差，然後加以去除。本研究考慮的空間特性包括面積和形狀指數 (shape index) 兩項。

四、模擬誤差實驗

本研究的第一項測試工作是模擬點圖時所可能產生的誤差，包括系統性和隨機性誤差。測試工作採用一個現成圖檔為標準，該圖是由五千分之一的相片基本圖上數化生產的土地利用圖檔，在 ARC/INFO 系統下產生。本研究利用 ARC/INFO 系統中的 ungenerate 指令，將

這個圖層中每一個多邊形的邊界坐標值輸出成一個 ASCII 檔，然後利用 S-Plus 這個軟體來產生數種不同大小、類型的亂數，再以自行開發的程式將這些亂數加到原始圖層中的 X、Y 坐標值，以產生數個包含不同誤差的 ASCII 圖檔。本研究所產生的亂數是以常態分佈（normal distribution）的形式產生。在這些常態分佈的亂數中，其平均值和標準差有六種不同的組合，平均值分別是地面上 5 公尺和 10 公尺，標準差分別是地面上 5 公尺、10 公尺、及 20 公尺。地面上 5 公尺的距離是五千分之一地圖上 1 mm 的長度，本研究認為總和各種可能的誤差來源之後，這個尺度的誤差並非少見。本部份測試總共產生六個 ASCII 圖檔，並將這些新的圖

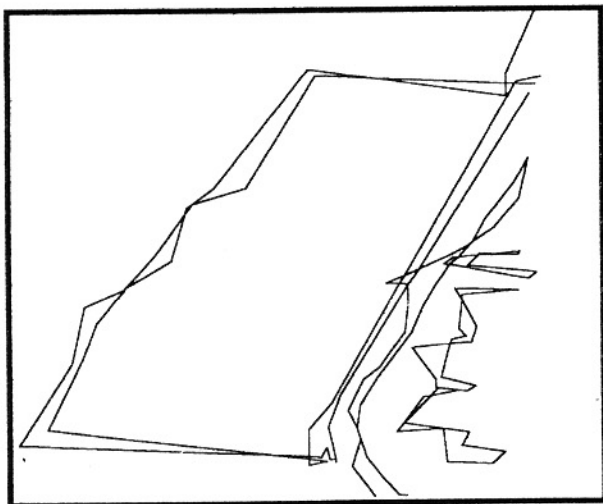


圖1. 模擬隨機誤差之圖層範例

檔傳回 ARC/INFO 系統，利用 generate 指令產生新的圖層，以模擬隨機誤差。由於亂數的給予完全是隨機的，是以會使本來應為相同坐標值的點，因為所賦予的亂數不同而成為不同的點。這個過程就好像是同一條線段被數化兩次，因數化時的不一致，使得兩張理應緊密吻合的圖層無法完全的重疊。整體而言，這個過程模擬了同一張地圖被重複數化兩次再加以套疊後所產生的誤差，圖1是測試圖檔中部份地區放大的範例之展示。這六個誤差檔案套疊後的統計數字，詳列如表1。

表1. 系統性誤差所產生的圖層之統計

檔案名稱	多邊形數	弧線數	點數	面積小於 100 m ² 之 多邊形數	面積小於 50 m ² 之 多邊形數	面積小於 10 m ² 之 多邊形數
原始圖層	83	157	1213	0	0	0
m5_sd5	636	1342	3016	360	294	141
m5_sd10	849	1707	3434	464	370	186
m10_sd5	630	1295	2976	360	293	130
m10_sd10	904	1814	3521	504	405	185
m20_sd5	535	1094	2817	260	205	93
m20_sd10	885	1775	3508	493	395	193

