農業工程學報　第 59 卷第 2 期
中華民國 102 年 6 月出版

Journal of Taiwan Agricultural Engineering
Vol. 59, No. 2, June 2013

應用多變量統計分析評估臺北水源特定區集水區水質

Watershed Water Quality Assessment in the Taipei Water Management District Using Multivariate Statistical Techniques

國立聯合大學土木與防災工程學系
碩士
吳 政 南
Cheng-Nan Wu

國立聯合大學土木與防災工程學系
教授
柳 文 成*
Wen-Cheng Liu

摘　要

台北水源特定區供應大台北地區近 500 萬人之自來水水源，係國內第一個經由都市計畫法設立之水源、水質與水量保護區，區域範圍涵蓋三個流域，包含新店溪流域及其上游之南勢溪流域及北勢溪流域。本研究以水源特定區管理局轄下的監測站之水質監測數據，配合氣象局雨量監測數據，合計 12 項物理與化學水質變數，包含溶氧量、生化需氧量、懸浮固體物、化學需氧量、pH 值、比導電度、水溫、大腸菌類數、濁度、總磷、氨氮及雨量等。分析 2008 年到 2010 年每月的流域測站之水質資料，以量化方式探討台北水源特定區集水區複雜的水質資料結構，並應用多變量統計分析探討水質的特性。

本研究利用判別分析可清楚判識三流域的特性有極大的差異，對照現況，可以驗證三個流域的水質變數各有其特性。利用主成份分析可得到台北水源特定區所包含的三個流域的水質主成份。北勢溪流域可萃取 5 個主成份，依序爲懸浮固體成份、營養鹽成份、pH 值成份、雨量成份及其他成份，可解釋的總變異量爲 67.260%；南勢溪流域也萃取出 5 個主成份，依序爲懸浮固體成份、人類活動成份、總磷成份、雨量成份及氨氮成份，可解釋的總變異量爲 66.624%；新店溪流域亦萃取出 5 個主成份，依序爲懸浮固體成份、總磷成份、雨量成份、水溫成份及比導電度成份，可解釋的總變異量爲 73.024%。依照分析結果並對照三個流域現況，北勢溪流域爲農業耕種盛行，但是污染負荷較低；南勢溪流域爲溫泉及森林遊樂區，遊憩觀光盛行，人類活動造成的污染負荷影響水體甚多；新店溪流域爲住宅區，人口稠密，雖有污

*通訊作者，國立聯合大學土木與防災工程學系教授，36003 苗栗市恭敬里聯大 1 號，E-mail: wcliu@nuu.edu.tw

水下水道處理，但是家庭廢水所造成的污染負荷也最爲嚴重。利用群集分析進行測站分群，可將北勢溪流域測站分成 4 群、南勢溪流域測站分成 3 群、新店溪流域測站分成 3 群。如果在經費考量或是希望能到更好的監測效果的前提下，可以嘗試增加臨時性的測站，或是對於差異性很小的測站進行位置的改變，以評估比較水質監測的效果。

關鍵詞：臺北水源特定區，水質，多變量統計分析，判別分析，主成分分析，群集分析。

ABSTRACT

The Taipei Water Management District, supplying the domestic water for 500 million people in Taipei city, is the first protection area of water sources, water quality and quantity according to the Urban Planning Law. It covers the three basins including the Pei-Shih Stream River, Nan-Shih Stream, and Hsin-Dian Stream. In the presennt study, the water quality monitoring data collected from the Water Management Office Stations and rainfall data collected from the Central Weather Bureau were adopetd and analyzed. A total of 12 physical and chemical water quality variables, including dissolved oxygen (DO), biochemical oxygen demand (BOD), suspended solids (SS), chemical oxygen demand (COD), pH, conductivity, water temperature, coliform fecal, turbidity, total phosphorus (TP), ammonia nitrogen (NH_3-N), and rainfall. The quantitative approaches were applied to analyze the long-term monitoring data from the water sampling stations located in the watershed during years 2008-2010. Also the multivariate analysis techniques were used to analyze the water quality characteristics.

