

應用MATLAB主成分分析於北、中、南三條流域水質特性之研究

陳鈞華¹ 蔡瀚德² 李煜基³ 鄭富尚⁴ 徐貴新⁵ 甘俊二⁶

1屏東科技大學土木工程系副教授、水利技師

2、3、4屏東科技大學土木所碩士班

5東南科技大學環境管理系副教授兼系主任、水利技師

6七星農田水利研究發展基金會董事長、水利技師

摘要

由於北、中、南氣候變遷差異大，水質流域上也有明顯的不同，藉由此原因來了解台灣西部北、中、南流域，水質參數上的分類情形，應用 Matlab 主成分分析分析方法對北、中、南 3 條流域(景美溪、大甲溪、東港溪)9 個水質監測站參數進行研究。結果顯示，根據 9 個水質測點聚為 3 類，由變量(Variiances)看出，共有 5 個主成分，但前面 2 個較為顯著，佔了總方差之 90%。而同流域上，其結果並沒有分為同一類，由於降雨時空不均，導致其同流域水質有不同差異性。然而用 Matlab 語言程式設計實現，簡化以往水質分類上的分析，為分析水質流域狀況提供參考。

關鍵字:水質特性、主成分分析、人口數

一、研究動機與目的

台灣地區雖然水資源豐沛，但受地形、地貌及降雨時空分佈不均影響，其北、中、南氣候變遷差異大水質差異性也大，相對水質分析及管理顯得非常重要，藉由此原因來了解台灣西部北、中、南流域，水質參數上的分類情形，本研究對北、中、南 3 條溪流流域各取 3 個共 9 個水質監測值，分別為景美溪(1.寶橋、2.萬壽橋、3.楓林橋)、大甲溪(4.大甲溪橋、5.迎賓橋、6.東勢橋(原名東豐大橋))、東港溪(7.潮州大橋、8.隴東橋(原為龍東橋)、9.興社大橋)，利用 3 條流域 5 個參數值(DO、 BOD_5 、COD、SS、 NO_3^-N)作為主成分分析之參數。讓水質分析上留有地區性分類

比較。

主要成分分析法(Principal component analysis, PCA)是向量空間變換常用減少多維數據集到更低的維度為分析。PCA 於 1901 由卡爾 Pearson 發明，現在它主要功能作為試探性數據分析並且做為預測模型。主成分分析分為兩個主要功能，一是精簡線性資料中的變項 (Saegusa et al., 2004)，亦即可以使用最精簡的主成分數來取得最大的解釋變異量，以簡化變項間分析的過程；另一個主要功能為線性轉換，以解決迴歸變項間共線性的問題，亦即變項間有高度相關的現象存在時，經過主成分分析就可使新的變項間成為零相關，然後再進一步將各個主成分對目標應變數作迴歸分析，即可得到個別主成分對應變數

之預測能力 (Liu et al., 2003) (陳與蔡, 2008)。

近年來，應用於水文上之研究者，(郭振泰, 1998) 對高屏溪與東港溪污染量調查及整治規劃來進行高屏溪總量管制之研究；(張乃斌等, 2000) 進行我國流域污染源削減最佳化策略與管理制度之研究；(陳振華, 2000) 建立我國河川流域環境品質之評估模式；(鄭詩音, 2003)，我國水污染防治制度之比較研究—高屏流域之應用；(張揚祺、鍾世強等人, 2003)，模糊聚類演算法應用於高雄海域污染範圍之判定；(邱昱禎、張雅婷等人, 2005)，模糊系統規劃理論於水資源經營之研究；(蔡瀚德、陳鈞華等人, 2011)，模糊綜合評估理論應用於濁水河流域之多指標水質評估。

由上述幾篇文獻，對水文相關研究之對象不少，但以主成份分析進行北、中、南水質上之研究則尚未看到。因此本研究利用應用主成份分析於北、中、南3條流域探討水質特性，簡化以往水質分類上的分析，分析水質流域狀況提供一條簡單有效的途徑。

二、研究材料與方法

2.1 研究區域調查

2.1.1 地理位置

本研究地理位置為台北淡水河流域的景美溪、台中大甲河流域的大甲溪、屏東東港流域的東港溪。

景美溪為台灣北部的河流，是新店溪支流，流域面積有 113.72 平方公里，全長約 28 公里。上游發源於台灣雪山山脈北端尾稜二格山系的火燒寮山，位於新北市石碇區、平溪區、坪林區交界處。過去原住民曾稱景美溪為霧里薛溪。在台灣北部早期發展歷史中，先民可從淡水搭船，沿著新店溪一路轉進景美溪與石碇鄉鄉民進行茶、煤、染料等交易。但近年來都市開發