說明：圖層名稱說明：m5_sd5 為平均誤差為5公尺，標準差為5公尺之隨機誤差圖層。

如表1所示，加入模擬誤差後產生的新圖層，其多邊形數目顯著地增加，約為原始圖層的6-11倍，平均面積則急遽減少，減少了約82-90%，而原始圖層的最小多邊形面積為134平方公尺，但新圖層中最小多邊形的面積則不到1平方公尺，且有半數以上的多邊形面積不到100平方公尺。這些數據顯示加入誤差後的圖層，在圖層套疊後產生許多細小多邊形，亦即所謂的slivers。

本部份模擬測試的控制因子分別為隨機誤差的平均值和標準差，測試結果顯示，slivers的數量和誤差的平均值並沒有直接的相關，在標準差為5公尺且平均誤差分別為5公尺、10公尺、和20公尺的三個圖層，其多邊形數量分別為636、630、535，並沒有隨平均值的增加而改變，相同的情形也出現在標準差為10公尺的三個圖層之中，隨平均誤差由5增加為20，其多邊形數量分別為849、904、885，並沒有隨平均值的增加而改變。在另一個控制因子，多邊形的數量和標準差大小呈現了有規律的變化，在平均誤差分別為5、10、20的三組測試資料中，多邊形的數量都隨標準差的增加而增加。以本研究有限的測試資料，尚難以斷定誤差的標準差大小和slivers數量的絕對關係，然而本研究中有測試組合所呈現的一致性變化，應當也絕非偶然。本研究在多邊形數量和誤差的標準差之間的發現，對於資料品質管制和誤差研究提供一個值得進一步探討的線索。

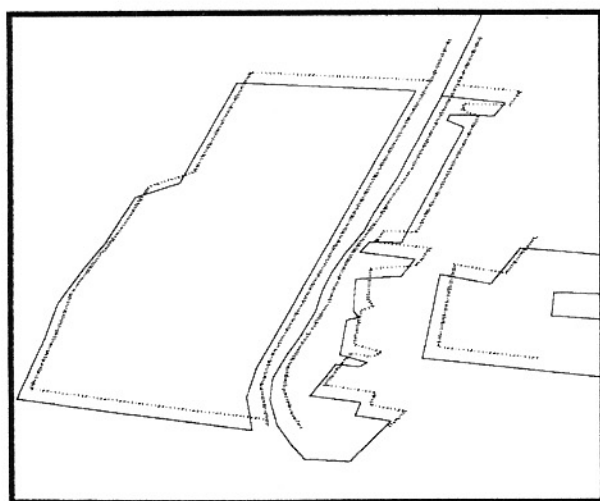


圖2. 含系統性誤差的圖層和原圖的套疊

為了單獨測試控制點錯誤所引起的誤差，本研究另外測試了系統性誤差對於疊圖作業的影響。利用電腦模擬產生了兩個圖檔，分別是將原圖的控制點（TIC points）坐標給予5公尺到15公尺不等的誤差，這些誤差同樣是由亂數產生，以模擬數化時的控制點輸入誤差。隨後，將含有控制點誤差的資料進行平移（transform）的工作，如此即產生兩個具有系統性誤差的圖層。最後，將這兩個有誤差的圖層和原始圖層套疊，得到兩個不同的疊圖結果。圖2是這種系統性誤差的範例，表2則顯示圖層套疊後的統計資料。

表2. 系統性誤差疊圖後之多所產生的新圖層的相關資訊

控制點平移之範圍	多邊形數	多邊形平均面積	最小多邊形面積(m^2)	面積小於 $100 m^2$ 之多邊形數	面積小於 $50 m^2$ 之多邊形數	面積小於 $10 m^2$ 之多邊形數
0-10 m	435	16381	1.1	118	70	8
15-25 m	416	16976	0.79	98	60	28

本測試的主要目的在於比較系統性誤差和前述的隨機性誤差之差異，而非不同尺度的系統誤差之間的比較。測試結果顯示，系統性誤差所產生的偽多邊形比起隨機誤差所產生的要來得少，相對而言，各個多邊形的面積也比較大。從圖2的範例顯示，系統性誤差所造成的 slivers，其形狀狹長，平均之形狀指數應該會比較小。有關於這兩類誤差之間的進一步比較，請見下節的空間因子分析。