The three basins were identified clearly by discriminant analysis (DA), indicating that the water quality variables in the three basins have their own characteristics. The analyzed results concurred with the actual situation of the three basins. By using the principal component analysis (PCA), principal components of three basins in the Taipei Water Management District are extracted. The Pei-Shih Stream basin extracted five principal components and the percentage of explained variance is 67.26%. The first principal component is suspended solids concentration, subsequently are nutrients, pH, rainfall components, and other components. The Nan-Shih Stream basin extracted five principal components and the percentage of explained variance is 66.624%. The first principal component is suspended solids concentration, subsequently are human activities, total phosphorus, rainfall components, and ammonia nitrogen components. The Hsin-Dian Stream basin extracted five principal components and the percentage of explained variance is 73.024%. The first principal component is suspended solids concentration, subsequently are total phosphorus, rainfall, water temperature components, and conductivity components. Based on the results above the current situation of the three basins, agriculture cultivation is more prevalent in the Pei-Shih Stream basin but the pollution load is the lower. There are hot spring and recreation areas in the Nan-Shih Stream basin, and pollution load of human activities has a great impact on the water quality. In the Hsin-Dian Stream basin, there are lots of residential areas, but pollution load of domestic wastewater are the most serious. Cluster analysis (CA) was used to cluster the measured stations. The

Pei-Shih Stream basin, Nan-Shih Stream basin, and Hsin-Dian Stream basin yield four clusters, three clusters, and three clusters, respectively. For financial or monitoring quality considerations, we may try to increase the temporarily measured stations or to change the measured locations for gathering similar monitoring data and then assess the effect of water quality monitoring.

Keywords: Taipei Water Management District, Water quality, Multivariate statistical techniques, Discriminant analysis, Principal components analysis, Cluster analysis.

一、前　言

1.1 研究背景

水資源對人類一直以來都是非常重要的資源，尤其是湖泊及河川都對人類活動具有多種的使用用途，像是飲用、灌溉、漁業及能源生產等(Iscen *et al*., 2008)，因此河川水質的優劣及多寡，對於人類生活有極大的影響。

在地狹人稠的臺灣，水資源的維護更加不易。在水資源的蓄存方面，台灣山脈走勢及氣候條件，造成水資源的蓄存不易；在水體品質方面，面對經濟快速發展下，人民的生活逐漸富裕，人口快速增加，然而在地狹的先天條件下，山坡地開始被大量的開發，而開發過程中森林的砍伐以及農業污染往往是造成水庫優養化的主要來源之一(Smith *et al*., 2005; Antikainen *et al*., 2005)，其影響範圍廣大且深遠。

台北水源特定區內有翡翠水庫，提供大台北地區的主要用水來源。台北水源特定區管理局在台北水源特定區中有設立水質採樣站，以監測水質變化。然而從原始數據中，並不易直接觀察出眞正的變化，自然及人爲的因素對於水質的變化都會有所影響，而且水質參數之間的物理及化學反應關係都是有所關連，不是獨立發生的。對於複雜的河川水質資料，以往分析的單變量統計分析過程太過繁複，因此藉由多變量統計分析，以獲得對於水質及水文資訊更好的詮釋。

1.2 文獻回顧

多變量統計分析，一般被用來分析大量且複雜的數據，其中包含了各種的參數及試驗單位(Perona *et al*., 1999; Beatriz *et al*., 2000; Ouyang, 2005)，運用多變量統計分析技術，往往能減少資料的維度，並能得到對資料更好的詮釋(Singh *et al*., 2005).

因爲複雜的水質監測數據，若是只使用單變量統計分析，僅能得到參數對參數的解釋，而且處理程序較爲複雜(Gikas *et al*., 2009)，運用多變量統計分析能夠幫助解釋複雜的資料矩陣，並用來描述及評估水體水質的變化。近年來，很多國內外學者已經採用多變量統計分析來進行水質評估，較常見的方法爲主成份分析、群集分析以及判別分析。各地方的水體環境不同，可以利用主成份分析來評估水質參數之間的相關性，以及分別占水質影響成份之多寡，藉此找出影響各區域水質的主要因素及條件。

於國內的參考文獻中，林淵文(1992)運用多變量統計分析，找出影響石門水庫流域水質的獨立因子，由此瞭解污染的緣由並比較地域間的差異。陳育偉(1995)運用多變量統計分析建立台灣地區水庫之優養評估模式。何勝惟(2002)運用因素分析來探討因素主要負荷、變異量並反映出土壤重金屬與地景指標等土地利用之相關性。簡宗儒(2003)運用主成份分析選出適合水庫的水質型式，再找出其代表性的水庫，另外用因素分析找出水質參數間不同性質的共同因子。劉瓊霦(2004)以主成份分析探討福山試驗林哈盆溪流域水化學的空間變異性，歸納其影響因子主要包

