

# 國立清華大學

## 碩士論文

題目：圓上實驗之群集分析與模型配適  
Cluster Analysis and Model Fitting  
for Experiments on Circle

所 別：統計學研究所 組別：工業統計組

指導教授：鄭少為 (Shao-Wei Cheng) 博士

姓 名：楊惇昱 (Dun-Yu Yang)

學 號：9524522

中華民國九十七年六月

## 摘要

一般我們在做實驗設計時，實驗區間通常為一正方體，但科技業的工程師經常在晶圓上做實驗設計，此時實驗區間就是一個圓。實驗區間為正方體的各種資料分析，可在大多數實驗設計的書找到，至於實驗區間是一個圓的資料分析，則很少在各種文獻中出現。本篇論文討論如何利用圓上的數據來依圓上的曲面相似程度做群集分析。這時會發生一個有趣的新問題，就是必須考慮曲面的旋轉。本論文提供了一套方法來解決曲面旋轉的問題，並將其套用在真實晶圓數據上，本論文還介紹了如何在圓上建立一個允許極座標之圓心位置改變之模型。

關鍵字：反應曲面，實驗區間，建模區間，極軸，多項式模型，群集分析。

## 致謝辭

首先非常的感謝指導教授鄭少為老師。在這一年裡我學到了非常多的東西，每周的報告我都覺得我收穫很多，讓我在做研究時，不會覺得是在浪費時間在做無意義的事。老師實力真的非常的好，而且又很有耐心的指導我，當老師的學生，讓我覺得自己真的很幸運。老師，真的很謝謝你。也感謝逢甲大學的教授，李燦銘與王迺聖老師，謝謝他們在我大學的時候，教了我非常多的東西，讓我有機會來到清大這個環境。感謝旺宏電子公司提供寶貴的數據，讓我有機會可以分析圓上的數據。感謝志群學長，明中，安國，佳煒，昱宏，靖元，盈甫在我遇到問題時，提供了非常多的幫助。謝謝班上所有的同學，這兩年我過得非常的愉快。謝謝你們大家。



# 目錄

1	緒論	1
2	文獻回顧：極座標建模與多變量群集分析	3
2.1	導論	3
2.2	配適適當的二階模型	3
2.3	群集分析	5
3	如何把圓上的曲面做群集分析	10
3.1	導論	10
3.2	晶圓旋轉	14
3.3	適合度的問題	21
4	真實數據的分析	26
4.1	導論	26
4.2	不轉換	30
4.3	觀測值減掉該片晶圓所有觀測值的平均	39
4.4	每個晶圓上的觀測值最大最小值調整成一樣	48

4.5	標準化 . . . . .	57
4.6	本章結論 . . . . .	66
<b>5</b>	<b>極座標模型</b>	<b>67</b>
5.1	導論 . . . . .	67
5.2	如何解決不同極軸的問題 . . . . .	67
5.3	允許極座標圓心位置改變 . . . . .	69
<b>6</b>	<b>總結</b>	<b>83</b>
	參考文獻	91



## 圖目錄

2.1	未改變極軸實驗區間與建模區間 . . . . .	4
2.2	改變極軸實驗區間與建模區間 . . . . .	5
2.3	單一聯結法 . . . . .	6
2.4	完全聯結法 . . . . .	6
2.5	平均聯結法 . . . . .	7
2.6	單一聯結法的樹狀圖 . . . . .	8
2.7	完全聯結法的樹狀圖 . . . . .	9
2.8	平均聯結法的樹狀圖 . . . . .	9
3.1	兩個反應曲面旋轉前的圖型 . . . . .	17
3.2	量測座標的位置與等高線圖 . . . . .	17
3.3	兩個反應曲面旋轉後的圖型 . . . . .	18
3.4	旋轉後量測座標的位置與等高線圖 . . . . .	18
3.5	把判定係數分三大塊 . . . . .	22
3.6	第一個方法的等高圖 . . . . .	23
3.7	當 $k = 1$ 第二個方法的等高圖 . . . . .	24

4.1	晶圓 1, 2, 3, 4的曲面 . . . . .	28
4.2	晶圓 5, 6, 7, 8的曲面 . . . . .	29
4.3	第一種轉換的盒狀圖 . . . . .	30
4.4	第一種轉換的等高圖 . . . . .	32
4.5	二階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	33
4.6	二階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	33
4.7	二階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	33
4.8	三階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	34
4.9	三階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	34
4.10	三階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	34
4.11	四階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	35
4.12	四階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	35
4.13	四階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	35
4.14	二階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	36
4.15	二階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	36
4.16	二階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	36
4.17	三階階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	37
4.18	三階階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	37
4.19	三階階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	37
4.20	四階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖 . . . . .	38
4.21	四階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖 . . . . .	38

4.22 四階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖 . . . . .	38
4.23 第二種轉換的盒狀圖 . . . . .	39
4.24 第二種轉換的等高圖 . . . . .	41
4.25 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單 一聯結法樹狀圖 . . . . .	42
4.26 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完 全聯結法樹狀圖 . . . . .	42
4.27 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平 均聯結法樹狀圖 . . . . .	42
4.28 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單 一聯結法樹狀圖 . . . . .	43
4.29 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完 全聯結法樹狀圖 . . . . .	43
4.30 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平 均聯結法樹狀圖 . . . . .	43
4.31 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單 一聯結法樹狀圖 . . . . .	44
4.32 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完 全聯結法樹狀圖 . . . . .	44
4.33 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平 均聯結法樹狀圖 . . . . .	44



4.34 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一聯結法樹狀圖 . . . . .	45
4.35 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全聯結法樹狀圖 . . . . .	45
4.36 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均聯結法樹狀圖 . . . . .	45
4.37 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一聯結法樹狀圖 . . . . .	46
4.38 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全聯結法樹狀圖 . . . . .	46
4.39 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均聯結法樹狀圖 . . . . .	46
4.40 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一聯結法樹狀圖 . . . . .	47
4.41 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全聯結法樹狀圖 . . . . .	47
4.42 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均聯結法樹狀圖 . . . . .	47
4.43 第三種轉換的盒狀圖 . . . . .	48
4.44 第三種轉換的等高圖 . . . . .	50
4.45 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖	51

4.46 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖	51
4.47 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖	51
4.48 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖	52
4.49 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖	52
4.50 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖	52
4.51 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖	53
4.52 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖	53
4.53 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖	53
4.54 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖	54
4.55 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖	54
4.56 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖	54
4.57 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖	55
4.58 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖	55
4.59 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖	55
4.60 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖	56
4.61 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖	56
4.62 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖	56
4.63 第四種轉換的盒狀圖 . . . . .	57
4.64 第四種轉換的等高圖 . . . . .	59
4.65 二階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	60
4.66 二階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	60

4.67	二階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	60
4.68	三階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	61
4.69	三階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	61
4.70	三階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	61
4.71	四階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	62
4.72	四階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	62
4.73	四階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	62
4.74	二階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	63
4.75	二階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	63
4.76	二階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	63
4.77	三階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	64
4.78	三階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	64
4.79	三階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	64
4.80	四階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖 . . . . .	65
4.81	四階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖 . . . . .	65
4.82	四階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖 . . . . .	65
5.1	不同極軸與對應判定係數的圖型 . . . . .	68
5.2	量測座標的位置與編號 . . . . .	73
5.3	圓心為編號 1 的量測座標 . . . . .	73
5.4	圓心為編號 8 的量測座標 . . . . .	74
5.5	圓心為編號 20 的量測座標 . . . . .	74

5.6	圓心爲編號 42 的量測座標 . . . . .	74
5.7	圓心爲編號 76 的量測座標 . . . . .	75
5.8	圓心爲編號 112 的量測座標 . . . . .	75
5.9	圓心爲編號 10 的量測座標 . . . . .	75
5.10	圓心爲編號 11 的量測座標 . . . . .	76
5.11	圓心爲編號 12 的量測座標 . . . . .	76
5.12	圓心爲編號 13 的量測座標 . . . . .	76
5.13	圓心爲編號 14 的量測座標 . . . . .	77
5.14	圓心爲編號 15 的量測座標 . . . . .	77
5.15	圓心爲編號 16 的量測座標 . . . . .	77
5.16	圓心爲編號 17 的量測座標 . . . . .	78
5.17	圓心爲編號 18 的量測座標 . . . . .	78
5.18	圓心爲編號 19 的量測座標 . . . . .	78
5.19	圓心爲編號 20 的量測座標 . . . . .	79
5.20	圓心爲編號 21 的量測座標 . . . . .	79
5.21	圓心爲編號 22 的量測座標 . . . . .	79
5.22	圓心爲編號 23 的量測座標 . . . . .	80
5.23	圓心爲編號 24 的量測座標 . . . . .	80
5.24	圓心爲編號 25 的量測座標 . . . . .	80
5.25	圓心爲 (0.1,0.1) . . . . .	81
5.26	圓心爲 (0.15,0.15) . . . . .	82

## 表目錄

3.1	群集分析資料形式 . . . . .	10
3.2	食品的營養資料 . . . . .	11
3.3	晶圓的數據 . . . . .	13
3.4	解決判定係數不高的第一個方法 . . . . .	23
3.5	解決判定係數不高的第二個方法 . . . . .	24
4.1	極軸當參數各種多項式模型的最佳判定係數 . . . . .	27
5.1	判定係數 . . . . .	70



# 第 1 章

## 緒論

隨著時代的進步，半導體相關產業也越來越重要了。積體電路 (integrated circuit, 簡稱 IC) 是資訊產品最重要的元件，許許多多的日常用品，上面皆佈滿 IC 半導體，而晶圓就是製造半導體最重要的材料。製造晶圓的原料主要是二氧化矽，工程師把二氧化矽經過一系列分解後就會提煉出多晶矽結晶，再將多晶矽拉成不同直徑大小的矽晶棒，矽晶棒經過研磨、拋光和切片後，就得到晶圓片。

每個晶圓片皆有許多的晶粒 (chip)，要判斷一片晶圓的好壞，通常會在晶粒上做實驗，利用實驗後的數據來判斷晶圓的好壞。現在某電子公司提供了晶圓的化學氣相沈積 (Chemical Vapor Deposition, 簡稱 CVD) 的真實數據，其收集數據的方法如下。因為一個晶圓內包含著許多小晶粒，每個晶粒都可以做實驗得到化學氣相沈積的數據，故我們可將整個晶圓想像成爲一個實驗區間，而將晶粒的化學氣相沈積的數據想像爲該實驗區間上的反應值，這時化學氣相沈積的數據與晶粒位置的關係便可用一個反應曲面來代表。當因爲成本的考量而無法在所有晶粒上做實驗時，工程師就會在機台的晶圓擺放區上挑選幾個具有代表性的位置，這些位置我們稱爲量測座標。當每片晶圓擺放在機台上，就會有某些晶粒對應到這些量測座標，工程師便會在這些晶

粒上做實驗得到反應值。在這個實驗裡，機台上的晶圓擺放區的 $X, Y$ 座標可視為兩個不同的因子，而每一個量測座標可視為一組設定值，所以此實驗可視為二因子的實驗設計。

關於實驗區間為圓的資料分析，文獻上很少討論，不過它在我們日常生活中是常見的。實驗區間為圓的資料分析通常都很特別，因為圓沒有上下左右的區別，所以分析時，常常要考慮圓是否要旋轉。在本論文中，我們利用某電子公司所提供的八片晶圓實驗數據，透過這些量測座標上晶粒實驗後的數據，就可以研究晶圓上的反應曲面如何隨著晶粒的幾何位置改變而變化。現在工程師有興趣的問題是，任兩片晶圓上的反應曲面是否有差異，有那些晶圓的反應曲面較類似，以及可分類出多少種不同型態的反應曲面。

要解決晶圓上反應曲面的分類問題，我們可使用統計上的群集分析 (cluster analysis) 法。群集分析是把一大堆物件歸納為少數幾群的一個方法。如果要對晶圓上的曲面作群集分析，必須解決曲面旋轉的問題，因為有些曲面旋轉至某個角度會很像另一個曲面。但是旋轉後會發現要比較差異的兩片晶圓，其做實驗的量測座標無法重疊的問題，這會造成定義兩個曲面的距離無法直接套用傳統群集分析裡的方法。我們在本論文中提出一套方法來解決曲面旋轉的問題，這些會在本論文第三章討論。本論文第四章把某電子公司提供的真實數據套用我們的新方法來做群集分析。曲面的旋轉就是改變極軸的觀念，Lin (2007) 有非常的清楚介紹改變極軸對迴歸模型的影響，但該篇論文討論的極座標模型圓心皆固定在 $(0, 0)$ 。在本論文中我們考慮了更一般的模型，其在圓上配適迴歸模型時要同時考慮不同的圓心與極軸，這些會在第五章介紹。



## 第 2 章

# 文獻回顧: 極座標建模與多變量群集分析

### 2.1 導論

資料的來源是工程師在機台上均勻的取點, 再把晶圓放在機台上做實驗後得到的數據。每筆資料皆有機台上  $x, y$  的直角座標, 與它所對應的反應值數據  $w$ 。下一節會介紹如何配適極座標上的二階模型。因為現在擁有的晶圓有 8 片, 而每個晶圓上的曲面各自有各自的特色, 因此決定對圓上的曲面作群集分析, 我們提出的做法會在下一章討論, 在本章第三節會先簡介層次群集法 (Hierarchical Clustering Methods), 以為下一章提供需要的基礎知識。

### 2.2 配適適當的二階模型

現在資料提供的是  $x, y$  直角座標, 與它所對應的反應值數據  $w$ 。所以一般情況下, 可配適一個直角座標的二階迴歸模型:  $w = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 y + \beta_4 y^2 + \beta_5 xy + \varepsilon$ 。但是當反應值具有圓效應(circular effect)時, 會發現此模型的判定係數 (coefficient of determinant, i.e.  $R^2$ ) 常不會很高。Lin (2007)提出了把  $x, y$  直角座標, 變為



$r, \theta$  的極座標。其中

$$r = x^2 + y^2$$

$$\theta = \begin{cases} \cos^{-1} \left( \frac{x}{\sqrt{x^2 + y^2}} \right) & \text{if } y > 0, \\ 2\pi - \cos^{-1} \left( \frac{x}{\sqrt{x^2 + y^2}} \right) & \text{if } y \leq 0 \end{cases}$$

此時若配適一個二階模型:  $w = \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 \theta + \beta_4 \theta^2 + \beta_5 r \theta + \varepsilon$  會發現判定係數變高許多。在Lin (2007)中, 討論了極軸改變對配適模型的影響。圖 2.1 和圖 2.2 呈現了使用不同的極軸時, 在 $(r, \theta)$ 極座標上的實驗點會如何改變, 更詳細的說明, 請見Lin (2007)。由圖2.1與2.2可發現, 當極軸改變後, 實驗點在 $(r, \theta)$ 這個建模區間上的位置也改變了, 因此所建立的迴歸模型之判定係數也會不同, 在第五章會詳述改變極軸對模型的影響, 與它的一些性質。

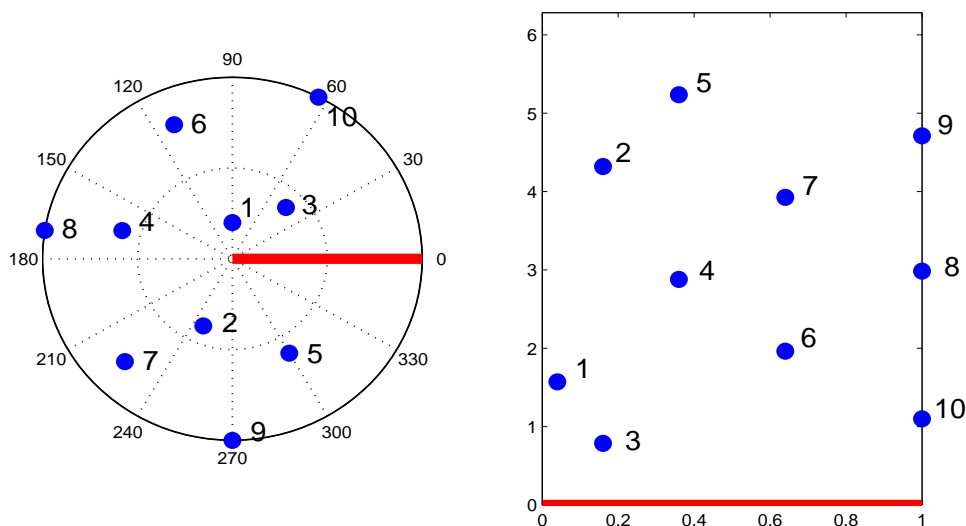


圖 2.1: 未改變極軸實驗區間與建模區間

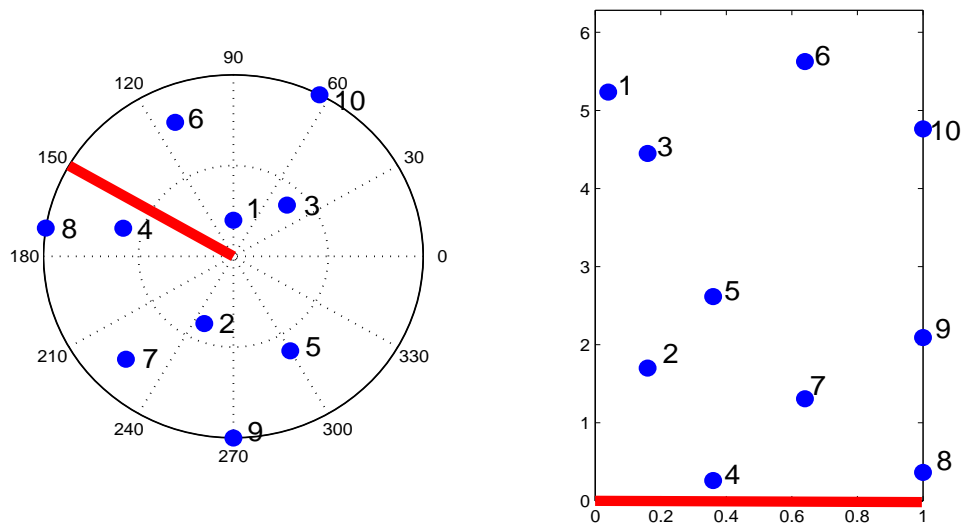


圖 2.2: 改變極軸實驗區間與建模區間

## 2.3 群集分析

群集分析是量測相似性的一種方法，它把所有的物件區分成幾個群體使得：

1：每一群內的物件彼此相似程度很高。

2：不同群間的物件相似程度低。

所謂相似性的高低依物件或研究目的的不同而不同。常見的分群的方法有：層次群集法 (hierarchical clustering), 非層次群集法 (Nonhierarchical Clustering Methods), 利用統計模型分群 (Clustering Based on Statistical Models)。因為本論文的目的要把圓上的曲面分群，並非討論那一種分群方式較好，因此本論文採用的分群方式為層次群集法。

所謂層次群集法就是, 距離最近的物件先為一群, 再將距離某群最近的物件或群, 歸為該群, 直到每一物件都歸至某群為止。常見的群與群之間的距離定義有下列三種: 單一聯結法, 完全聯結法, 平均聯結法。所謂單一聯結法為某群內所有物件至另一群內所有物件的最短距離; 完全聯結法為某群內所有物件至另一群內所有物件的最遠距離; 平均聯結法為某群內所有物件至另一群內所有物件的平均距離。

例子: 現在物件 1,2,3 為一群, 物件 4,5 為一群, 物件分佈的狀況為下圖, 單一聯結法就是物件 3 與 4 的距離為這兩群的距離。

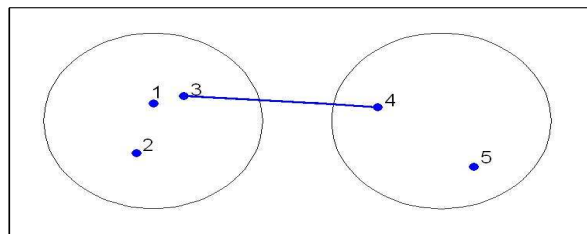


圖 2.3: 單一聯結法

例子: 物件 1,2,3 為一群, 物件 4,5 為一群, 物件分佈的狀況為下圖, 完全聯結法就是物件 2 與 5 的距離為這兩群的距離。

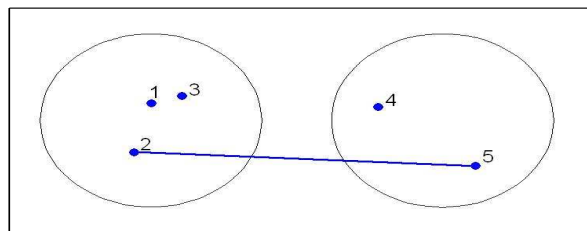


圖 2.4: 完全聯結法

例子：物件1,2,3爲一群，物件4,5爲一群，物件分佈的狀況爲下圖，平均聯結法就是物件1,2,3與4,5的距離平均爲這兩群的距離。

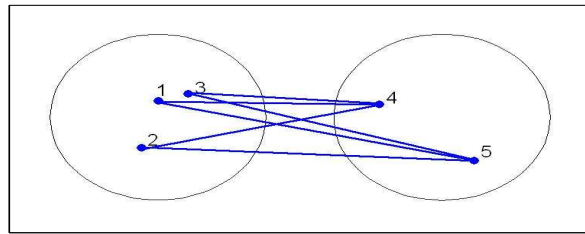


圖 2.5：平均聯結法

現在用一個Johnson and Wichern (2007)的例子來說明層次群集法，本例子有5個物件要分群。首先會先有一個量測每個物件間距離的距離定義，兩兩間物件的差異由距離定義得到後，就可以寫出距離矩陣。以下爲某個距離定義得到的距離矩陣。

	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	8
5	11	10	2	8	0

因爲物件3,5的距離最小，所以它們最先被歸爲一群。現在 (35) 已經是一群了，要衡量物件1,2,4對 (35) 的距離方式有三個：

1：單一聯結法 (single-linkage method)，此法就是取物件 1,2,4對 (35) 的

距離最小值當作距離，以物件 1 對 (35) 為例， $d_{(35)1} = \min\{d_{31}, d_{51}\} = \min\{3, 11\} = 3$

2：完全聯結法 (complete-linkage method)，此法就是取物件 1,2,4 對 (35) 的距離最遠值當作距離，以物件 1 對 (35) 為例， $d_{(35)1} = \max\{d_{31}, d_{51}\} = \max\{3, 11\} = 11$

3：平均聯結法 (average-linkage method)，此法就是取物件 1,2,4 對 (35) 的距離平均值當作距離，以物件 1 對 (35) 為例， $d_{(35)1} = \text{avg}\{d_{31}, d_{51}\} = \text{avg}\{3, 11\} = 7$

重覆上述步驟直到五個物件皆連結起來才停止，這就是層次群集法。由層次群集法我們可以得到，單一，完全，平均聯結法的樹狀圖，其分別列於圖 2.3, 2.4, 及 2.5。

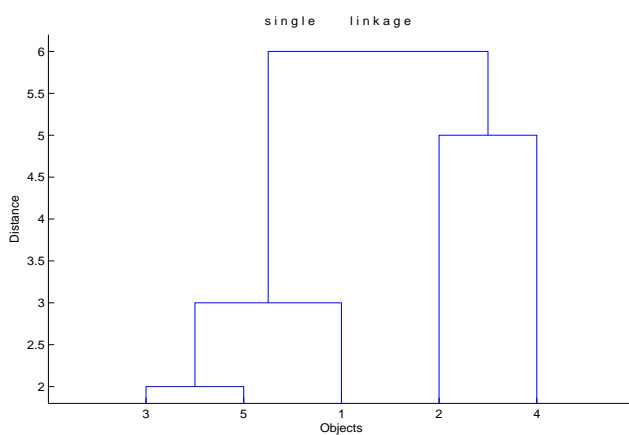


圖 2.6: 單一聯結法的樹狀圖

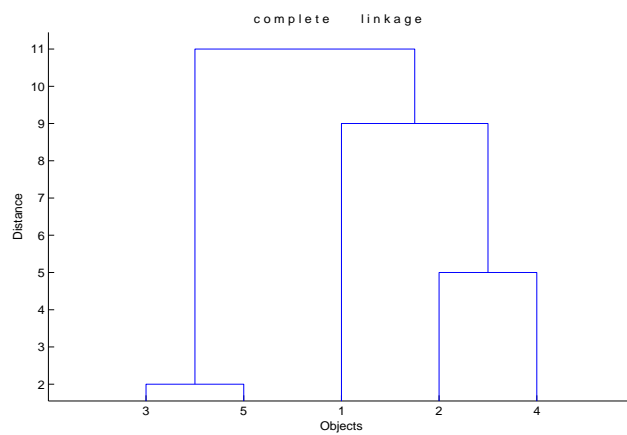


圖 2.7: 完全聯結法的樹狀圖

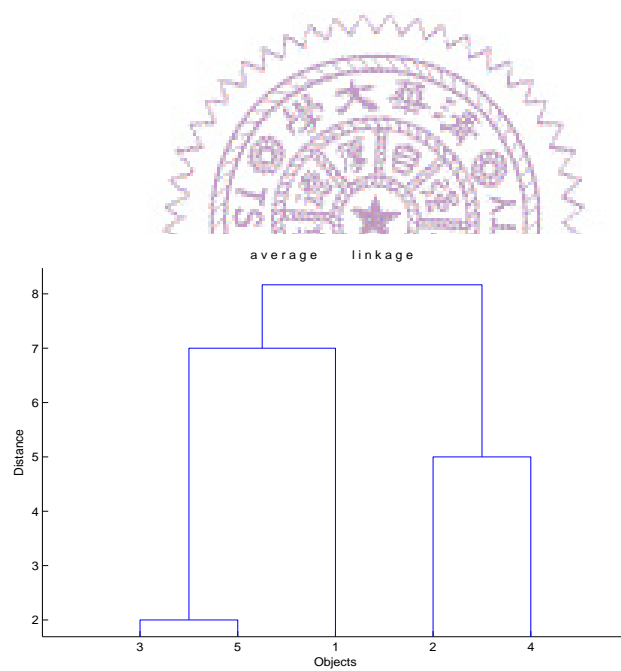


圖 2.8: 平均聯結法的樹狀圖

## 第 3 章

# 如何把圓上的曲面做群集分析

### 3.1 導論

一般要對幾個物件做群集分析，通常會量測物件的幾個重要特徵，再利用那些特徵的差異來定義物件間的距離，假設現在有  $n$  個物件要分群，每個物件觀測了  $m$  種不同的特徵。令  $w_{i,j}$  代表物件  $i$  第  $j$  個特徵的觀測值。則資料常會以下列形式呈現：

表 3.1: 群集分析資料形式

資料名稱	特徵 1	特徵 2	特徵 3	特徵 4	特徵 5	特徵 6	...	特徵 $m$
物件 1	$w_{1,1}$	$w_{1,2}$	$w_{1,3}$	$w_{1,4}$	$w_{1,5}$	$w_{1,6}$	...	$w_{1,m}$
物件 2	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	$w_{2,4}$	$w_{2,5}$	$w_{2,6}$	...	$w_{2,m}$
物件 3	$w_{3,1}$	$w_{3,2}$	$w_{3,3}$	$w_{3,4}$	$w_{3,5}$	$w_{3,6}$	...	$w_{3,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
物件 $n$	$w_{n,1}$	$w_{n,2}$	$w_{n,3}$	$w_{n,4}$	$w_{n,5}$	$w_{n,6}$	...	$w_{n,m}$