快速，景美溪沿岸多已築堤、靠近大台北市區的河床幾乎都渠道化，兩側山地也多蓋起大樓造成溪中土石淤積嚴重，其自然風貌已不落往常。

大甲溪是台灣中部重要的河川，屬於中央管河川，主流上游為南湖溪，其源流中央尖溪發源於南湖大山東峰(標高 3,632 公尺)，流域主要分佈於台中市，並包括南投縣、宜蘭縣之一小部分。南湖溪流至 730 林道環山檢查哨附近與發源自南投縣仁愛鄉的大支流合歡溪匯合後，續流至台七甲線 65.5 公里處(清泉橋與太保久間)與另一大支流伊卡丸溪匯流後，始稱大甲溪。流經梨山、佳陽、德基、谷關、白冷、馬鞍等聚落，流入東勢區、新社區後，逐漸進入平地，後流經石岡區、豐原區、后里區、神岡區、外埔區、清水區、大甲區及大安區，最後注入台灣海峽。

東港溪位於台灣南部，屬於中央管河川，全長約有 44.00 公里，流經屏東縣的六個鄉鎮，分別是內埔鄉、萬巒鄉、竹田鄉、潮州鎮、崁頂鄉、東港鎮。流域面積約有 472.20 平方公里。東港溪之名是由屏東縣的東港鎮而來，東港之意，原為「位於高屏溪的東邊」。過去東港溪河口是相當繁榮的貿易地與物資交換地點，清朝時就在東港溪沿岸形了三個街市，分別是萬丹街，新園街和崁頂街。東港溪主流上游為萬安溪，發源於屏東縣泰武鄉的日湯真山。

本研究探討圖 1 北、中、南 3 條溪流流域各取 3 個共 9 個水質監測值，分別為景美溪(1.寶橋、2.萬壽橋、3.楓林橋)、大甲溪(4.大甲溪橋、5.迎賓橋、6.東勢橋(原名東豐大橋))、東港溪(7.潮州大橋、8.隴東橋(原為龍東橋)、9.興社大橋)，利用 3 條流域 5 個參數值(DO、 BOD_5 、COD、SS、 NO_3^-N)作為主成分分析之參數。(資料來源:維基百科)

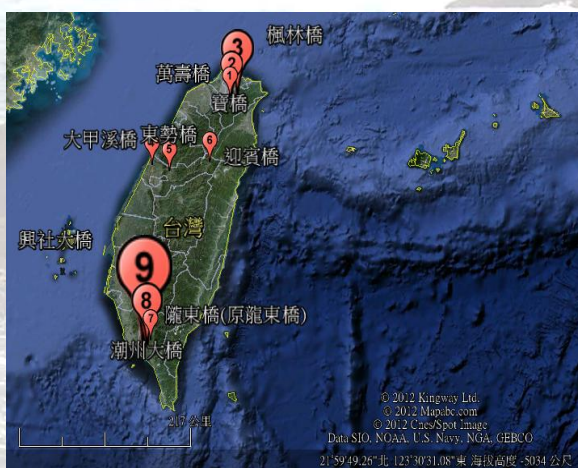


圖 1 地理位置圖

(資料來源 Google 地球)

2.2 水質觀測數據

本研究所利用水質參數監測值，資料來自行政院環境保護署-全國環境水質監測資訊網，利用表 2 中的水質監測數據，採用 MATLAB 主成分分析來了解統計分析相關訊息。

2.3 主成分分析

主成分分析把多個指標化為少數幾個綜合指標的一種統計分析方法。

在多變量的研究中，往往由於變量個數太多，並且彼此之間存在一定的相關性，使得所觀察的數據在一定程度上反應的信息有所重疊。利用主成分分析則可以將這些問題化簡，即通過降維，找到幾個綜合因子來代表原來眾多的變量，使這些綜合因子能盡量反映原來變量的信息，而且彼此之間不相關(許國根與許萍萍，2005)。

2.3.1 計算相關係數矩陣

$$R = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pp} \end{bmatrix} \quad (1)$$

r_{ij} ($i, j=1, 2, \dots, p$) 為原變數的 x_i 與 x_j 之間的相關係數，其計算公式為

$$r_{ij} = \frac{\sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ki} - \bar{X}_i)^2 (X_{kj} - \bar{X}_j)^2}} \quad (2)$$