括四個：(1)大氣沉降、化學風化作用及海鹽飛沫；(2)植生影響因子，包括養分的吸收、淋洗及淋溶；(3)集水區的土壤及水文的影響；(4)其他因子。林炤映(2004)運用因素分析將水質變項分類，即可將共同因素分組命名，而以具體的變項展現抽象的水質現況。葉春國等(2007)利用主成份分析跟群集分析，探討水里溪上游集水區溪流水質的空間變異分析。許東鳴(2009)分析淡水河流域的水文資料在時間和空間上之變異，先利用線性迴歸與無母數趨勢檢定來評估時序資料的變化趨勢，並利用主成份分析跟群集分析進行空間的變異分析。廖少威(2010)利用主成份分析屏東大鵬灣海域暨鄰近河川底質特性以說明顯著的潛在因子。結果顯示最主要的潛在因子爲「土壤質地影響因子」、「多環芳香烴類影響因子」及「營養鹽影響因子」。進而推估高屏溪、東港溪、林邊溪三條溪流的底質特性主要係受到工業廢水、城鎮污水、農業禽畜等養殖污水排入，而大鵬灣海域囊袋入口處及末端處群集除城鎮污水及養殖污水的排入外又受到海域複雜的生物化學反應及地質影響。

於國外的參考文獻中，Viganò *et al*. (2003)運用主成份分析探討湖泊及底泥兩者的相互關係及影響來源。Zeng and Rasmussen (2005)運用因素分析找出影響河川的水質因子，以及使用群集分析對變數及試驗時間進行分群，得到時間與空間上的解釋。Jiang *et al*. (2007)運用群集分析將測站以時間做分群，探討時間上水質狀況的異同，並使用主成份分析探討影響水質的主要成份。Sherstha and Kazama (2007)運用群集分析先將測站資料進行分群，再使用判別分析進行驗證計算，以及主成份分析評估其分群的主要成份。Heikka (2008)將計算所得之主成份分數繪製成二維空間座標散佈圖，並以此結果探討各測站所代表的主成份。Zhou *et al*. (2008)運用群集分析及主成份分析探討影響底泥沉積物與水體性質的主要參數。Babu *et al*. (2009)運用主成份分析探討在雨季及非雨季時期，分別影響水質的主要成份。Dalal (2009)及 Joseph and Ouseph (2009)運用主成份分析探討影響海口水質的主要成份。Kazi (2009)以主成份分數爲基礎進行分群，後計算得出各測站之主成份分數，並繪製成二維空間座標散佈圖，探討兩種分析結果上的差異。Koklu *et al*. (2009)以主成份分析探討影響河川高流量時期與低流量時期水質變化之主要成份，並運用判別分析進行時間與空間上的驗證。Malik and Zeb (2009)運用群集分析，透過將微量元素分群來找出性質異於其他者的微量元素。Rodríguez-Barroso *et al*. (2009)探討底泥粒徑與微量元素間的相互關係，以及使用因素分析探討影響當地水質狀況之主要因素爲何。Shahrban and Etemad-Shahid (2009)運用群集分析將測站資料進行分群，使擁有相同性質的測站同屬一群，而不同群組的測站其性質差異大，隨後使用判別分析進行驗證計算。Juahir *et al*. (2010)在馬來西亞的 Langat River 選定 7 個測站的 23 個水質參數進行分析，包含群集分析(CA)、判別分析(DA)、主成份分析(PCA)及因素分析(FA)，得到 6 個主成份並將測站分成三群，『低污染區』、『中度污染區』及『高度污染區』，而中度及高度污染區的汙染是來自於人爲活動，包含工業、都市垃圾及農業逕流；低污染區的汙染來源則是來自家庭及農業逕流。Simeonov *et al*. (2010)利用保加利亞的 Pirin Mountain 湖泊水質監測數據的評估，證明可以使用群集分析、主成份分析以及神經網絡分類進行分類及預測，不過需要對數據的做標準化處理以避免產生數據的維度問題。Liu *et al*. (2011)利用多變量統計分析與地理統計法探討台灣鴛鴦湖的水質狀況。

1.3 研究方法與目的

台北水源特定區供應大台北地區近 500 萬人之自來水水源，係國內第一個經由都市計畫法設立之水源、水質、水量保護區，負責管理新店溪青潭堰上游集水區之水源、水質、水量之安全與潔淨，其區域係指新店青潭堰以上地區，範圍涵蓋三個流域，包含新店溪及其上游之南勢溪及北勢溪流域。