接著我們可以對  $w_{i,j}$  引入適當的距離定義來衡量物件之間的差異程度。比如說通

常使用的距離定義為：

$$D_{i_1, i_2}^n = \left( \sum_{j=1}^m (|w_{i_1, j} - w_{i_2, j}|)^n \right)^{1/n}, \quad (3.1)$$

當  $n=2$  時，此距離叫歐式距離 (euclidean distance)，這是最常被大家拿來當作距離的定義。另外還有一個常用的距離叫統計距離 (statistical distance)，定義為：

$$D_{i_1, i_2} = \sqrt{(\mathbf{w}_{i_1} - \mathbf{w}_{i_2})' A (\mathbf{w}_{i_1} - \mathbf{w}_{i_2})},$$

其中  $\mathbf{w}_{i_1} = [w_{i_1, 1}, w_{i_1, 2}, \dots, w_{i_1, m}]'$ ， $\mathbf{w}_{i_2} = [w_{i_2, 1}, w_{i_2, 2}, \dots, w_{i_2, m}]'$ ， $A = S^{-1}$ ， $S$  為樣本共變異數矩陣。這些定義距離的方式，皆有一個特色，就是同一個特徵，每個物件皆有觀測值。以一個 Sharma (1996) 中提及的食物的營養資料為例，在這組數據中，共有 6 個食品項目要分群，分別為：燉牛肉，漢堡，烤牛肉，牛排，罐裝牛肉，烤雞。而每個物件共觀察了 5 個不同的特徵，分別為：卡路里，蛋白質，脂肪，鈣，鐵，以下為它的資料形式：

表 3.2: 食品的營養資料

食品項目	卡路里	蛋白質	脂肪	鈣	鐵
燉牛肉	340	20	28	9	2.6
漢堡	245	21	17	9	2.7
烤牛肉	420	15	39	7	2.0
牛排	375	19	32	9	2.6
罐裝牛肉	180	22	10	17	3.7
烤雞	115	20	3	8	1.4



如果要比較燉牛肉與漢堡之間的差異，且使用的距離定義為歐式距離，同一個特徵每個物件皆有觀測值的意思是：燉牛肉與漢堡在卡路里，蛋白質，脂肪，鈣，鐵，都必須有觀測值。最後會算出9147為它們之間的距離。算法為：

$$\sqrt{(340 - 245)^2 + (20 - 21)^2 + (28 - 17)^2 + (9 - 9)^2 + (2.6 - 2.7)^2},$$

現在某電子公司提供了晶圓的化學氣相沈積 (Chemical Vapor Deposition, 簡稱 CVD) 的真實數據，其收集數據的方法如下。因為一個晶圓內包含著許多小晶粒 (chip)，每個晶粒都可以做實驗得到化學氣相沈積的數據，故我們可將整個晶圓想像成一個實驗區間，而將晶粒的化學氣相沈積的數據想像為該實驗區間上的反應值，這時化學氣相沈積的數據與晶粒位置的關係便可用一個反應曲面來代表。當因為成本的考量而無法在所有晶粒上做實驗時，工程師就會在機台的晶圓擺放區上挑選幾個具有代表性的位置，這些位置我們稱為量測座標。當每片晶圓擺放在機台上，就會有某些晶粒對應到這些量測座標，工程師便會在這些晶粒上做實驗得到反應值。在這個實驗裡，機台上的晶圓擺放區的X,Y座標可視為兩個不同的因子，而每一個量測座標可視為一組設定值，所以此實驗可視為二因子的實驗設計。

透過這些量測座標上晶粒實驗後的數據，就可以研究晶圓上的反應曲面如何隨著晶粒的幾何位置改變而變化。現在工程師有興趣的問題是，任兩片晶圓上的反應曲面是否有差異，有那些晶圓的反應曲面較類似，以及可分類出多少種不同型態的反應曲面。雖然每片晶圓放置在機台上後，工程師便會測量某些固定的座標上的晶粒的反應值。不過晶圓沒有上下左右的區別，因此工程師把晶圓擺放在機台上時，其擺放位置並不是固定的。這正如我們把光碟片放入光碟機的情況是相同的，每一次我們將同一片光碟放入光碟機時，其擺入位置與前一次擺入位置未必會相同，但這兩個擺入位置

經過某個角度的旋轉後，便會一致。故即便是同一片晶圓被重覆擺放在機台上，且使用相同的量測座標來做實驗，但是因為晶圓每次擺放在機台上的擺入位置未必會相同，所以即使該晶圓被重覆量測時的量測座標是完全相同的，也不會每次皆在相同的晶粒上做實驗。表 3.3 代表 8 片晶圓在  $n$  個量測座標下做實驗得到的數據，其中每片晶圓皆被任意擺放在機台上做一次實驗。 $w_{i,j}$ 代表第 $i$ 片晶圓的第 $j$ 個量測座標上所得到的反應值。

表 3.3: 晶圓的數據

晶圓	座標 1	座標 2	座標 3	座標 4	座標 5	座標 6	...	座標 $n$
晶圓 1	$w_{1,1}$	$w_{1,2}$	$w_{1,3}$	$w_{1,4}$	$w_{1,5}$	$w_{1,6}$	...	$w_{1,n}$
晶圓 2	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	$w_{2,4}$	$w_{2,5}$	$w_{2,6}$	...	$w_{2,n}$
晶圓 3	$w_{3,1}$	$w_{3,2}$	$w_{3,3}$	$w_{3,4}$	$w_{3,5}$	$w_{3,6}$	...	$w_{3,n}$
晶圓 4	$w_{4,1}$	$w_{4,2}$	$w_{4,3}$	$w_{4,4}$	$w_{4,5}$	$w_{4,6}$	...	$w_{4,n}$
晶圓 5	$w_{5,1}$	$w_{5,2}$	$w_{5,3}$	$w_{5,4}$	$w_{5,5}$	$w_{5,6}$	...	$w_{5,n}$
晶圓 6	$w_{6,1}$	$w_{6,2}$	$w_{6,3}$	$w_{6,4}$	$w_{6,5}$	$w_{6,6}$	...	$w_{6,n}$
晶圓 7	$w_{7,1}$	$w_{7,2}$	$w_{7,3}$	$w_{7,4}$	$w_{7,5}$	$w_{7,6}$	...	$w_{7,n}$
晶圓 8	$w_{8,1}$	$w_{8,2}$	$w_{8,3}$	$w_{8,4}$	$w_{8,5}$	$w_{8,6}$	...	$w_{8,n}$

針對表 3.3 晶圓的數據，我們有興趣做群集分析來對其分類，若對照之前關於群集分析數據的描述，一個很直覺的想法便是把不同的量測座標當成不同的特徵，比如表 3.3 的座標  $1 \sim n$ 可視為 $n$ 個不同的特徵變數，在這樣的處理下，兩片晶圓上反應曲面的差異便可套用等式 (3.1) (或其它合理的距離定義) 而以下列等式得到的值來

衡量。

$$D_{i_1, i_2}^2 = \left( \sum_{j=1}^n (w_{i_1, j} - w_{i_2, j})^2 \right)^{1/2}, \quad (3.2)$$

用此距離定義來衡量兩片晶圓上反應曲面的差異，這做法看似合理，不過之前有介紹過，雖然量測座標是完全相同的，但即使同一片晶圓重覆做實驗，也不會每次皆在晶圓上相同的晶粒上做實驗。不過比較兩個曲面像不像，應該要比較相同位置的晶粒。因為不知道兩片晶圓相同位置的晶粒在那裡，所以比較時，必須一片晶圓位置固定，另一片晶圓不停的旋轉，這樣才合理。上述距離的定義已經不適用了，下一節我們就介紹如何定義此群集分析的距離。

## 3.2 晶圓旋轉

針對上一節晶圓的資料，我們可以用下面的模型來描述它：

$$w_k(x, y) = f_k(x, y) + \varepsilon,$$

其中  $f_k(x, y)$  為編號  $k$  的晶圓上一個未知的真實反應曲面。若我們在某個量測座標  $(x^*, y^*)$  上觀測到數據  $w_k(x^*, y^*)$ ，則  $w_k(x^*, y^*)$  為反應曲面的值  $f_k(x^*, y^*)$  再加上一個量測誤差  $\varepsilon$ 。

以下我們先舉一個例子來說明我們的想法，若圖 3.1 為某兩片晶圓上的真實反應曲面，左邊的曲面為晶圓 A 擺放於機台上量測時的反應曲面  $f_A(x, y)$ ，則在某些量測座標下獲得的量測數據  $w_A(x, y)$  為反應曲面的值  $f_A(x, y)$  再加上一個量測誤差  $\varepsilon$ 。故此數據之模型為：

$$w_A(x, y) = f_A(x, y) + \varepsilon,$$

同理右邊的曲面為晶圓 B 擺放於機台上量測時的反應曲面 $f_B(x, y)$ 。其在相同的量測座標下獲得的數據之模型為：

$$w_B(x, y) = f_B(x, y) + \varepsilon,$$

假設工程師挑選的實驗點之量測座標為圖 3.2 中編號1 ~ 10的點, 我們並將 A 曲面的等高線圖畫於圖 3.2 左邊, 由此等高圖我們可以發現,A 曲面是由量測座標 5 及 9 的位置一路向量測座標 6,8 的位置下降; 同理我們將 B 曲面的等高線圖畫於圖 3.2 右邊, 由此等高線圖我們可以發現,B 曲面是由量測座標 7 的位置一路向量測座標 10 的位置下降。假設量測誤差 $\varepsilon$ 並不大, 則此時若利用等式 (3.2) 來計算 A,B 曲面的差異, 最後算出來的距離一定不小, 因為圖 3.2 的等高線告訴我們 A,B 兩曲面是往不同的兩個方向遞減的。

但若我們將其中一片晶圓位置固定, 而讓另一片晶圓不停的旋轉, 再不斷比較這兩個曲面是否相似。比如固定晶圓 A, 讓晶圓 B 不停的旋轉, 圖 3.3 右邊的曲面是晶圓 B 逆時鐘方向旋轉 $\frac{25}{36}\pi$ 後得到的, 這時若我們再比較曲面 A 與旋轉後的曲面 B 之等高線 (圖 3.4), 則我們可發現這兩曲面皆是從同一個方向向另一個方向遞減, 亦即 A,B 兩曲面比之前像多了。故我們可得知對此晶圓實驗, 若要比較其曲面是否有差異, 應將晶圓旋轉納入考量。

不過 B 曲面旋轉後就會發生一個問題, 就是工程師挑選的實驗點之量測座標的位置也跟著改變。我們可由圖 3.4 中左邊晶圓上的實驗點 1 不再與右邊晶圓實驗點 1 位在相同的位置便可發現, 相同的問題也發生在其他實驗點身上, 這使得之前利用等式 (3.1) 來算 A,B 兩曲面的差異, 產生了問題。由圖 3.4 我們可以很清楚地發現, 晶圓 A 上的第  $j$  個量測座標上的反應值  $w_{A,j}$ 不應該與晶圓 B 上的第  $j$  個量測座

標上的反應值  $w_{B,j}$  相比較。因此在比較兩個反應曲面的差異時, 我們必須找個方法來解決旋轉後, 兩個晶圓的量測座標不再重疊在一起的問題。

在此我們提出了使用配適值取代觀測數據的概念。因為有了量測座標上晶粒的反應值, 我們便可以建立配適模型來得到配適的反應曲面。因為配適的反應曲面在晶圓上任何地方都有配適值, 因此即使晶圓旋轉了, 晶圓 A 與旋轉後的晶圓 B, 在任何一个位置上, 仍各有配適值可用來比較曲面是否相似。故使用配適值便可避免量測座標因旋轉而無法重疊, 使得晶圓 A 量測座標上的觀測值在晶圓 B 上找不到相對應的觀測值與之比較的問題。

以下令  $\hat{f}_k(x, y)$  為利用量測座標上的數據  $w_k(x, y)$  所建立的一個配適反應曲面, 有關如何建立  $\hat{f}_k(x, y)$  的討論, 我們將在後面的章節再討論。令  $\hat{f}_A$  與  $\hat{f}_B$  分別為晶圓 A, B 上的配適反應曲面。若要在考慮晶圓旋轉的情況下比較  $\hat{f}_A$  與  $\hat{f}_B$  的差異, 則我們可使用下面的距離定義:

$$D(\hat{f}_A, \hat{f}_B) = \min_{0 < \Delta\theta < 2\pi} \left\{ \int \int_{x^2+y^2 < 1} [\hat{f}_A(x, y) - \hat{f}_B(x', y')]^2 dx dy \right\}, \quad (3.3)$$

其中  $\Delta\theta$  代表晶圓 B 旋轉的角度,  $(x', y')$  為  $(x, y)$  旋轉  $\Delta\theta$  後的座標, 故  $x', y'$  皆為  $x, y$  與  $\Delta\theta$  的函數, 即:

$$\begin{aligned} x' &= \sqrt{x^2 + y^2} \cos(\arctan(\frac{x}{y}) + \Delta\theta), \\ y' &= \sqrt{x^2 + y^2} \sin(\arctan(\frac{x}{y}) + \Delta\theta), \end{aligned}$$

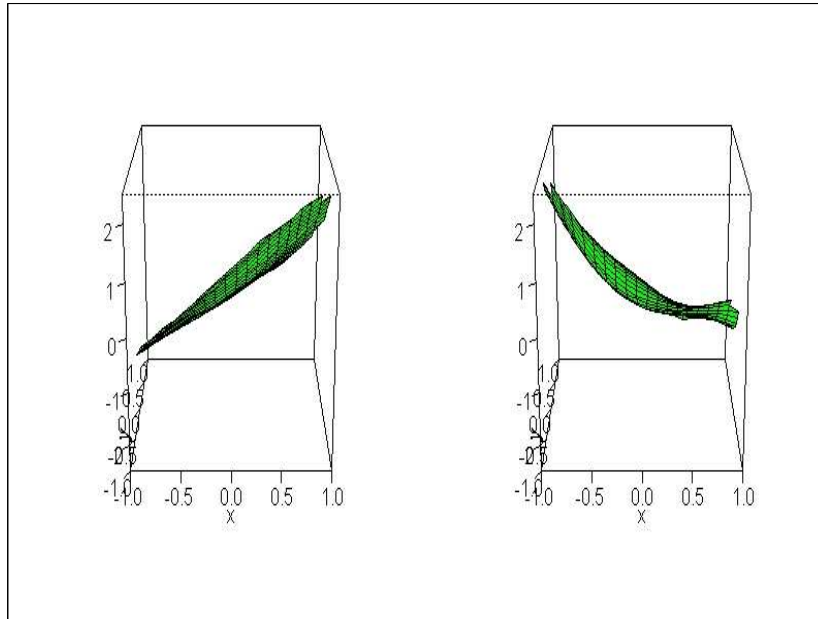


圖 3.1: 兩個反應曲面旋轉前的圖型

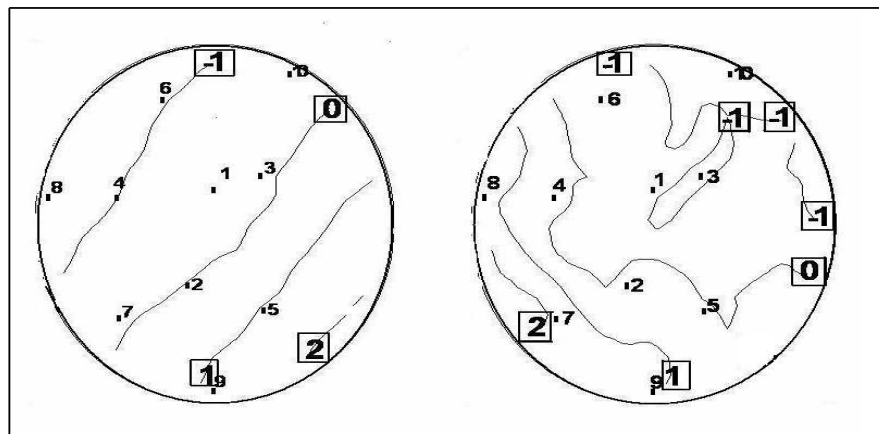


圖 3.2: 量測座標的位置與等高線圖

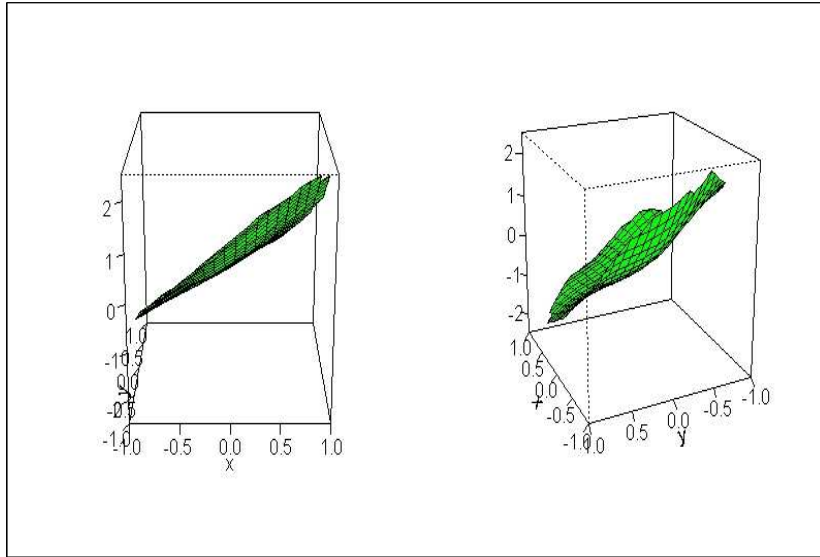


圖 3.3: 兩個反應曲面旋轉後的圖型

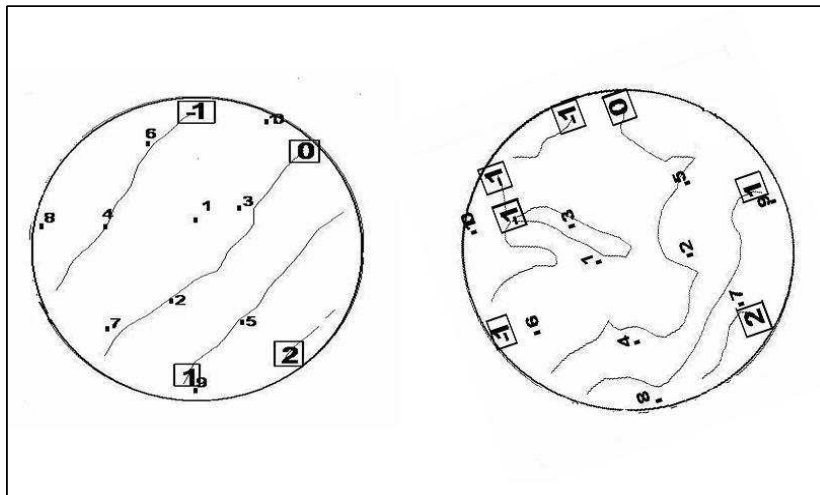


圖 3.4: 旋轉後量測座標的位置與等高線圖



使 (3.3) 最小的  $\Delta\theta$  值代表曲面  $\hat{f}_B$  在旋轉  $\Delta\theta$  的角度後和曲面  $\hat{f}_A$  最像。在極座

標的表示法下, (3.3) 式亦可表示為:

$$D(\hat{f}_A, \hat{f}_B) = \min_{0 < \Delta\theta < 2\pi} \left\{ \int_0^{2\pi} \int_0^1 [\hat{f}_A(r\cos(\theta), r\sin(\theta)) - \hat{f}_B(r\cos(\theta + \Delta\theta), r\sin(\theta + \Delta\theta))]^2 |J| dr d\theta \right\} \quad (3.4)$$

其中  $|J| = r$ 。

以下我們令

$$\hat{g}_A(r, \theta) = \hat{f}_A(r\cos(\theta), r\sin(\theta)),$$

$$\hat{g}_B(r, \theta) = \hat{f}_B(r\cos(\theta), r\sin(\theta)),$$

若我們使用極座標所形成的一個一階模型 (1st-order model) 來當配適模型, 亦即:

$$\hat{g}_A(r, \theta) = \hat{\beta}_{A0} + \hat{\beta}_{A1}\gamma + \hat{\beta}_{A2}\theta,$$

$$\hat{g}_B(r, \theta) = \hat{\beta}_{B0} + \hat{\beta}_{B1}\gamma + \hat{\beta}_{B2}\theta,$$

則 (3.4) 式變為:

$$\begin{aligned} D(\hat{f}_A, \hat{f}_B) &= \min_{0 < \Delta\theta < 2\pi} \left\{ \int_0^{2\pi} \int_0^1 [\hat{g}_A(r, \theta) - \hat{g}_B(r, \theta + \Delta\theta)]^2 |J| dr d\theta \right\}, \\ &= \min_{0 < \Delta\theta < 2\pi} \left\{ \int_0^{2\pi} \int_0^1 [(\hat{\beta}_{A0} - \hat{\beta}_{B0}) + (\hat{\beta}_{A1} - \hat{\beta}_{B1})r + (\hat{\beta}_{A2} - \hat{\beta}_{B2})\theta - \hat{\beta}_{B2}\Delta\theta]^2 |J| dr d\theta \right\}, \end{aligned} \quad (3.5)$$



我們可發現 (3.5) 大括號內的部份為 $\Delta\theta$ 的二次多項式, 若令其為  $h(\Delta\theta)$ , 則將

$h(\Delta\theta)$  對  $\Delta\theta$  微分可得:

$$\begin{aligned} h'(\Delta\theta) &= \int_0^{2\pi} \int_0^1 2[(\hat{\beta}_{A0} - \hat{\beta}_{B0}) + (\hat{\beta}_{A1} - \hat{\beta}_{B1})r + (\hat{\beta}_{A2} - \hat{\beta}_{B2})\theta - \hat{\beta}_{B2}\Delta\theta](-\hat{\beta}_{B2}) \times r dr d\theta, \end{aligned}$$

令  $h'(\Delta\theta) = 0$ , 得到的解為:

$$\begin{aligned} \Delta\theta^* &= \frac{2\hat{\beta}_{B2}(\hat{\beta}_{A0} - \hat{\beta}_{B0}) \int_0^{2\pi} \int_0^1 r dr d\theta}{2\hat{\beta}_{B2}^2 \int_0^{2\pi} \int_0^1 r dr d\theta} + \frac{2\hat{\beta}_{B2}(\hat{\beta}_{A1} - \hat{\beta}_{B1}) \int_0^{2\pi} \int_0^1 r^2 dr d\theta}{2\hat{\beta}_{B2}^2 \int_0^{2\pi} \int_0^1 r dr d\theta} \\ &\quad + \frac{2\hat{\beta}_{B2}(\hat{\beta}_{A2} - \hat{\beta}_{B2}) \int_0^{2\pi} \int_0^1 r\theta dr d\theta}{2\hat{\beta}_{B2}^2 \int_0^{2\pi} \int_0^1 r dr d\theta} \\ &= \frac{2\pi\hat{\beta}_{B2}(\hat{\beta}_{A0} - \hat{\beta}_{B0}) + \frac{2}{3}\pi\hat{\beta}_{B2}(\hat{\beta}_{A1} - \hat{\beta}_{B1}) + 2\pi^2\hat{\beta}_{B2}(\hat{\beta}_{A2} - \hat{\beta}_{B2})}{2\pi\hat{\beta}_{B2}^2}, \\ &= \frac{(\hat{\beta}_{A0} - \hat{\beta}_{B0}) + \frac{1}{3}(\hat{\beta}_{A1} - \hat{\beta}_{B1}) + \pi(\hat{\beta}_{A2} - \hat{\beta}_{B2})}{\hat{\beta}_{B2}}, \end{aligned}$$

若對 $h(\Delta\theta)$ 二次微分再把上式 $\Delta\theta^*$ 代入, 得到:  $h''(\Delta\theta^*) = 2\pi\hat{\beta}_B^2 > 0$ , 所以使 (3.5)

最小的  $\Delta\theta$  為  $\Delta\theta^*$ , 這代表 A,B 兩個配適曲面最像的時候是晶圓 B 旋轉  $\Delta\theta^*$  度。

同理若  $\hat{g}_A, \hat{g}_B$  使用二階, 三階, 或更高階的配適模型, 其仍可使用上述的步驟來求得

$\Delta\theta^*$ , 唯一不同之處為當  $\hat{g}_A, \hat{g}_B$  為二階模型時,  $h(\Delta\theta)$  為  $\Delta\theta$  的四次多項式;  $\hat{g}_A, \hat{g}_B$  為

三階模型時,  $h(\Delta\theta)$  為  $\Delta\theta$  的六次多項式;  $\hat{g}_A, \hat{g}_B$  為  $n$  階模型時,  $h(\Delta\theta)$  為  $\Delta\theta$  的  $2n$

次多項式。

### 3.3 適合度的問題

在3.2節, 我們提出了用配適值來取代觀測數據的方法, 用以解決晶圓旋轉後, 量測座標無法再重疊在一起的問題。但此時會產生一個新問題, 就是配適曲面是否可以完全代表觀測數據背後的真實曲面。要衡量配適曲面與真實曲面的相似程度, 一個可用的指標便是配適模型的判定係數。若判定係數很小, 那代表配適曲面跟真實曲面的差異很大, 故可合理懷疑配適曲面會與真實曲面差異頗大, 在這種情形下, 即使使用等式 (3.3) 計算出來的距離很小, 我們也只能夠說, 它們配適曲面很像, 而不應推論它們的真實曲面很像。故若僅用 (3.3) 式的距離大小來對晶圓分類是不夠的, 我們尚需適當的將配適的好壞納入距離定義內考量。當然我們可嘗試各種不同的配適模型來提高判定係數, 但即使如此, 任何合理的迴歸模型它的判定係數皆不應為1。以下我們就幾種情況討論該如何讓判定係數在分類的距離中發揮其影響。

- 1 : 若要比較的兩個配適曲面, 其判定係數皆很高, 則我們可以相信配適曲面與真實曲面頗相似, 此時使用 (3.3) 式的距離來做群集分析便會比較適合。
- 2 : 若要比較的兩個配適曲面, 判定係數皆很小, 則配適曲面皆很像水平面, 此時若兩水平面很接近則 (3.3) 式的距離便容易小, 但此時真實曲面是否相似是值得懷疑的。
- 3 : 若要比較的兩個配適曲面, 判定係數一高一低, 則一個配適曲面與真實曲面頗相似, 另一個配適曲面則否, 則我們認定兩個真實曲面應該是不像的。因為若兩個真實曲面頗相似, 則在相同的量測誤差下, 使用相同的配適模型所得到的判定