因為 R 是實對稱矩陣 (即 $r_{ij}=r_{ji}$)，所以只需計算上三角元素或下三角元素即可。

2.3.2 計算特徵值與特徵向量

首先解特徵方程 $|\lambda I - R| = 0$ ，通常用雅可比法 (Jacobi) 求出特徵值 λ_i ($i=1, 2, \dots, p$)，並使其按大小順序排列，即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ；然後分別求出對應於特徵值 λ_i 的特徵向量 e_i ($i=1, 2, \dots, p$)。這裡要求 $\|e_i\|=1$ ，即

$$\sum_{j=1}^p e_{ij}^2 = 1, \text{ 其中 } e_{ij} \text{ 表示向量 } e_i \text{ 的第 } j \text{ 個分量。}$$

2.3.3 計算主成分貢獻率及累計貢獻率

主成分 z_i 的貢獻率為

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (i=1, 2, \dots, p) \quad (3)$$

累計貢獻率為

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} \quad (i=1, 2, \dots, p) \quad (4)$$

一般取累計貢獻率達 85~95% 的特徵值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 所對應的第一、第二、...、第 m ($m \leq p$) 個主成分。

2.3.4 計算主成分載荷

其計算公式為

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} \quad (i, j=1, 2, \dots, p)$$

得到各主成分的載荷後，還可以進一步計算，得到各主成分的得分

$$Z = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (5)$$

2.4 地區人口調查數

本研究為了解析水質是否優良，將地區人口數列表來觀察水質品質。
(資料來源:各縣市政府全球資訊網)

表1 北、中、南3條溪流
流域人口調查

溪流名稱	水質測站	地區名稱	人口數(人)	河流的流路
景美溪	1.寶橋	新店區	296513	下
	2.萬壽橋	文山區	267325	中
	3.楓林橋	石碇區	7890	上
大甲溪	4.大甲溪橋	大甲區	78178	下
	5.迎賓橋	和平區	10628	上
	6.東勢橋	東勢區	52684	中
東港溪	7.潮州大橋	潮州鎮	55267	中
	8.隴東橋	萬巒鄉	21564	上
	9.興社大橋	萬丹鄉	53437	下

三、案例應用

3.1 分析函數

由全國環境水質監測資訊網取北、中、南3條溪流流域各取3個共9個水質監測值，分別為景美溪(1.寶橋、2.萬壽橋、3.楓林橋)、大甲溪(4.大甲溪橋、5.迎賓橋、6.東勢橋(原名東豐大橋))、東港溪(7.潮州大橋、8.隴東橋(原為龍東橋)、9.興社大橋)，2007到2012年2月3條流域5個水質參數值(DO、 BOD_5 、COD、SS、 NO_3^-N)，並取平均值，試根據表2數據作主成分分析。

表2 北、中、南3條溪流
流域水質參數監測值

參數 地區	DO(mg/L)	BOD_5 (mg/L)	COD(mg/L)	SS(mg/L)	NO_3^-N (mg/L)
淡水河流域之景美溪					
1.寶橋	6.94	5.84	16.58	17.32	1.61
2.萬壽橋	7.18	2.24	8.86	10.24	0.836
3.楓林橋	10.24	0.98	4.88	3.18	0.11
大甲河流域之大甲溪					
4.大甲溪橋	9.96	3.06	10.94	135.18	0.172
5.迎賓橋	9.66	2.76	9.7	1.28	0.038
6.東勢橋	9.6	1	3.9	11.22	0.034
東港河流域之東港溪					
7.潮州大橋	6.08	1.46	6.86	22.74	1.34
8.隴東橋	7.14	1.14	6.04	9.56	1.75
9.興社大橋	2.66	9.5	36.46	51.96	7.256

3.2 主成分分析模式建立

主成分分析程式碼:

```
x=[6.94 5.84 16.58 17.32 1.61 ;
    7.18 2.24 8.86 10.24 0.836 ;
    10.24 0.98 4.88 3.18 0.11 ;
    9.96 3.06 10.94 135.18 0.172 ;
    9.66 2.76 9.7 1.28 0.038 ;
    9.6 1 3.9 11.22 0.034 ;
    6.08 1.46 6.86 22.74 1.34 ;
    7.14 1.14 6.04 9.56 1.75 ;
    2.66 9.5 36.46 51.96 7.256];
```

```
stdr=std(x);
jo=x./stdr(ones(9,1),:); %
數據標準化
[pcs,newdata,variances,t2]=
princomp(jo); %主成分分析
```