本研究方法係先選擇台北水源特定區，以水源特定區管理局轄下的監測站之水質監測數

據，配合氣象局雨量監測數據，合計 12 項參數，並採用多變量統計分析的主成份分析應用於水文及水質參數，以探討水質變化的主要成份因子；利用群集分析以數學方式將測站進行客觀的同質性分群；判別分析則是挑選出分群中最具影響性的參數，建立數值分群的模式，並進一步進行再次的驗證。

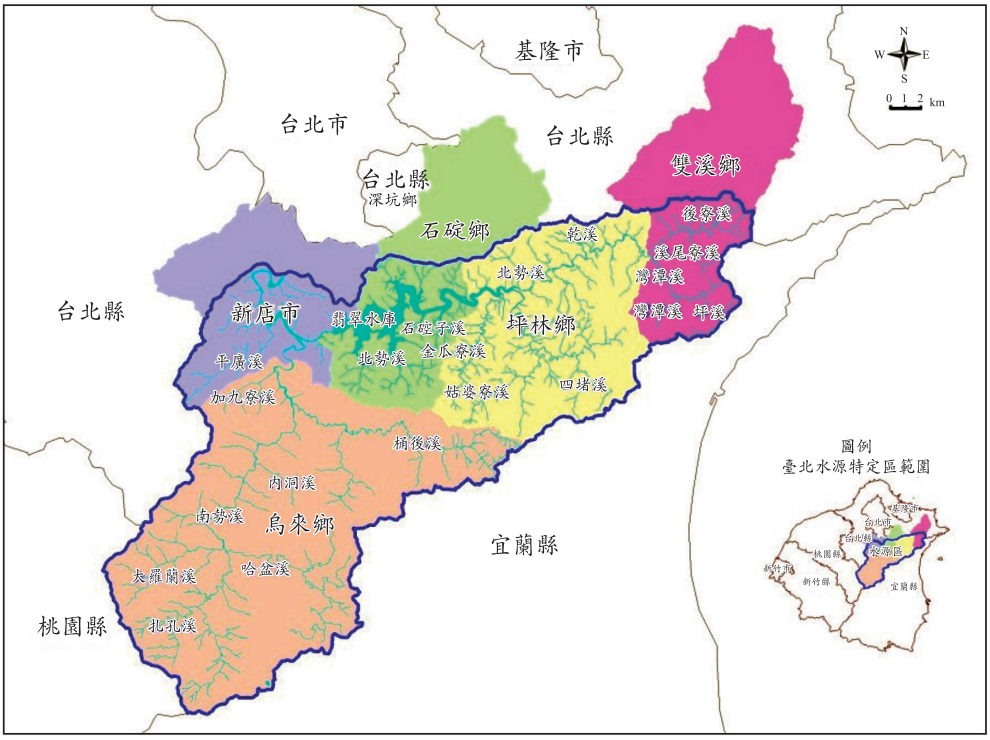
本研究目的期望藉由多變量統計分析以量化方式探討台北水源特定區之水質變化，經由多變量統計法解釋獲得的成份資料，評估影響河川水質的主要因子；對於流域中的測站做群集分析，以瞭解測站之間的特性，對於同質性高的測站，可以做刪減或是進行位置的改變，以達到經費節約或是得到更加的水質監測效果。

二、研究區域概述

台北水源特定區位於臺灣東北部、台北盆地東南方，行政區涵蓋台北縣之新店市、烏來鄉(不含烏來水源特定區計畫區)、石碇鄉之永安村、格頭村、坪林鄉(不含坪林水源特定區計畫區)、雙溪鄉泰平村等『四鄉一市』，面積約 690 平方公里，約佔全台北縣行政區域面積的三分之一(圖 1)。