係數大小應相近。

上述的三種狀況可用圖 3.5 表示，圖 3.5 的橫軸為第一個配適模型的判定係數，縱軸為第二個配適模型的判定係數，圖 3.5 分了三個區塊，在區域I 裡，就是上述 1 的情況，此時我們希望 (3.3) 式維持不變；區域II，就是上述 3 的情況，在區域II裡，無足夠的資訊足以判定兩個真實曲面是否相似，此時一個可行的做法便是，增加配適模型的複雜度以讓判定係數變大，否則我們寧願採取較保守的做法，認定此兩真實曲面不像；區域III，就是上述 2 的情況，此時我們希望認定此兩真實曲面它們不像。以下我們提出兩個做法，其利用將 (3.3) 式乘上一個懲罰項來將判定係數的影響帶入分類的距離定義中。

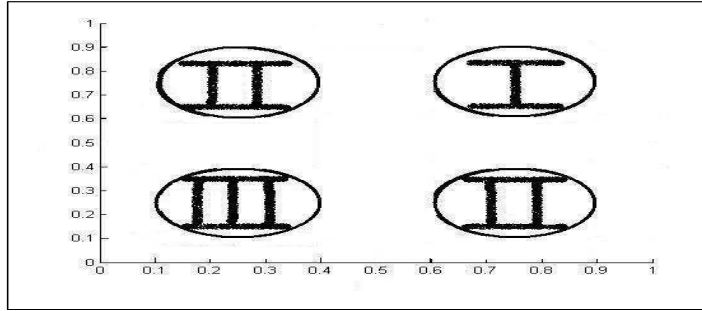


圖 3.5: 把判定係數分三大塊

第一個做法採用相對風險 (relative risk) 的觀念，若令  $R_A^2, R_B^2$  分別為配適曲面  $\hat{f}_A$  與  $\hat{f}_B$  的判定係數，則懲罰項定義為：

$$\frac{\max(R_A^2, R_B^2)}{\min(R_A^2, R_B^2)},$$

此公式的等高圖為圖 3.6, 由等高圖可看出，如果其中有一個判定係數小於 0.1，則值

會非常的大；兩個判定係數很接近，則值會接近 1；一高一低，則值會很大；皆低，則值介於判定係數很接近與一高一低之間。換成群集分析的用語：判定係數其中有一個小於 0.1，懲罰最大；兩個判定係數很接近懲罰小；一高一低則懲罰很大；皆低則懲罰介於兩者之間，可以用表 3.4 來表示：

表 3.4: 解決判定係數不高的第一個方法

值	$R_A^2 < 0.1$ 或 $R_B^2 < 0.1$	$R_A^2, R_B^2$ 一高一低	$R_A^2, R_B^2$ 皆低	$R_A^2 \approx R_B^2$
懲罰	最大	第二大	第三大	最小

所以距離的定義應該是 (3.3) 式再乘上此懲罰項。

$$D(\hat{f}_A, \hat{f}_B) \times \frac{\max(R_A^2, R_B^2)}{\min(R_A^2, R_B^2)}, \quad (3.6)$$

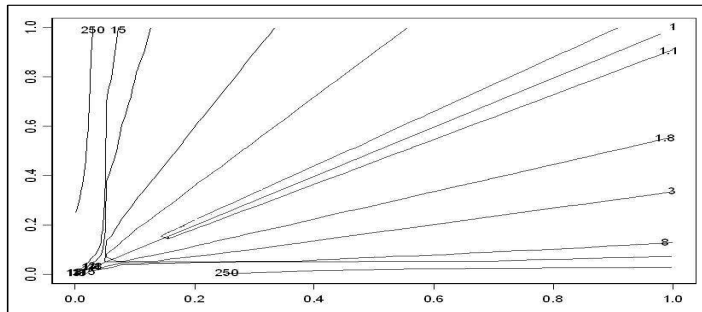


圖 3.6: 第一個方法的等高圖

第二個方法之懲罰項為：

$$1 + (1 - R_A^2)^k \times (1 - R_B^2)^k,$$

其中 $k$ 爲一正整數, 則 $k$ 變大, 則判定係數對該懲罰項的影響會變小。此公式的等高圖爲圖 3.7, 由等高圖可看出, 如果判定係數皆很高, 值會接近 1 ; 皆低則值偏大; 一高一低則值介於兩者之間。換成群集分析的用語: 判定係數皆很高懲罰小; 皆低則懲罰很大; 一高一低則懲罰介於兩者之間, 可以用表 3.5 來表示:

表 3.5: 解決判定係數不高的第二個方法

值	$R_A^2, R_B^2$ 皆低	$R_A^2, R_B^2$ 一高一低	$R_A^2 \approx R_B^2 \approx 1$
懲罰	最大	第二大	最小

所以距離的定義應該是 (3.3) 式再乘上此懲罰項。

$$D(\hat{f}_A, \hat{f}_B) \times [1 + (1 - R_A^2)^k \times (1 - R_B^2)^k], \quad (3.7)$$

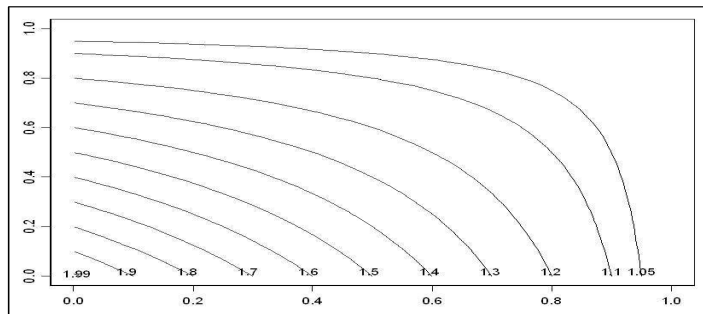


圖 3.7: 當  $k = 1$  第二個方法的等高圖

上述解決配適模型所得到的判定係數不高的兩個方法, 基本上是合理的。不過因爲真實曲面的樣貌是我們有興趣的, 所以我們還是會希望配適的模型判定係數能高。以下爲提高判定係數的方法:

- 1：在Lin (2007)中曾提出, 若真實曲面有所謂的圓效應(circular effect) 時, 在極座標上配適多項式模型會比在直角座標上配適多項式模型, 常能得到更高的判定係數。並且在極座標上建模時, 我們可考慮使用不同的極軸當參數來建模, 這會使得配適模型的 判定係數更容易提高。見本論文第五章。
- 2：迴歸分析有一個挑選模型的準則叫 PRESS (predicted residual sum of square)。
- 這個準則的主要目的是用來做預測, 它希望觀測值與配適值的差異越小越好, 因此它的配適值最具代表性, 而它的缺點就是配適的模型常會太大。而現在作群集分析希望的是, 找一個配適曲面可以取代真實曲面, 這和 PRESS 的理念一致, 所以可以考慮 PRESS 所選出的大模型。



## 第 4 章

# 真實數據的分析

### 4.1 導論

上一章已經提出了一套方法，定義了兩片晶圓的反應曲面是否相似的距離測度，並解決了比較兩片晶圓上的曲面像不像時，要考慮晶圓旋轉的問題。現在就拿某電子公司提供的晶圓觀測數據，來依晶圓上的曲面相似程度做群集分析，圖 4.1, 圖 4.2 為我們把觀測數據利用無母數配適得到的曲面，我們要把這 8 個曲面做群集分析。令第  $i$  片晶圓上的觀測數據為  $w_{i,1}, w_{i,2}, \dots, w_{i,n}$ ，以下我們考慮將晶圓上的觀測數據  $w_{i,j}$  做幾種轉換，之後再對轉換後的數據  $z_{i,j}$  套用第三章的方法來做群集分析。

1：觀測值不轉換，即  $z_{i,j} = w_{i,j}$ 。

2：每片晶圓上的觀測值減掉該片晶圓所有觀測值的平均，即  $z_{i,j} = w_{i,j} - \overline{w_i}$ ，其

中  $\overline{w_i} = \sum_{j=1}^n w_{i,j}/n$ 。

3：每片晶圓上的觀測值最大最小值調整成一樣，即  $z_{i,j} = (w_{i,j} - \min_j(w_{i,j})) / (\max_j(w_{i,j}) - \min_j(w_{i,j}))$

4：觀測值做標準化，即  $z_{i,j} = (w_{i,j} - \overline{w_i})/s_i$ ，其中  $\overline{w_i} = \sum_{j=1}^n w_{i,j}/n$ ,  $s_i = \sqrt{\sum_{j=1}^n (w_{i,j} - \overline{w_i})^2 / (n - 1)}$ 。

其中第一種轉換代表原始觀測值的單位很重要，所以進行分析時，很容易受位置與尺度的影響，亦即，一個晶圓上的曲面像是另一個晶圓上的曲面放大或縮小，是晶圓本身的問題，不是工程師記錄觀測數據時，使用了不同的單位量測造成的，所以它們不應該分在同一群。第二種轉換，是對晶圓上的曲面做平移，此時每片晶圓的觀測值平均數皆為零，所以進行分析時，已經把位置的影響去除，只剩下尺度的影響。因為尺度的問題沒有解決，所以此種轉換沒有解決曲面放大或縮小的問題。第三與第四種的資料轉換，皆已經解決了因為位置與尺度的不同所造成的影響，亦即，一個晶圓上的曲面像是另一個晶圓上的曲面放大或縮小，我們把它們分為很像的一類，兩者轉換的不同處，由之後提供盒狀圖 (Boxplot) 可以很明顯的看出來。以上對觀測值的轉換，皆是作位置或尺度的轉換，所以配適  $t$ ,  $t$  為正整數，階多項式模型後的判定係數皆不會改變。下列為八片晶圓，考慮以極座標的  $r, \theta$  構成的二階，三階，四階多項式模型下的最佳判定係數 (見第五章)。

表 4.1: 極軸當參數各種多項式模型的最佳判定係數

R-square	晶圓 1	晶圓 2	晶圓 3	晶圓 4	晶圓 5	晶圓 6	晶圓 7	晶圓 8
二階模型	0.96	0.979	0.951	0.984	0.749	0.723	0.705	0.861
三階模型	0.977	0.985	0.98	0.988	0.877	0.887	0.891	0.962
四階模型	0.987	0.99	0.983	0.989	0.908	0.904	0.912	0.994



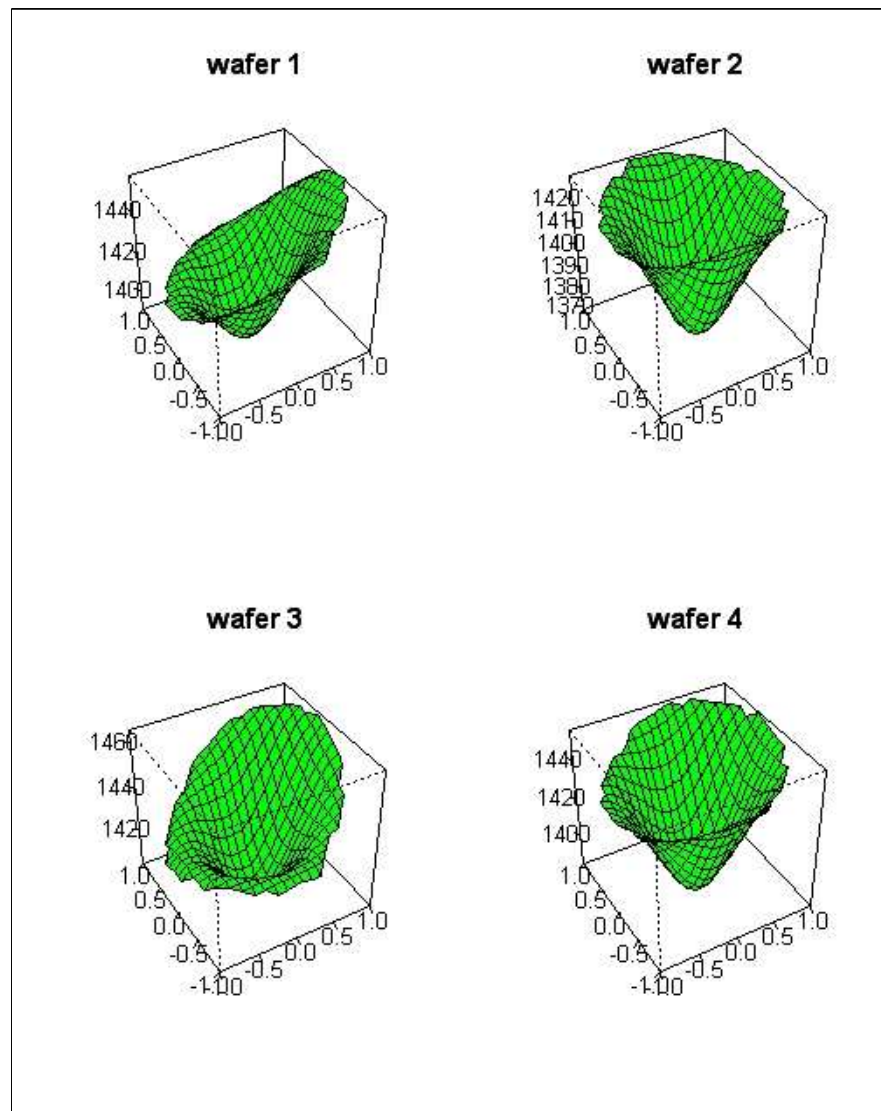


圖 4.1: 晶圓 1, 2, 3, 4 的曲面

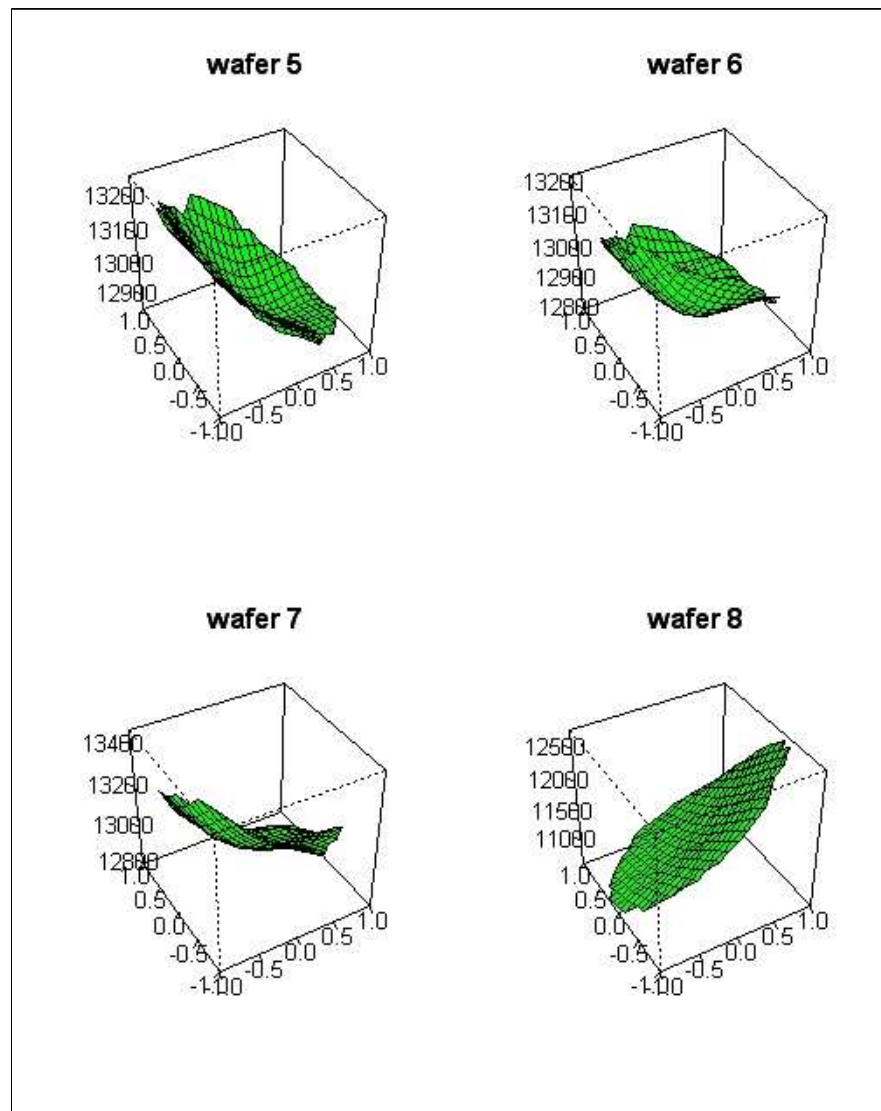


圖 4.2: 晶圓 5, 6, 7, 8的曲面

## 4.2 不轉換

若對觀測值不做任何轉換時, 則  $z_{i,j} = w_{i,j}$ , 圖 4.3 為  $z_{i,j}$  依不同晶圓畫出的盒狀圖。

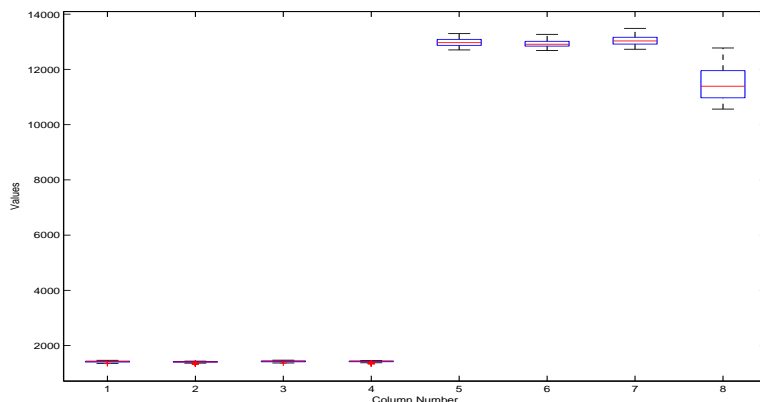


圖 4.3: 第一種轉換的盒狀圖

由圖 4.3, 可以很明顯的看出有三群, 晶圓 1,2,3,4 的觀測值分布範圍很接近; 晶圓 5,6,7 上數據的分布範圍較相近, 而晶圓 8 則與其他晶圓較不像, 而且晶圓 8 的四分位距 (interquartile range) 也比其它晶圓來的大。而且晶圓 5,6,7,8 的觀測值比晶圓 1,2,3,4 的觀測值大了約 5 倍以上。

圖 4.4 為每片晶圓的無母數配適曲面之等高圖, 由等高圖看出晶圓 1,2,3,4 中間位置的觀測值最小, 隨著半徑的增加, 觀測值也漸漸變大, 而觀測值分佈的範圍為:1360 ~ 1500 ; 晶圓 5,6,7 的觀測值有左下一路向右上遞減的現象, 而觀測值分佈的範圍為:12800 ~ 13400 ; 晶圓 8 觀測值有右下一路向左上遞減的現象, 而觀測值分佈

的範圍為:11000 ~ 12500。

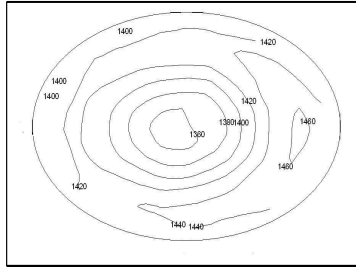
若在 $z_{i,j}$ 數據上使用 (3.6) 或 (3.7) 式, 求任兩片晶圓上曲面的距離, 就可以得到距離矩陣。即可利用此距離矩陣, 再套用第二章的單一聯結法, 完全聯結法, 平均聯結法, 則可畫出它們的樹狀圖。

距離的定義為 (3.6) 式, $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.5, 圖 4.6, 圖 4.7 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式, $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.8, 圖 4.9, 圖 4.10 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式, $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.11, 圖 4.12, 圖 4.13 為各種聯結法的樹狀圖。

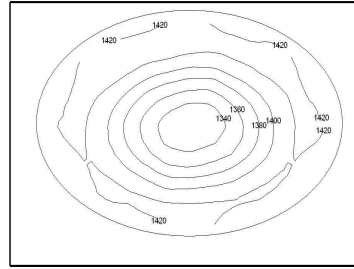
距離的定義為 (3.7) 式且 $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.14, 圖 4.15, 圖 4.16 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且 $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.17, 圖 4.18, 圖 4.19 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且 $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.20, 圖 4.21, 圖 4.22 為各種聯結法的樹狀圖。

結論: 無論是何種的距離定義, 幾階的多項式模型, 不同的分群方式, 最後的結果都是一樣的。如果分三群, 晶圓 1,2,3,4 一群; 晶圓 5,6,7 一群; 晶圓 8 一群。如果要再多分一群, 晶圓 7 會與晶圓 5,6 分開。晶圓 1,2,3,4 皆是以圓中心向四周先遞增再遞減。晶圓 5,6,7,8 比較像從某個方向向另一方向遞減。