五、多邊形的空間因子分析

從圖1和圖2的範例顯示，資料數化時所造成的系統性誤差和隨機性誤差，在疊圖後都是以 slivers 的形式出現。吾人若可以偵測出 slivers 的特性，就可以進一步加以刪除。面積的大小是區分 slivers 和正常多邊形的第一個空間特性。以本研究的測試資料而言，原始圖層中最小的多邊形之面積為地面上 134 平方公尺，換算到五千分之一的地圖上約為 $2\text{mm} \times 2\text{mm}$ 的一小塊地區，以人工判圖和數化工作的解析度而言，這個面積已經是趨近於可辨識的極限，小於這個面積的多邊形即可以視為非正常產生的多邊形。根據點圖工作的精密度和原始地圖的比例尺大小，可以概略訂出 slivers 和正常多邊形之間面積大小的分際。然而，我們也可以預期兩者之間的面積區分並無明確的劃分，中間存有一個模糊帶，這是誤差偵測和管理上所需要面對的難題。

爲了增加誤差偵測的準確性，本研究嘗試納入形狀指數以作為區分 slivers 的另一項空間特性指標。由 sliver 這個英文字的原意來看，它應該是具有狹長的形狀，所以理應可以由多邊形形狀的扁平與否來偵測 slivers。本研究以 Miller's Index 來計算形狀指數，此一指數的範圍在 0 與 1 之間，越扁平的多邊形其指數將趨近於 0，越近似於正圓的多邊形則指數趨近於 1 (Campbell, 1991, p. 116)。表三標示了原始圖層和各個誤差圖層的形狀指數之統計。由數據顯示，原始圖層中多邊形的平均形狀指數要大於各個誤差圖層的平均形狀指數。然而，就同一個圖層所有多邊形的平均形狀指數和面積小於 200m^2 的多邊形之平均形狀指數來比較（表 3 之第二欄與第四欄），兩者之間並沒有明顯差距。這個結果和歷來普遍認定 slivers 是狹長多邊形的概念並不相符，顯示形狀指數本身無法單獨作為一個區分正常多邊形和 slivers 的指標。

各個圖層中，多邊形的面積和形狀指數的統計詳列如表 4，這些數字可以作為 slivers 數量多寡的一個參考。如果僅保留面積和形狀指數大於特定門檻的多邊形（如表 4 的第六、七欄數字），而將其它多邊形予以合併，則各圖層中的多邊形數量將大幅減少，得以比較接近於原始圖層中的多邊形數。



表3. 各圖層形狀指數之統計數字

圖層名稱	所 有 多 邊 形		面積小於200 m ² 之多邊形	
	平均值	標準差	平 均 值	標 準 差
原始圖層	0.52	0.22	0.48	無
m5_sd5	0.36	0.17	0.36	0.18
m5_sd10	0.39	0.19	0.41	0.18
m10_sd5	0.33	0.20	0.34	0.18
m10_sd10	0.40	0.19	0.41	0.17
m20_sd5	0.35	0.20	0.36	0.18
m20_sd10	0.40	0.18	0.41	0.17
平移 0-10m	0.36	0.23	0.46	0.20
平移15-25m	0.45	0.22	0.49	0.21

表4. 各圖層多邊形之面積和形狀指數（SI）統計數字

圖層名稱	面積 < 200m ² 之多邊形數	面積 < 100m ² 之多邊形數	SI < 0.3之多邊形數	SI < 0.52之多邊形數	面積 > 200m ² 且SI > 0.3之多邊形數	面積 > 100m ² 且SI > 0.3之多邊形數
m5_sd5	431	360	289	499	84	120
m5_sd10	563	464	267	614	156	233
m10_sd5	431	360	302	493	84	117
m10_sd10	624	504	301	631	152	234
m20_sd5	329	260	241	416	96	136
m20_sd10	588	493	276	618	186	252
平移 0-10m	195	120	191	322	87	142
平移15-25m	138	98	104	239	191	227