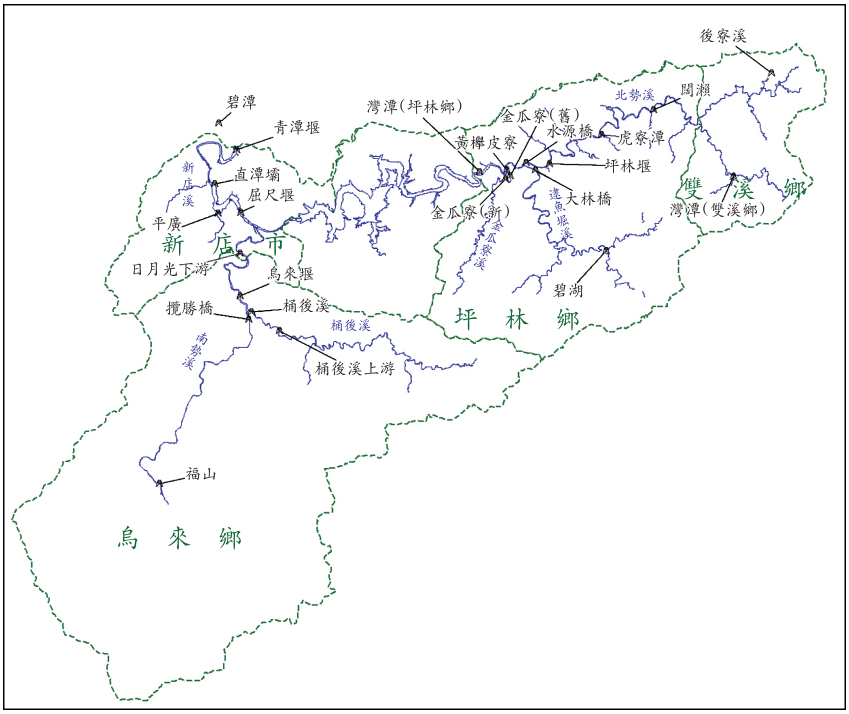
台北水源特定區係國內第一個經由都市計畫法設立之水源、水質、水量保護區，負責管理新店溪青潭堰上游集水區之水源、水質、水量之安全與潔淨，其區域係指新店青潭堰以上地區，範圍涵蓋三個流域，包含新店溪及其上游之南勢溪及北勢溪流域。

台北水源特定區供應大台北地區近 500 萬人之自來水水源，台北水源特定區管理局歷年皆依據行政院環境保護署所規劃訂定之相關監測規定及項目進行水質監測，並配合水位流量之長期觀測，持續建立完整之水質水量資料庫。目前台北水源特定區水質監測站位置如圖 2 所示，測站之經緯度位置資料如表 1。



引用自台北水源特定管理局網站

圖 1　台北水源特定區範圍圖



引用自台北水源特定管理局網站

圖 2　台北水源特定區水質測站位置圖

表 1　測站緯度表

流域	測站代碼	測站名稱	緯度	經度
北勢溪	A001	闊瀨	24° 57'42.54"北	121° 46'80.08"東
北勢溪	A002	坪林堰	24° 56'70.05"北	121° 42'44.50"東
北勢溪	A003	碧湖	24° 53'27.15"北	121° 44'43.86"東
北勢溪	A004	大林橋	24° 55'56.15"北	121° 42'18.15"東
北勢溪	A005	水源橋	24° 56'10.36"北	121° 42'10.82"東
北勢溪	A006	金瓜寮溪(舊)	24° 55'51.76"北	121° 41'29.95"東
北勢溪	A007	黃櫸皮寮	24° 55'53.27"北	121° 41'13.12"東
北勢溪	A008	灣潭(坪林)	24° 55'52.58"北	121° 40'29.79"東
北勢溪	A009	後察溪(雙溪)	24° 58'58.85"北	121° 49'56.11"東
北勢溪	A010	灣潭(雙溪)	24° 55'46.11"北	121° 48'43.24"東
北勢溪	A011	虎寮潭	24° 56'58.55"北	121° 44'27.78"東
北勢溪	A012	金瓜寮溪(新)	24° 55'46.95"北	121° 41'29.63"東
南勢溪	B001	福山	24° 46'47.27"北	121° 29'50.53"東
南勢溪	B002	攬勝橋	24° 51'42.87"北	121° 33'30.75"東
南勢溪	B003	桶後溪上游	24° 51'18.55"北	121° 34'30.21"東
南勢溪	B004	桶後溪	24° 51'51.56"北	121° 33'60.53"東
南勢溪	B005	烏來堰	24° 55'36.42"北	121° 31'55.71"東
新店溪	C001	平廣	24° 54'43.64"北	121° 32'70.48"東
新店溪	C002	屈尺堰	24° 55'90.93"北	121° 32'28.49"東
新店溪	C003	直潭壩	24° 56'22.27"北	121° 31'35.62"東
新店溪	C004	青潭堰	24° 56'36.91"北	121° 32'43.36"東
新店溪	C005	碧潭	24° 57'31.69"北	121° 32'80.80"東

表 2　各水質變數之縮寫及單位

水質變數	縮寫	單位	水質變數	縮寫	單位
溶氧量	DO	mg/L	大腸菌類數	MPN	CFU/100ml
生化需氧量	BOD	mg/L	濁度	Tur.	NTU
化學需氧量	COD	mg/L	總磷	TP	mg/L
pH 值	pH	—	氨氮	NH_3-N	mg/L
水溫	WTemp.	℃	懸浮固體物	TSS	mg/L
比導電度	Cond.	μmho/cm	雨量	Rain	mm