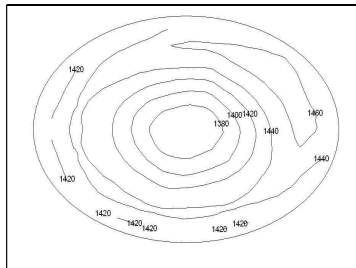
water 1



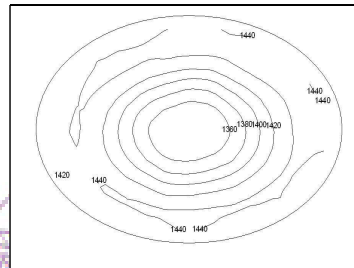
water 2



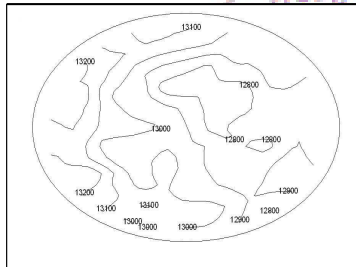
water 3



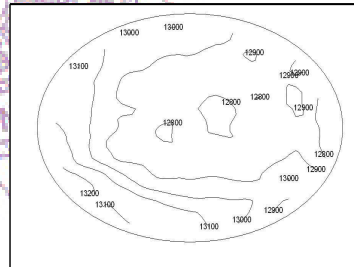
water 4



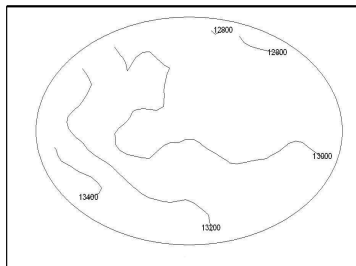
water 5



water 6



water 7



water 8

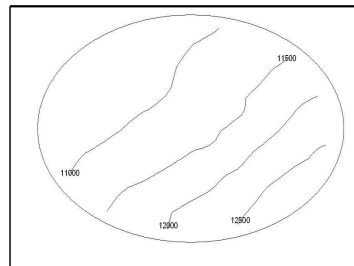


圖 4.4: 第一種轉換的等高圖

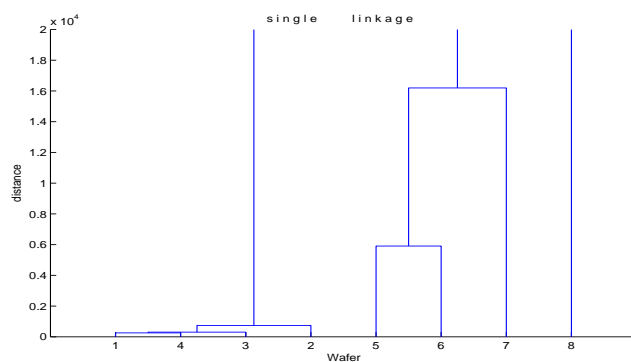


圖 4.5: 二階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖

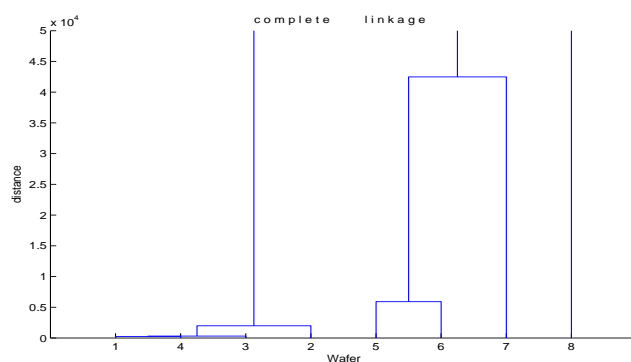


圖 4.6: 二階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖

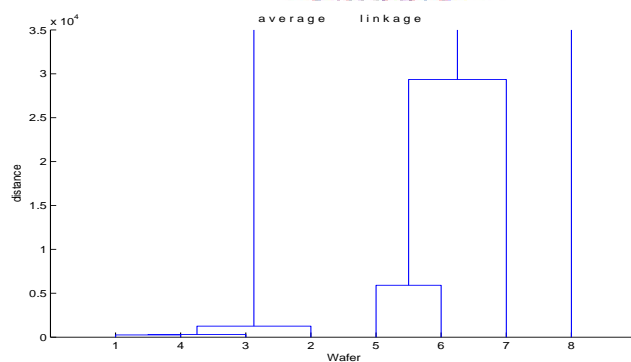


圖 4.7: 二階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖

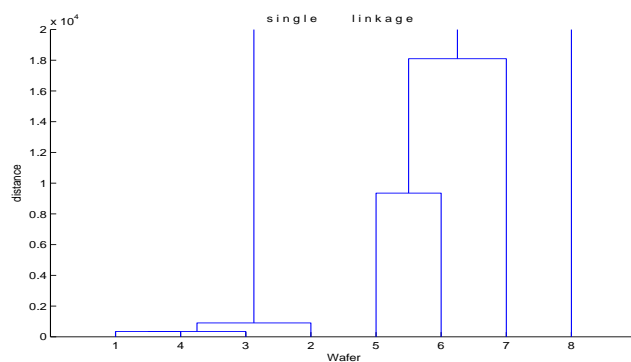


圖 4.8: 三階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖

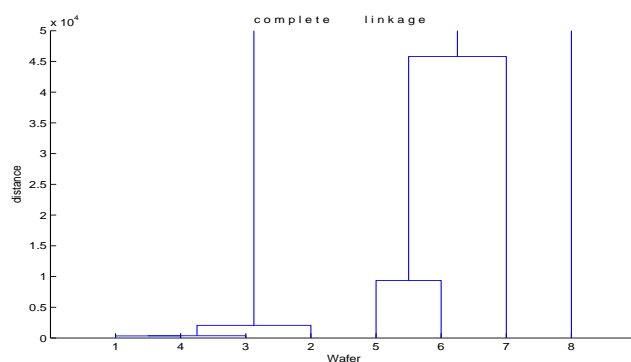


圖 4.9: 三階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖

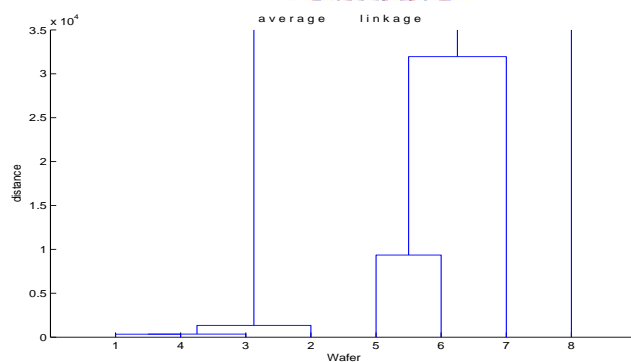


圖 4.10: 三階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖

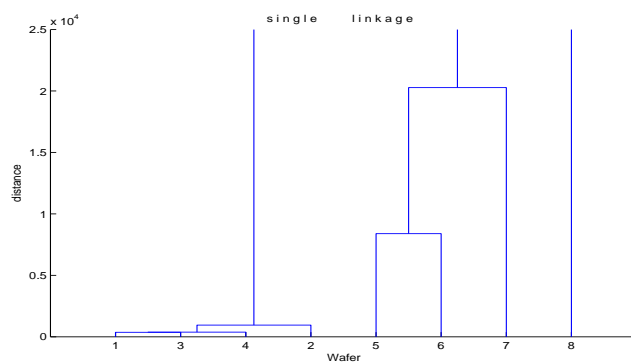


圖 4.11: 四階模型距離定義為(3.6) 式觀測值不轉換的單一聯結法樹狀圖

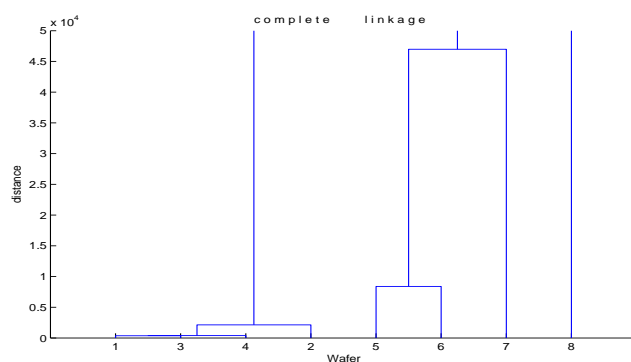


圖 4.12: 四階模型距離定義為(3.6) 式觀測值不轉換的完全聯結法樹狀圖

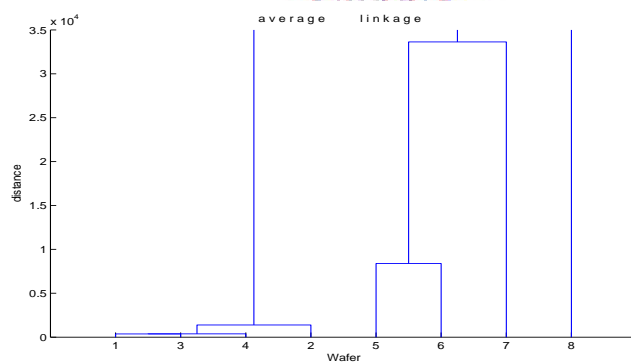


圖 4.13: 四階模型距離定義為(3.6) 式觀測值不轉換的平均聯結法樹狀圖



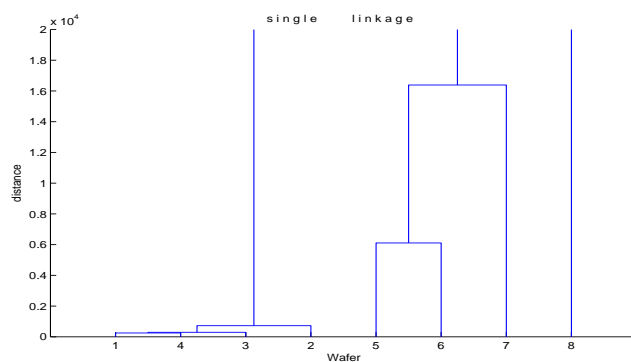


圖 4.14: 二階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖

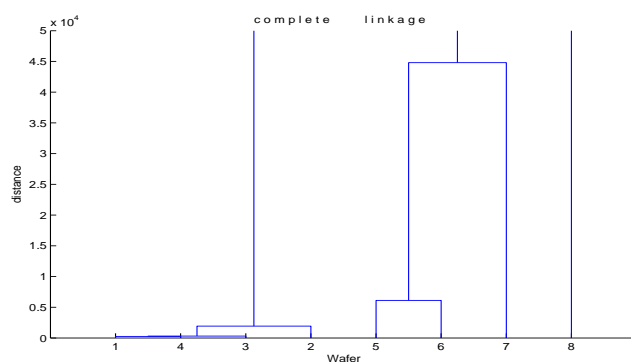


圖 4.15: 二階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖

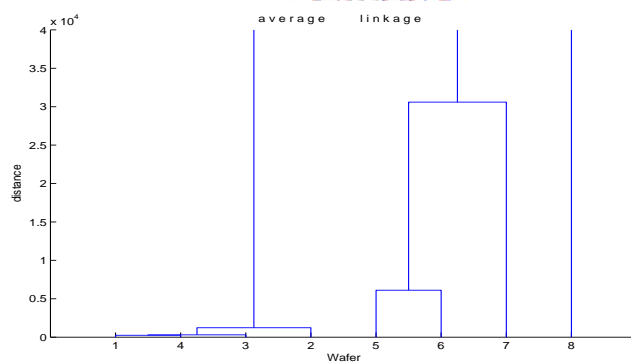


圖 4.16: 二階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖

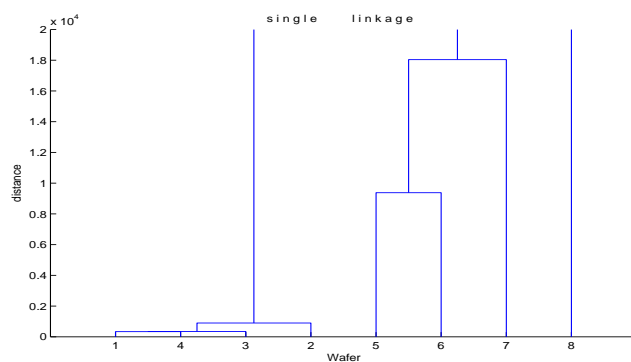


圖 4.17: 三階階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖

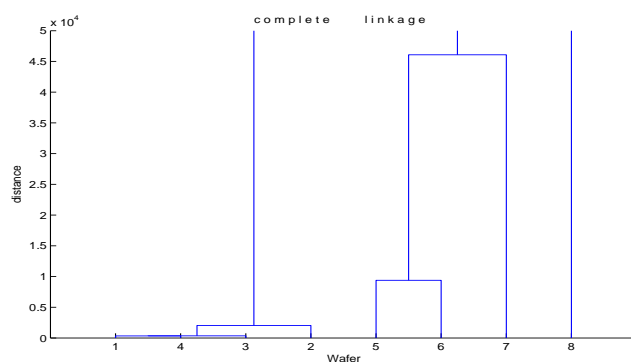


圖 4.18: 三階階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖

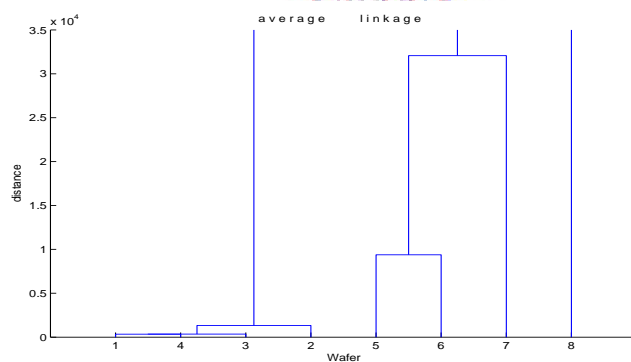


圖 4.19: 三階階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖

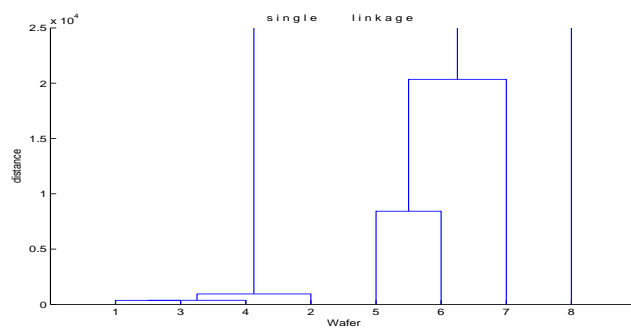


圖 4.20: 四階模型距離定義為(3.7) 式觀測值不轉換的單一聯結法樹狀圖

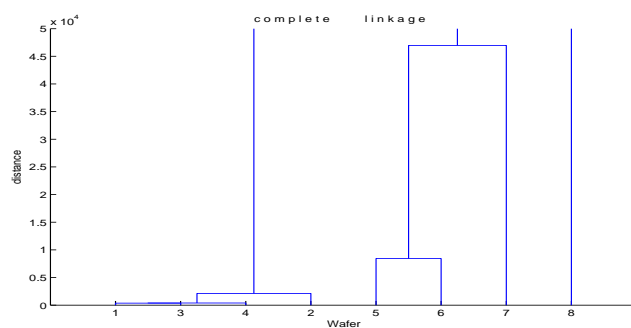


圖 4.21: 四階模型距離定義為(3.7) 式觀測值不轉換的完全聯結法樹狀圖

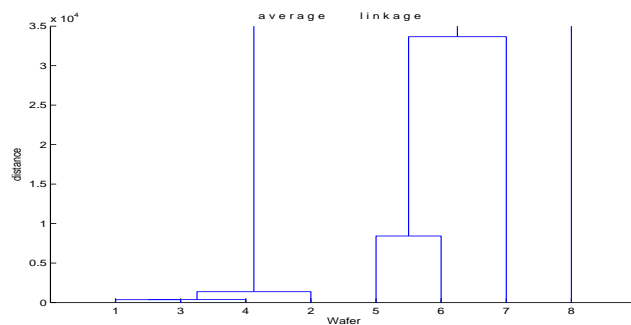


圖 4.22: 四階模型距離定義為(3.7) 式觀測值不轉換的平均聯結法樹狀圖

### 4.3 觀測值減掉該片晶圓所有觀測值的平均

若每片晶圓上的觀測值減掉該片晶圓所有觀測值的平均時, 則  $z_{i,j} = w_{i,j} - \overline{w_i}$ ,

其中  $\overline{w_i} = \sum_{j=1}^n w_{i,j}/n$ 。圖 4.23 為  $z_{i,j}$  依不同晶圓畫出的盒狀圖。

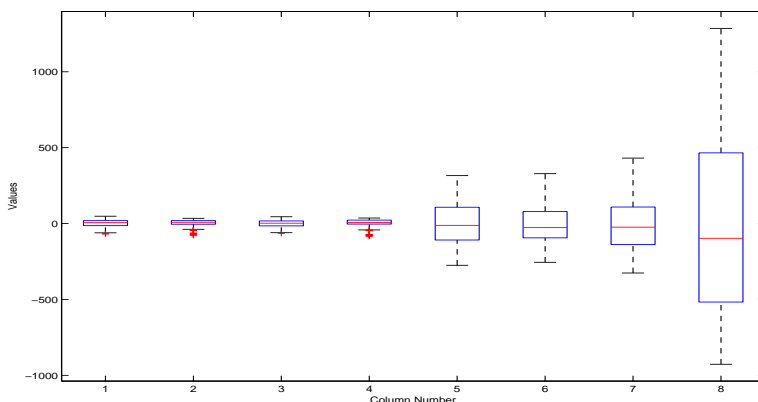


圖 4.23: 第二種轉換的盒狀圖

由圖 4.23, 還是很明顯的可以看出有三群, 晶圓 8 觀測值的變異數最大, 其次是晶圓 5,6,7, 晶圓 1,2,3,4 的變異數最小。由變異數得知, 此種轉換沒有解決晶圓上的曲面放大縮小的功能。

圖 4.24 為每片晶圓做第二種轉換後的無母數配適曲面之等高圖, 等高圖的樣貌會和第一種轉換的等高圖很像。晶圓 1,2,3,4 觀測值分佈的範圍為:  $-60 \sim 40$ ; 晶圓 5,6,7 觀測值分佈的範圍為:  $-200 \sim 500$ ; 晶圓 8 觀測值分佈的範圍為:  $-500 \sim 1000$ 。

若在  $z_{i,j}$  數據上使用 (3.6) 或 (3.7) 式, 求任兩片晶圓上曲面的距離, 就可以得到

距離矩陣。即可利用此距離矩陣, 再套用第二章的單一聯結法, 完全聯結法, 平均聯結法, 則可畫出它們的樹狀圖。

距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.25, 圖 4.26, 圖 4.27 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.28, 圖 4.29, 圖 4.30 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.31, 圖 4.32, 圖 4.33 為各種聯結法的樹狀圖。

距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.34, 圖 4.35, 圖 4.36 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.37, 圖 4.38, 圖 4.39 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.40, 圖 4.41, 圖 4.42 為各種聯結法的樹狀圖。

結論: 距離的定義為 (3.6), (3.7) 式對結論沒影響。分群的方式對結論沒影響。如果分三群, 晶圓 1,2,3,4 一群; 晶圓 5,6,7 一群; 晶圓 8 一群。如果要再多分一群, 在二階的情況下, 晶圓 6 會與晶圓 5,7 分開。在三階, 四階的情況下, 晶圓 7 會與晶圓 5,6 分開。

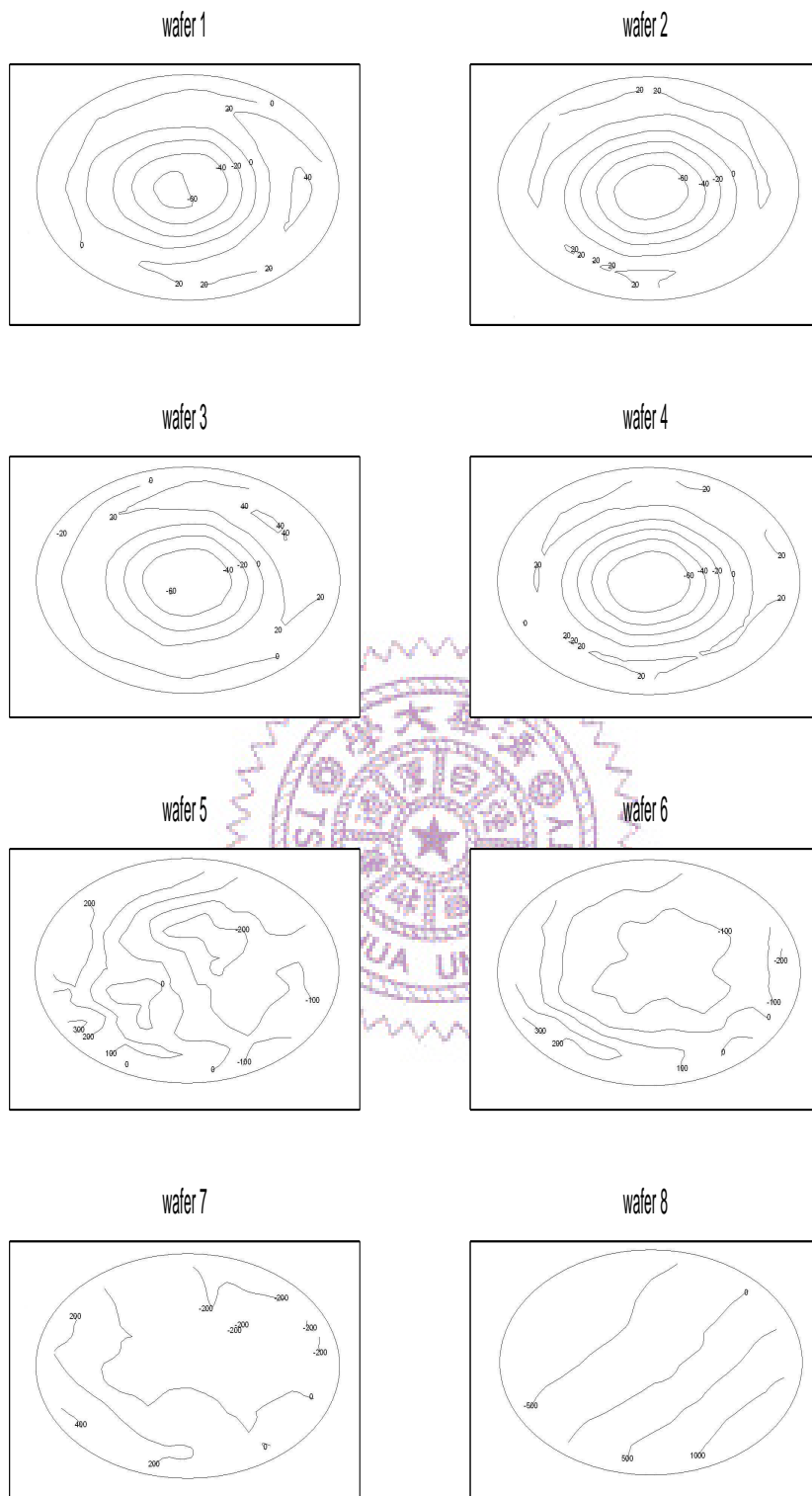


圖 4.24: 第二種轉換的等高圖

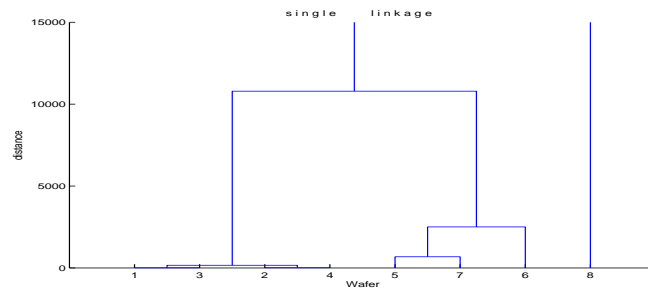


圖 4.25: 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單一  
聯結法樹狀圖

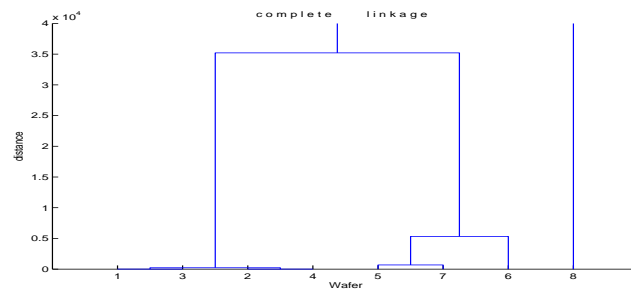


圖 4.26: 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完全  
聯結法樹狀圖

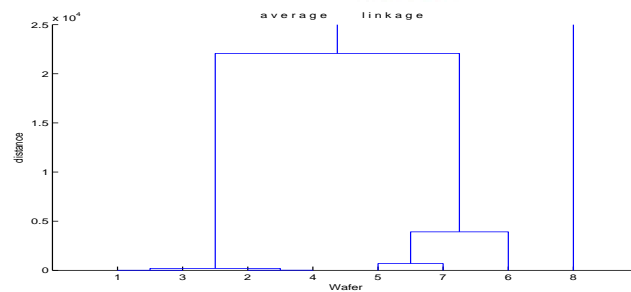


圖 4.27: 二階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平均  
聯結法樹狀圖

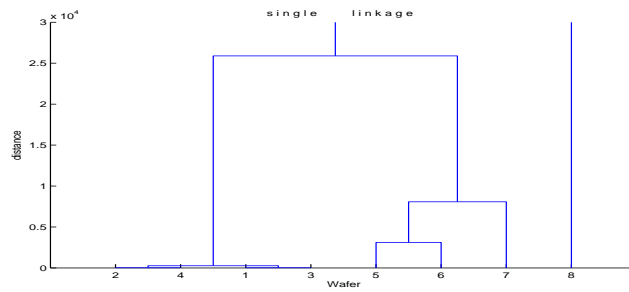


圖 4.28: 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單一  
聯結法樹狀圖

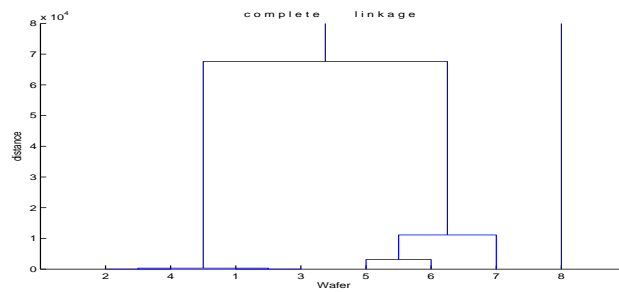


圖 4.29: 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完全  
聯結法樹狀圖

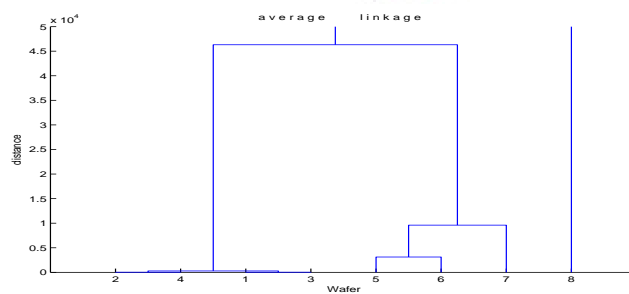


圖 4.30: 三階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平均  
聯結法樹狀圖



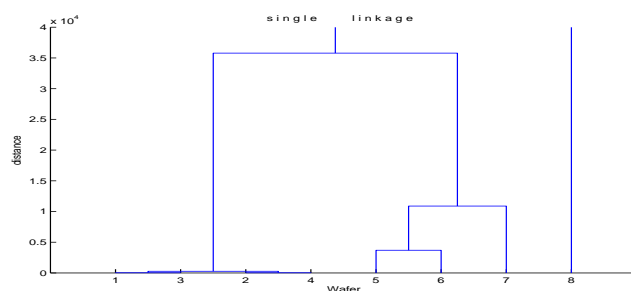


圖 4.31: 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的單一  
聯結法樹狀圖

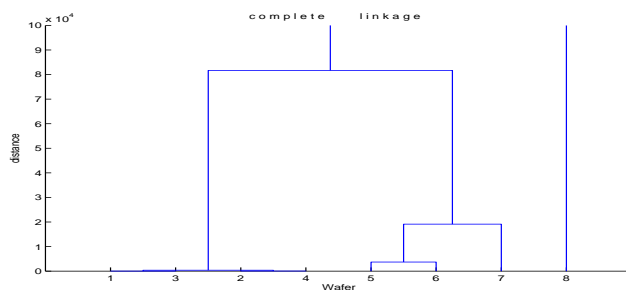


圖 4.32: 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的完全  
聯結法樹狀圖

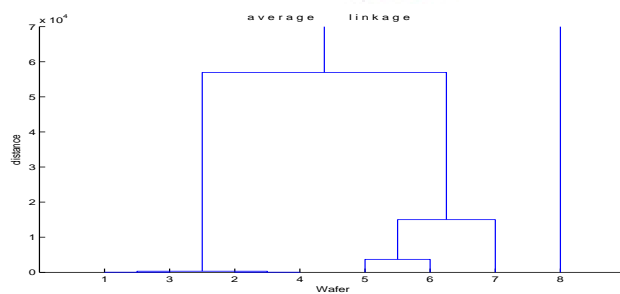


圖 4.33: 四階模型距離定義為(3.6) 式觀測值減掉該片晶圓所有觀測值的平均的平均  
聯結法樹狀圖

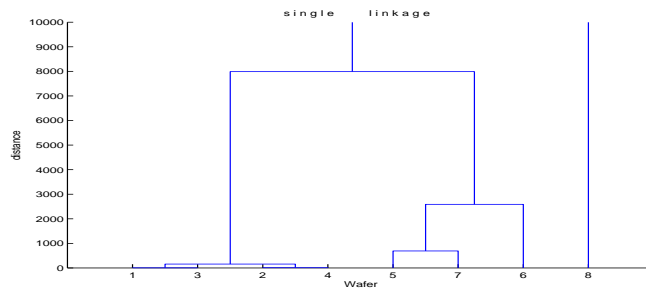


圖 4.34: 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一  
聯結法樹狀圖

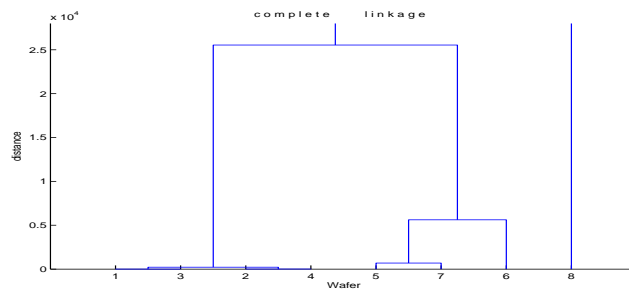


圖 4.35: 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全  
聯結法樹狀圖

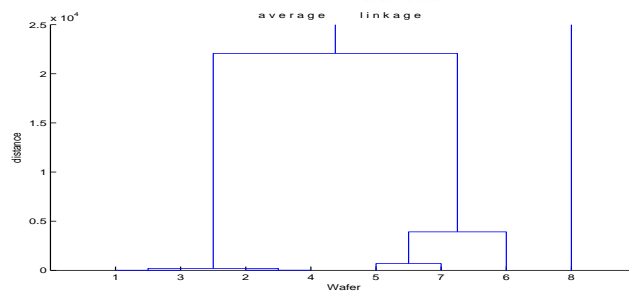


圖 4.36: 二階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均  
聯結法樹狀圖

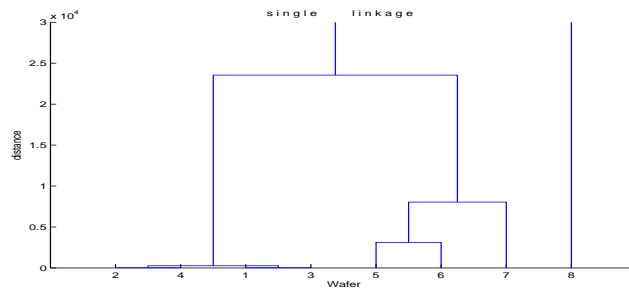


圖 4.37: 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一  
聯結法樹狀圖

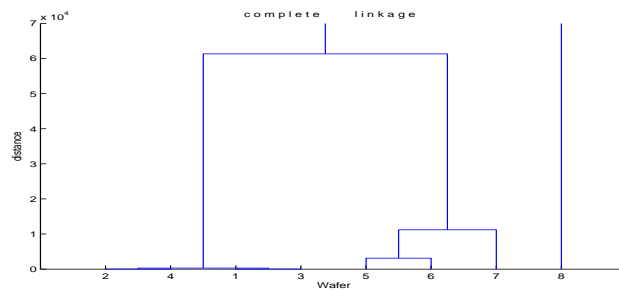


圖 4.38: 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全  
聯結法樹狀圖

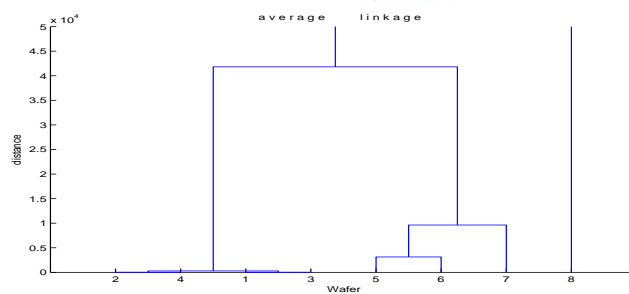


圖 4.39: 三階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均  
聯結法樹狀圖

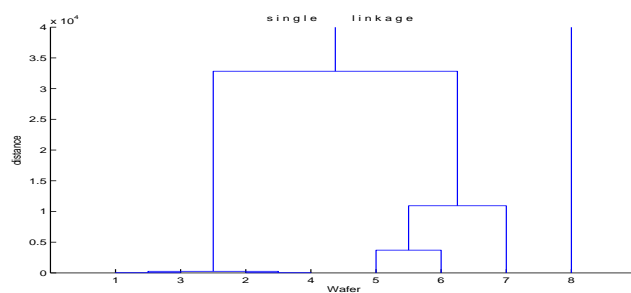


圖 4.40: 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的單一聯結法樹狀圖

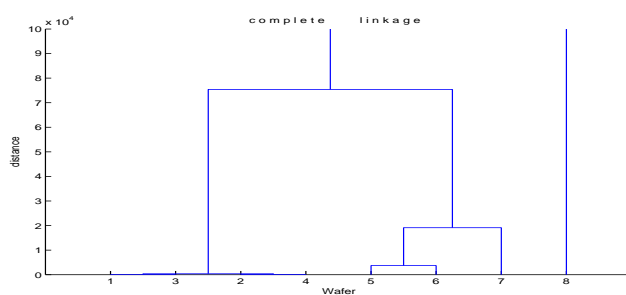


圖 4.41: 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的完全聯結法樹狀圖

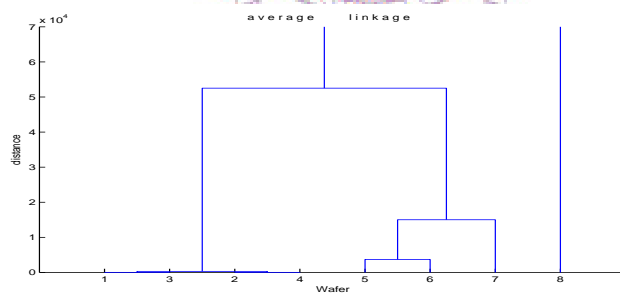


圖 4.42: 四階模型距離定義為(3.7) 式觀測值減掉該片晶圓所有觀測值的平均的平均聯結法樹狀圖

## 4.4 每個晶圓上的觀測值最大最小值調整成一樣

若每片晶圓上的觀測值最大最小值調整成一樣時。則  $z_{i,j} = (w_{i,j} - \min_j(w_{i,j})) / (\max_j(w_{i,j}) - \min_j(w_{i,j}))$ , 圖 4.43 為  $z_{i,j}$  依不同晶圓畫出的盒狀圖。