表3 主成分分析函數

函數	功能	適用格式
barttest	進行 Barttest 維數檢驗	ndim=barttest(x,alpha) 在 alpha 水平下檢驗數據矩陣 X 的非隨機變化特征。Ndlim=1 的檢驗室檢驗與每個因子的方差是否相等。ndim=2 是檢驗與第二個因子至最後一個因子的方差是否相等，以此類推。 [ndim, prob, chisquare]= barttest(x,alpha)
pcacov	使用協方差矩陣進行主成分分析	pc=pcacov(x) [pc,latent,explained]= pcacov(x) 返回主要因子 pc, 協方差矩陣 X 的特征值 latent, 觀察量中由每一個特征向量所解釋的總方差的百分比 explained。
pcares	主成分分析殘差	residuals=pcares(x,ndim) 通過保留 X 的 ndim 個因子成分來獲得殘差。
princomp	主成分分析	pc=princomp(x)[pc,score,latent,tsquare]=princomp(x) 根據矩陣 x 返回因子成分 pc, z 分數 score, x 的協方差矩陣的特征值 latent 和 Hotelling's T^2 統計量 tsquarel

四、結果與討論

4.1 主成分分析結果

```

pcs =
    %主成分
    0.4630 -0.3030 0.6404
    0.5224 0.1038
    -0.4966 -0.0961 0.5600
    -0.4048 0.5164
    -0.5140 -0.0736 0.3444
    0.1435 -0.7689
    -0.1337 -0.9351 -0.3258
    -0.0299 0.0274
    -0.5069 0.1386 -0.2269
    0.7361 0.3613
newdata =
    -0.8743 0.2374 0.5816
    -0.4264 0.0973
    0.3874 0.4944 -0.1970
    -0.2178 -0.0840
    1.5710 0.2994 0.3438
    0.3268 0.0009
    0.4171 -2.6801 -0.1306
    -0.0153 -0.0011
    0.9270 0.3132 0.7292
    -0.0031 -0.0810
    1.4884 0.2035 0.0932
    0.1436 0.0454
    0.2667 0.4268 -0.8518
    -0.2156 -0.0316
    0.5147 0.6275 -0.6060
    0.1854 0.0741
    -4.6980 0.0780 0.0375
    0.2223 -0.0201
variances =
    %
    方差
    3.6299
    1.0369
    0.2684
    0.0609
    0.0040
t2 =
    6.8735
    2.9622
    2.9610
    7.0432
    3.9503
    1.5360
    3.9117
    3.7583
    7.0038

```

```
plot(newdata(:,1),newdata(:,2),'+')
```

以圖 2 所示

gname %獲取各點代表的樣本

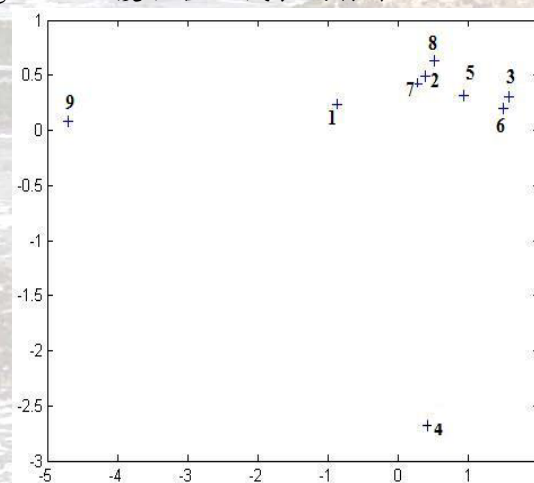


圖 2 主成分得分圖

- (1)從主成分分析結果中，第一排 newdata 數據，可找出圖 2 代表的各點點位，由這些點來探討各點位差異性，且從變量 (Variances) 結果可看出，共有 5 個主成分，但前面 2 個主成分較為顯著，佔了總方差之 90%。
- (2)從北、中、南 9 個水質監測站，樣本聚為 3 類；分別為 4.大甲溪橋、9.興社大橋各聚為一類； 2.萬壽橋、3.楓林橋、5.迎賓橋、6.東勢橋、7.潮州大橋、8.隴東橋為一類，由此分類可看出 1、4、9 號水質較為不同，由於河流的流路都於下游路段 SS 參數值都有稍為偏高狀態，其造成分析上與其他參數有差異大小之區分。
- (3)由景美溪、大甲溪與東港河流域為探討對象，其最後分析結果以大甲溪為例，雖大甲溪橋、迎賓橋與東勢橋都在同一流域上，但結果並沒有將 3 個水質監測站分為同一類，由於水質上的變化，同一流域可能因降雨時空不同或周遭環境所受影響，導致其流域於上、中、下游都有不同差異性。