六、結 論

資料品質優劣攸關地理資訊系統發展的成敗，所以資料誤差是學術研究者和系統應用者所共同關心的課題。本研究首先利用電腦模擬產生數種不同類型的資料誤差，再據以分析各種不同誤差在疊圖作業中所呈顯的影響。其次，本研究採用空間因子的分析以分辨正常多邊形和 slivers 的差異，以之作爲誤差刪除的依據。本研究的具體結論如下：

- (1)在點圖數化的過程中，隨機性位置誤差之平均值對於疊圖過程所產生的 slivers 多寡沒有明顯影響，誤差的標準差大小和 slivers 的數量則呈現規則性的變化，隨標準差變大而增加。此一現象顯示，在資料品質管理和誤差研究中，點圖工作的精確性是一項重要因子。
- (2)疊圖過程所產生的 slivers 多寡，是資料品質優劣的一項指標。如何消除 slivers 是資料品質管理上的一項工作，如何區分 slivers 和正常多邊形也就成爲一個重要工作。面積大小是偵測 slivers 的首要因子。爲了增加偵錯能力，本研究另外引進形狀指數，試圖以多邊形的形狀扁平與否來偵測 slivers，研究者意外地發現 slivers 不必然要比正常多邊形來得細長。此一現象是否乃是因爲本研究所模擬的誤差之特性所引起，或同樣出現在正常點圖作業所產生的誤差中，是研究者所急於知道，將進一步加以研究探討。就 slivers 的偵測而言，目前所使用的兩個空間分析因子都有其無法突破的模糊帶問題，未來應考慮納入更多的空間因子，以增加偵測的準確性。

本研究結合電腦模擬與空間特性分析，提昇了電腦系統在研究上的功能。研究本身尚屬初步測試，未來研究將進行更完整的誤差模擬，並進行更有系統的測試，希望藉由這種大量的模擬和分析，可以提昇吾人對於資料誤差相關議題的瞭解，並提供系統管理者在資料品質管理上有具體的作法。

參考文獻

- 1.賴進貴（1995）：數值地理資料品質之研究：以土地利用資料爲例，國立台灣大學地理學報，第19期，第91-103頁。
- 2.台北市政府地政處測量大隊（1991）：台北市圖解地籍圖數值化工作手冊。
- 3.Campbell, J. (1991), Map Use and Analysis, Dubuque, IA: Wm. C. Brown Publishers .
- 4.Dunn, R., Harrison, H.R. & White, J.C. (1990), Positional Accuracy and Measurement Error in Digital Databases of Land Use: An Empirical Study, INT. J. of Geographical Information Systems, Vol. 4, No. 4, pp. 385-397.
- 5.Goodchild, M. F. & Gopal, S. (Ed.) (1989), Preface, Accuracy of Spatial Databases, p. xiv, New York: Taylor & Francis.
- 6.Lay, J.G. (1995), Data Accuracy in A Landuse Study, Proceedings of GeoInfomatics '95 Hong Kong, pp. 713-720.
- 7.Maffini, G. , Aron, M. & Bitterlich, W. (1989), Observations and comments on the generation and treatment of error in digital GIS data, in M. F. Goodchild & S. Gopal (Ed.) : Accuracy of

Spatial Databases, New York: Taylor & Francis. pp. 55-68.

8. Thapa, K. & Bossler, J. (1992), Accuracy of Spatial Data Used in Geographic Information Systems, PE&RS, Vol. 58, No. 6, pp. 835-841.

9. Warner, W. S. & Carson, W. (1991), Errors Associated with a Standard Digitizing Tablet, ITC Journal, pp. 82-89.