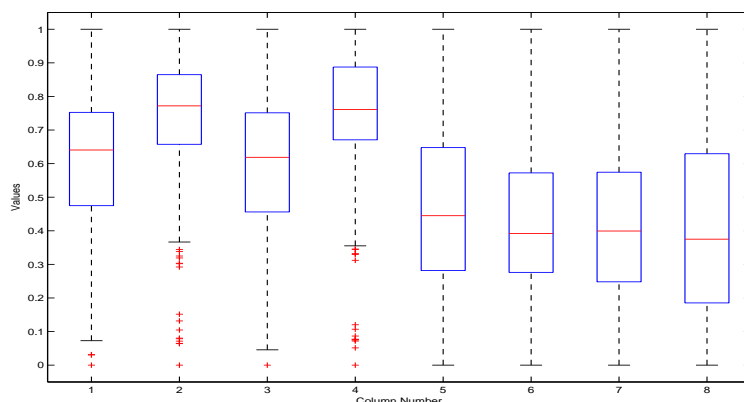


圖 4.43: 第三種轉換的盒狀圖

由盒狀圖看出晶圓 1,3 的觀測值分布範圍很像; 晶圓 2,4 分布範圍很像; 而晶圓 5,6,7,8 分布範圍很像。現在這八片晶圓上的觀測值變異數的差異, 已經由轉換讓它們變的差不多, 所以已經沒有曲面放大與縮小的問題。此種轉換存在一個問題, 就是容易受離群值的影響, 由圖可看出晶圓 2,4 深受離群值影響, 不過有些人會認為, 有離群值的資料要歸為一類, 因此偏好此種轉換。

圖 4.44 為每片晶圓做第三種轉換後的無母數配適曲面之等高圖, 等高圖的樣貌會和第一種轉換的等高圖很像。所有晶圓觀測值分佈的範圍為: 0 ~ 1。

若在  $z_{i,j}$  數據上使用 (3.6) 或 (3.7) 式, 求任兩片晶圓上曲面的距離, 就可以得到距離矩陣。即可利用此距離矩陣, 再套用第二章的單一聯結法, 完全聯結法, 平均聯結

法, 則可畫出它們的樹狀圖。

距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.45, 圖 4.46, 圖 4.47 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.48, 圖 4.49, 圖 4.50 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.51, 圖 4.52, 圖 4.53 為各種聯結法的樹狀圖。

距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.54, 圖 4.55, 圖 4.56 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.57, 圖 4.58, 圖 4.59 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1, \hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.60, 圖 4.61, 圖 4.62 為各種聯結法的樹狀圖。

結論: 距離的定義為 (3.6), (3.7) 式對結論沒影響。分群的方式對結論沒影響。如果分三群, 晶圓 1,3 一群; 晶圓 2,4 一群; 晶圓 5,6,7,8 一群。如果要再多分一群, 在二階的情況下, 晶圓 8 會與晶圓 5,6,7 分開。在三階, 四階的情況下, 晶圓 5,6 會與晶圓 7,8 分開。

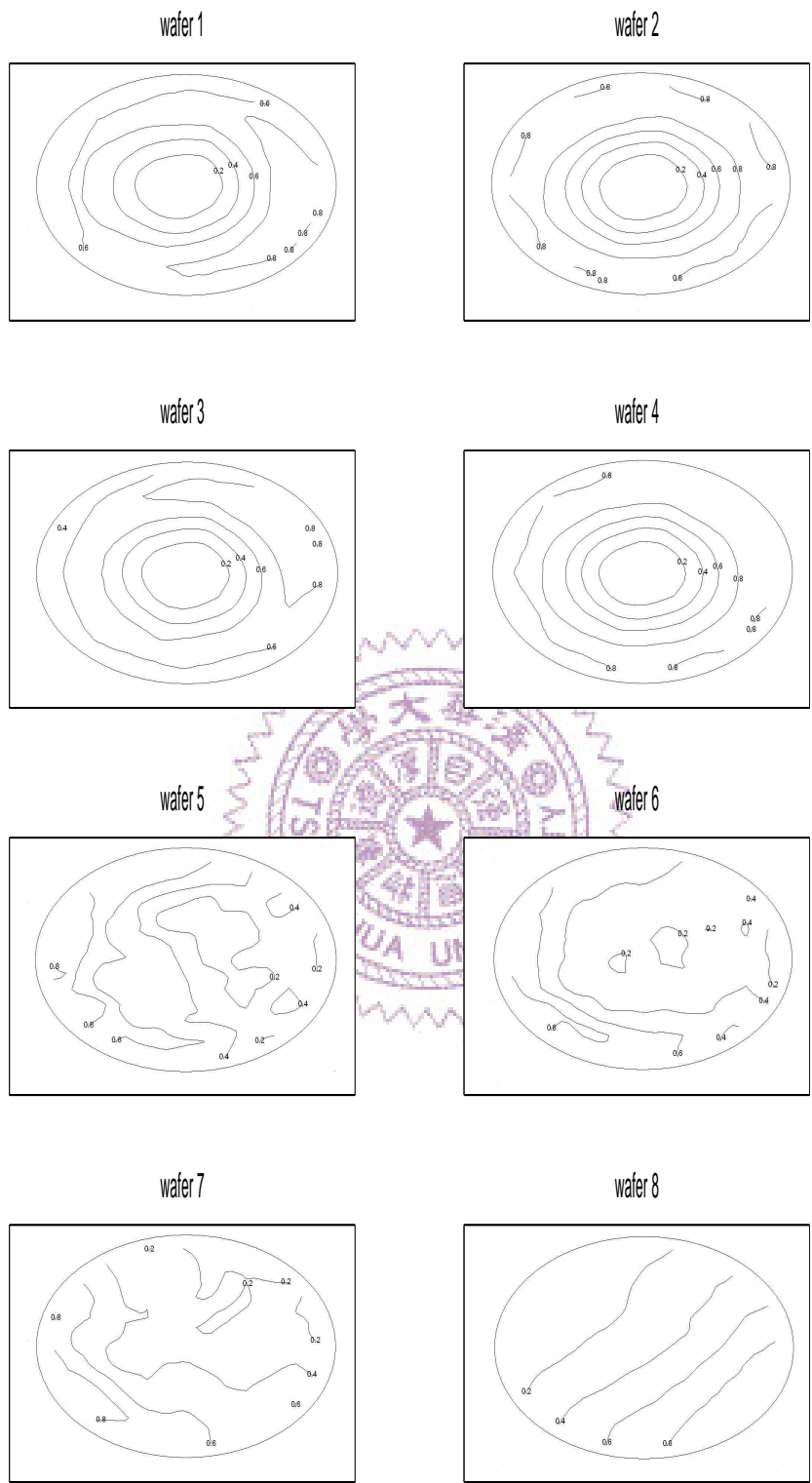


圖 4.44: 第三種轉換的等高圖

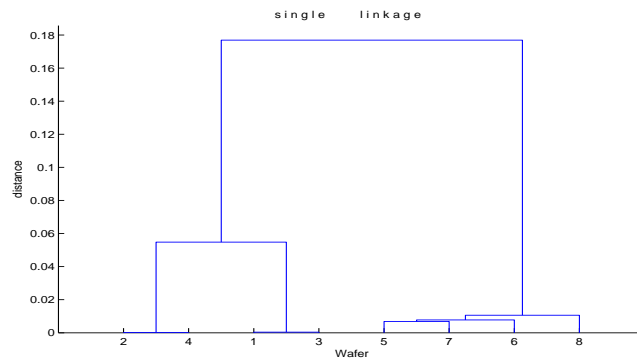


圖 4.45: 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖

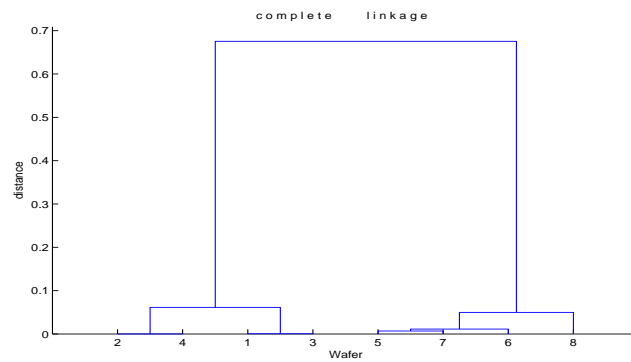


圖 4.46: 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖

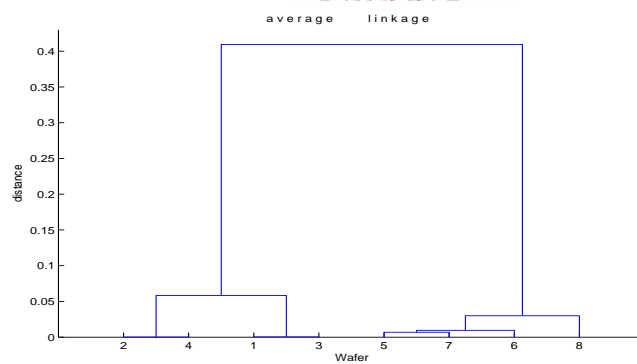


圖 4.47: 二階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖



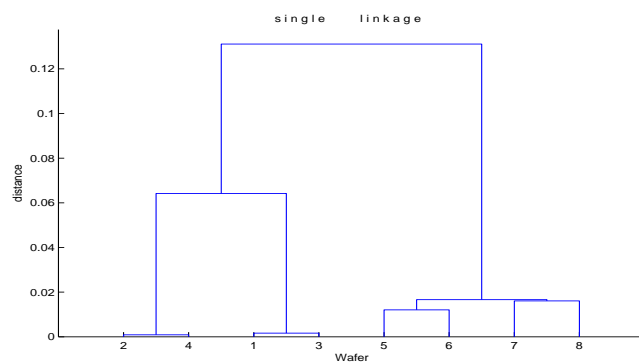


圖 4.48: 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖

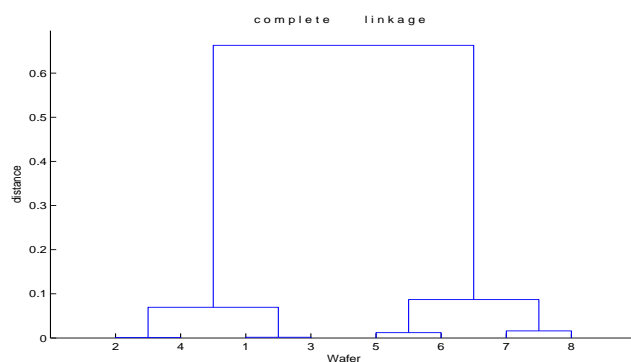


圖 4.49: 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖

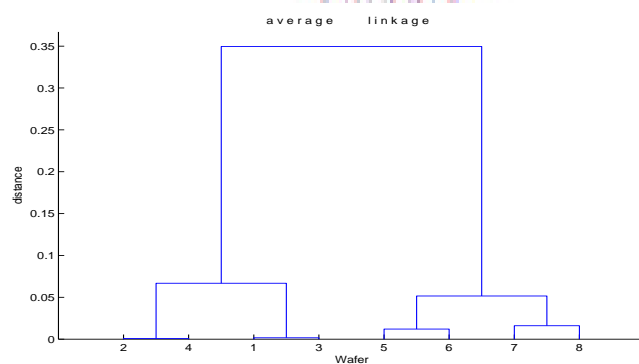


圖 4.50: 三階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖

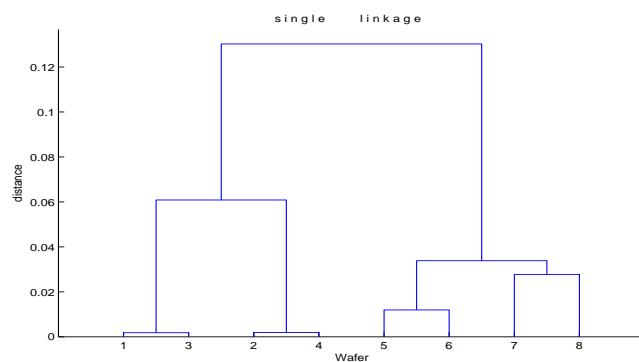


圖 4.51: 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的單一聯結法樹狀圖

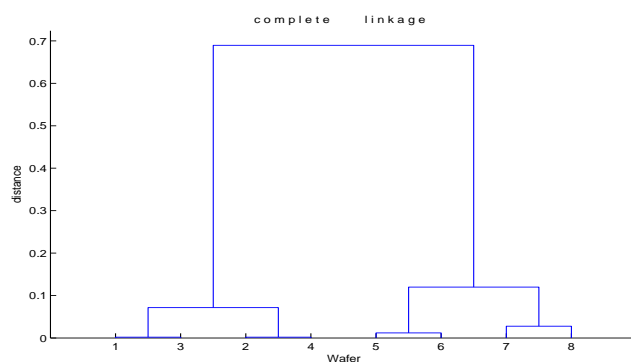


圖 4.52: 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的完全聯結法樹狀圖

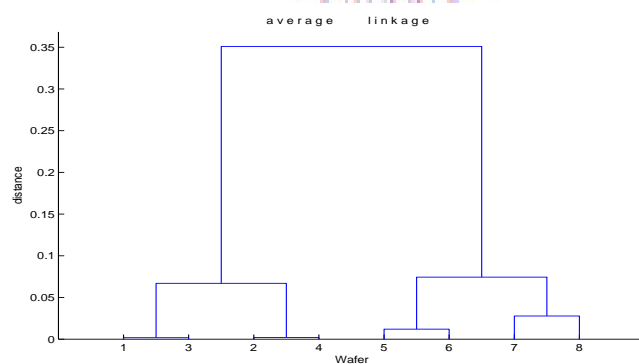


圖 4.53: 四階模型距離定義為(3.6) 式觀測值全距調整成一樣的平均聯結法樹狀圖

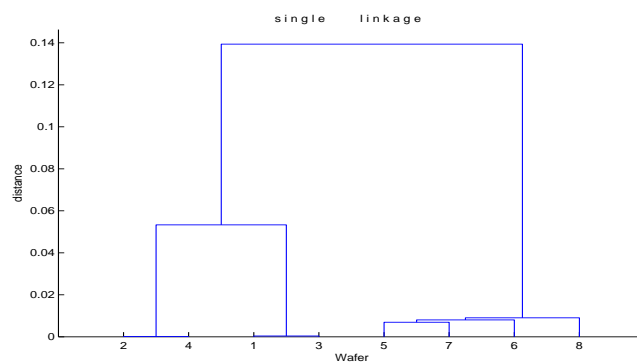


圖 4.54: 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖

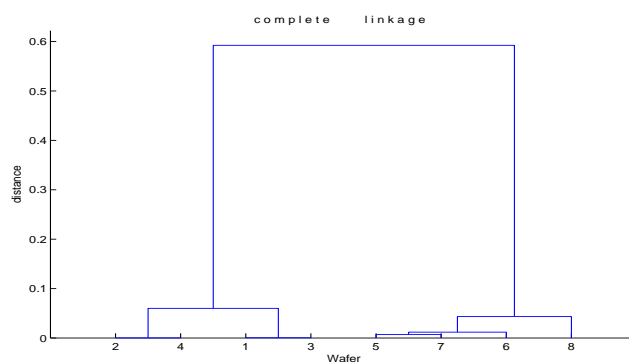


圖 4.55: 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖

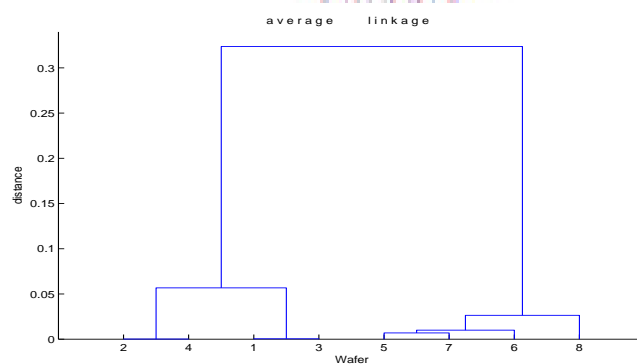


圖 4.56: 二階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖

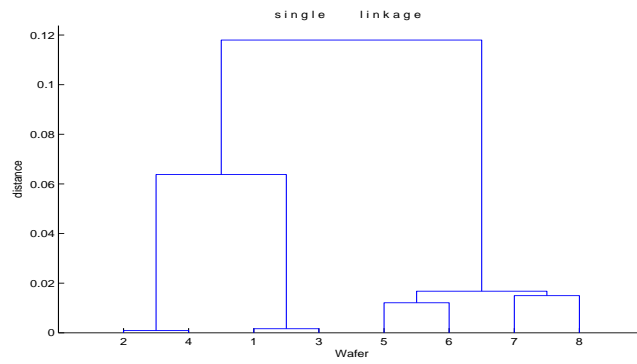


圖 4.57: 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖

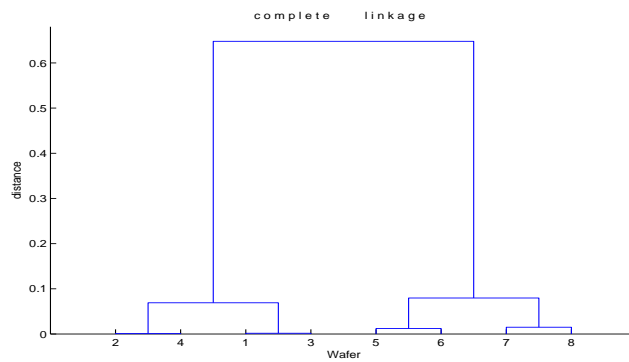


圖 4.58: 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖

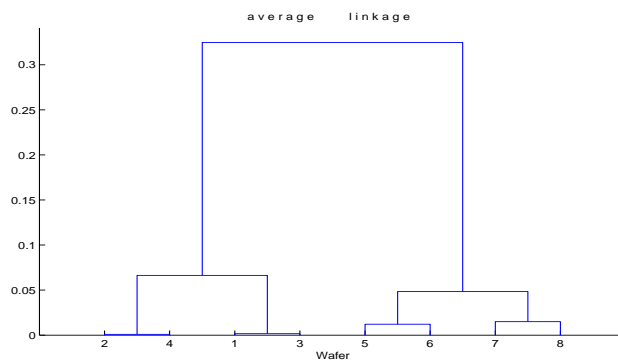


圖 4.59: 三階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖

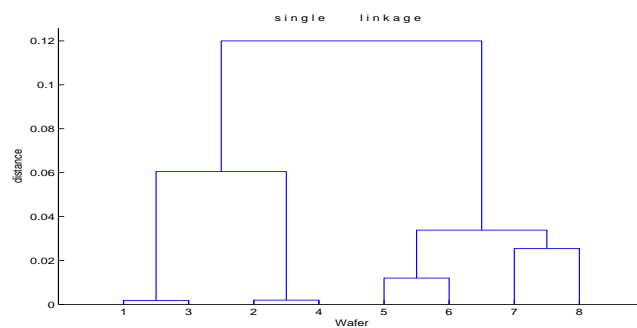


圖 4.60: 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的單一聯結法樹狀圖

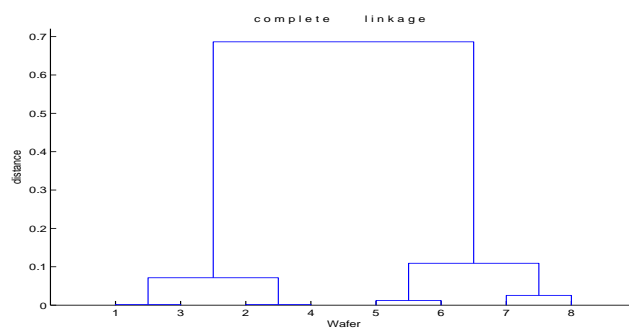


圖 4.61: 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的完全聯結法樹狀圖

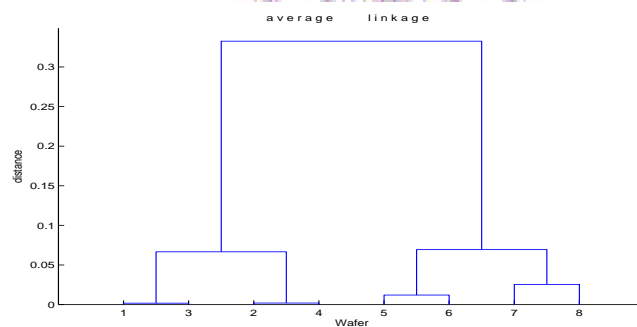


圖 4.62: 四階模型距離定義為(3.7) 式觀測值全距調整成一樣的平均聯結法樹狀圖

## 4.5 標準化

若觀測值做標準化時。則  $z_{i,j} = (w_{i,j} - \bar{w}_{i\cdot})/s_i$  , 其中  $\bar{w}_{i\cdot} = \sum_{j=1}^n w_{i,j}/n, s_i = \sqrt{\sum_{j=1}^n (w_{i,j} - \bar{w}_{i\cdot})^2 / (n - 1)}$ , 圖 4.63 為  $z_{i,j}$  依不同晶圓畫出的盒狀圖。

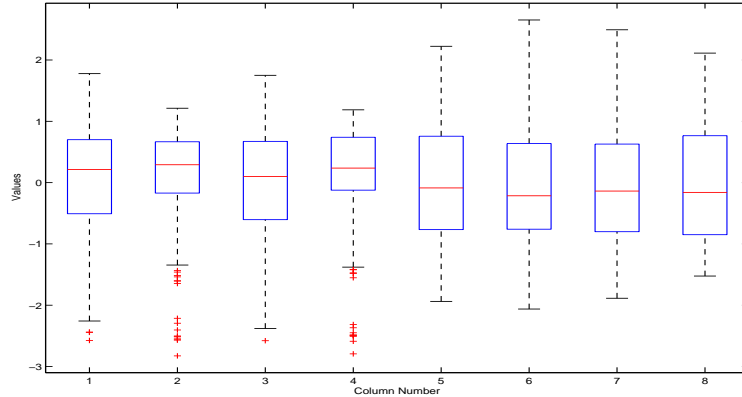


圖 4.63: 第四種轉換的盒狀圖

現在每片晶圓上的觀測值位置與尺度已經差不多了, 此種轉換與先前把觀測值範圍調整為 0 ~ 1 區間比較不受離群值的影響。一般統計學家分析資料時, 解決位置與尺度的不同, 就是把資料標準化。比較特別的為, 晶圓 2,4 的四分位距較小且離群值較多。而圖 4.64 為它們的等高圖。

圖 4.44 為每片晶圓做第三種轉換後的無母數配適曲面之等高圖, 等高圖的樣貌會和第一種轉換的等高圖很像。所有晶圓觀測值分佈的範圍為: -2 ~ 2。

若在  $z_{i,j}$  數據上使用 (3.6) 或 (3.7) 式, 求任兩片晶圓上曲面的距離, 就可以得到距離矩陣。即可利用此距離矩陣, 再套用第二章的單一聯結法, 完全聯結法, 平均聯結法, 則可畫出它們的樹狀圖。

距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.65, 圖 4.66, 圖 4.67 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.68, 圖 4.69, 圖 4.70 為各種聯結法的樹狀圖。距離的定義為 (3.6) 式,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.71, 圖 4.72, 圖 4.73 為各種聯結法的樹狀圖。

距離的定義為 (3.7) 式且  $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個二階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.74, 圖 4.75, 圖 4.76 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個三階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.77, 圖 4.78, 圖 4.79 為各種聯結法的樹狀圖。距離的定義為 (3.7) 式且  $k = 1$ ,  $\hat{f}_A, \hat{f}_B$  為  $x, y$  的一個四階模型。 $R_A^2, R_B^2$  的值為表 4.1 要比較的那兩個, 圖 4.80, 圖 4.81, 圖 4.82 為各種聯結法的樹狀圖。

結論: 距離的定義為 (3.6), (3.7) 式對結論沒影響。如果分三群, 晶圓 1,3 一群; 晶圓 2,4 一群; 晶圓 5,6,7,8 一群。如果要再多分一群, 在二階的情況下, 晶圓 8 會與晶圓 5,6,7 分開。在三階, 四階的情況下, 晶圓 5,6 會與晶圓 7,8 分開。