(4)由人口數觀測水質程度好壞，上游水質由於人口居住數較少水質較為乾淨，以景美溪楓林橋、大甲溪迎賓橋、東港溪隴東橋水質比較，發現人口數越少水質必定乾淨，因大甲溪迎賓橋和平區人口約1萬人，水質確比景美溪楓林橋石碇區8千人乾淨，由於大甲溪發源地標高較高、流域長度較長，有沖淡淨化水質之功能，因此判斷水質好壞程度，須參考溪流的發源地標高、流域長度及上游產業。

五、結論與建議

- 1.本研究從北、中、南3條流域(景美溪、大甲溪、東港溪)9個水質監測站(1.寶橋、2.萬壽橋、3.楓林橋、4.大甲溪橋、5.迎賓橋、6.東勢橋、7.潮州大橋、8.隴東橋、9.興社大橋)，進行主成分分析，結果顯示從變量(Variiances)可看出，共有5個主成分，但前面2個主成分較為顯著，佔了總方差之90%。
- 2.本研究從北、中、南9個水質監測站，樣本聚為3類，由此分類可看出編號1、4、9水質參數較為不同，其位於下游路段SS參數值稍為偏高，造成分析上與其他參數有差異大小之區別。
- 3.本研究以景美溪、大甲溪與東港溪流域為探討對象，其分析結果以大甲溪為例，雖然都於同一流域上，其結果並沒有分為同一類，由於同流域可能降雨時空不同或周遭環境所受影響，導致其流域上、中、下游水質有不同差異性。
- 4.本研究以人口數觀測上游水質好壞，結果顯示人口數越少水質必定乾淨，須參考溪流的發源地標高、流域長度及上游

產業，才能有效的判斷水質好壞程度。

- 5.本研究利用景美溪、大甲溪與東港溪之水質參數進行主成分分析，尚未探討東部地區之流域水質參數，此相關課題還需進一步探討及研究。

參考文獻

- 1.Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine* 2(6):559-572.
- 2.Liu, R. X., J. Kuang, Q. Gong and X. L.Hou .(2003). "Principal Component Regression Analysis with SPSS", *Computer Methods and Programs in Biomedicine*, Vol.71, pp.141-147.
- 3.Saegusa, R., H. Sakano and S. Hashimoto.(2004). "Nonlinear Principal Component Analysis to Preserve the Order of Principal Components", *Neurocomputing*, Vol.61,p.p.57-70.
- 4.陳鴻烈、蔡大偉，2008，以主成分分析法探討水庫優養化之動力研究，*水土保持學報* 40(2)，第137~162頁。
- 5.蔡瀚德、李煜基、陳鈞華、宋建明、梁惠儀，2011，模糊綜合評估理論應用於濁水河流域之多指標水質評估，*國立嘉義大學第20屆水利研討會*。
- 6.郭振泰，1998，高屏溪、東港溪污染量調查及整治規劃之高屏溪總量管制，*環保署委託計畫，臺灣大學土木工程所*。
- 7.張乃斌，2000，*事業廢水管理制度研究—流域污染源削減最佳化策略與管理制度之研究*，*環保署委託計畫，計畫編號：EPA-89-U1G1-03-117*，成大研究發展基

- 金會。
- 8.李嗣瑞，2000，從環境生態的觀點探討河川流域管理機關之組織與功能—以高屏溪流域為例，國立成功大學都市計劃研究所碩士論文。
 - 9.陳振華，2000，河川綜合環境品質評估模式之建立與應用—以高屏溪為例，國立東華大學自然資源管理研究所碩士論文。
 - 10.葛健群，2002，高屏溪水生生態風險評估，國立高雄師範大學生物科學研究所碩士論文。
 - 11.鄭詩音，2003，我國水污染防治制度之比較研究—高屏溪流域之應用，國立中山大學經濟學碩士論文。
 - 12.邱昱禎、張雅婷、張斐章，2005，模糊系統規劃理論於水資源經營之研究，台灣水利 53(1)，第 11~22 頁。
 - 13.張揚祺、張敦程、鍾世強，2003，模糊聚類演算法應用於高雄海域污染範圍之判定，國立台灣海洋大學第 25 屆海洋工程研討會論文集。
 - 14.許國根、許萍萍，2005，MATLAB 在化學中應用，西安交通大學出版社。
 - 15.徐建華，2005，計量地理學，高等教育出版社。

收稿：101 年 4 月 30 日
修改：101 年 5 月 14 日
接受：101 年 5 月 22 日