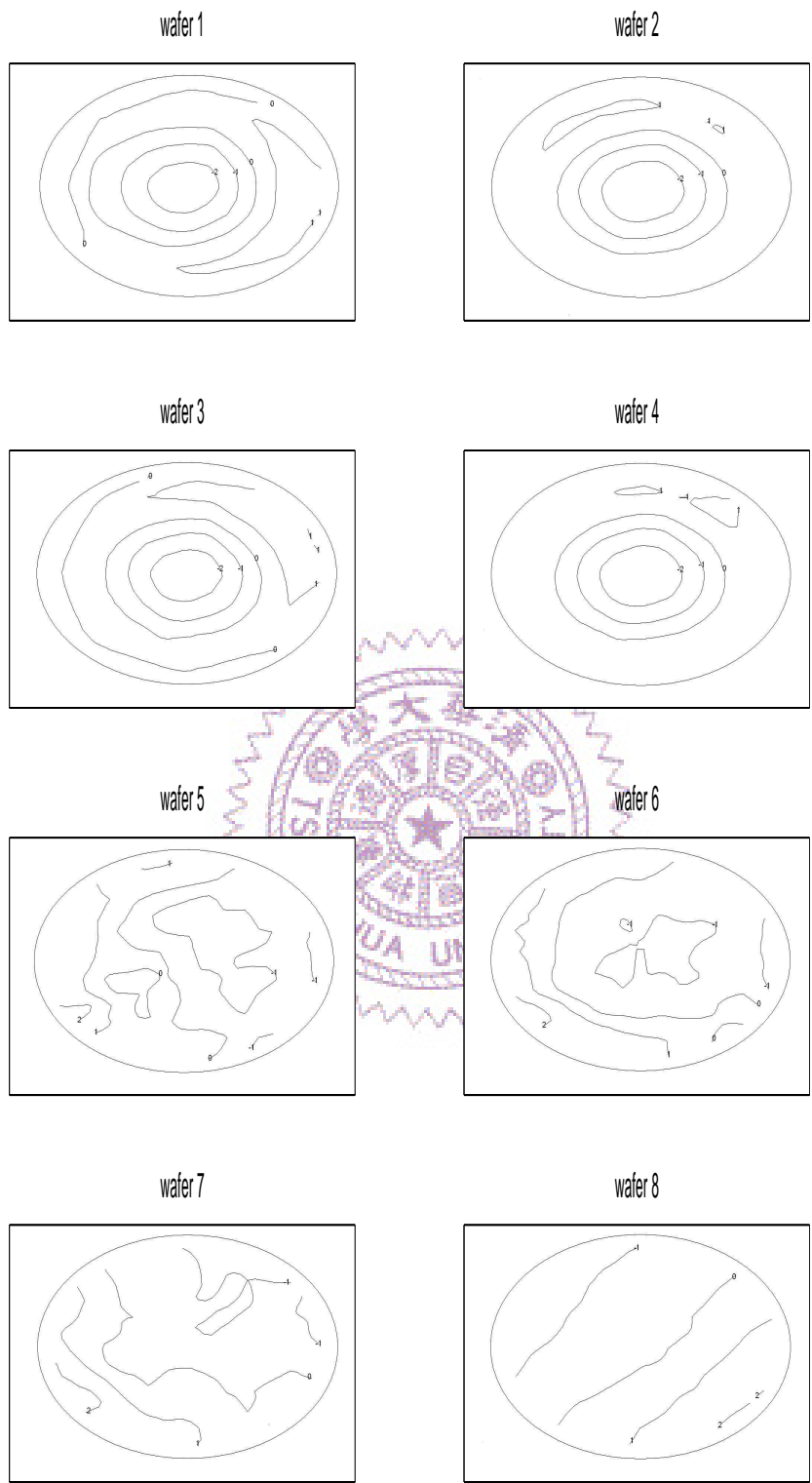


圖 4.64: 第四種轉換的等高圖



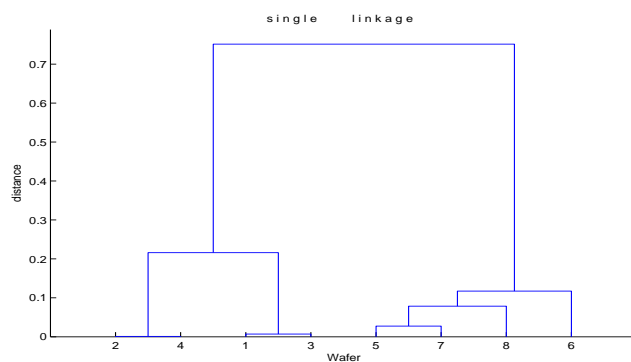


圖 4.65: 二階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖

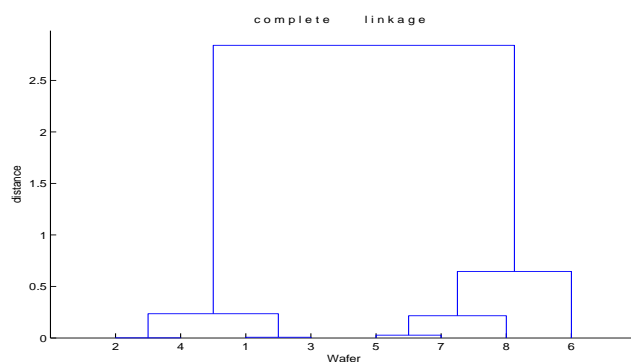


圖 4.66: 二階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖

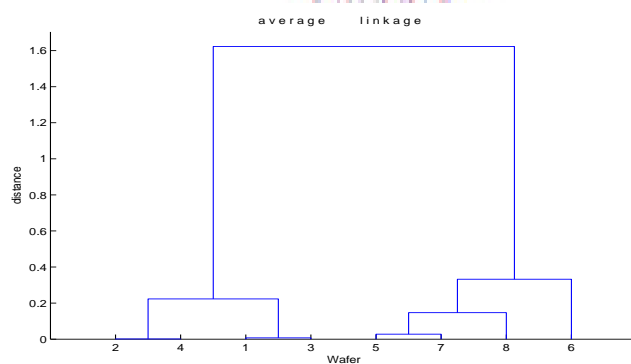


圖 4.67: 二階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖

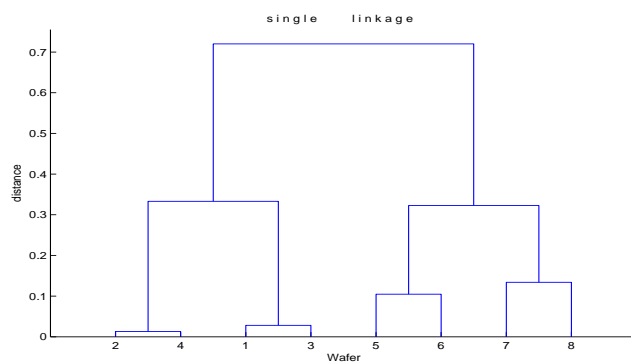


圖 4.68: 三階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖

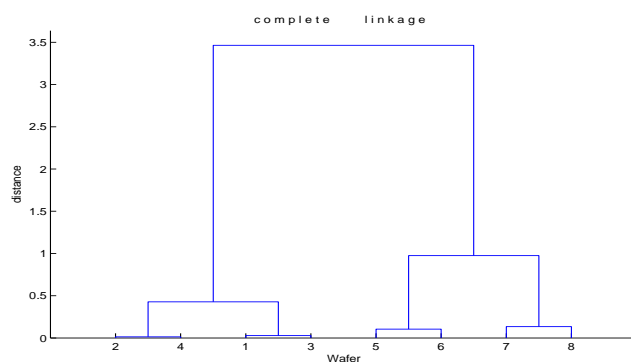


圖 4.69: 三階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖

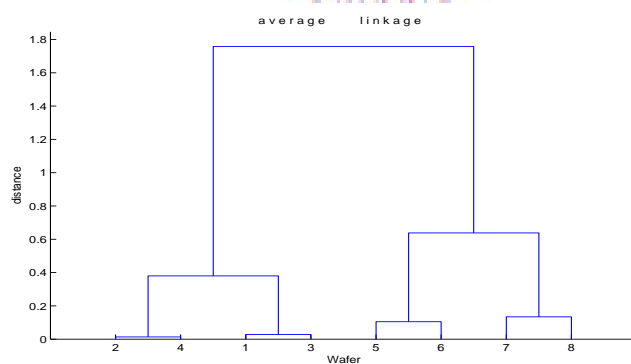


圖 4.70: 三階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖

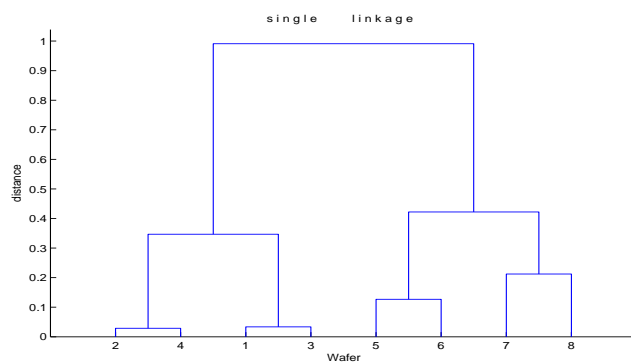


圖 4.71: 四階模型距離定義為(3.6) 式觀測值標準化的單一聯結法樹狀圖

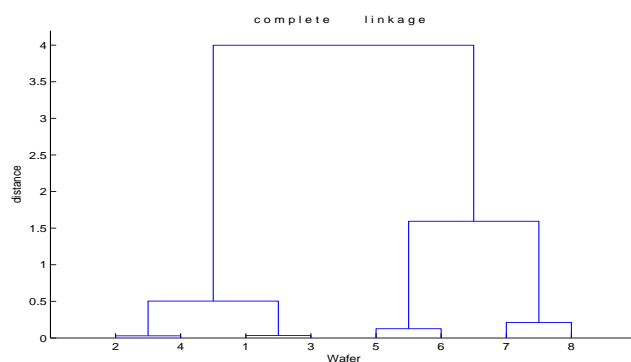


圖 4.72: 四階模型距離定義為(3.6) 式觀測值標準化的完全聯結法樹狀圖

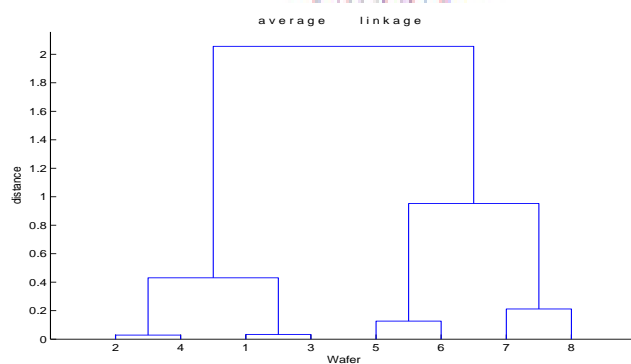


圖 4.73: 四階模型距離定義為(3.6) 式觀測值標準化的平均聯結法樹狀圖

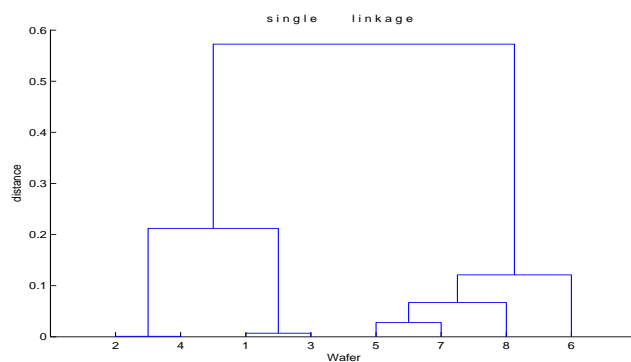


圖 4.74: 二階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖

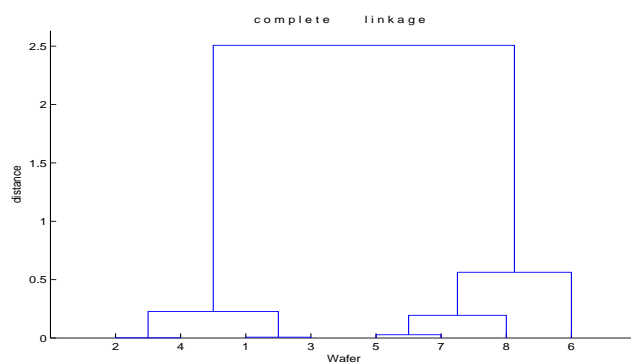


圖 4.75: 二階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖

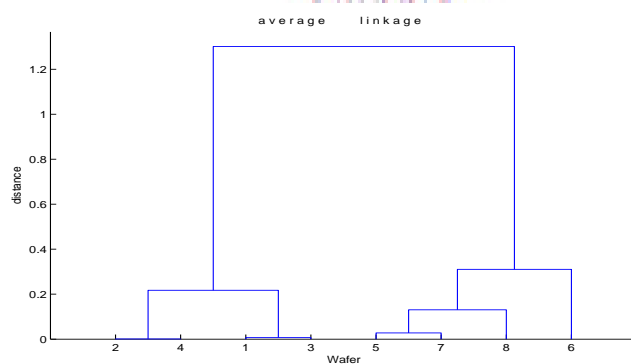


圖 4.76: 二階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖

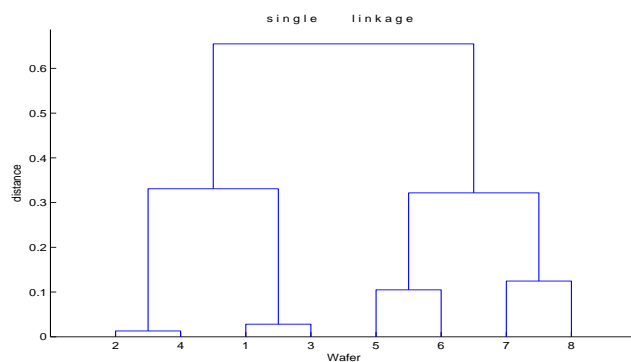


圖 4.77: 三階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖

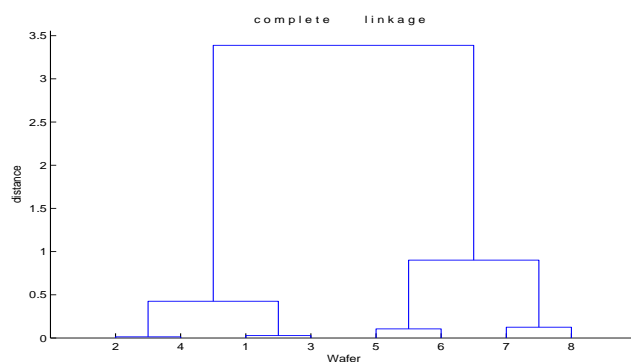


圖 4.78: 三階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖

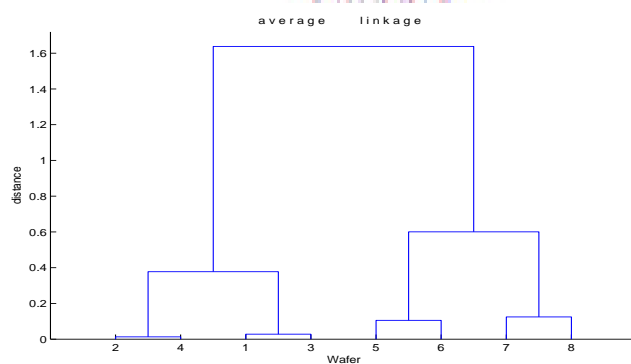


圖 4.79: 三階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖

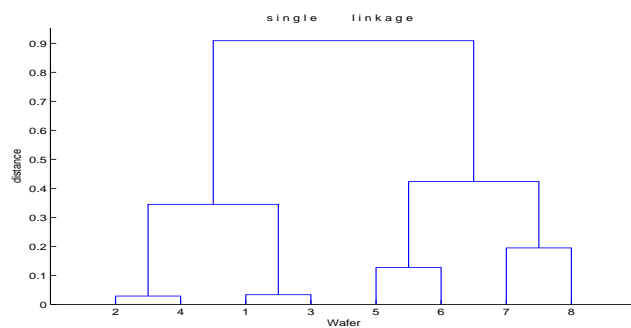


圖 4.80: 四階模型距離定義為(3.7) 式觀測值標準化的單一聯結法樹狀圖

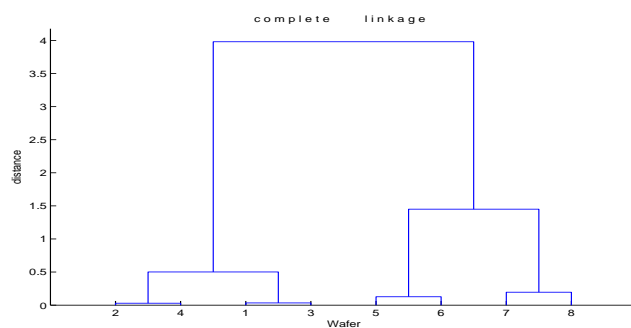


圖 4.81: 四階模型距離定義為(3.7) 式觀測值標準化的完全聯結法樹狀圖

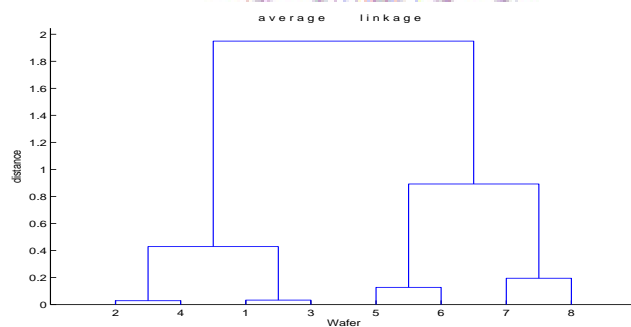


圖 4.82: 四階模型距離定義為(3.7) 式觀測值標準化的平均聯結法樹狀圖

## 4.6 本章結論

本章使用的距離定義有兩個，觀測值作四種轉換，分群方式有三種，配適的模型大小有三個，所以總共有 72 張樹狀圖。因為我們配適模型時，希望模型的判定係數可以很高，因此我們在極座標建模，並考慮極軸當參數（見第五章），所以得到的判定係數最小也有 0.7 以上，以致於我們提出對判定係數不高的那兩個懲罰項，在這裡算出來的值會差不多，所以用 (3.6) 或 (3.7) 式當距離的定義，結果是一樣的。三階模型與四階模型的判定係數差不多，所以分群的結果一致性很高。晶圓 1,2,3,4 在二階模型的時候，判定係數已經高達 0.95 以上，所以對晶圓 1,2,3,4，使用幾階模型，結果是一樣。由原始觀測值的盒狀圖，圖 4.3 看出晶圓 8 的數據分佈範圍很廣，所以它的曲面很容易是其它晶圓的曲面放大了幾倍，對觀測值作的第一種與第二種轉換沒有解決曲面放大縮小的問題，所以晶圓 8 的曲面表現的與其它晶圓的曲面最不像。由圖 4.1，圖 4.2 看出如果不考慮曲面的放大縮小，單單只是看曲面像不像，會發現晶圓 7 的曲面旋轉後應該和晶圓 8 的曲面很像。對觀測值作的第三種與第四種轉換，解決曲面放大縮小的問題，結果得到晶圓 7 與晶圓 8 的距離很小，它們是非常像的，這也說明了我們上一章所介紹的那一套方法，是有效的。

在附錄裡會提供 4.5 節把觀測值標準化後，經過分析找到旋轉至兩兩最像的角度，再畫出兩兩最像的曲面圖。

## 第 5 章

### 極座標模型

#### 5.1 導論

在圓上配適迴歸模型，要考慮極軸與圓心的位置，不同的極軸與圓心皆會造成判定係數的不同，所以要得到一個圓上最佳的二階模型，就必須考慮不同的極軸與圓心。Lin (2007) 提到圓上的數據有圓效應(circular effect) 時，利用直角座標  $X, Y$  當自變數建模，通常結果不會太好。該論文提出了把直角座標轉換至極座標，利用極座標的  $r, \theta$  當自變數再建模，此時模型的判定係數會大很多。Lin (2007)也提出了極軸改變是如何影響 $r, \theta$  區間的實驗點位置，這在第三章已經有簡述了。那篇論文對找一個圓上最佳模型的結論是：要在極座標空間建模，且固定圓心，相鄰兩實驗點間的極軸，判定係數的值並不會有太大的改變。

#### 5.2 如何解決不同極軸的問題

圖5.1為某電子公司提供的晶圓 7，圓心為  $(0,0)$ ，在圓上均勻的取 120 個極軸後，不同極軸與對應判定係數的圖形。由圖 5.1 可看出相鄰的極軸，判定係數的值並



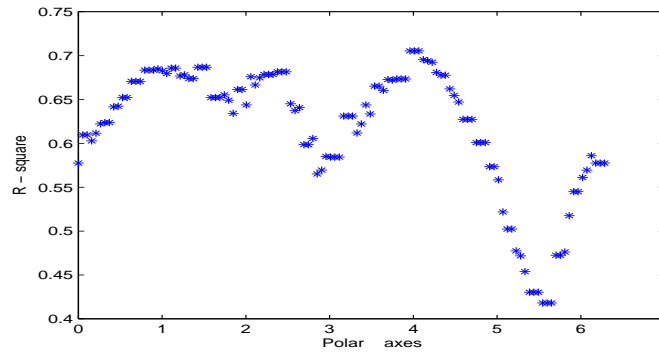


圖 5.1: 不同極軸與對應判定係數的圖型

不會有太大的改變, 而且有些判定係數還一模一樣, 這個觀念是Lin (2007)提出來的。它說明了固定圓心在(0,0), 即使極軸範圍在  $0 \sim 2\pi$ 之間變化, 還是可以找到逼近最佳的迴歸模型。晶圓 7 的最佳模型應該發生在極軸為 3 跟 4 之間, 它的判定係數約為 0.7。我們由圖 5.1 發現了, 相鄰的極軸, 判定係數有時還會一模一樣, 因此現在就有興趣想知道為什麼極軸的不同, 判定係數有時還會一樣。最後我們由第二章的圖 2.1 與圖 2.2 可發現, 如果任兩個極軸間沒有觀測值, 所有觀測值的位置在這兩個極軸的表現是:  $r$  值不變,  $\theta$  值減掉一個固定常數。這個現象就是迴歸模型的自變數  $\theta$  做位置的轉換。在迴歸分析中, 若配適一個  $t$  階多項式模型,  $t=1, 2, \dots$  時, 不管因變數或自變數作任何位置或尺度的轉換, 判定係數皆不會變。所以解決了極軸的不同, 有時判定係數還是不會變的問題。

**定理 1.** 在  $r, \theta$  的極座標空間, 若任兩個極軸之間沒有觀測值, 則配適一個  $t$  階多項式模型的判定係數是一樣的。

**証明.** 以下我們以配適  $r, \theta$  的二階多項式模型來說明, 其可令  $A, B$  為任兩個建模區間上的極軸 ( $A > B$ ), 且它們之間沒有任何觀測值。則  $n$  個量測座標在以  $A$  為極軸的

值爲  $(\theta_1, \theta_2, \dots, \theta_n)$ 。令  $d$  爲 A,B 兩極軸的夾角, 則此  $n$  個量測座標以 B 爲極軸時的  $\theta$  值爲  $(\theta_1 + d, \theta_2 + d, \dots, \theta_n + d)$ 。故極軸爲 A 的模型爲:

$$w = \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 \theta + \beta_4 \theta^2 + \beta_5 r \theta + \varepsilon,$$

極軸爲 B 的模型爲:

$$\begin{aligned} w' &= \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3(\theta + d) + \alpha_4(\theta + d)^2 + \alpha_5 r(\theta + d) + \varepsilon, \\ &= (\alpha_0 + \alpha_3 d + \alpha_4 d^2) + (\alpha_1 + \alpha_5 d)r + \alpha_2 r^2 + (\alpha_3 + 2d\alpha_4)\theta + \alpha_4 \theta^2 + \alpha_5 r \theta + \varepsilon, \\ &= \beta'_0 + \beta'_1 r + \beta'_2 r^2 + \beta'_3 \theta + \beta'_4 \theta^2 + \beta'_5 r \theta + \varepsilon', \end{aligned}$$

故極軸 A 與 B 的模型矩陣是一樣的, 因此極軸 A 與 B 的誤差平方和一樣。  $\square$

因此固定圓心時, 我們可以找到最佳模型, 方法爲: 任兩個實驗點間, 我們只要任意選一個極軸當極軸再建模, 因爲實驗點一定是有限的, 所以模型數也是有限的, 因此判定係數是有限的, 最佳模型就找到了。

### 5.3 允許極座標圓心位置改變

現在允許極座標圓心位置改變, 假設挑選  $n$  個量測座標作實驗, 現在令圓上某一個量測座標  $(x^*, y^*)$  爲新圓心, 某個  $\theta^*$  當極軸, 此時  $n$  個量測座標的位置在極座標空間的位置爲  $(r_1, \theta_1), (r_2, \theta_2), \dots, (r_n, \theta_n)$ , 而  $(r, \theta)$  由下列公式得到。令:

$$\theta_A = \begin{cases} \cos^{-1} \left( \frac{x-x^*}{\sqrt{(x-x^*)^2 + (y-y^*)^2}} \right) & \text{if } (y - y^*) > 0, \\ 2\pi - \cos^{-1} \left( \frac{x-x^*}{\sqrt{(x-x^*)^2 + (y-y^*)^2}} \right) & \text{if } (y - y^*) \leq 0 \end{cases}$$

if  $\theta_A \geq \theta^*$ ,

$$r = (x - x^*)^2 + (y - y^*)^2$$

$$\theta = \theta_A - \theta^*$$

if  $\theta_A < \theta^*$ ,

$$r = (x - x^*)^2 + (y - y^*)^2$$

$$\theta = 2\pi + \theta_A - \theta^*$$

令  $w_i$  為第  $i$  個量測座標上的觀測值。現在可以以  $(r_i, \theta_i)$  對  $w_i$  配適一個  $t$  階多項式模型, 即可得到判定係數  $R_{x^*, y^*, \theta^*}^2$ 。如果要找一個任意量測座標當新圓心, 極軸在任意位置的最大判定係數, 亦即找  $\max_{x^*, y^*, \theta^*} R_{x^*, y^*, \theta^*}^2$ 。因為由上一節的定理得知, 在  $r, \theta$  的極座標空間, 若任兩個極軸之間沒有觀測值, 則配適一個  $t$  階多項式模型的判定係數是一樣的。所以每個量測座標當新圓心時的判定係數皆是有限的, 又量測座標個數是有限的, 所以可以找到  $\max_{x^*, y^*, \theta^*} R_{x^*, y^*, \theta^*}^2$ 。表 5.1 為某電子公司提供的晶圓, 配適二階多項式模型, 以  $(0,0)$  為圓心及以任意量測座標為新圓心上的最大判定係數。

表 5.1: 判定係數

R-square	晶圓 1	晶圓 2	晶圓 3	晶圓 4	晶圓 5	晶圓 6	晶圓 7	晶圓 8
$(0,0)$	0.96	0.979	0.951	0.984	0.749	0.723	0.705	0.861
量測座標	0.96	0.979	0.951	0.984	0.816	0.829	0.890	0.989

如果更進一步的推廣, 考慮新圓心可以在圓上的任意位置, 亦即新圓心的位置不一定是在量測座標上時, 我們發現了兩個特殊的狀況。

1 : 若量測座標的分布如圖 5.2 所示。首先我們把所有量測座標依距離晶圓中心 (0,0) 的遠近分五大類, 第一類為編號2 ~ 9, 第二類為編號10 ~ 25, 第三類為編號26 ~ 49, 第四類為編號50 ~ 81, 第五類為編號82 ~ 121。如果以第  $s, 1 \leq s \leq 5$ , 類的某個點為新圓心, 則這些量測座標在建模區間的分布, 不再如圖 5.3 一般均勻地分布在正方形區間內。其分布形式簡述如下: 對第  $1 \sim s - 1$  類, 每一類的量測座標會圍成一個類似橢圓的形狀; 第  $s \sim 5$  類, 每一類的量測座標會形成一個波。見圖 5.4 至 5.8。圖 5.4 是以第1類的某個點當新圓心, 所以有五個波; 圖 5.5 是以第2類的某個點當新圓心, 所以有四個波一個橢圓; 圖 5.6 是以第3類的某個點當新圓心, 所以有三個波兩個橢圓; 圖 5.7 是以第4類的某個點當新圓心, 所以有兩個波三個橢圓; 圖 5.8 是以第5類的某個點當新圓心, 所以有一個波四個橢圓。若新圓心位在距離晶圓中心 (0,0) 相同距離的位置, 比如先以某個位置為新圓心, 再逆時針依序取離 (0,0) 相同距離的點當作新圓心。此時若將極軸固定為一常數, 對這些新圓心, 量測座標在建模區間上的位置會有規則的變化。見圖 5.9 至 5.24。圖 5.9 至 5.24 是極軸為 0 且第二類的量測座標依逆時針方向依序為新圓心。由圖形可以看出, 它們的變化有點類似圖 2.1 及 2.2 那樣, 前一個圖上方的點在下一個圖時會移動至下方且不斷循環。

2 : 若兩新圓心的座標相當接近, 當其使用相同的極軸時, 它們的判定係數會差不多。

見圖 5.25,5.26。由圖 5.25,5.26 看出極軸對判定係數的圖形樣貌與值的大小很接近。

第一個特殊的狀況對尋找最佳的判定係數幫助不大, 因為量測座標移動的方式並非跟

圖 2.1 及 2.2 完全一樣, 所有量測座標作一個固定常數的位移, 這裡只是陳述會有這種有趣的狀況。第二個特殊的狀況對尋找最佳的判定係數有少許的幫助, 兩個相近的新圓心, 因為量測座標在建模區間的散佈情形會差不多, 所以即使極軸為參數, 它們的判定係數還是會很像。雖然有上述兩個特殊的狀況, 不過所給予的訊息還是不足以找到最佳的判定係數, 但是找到近似最佳的判定係數是辦的到的, 以下是找近似最佳判定係數的步驟:

- 1 : 無論工程師選了幾個位置當量測座標作實驗, 我們可在圓上均勻的取 $a$ 個點當作分析時的新圓心。例如可把圖 5.2 的那些量測座標當作新圓心。
- 2 : 每個挑選出來的新圓心, 它們所考慮的極軸皆是  $0 \sim 2\pi$  之間均勻的取 $b$ 個位置。
- 3 : 由步驟 1,2 的新圓心與極軸建立的不同之配適模型, 共有 $a \times b$ 個判定係數, 其中最大的判定係數就會接近圓上的最佳判定係數。

在第一步驟, 我們要在圓上均勻的取點當作分析時的新圓心原因是: 上述第二個特殊的狀況陳述了, 只要是相近的新圓心, 無論極軸在那個位置, 它們的判定係數差不多。所以圓上的某一區塊, 我們只要挑選一個點當新圓心, 那個新圓心就可以代表那個區塊的所有點當新圓心。在第二步驟均勻的取  $b$  個位置當作極軸的原因為: Lin (2007) 提出了固定圓心, 相鄰兩實驗點間的極軸, 判定係數的值並不會有太大的改變。而且本篇論文也發現了, 固定圓心即使把極軸當參數, 判定係數的個數會小於或等於量測座標數(由定理 1 得到)。所以為了分析時的方便性, 只要考慮均勻的選取足夠多的極軸就行了。

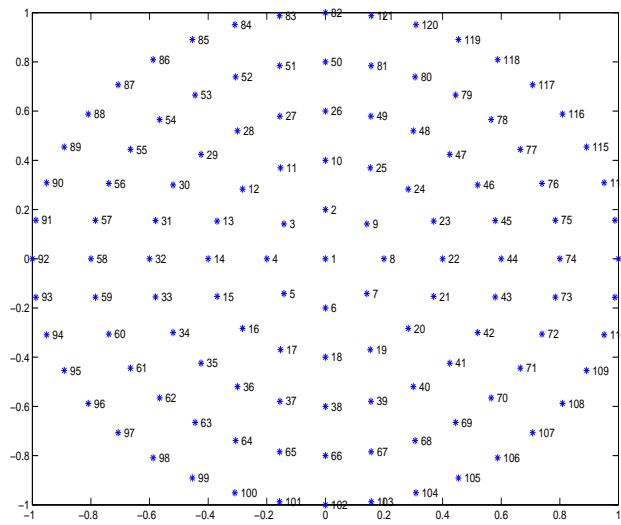


圖 5.2: 量測座標的位置與編號

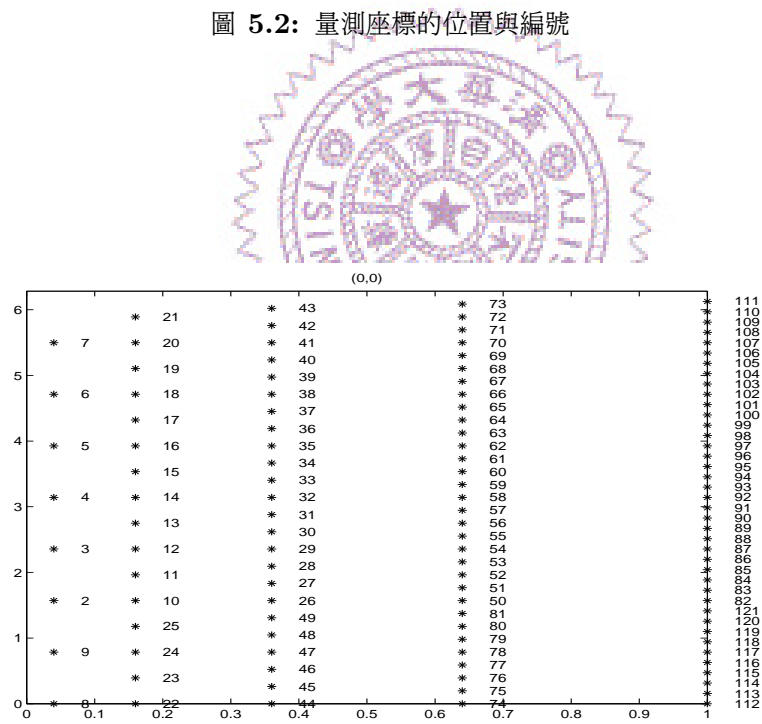


圖 5.3: 圓心為編號 1 的量測座標

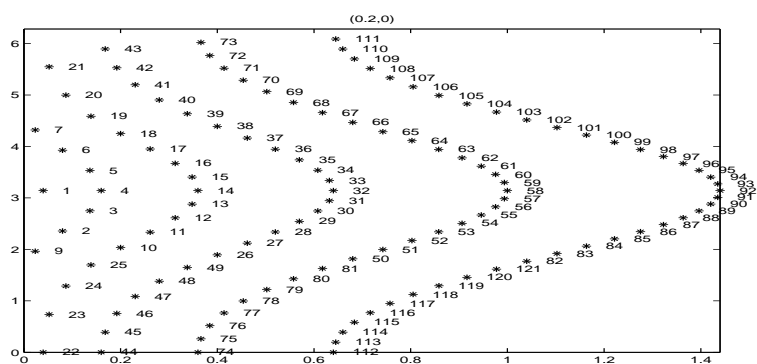


圖 5.4: 圓心為編號 8 的量測座標

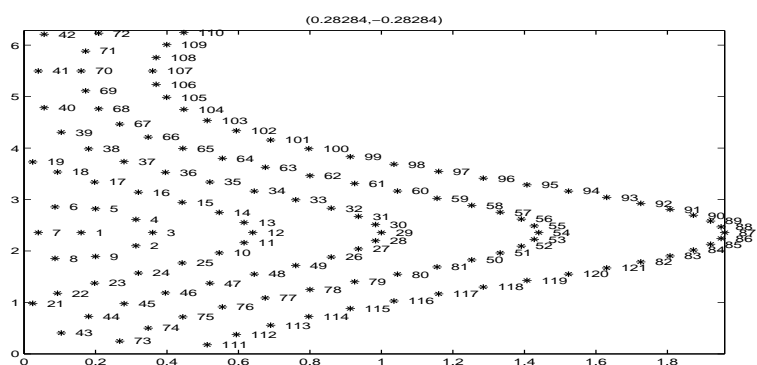


圖 5.5: 圓心為編號 20 的量測座標

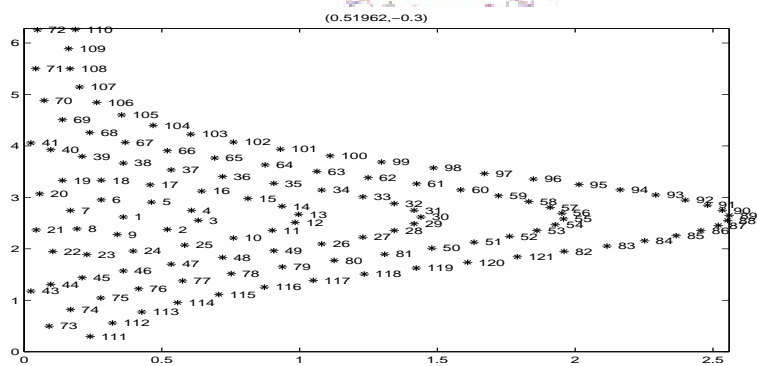


圖 5.6: 圓心為編號 42 的量測座標

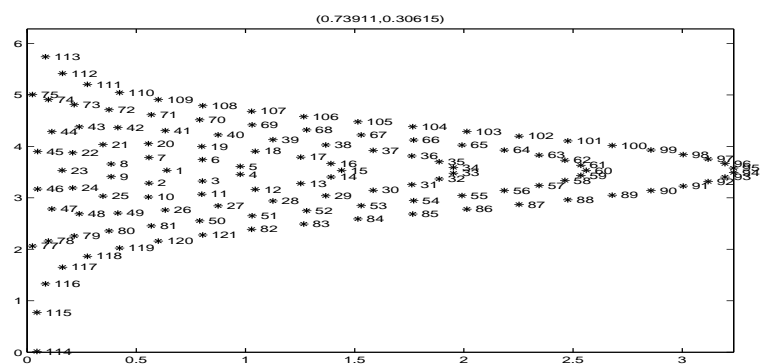


圖 5.7: 圓心為編號 76 的量測座標

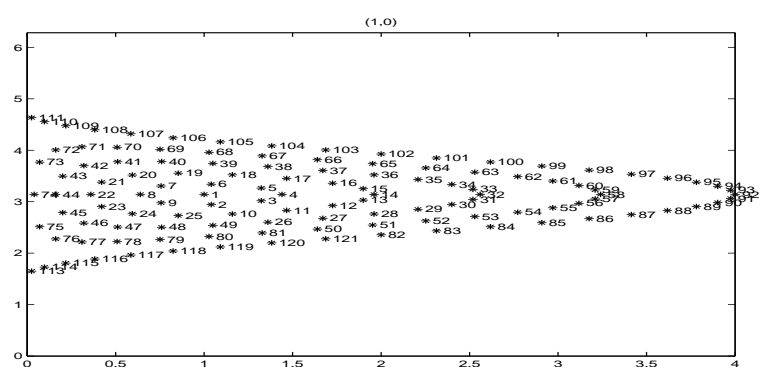


圖 5.8: 圓心為編號 112 的量測座標

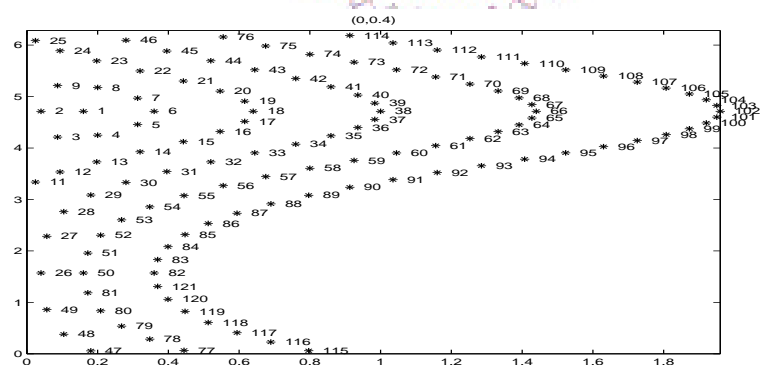


圖 5.9: 圓心為編號 10 的量測座標



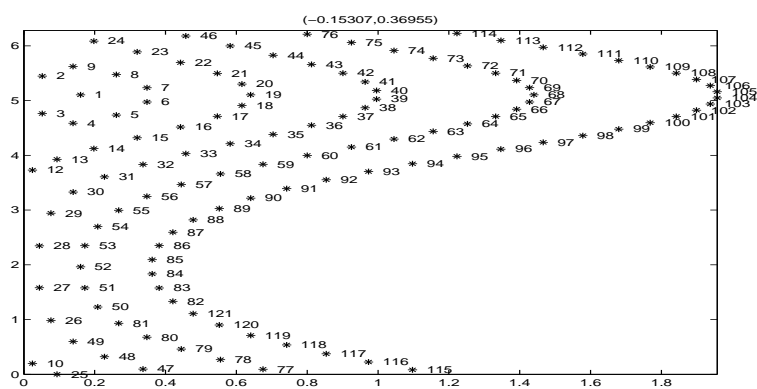


圖 5.10: 圓心為編號 11 的量測座標

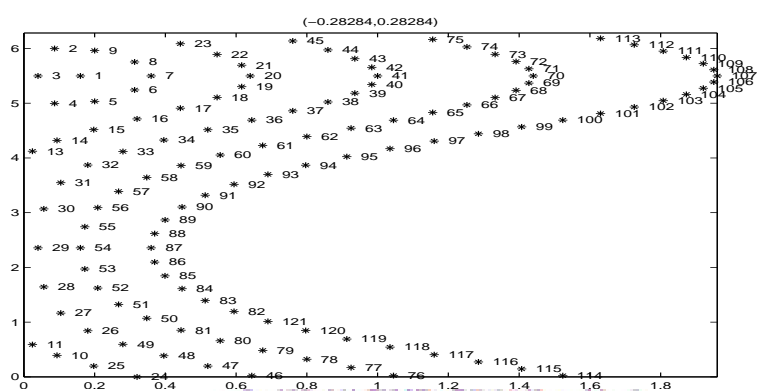


圖 5.11: 圓心為編號 12 的量測座標

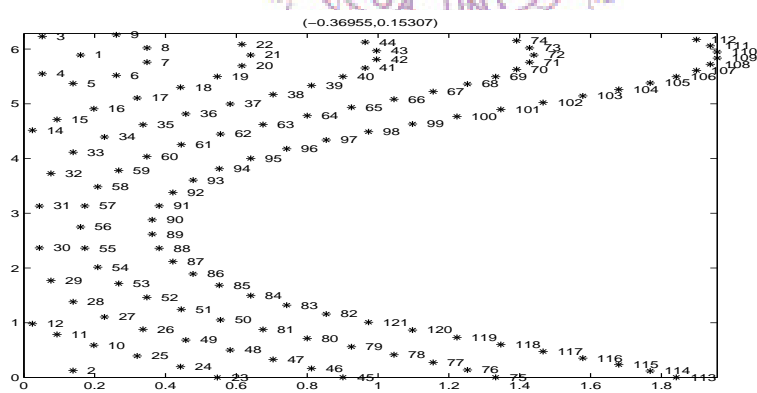


圖 5.12: 圓心為編號 13 的量測座標

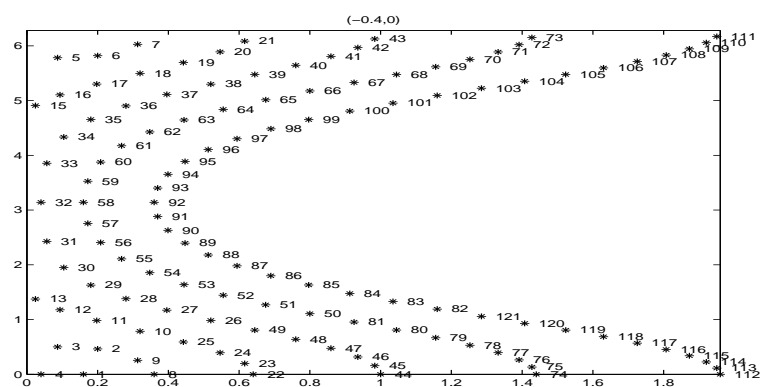


圖 5.13: 圓心為編號 14 的量測座標

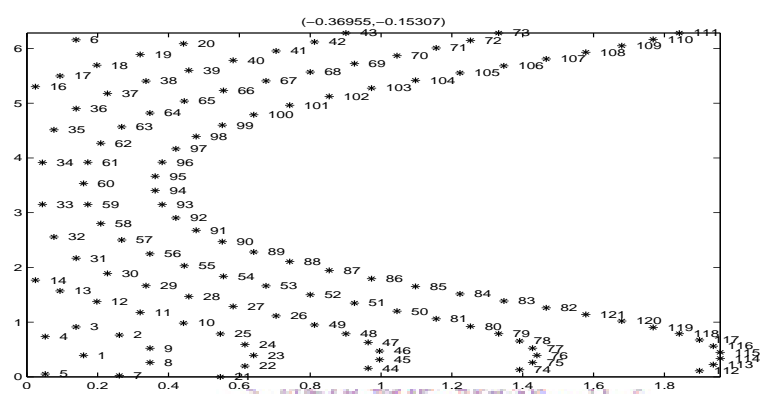


圖 5.14: 圓心為編號 15 的量測座標

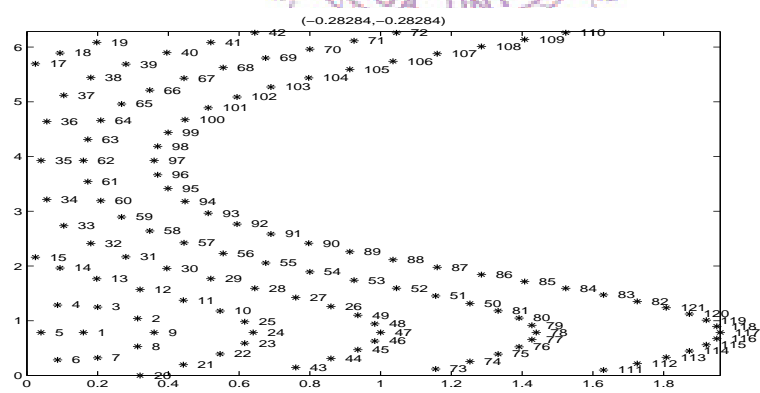


圖 5.15: 圓心為編號 16 的量測座標

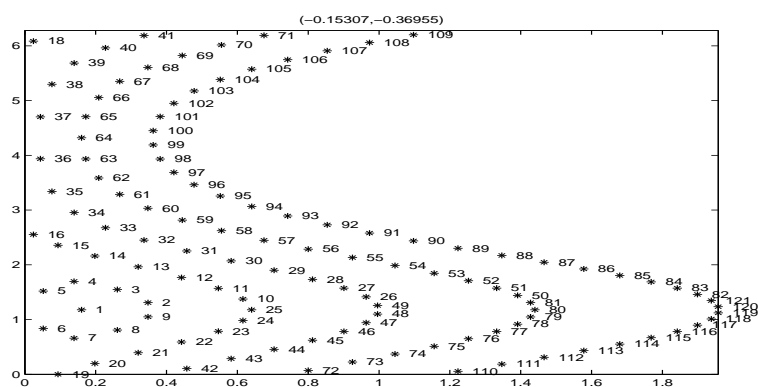


圖 5.16: 圓心為編號 17 的量測座標

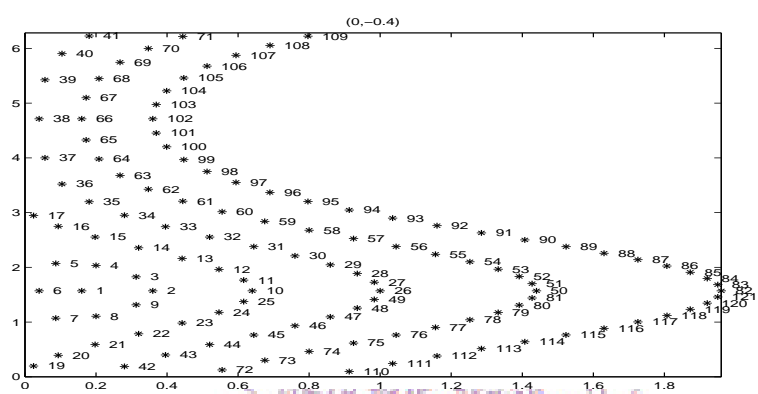


圖 5.17: 圓心為編號 18 的量測座標

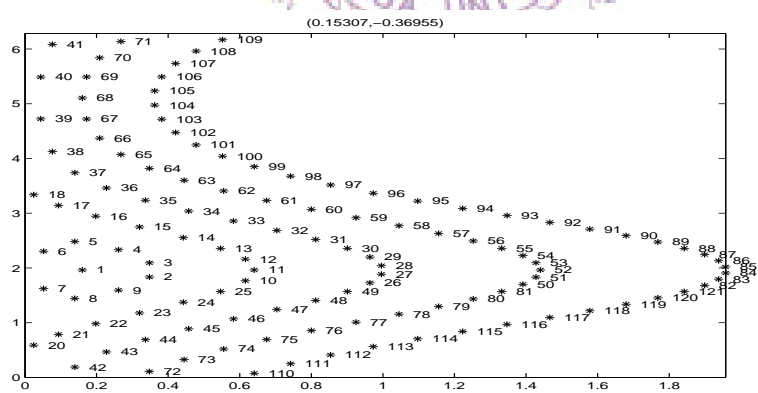


圖 5.18: 圓心為編號 19 的量測座標

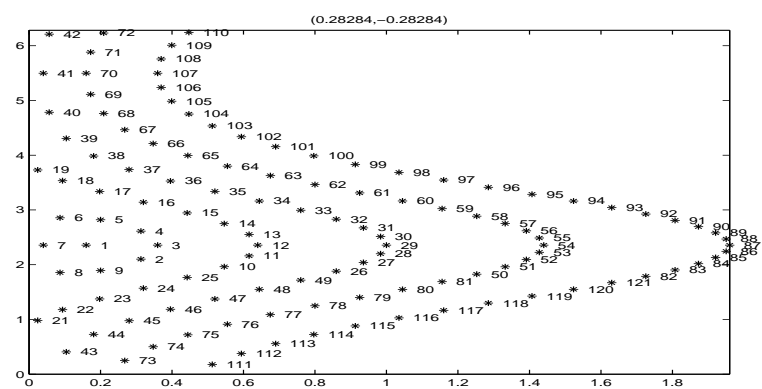


圖 5.19: 圓心為編號 20 的量測座標

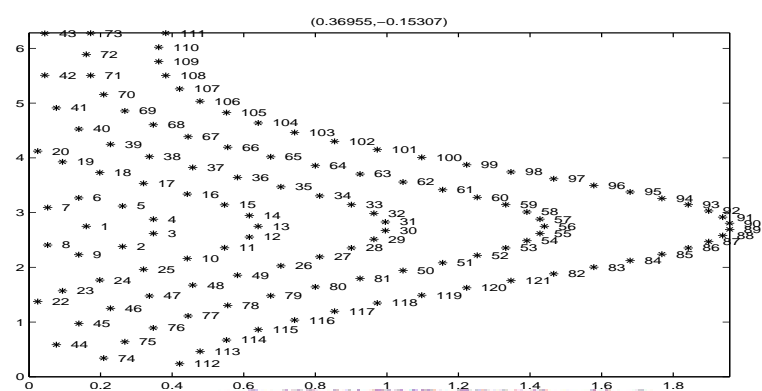


圖 5.20: 圓心為編號 21 的量測座標

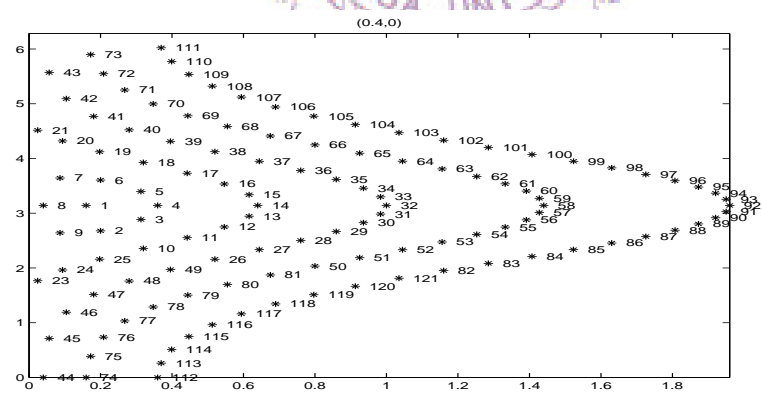


圖 5.21: 圓心為編號 22 的量測座標

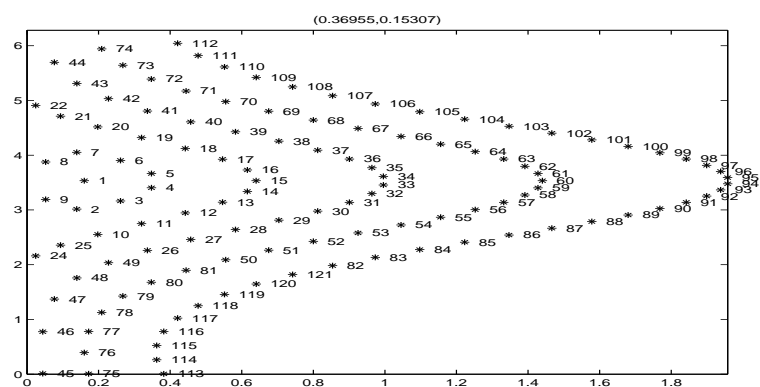


圖 5.22: 圓心為編號 23 的量測座標

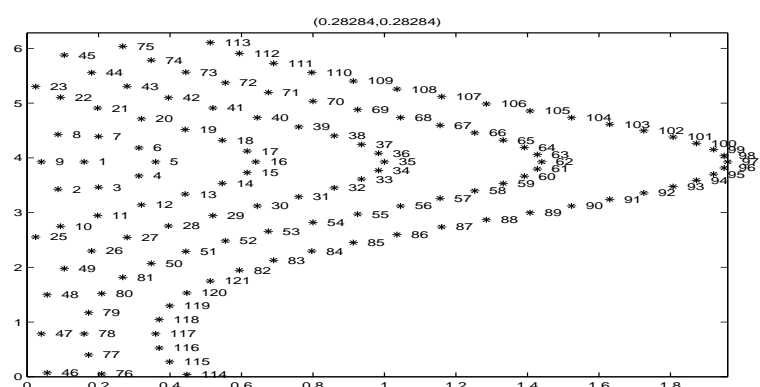


圖 5.23: 圓心為編號 24 的量測座標

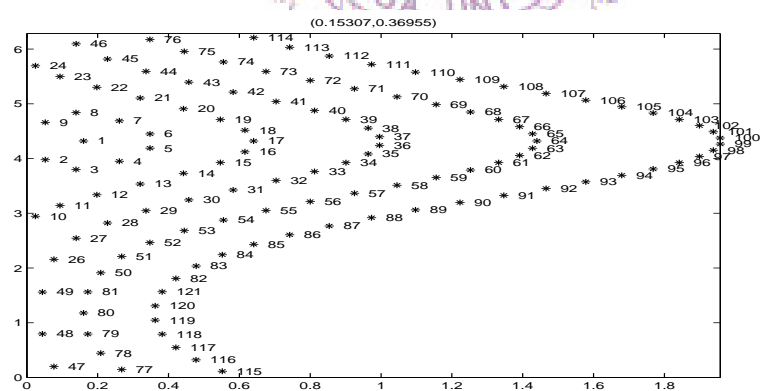


圖 5.24: 圓心為編號 25 的量測座標

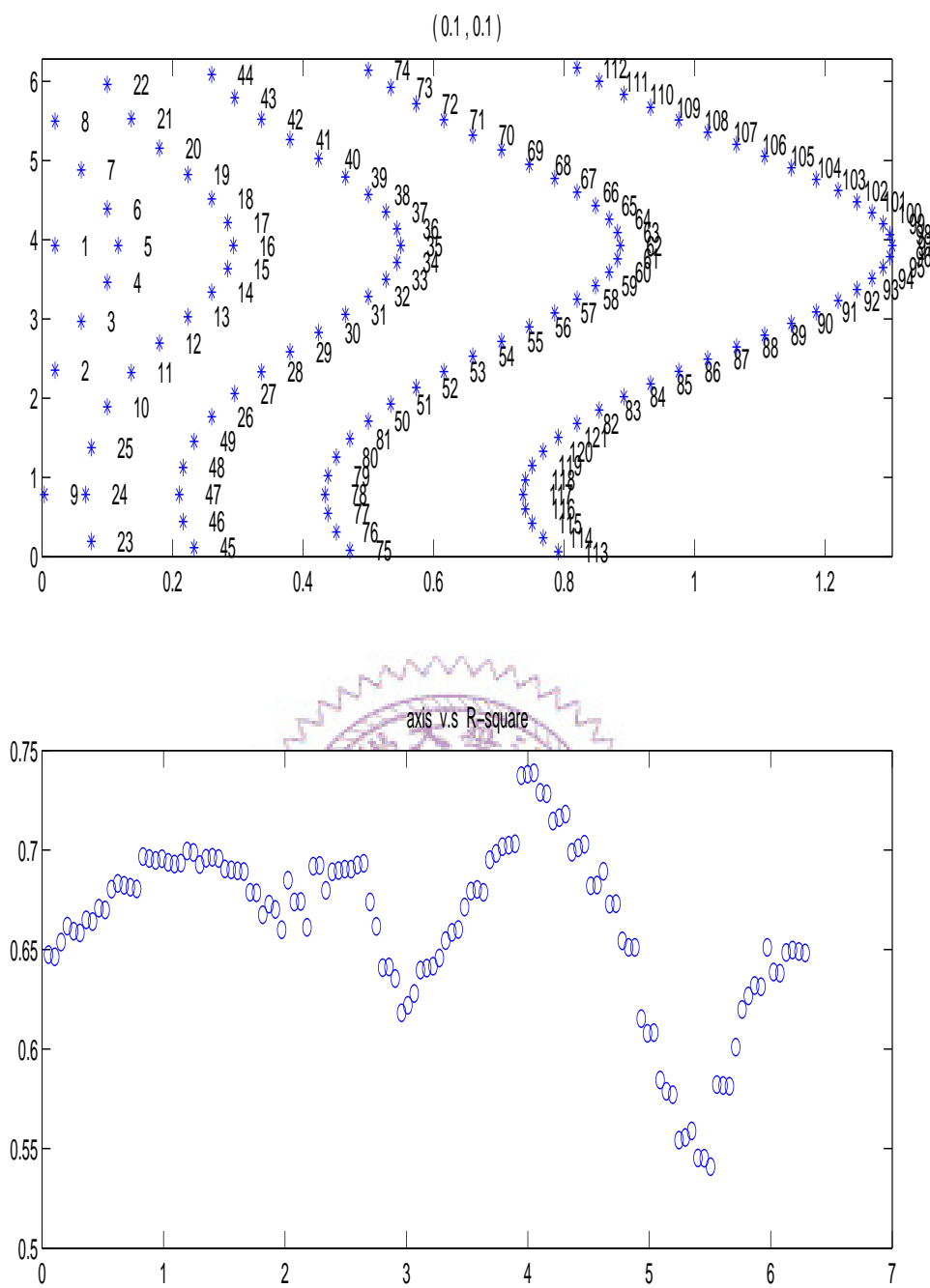


圖 5.25: 圓心為 (0.1,0.1)

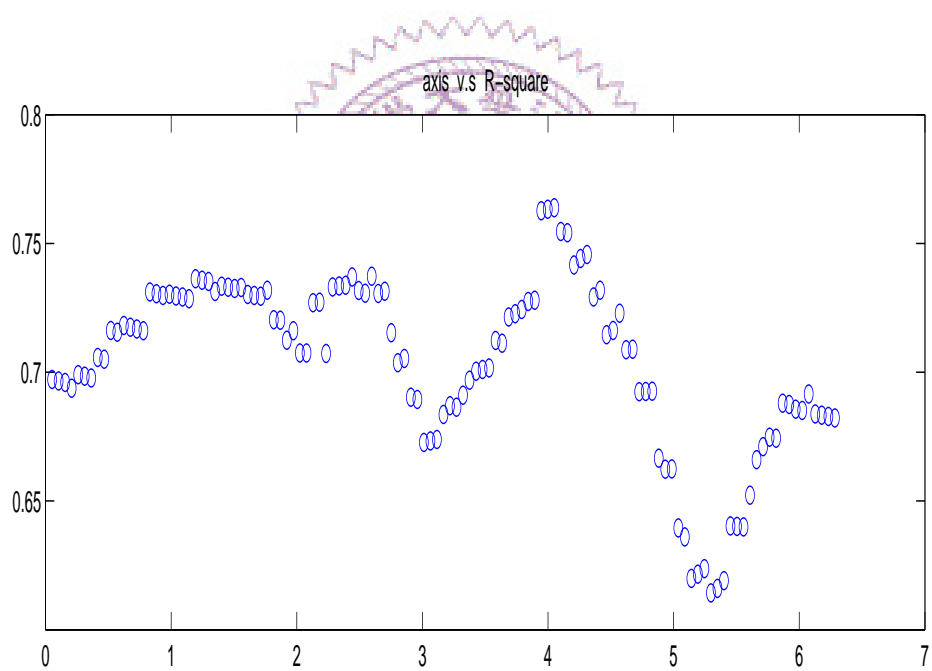
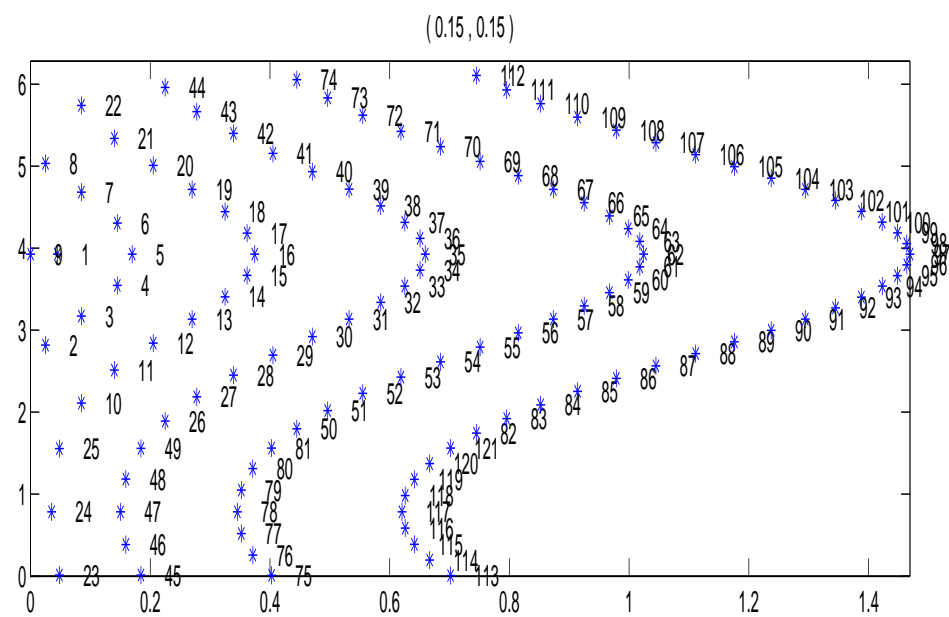


圖 5.26: 圓心為  $(0.15, 0.15)$

## 第 6 章

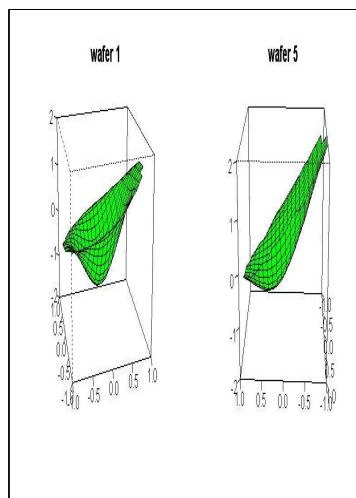
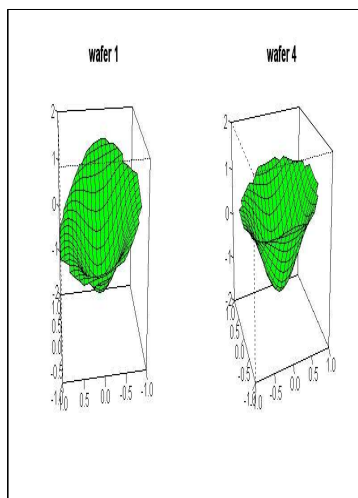
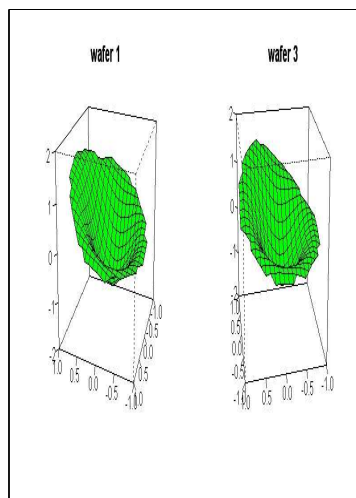
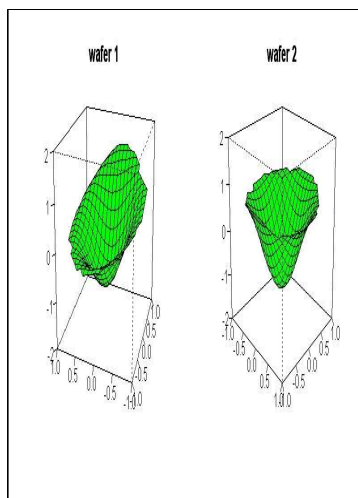
### 總結

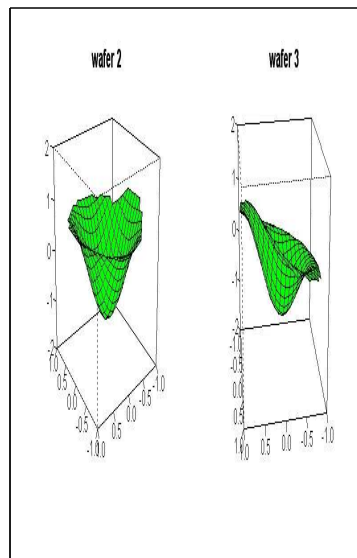
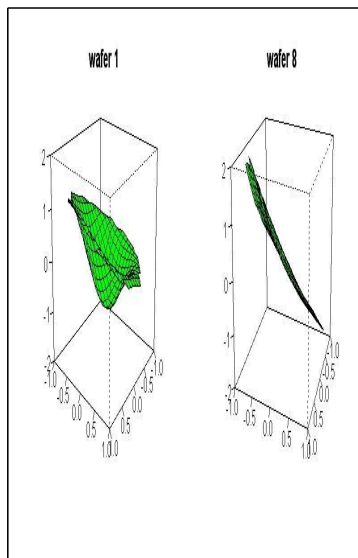
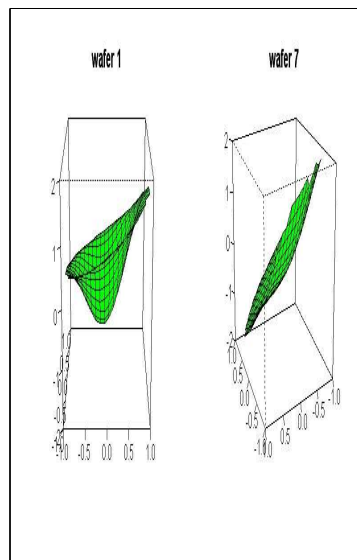
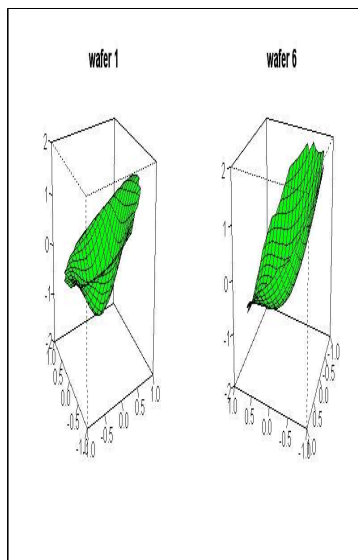
本論文主要在解決如何把圓上的曲面做群集分析，本論文介紹了，做此問題的群集分析，最重要的是要解決曲面的旋轉，在第三章我們提出了一套完整的方法來解決旋轉的問題。在第四章我們利用真實數據套用第三章提出的距離定義來分析。另外我們亦探討了如何在圓上配適一個最佳的二階模型。不同的圓心與極軸皆會造成判定係數的不同，本論文已經把極軸與判定係數的關係找出來了，並有辦法利用數值法，找到圓上最佳的判定係數。

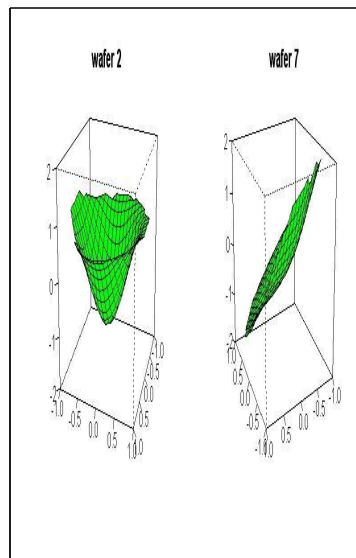
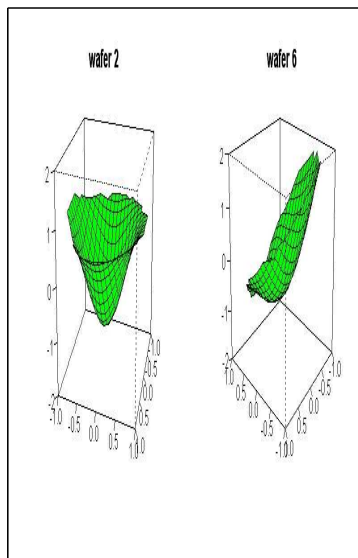
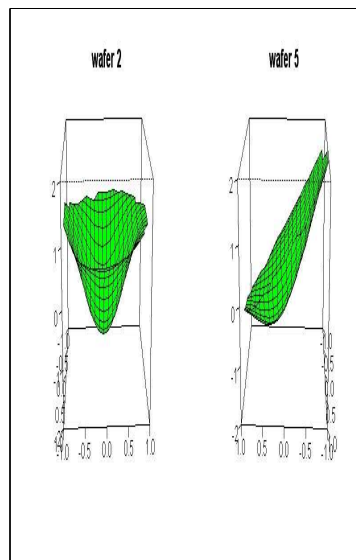
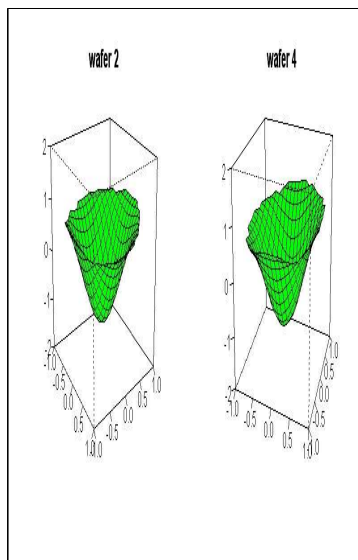


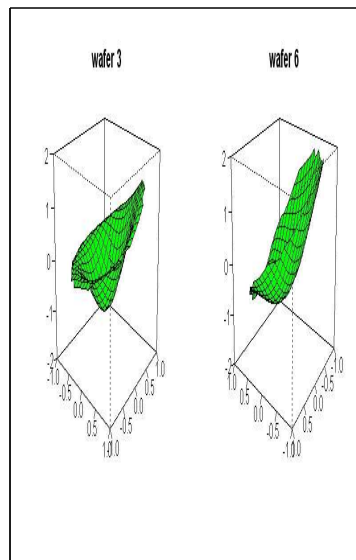
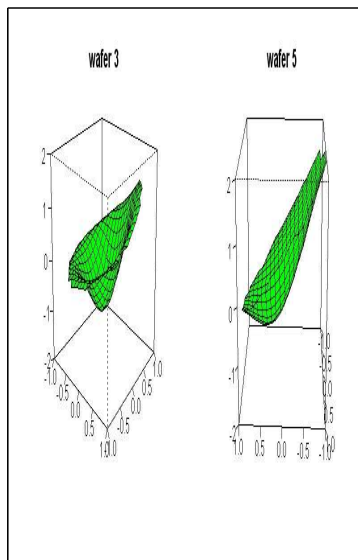
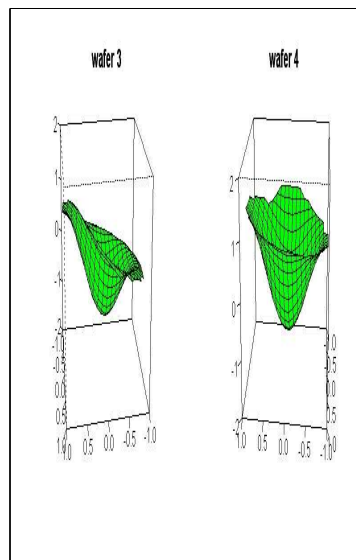
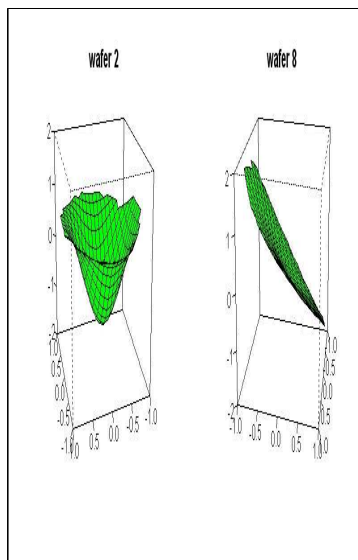
## 附錄

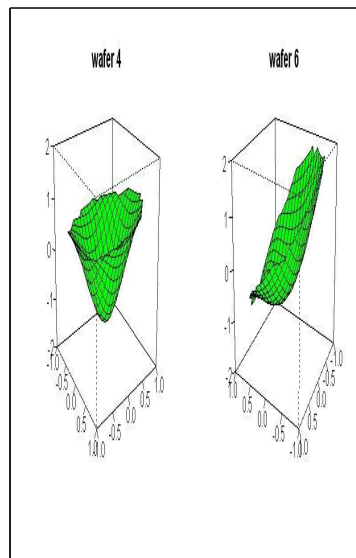
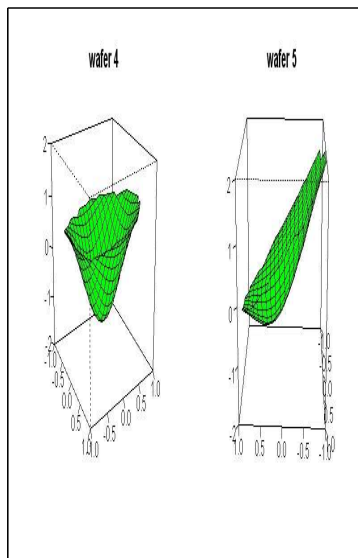
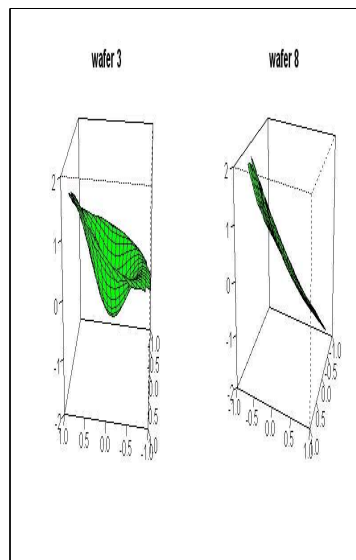
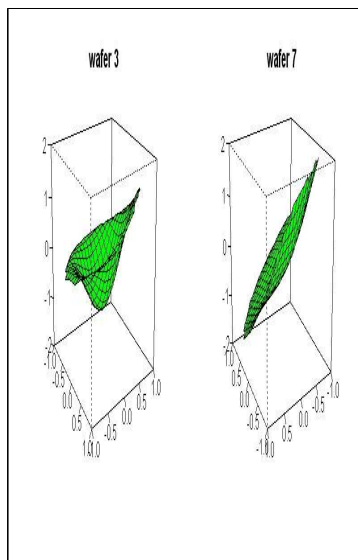
4.5 節, 把觀測數據做標準化, 旋轉至兩兩最像的曲面圖。

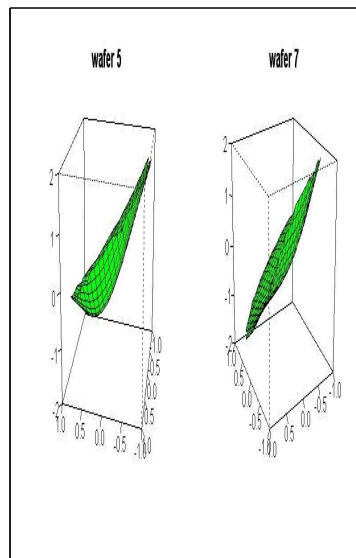
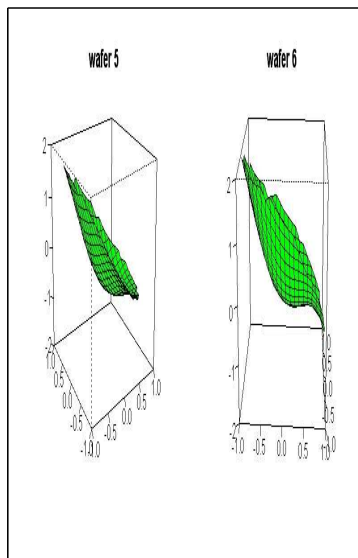
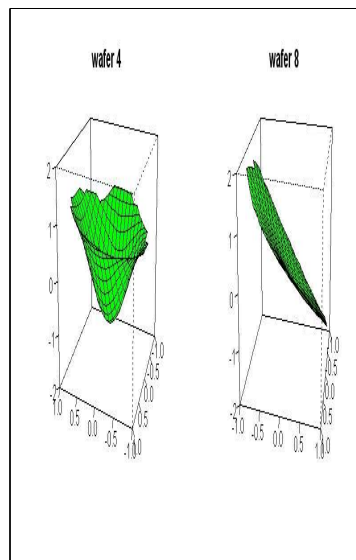
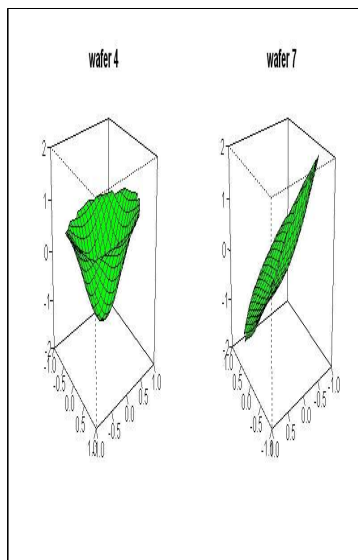


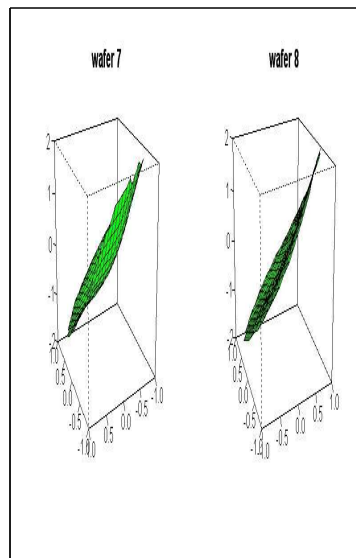
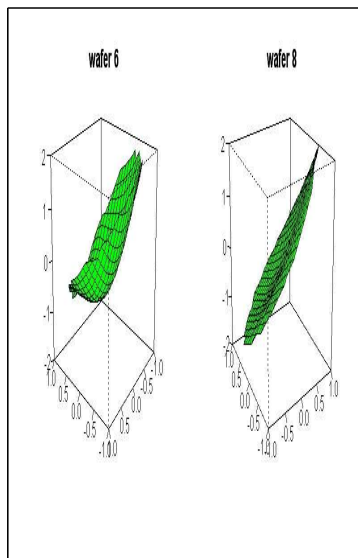
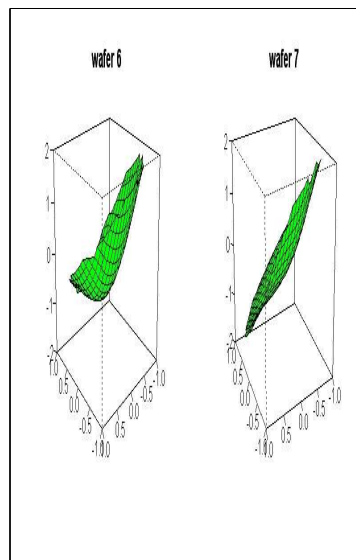
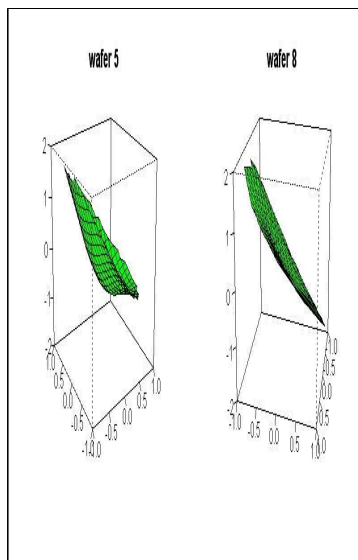












## 參考文獻

- Johnson, R.A. and Wichern, D.W. (2007), *Applied Multivariate Statistical Analysis*, Pearson.
- Lin, L.K (2007), “Design and analysis of experiments on circle”, *Master Thesis, Institute of Statistics, National Tsing Hua University, Taiwan, ROC.*
- Sharma, S. (1996), *Applied Multivariate Techniques*, Wiley, New York.