表 3　測站名稱及代碼對照表

北勢溪流域		南勢溪流域		新店溪流域	
測站代碼	測站名稱	測站代碼	測站名稱	測站代碼	測站名稱
A001	闊瀨	B001	福山	C001	平廣
A002	坪林堰	B002	攬勝橋	C002	屈尺堰
A003	碧湖	B003	桶後溪上游	C003	直潭壩
A004	大林橋	B004	桶後溪	C004	青潭堰
A005	水源橋	B005	烏來堰	C005	碧潭
A006	金瓜寮溪(舊)				
A007	黃櫸皮寮				
A008	灣潭(坪林)				
A009	後寮溪(雙溪)				
A010	灣潭(雙溪)				
A011	虎寮潭				
A012	金瓜寮溪(新)				

三、研究方法與理論

3.1 觀察值與變數

採樣期間自 2008 年 1 月至 2010 年 12 月，每月進行採樣一次，合計共 828 筆，每筆監測值均包含 12 項水質變數。監測的水質變數包括溶氧量、生化需氧量、化學需氧量、pH 值、水溫、大腸菌類數、濁度、總磷、氨氮、懸浮固體物、雨量，其名稱縮寫及使用單位如表 2 所示。

本研究數據採用台北水源特定區所涵蓋的三個流域的，包括北勢溪流域、南勢溪流域及新店溪流域，合計共 23 個測站，流域測站名稱及代碼如表 3 所示。

3.2 主成份分析

主成份分析是由皮爾森(Pearson)於 1901 年提出，再由侯特齡(Hotelling)於 1933 年加以發展，其主要的目的是希望用較少的變數去解釋原始資料大部分的變異(陳順宇，2005；Hair *et al.*, 2010)，亦即將原本大量的變數予以減少，使其變爲少數幾個互相獨立的線性組合變數(即主成份)，這些新變數便是利用主成份分析法整理而得的整體性指標。

假設有 m 個原始變數，有 n 個觀測樣本，則可構成一個 n × m 的原始數據矩陣 X：

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

爲了減少因單位不同而產生的計算誤差，首先必須將原始數據進行標準化處理，使其平均值爲 0，而變異數爲 1，即

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i} \quad \text{.....(2)}$$

其中 $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ ，爲變數的平均值。 .....(3)

$$s_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2}$$ ，爲變數的標準差。 ...(4)

經標準化後，可求出相關相關矩陣 L，並計算出此相關矩陣之特徵值(Eigen Values) λ_i 及特徵向量(Eigen Vectors)，即

$$\left|L - \lambda I\right| = 0 \quad \text{.....(5)}$$

其中 L 爲相關矩陣，I 爲單位矩陣。

特徵值分別爲

$\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$ ，且 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m$ 。

.....(6)

特徵向量爲

$$a_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}, a_2 = \begin{pmatrix} a_{21} \\ \vdots \\ a_{2m} \end{pmatrix}, \ldots, a_p = \begin{pmatrix} a_{p1} \\ \vdots \\ a_{mm} \end{pmatrix}$$ 。 .....(7)

因爲原始變數共有 m 個，故運算後可得到 m 個主成份個數，但爲了達到簡化的目標，只選取最大的 q 個新變數，來代替原來的 m 個變數。而這 q 個主成份解釋原來 m 個變數的變異數比例稱爲解釋變異百分比，即

$$R^2 = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_q}{\lambda_1 + \lambda_2 + \cdots + \lambda_m} \quad \text{.....(8)}$$

一般選取主成份之選取方式有下列幾種：

1. 解釋變異百分比 R^2 超過某一水準。如果已萃取出的主成份能解釋超過 75%的變異數，則後續萃取的主成份對變異數的解釋如果小於 5%，就不再選取。
2. 選取特徵值大於 1 的主成份，藉由特徵值對主成份個數繪製陡坡圖(Scree Plot)，其斜率明顯變化時之點位，即爲所求。

特徵向量即爲主成份負荷(Component Loading)，其數值若大於 0.75，代表相關性較強；若在 0.50 到 0.75 間，代表相關性普通；若在 0.30 到 0.50 間，則代表相關性較弱(Liu *et al*., 2003)。

將式(7)得到之特徵向量帶入原始變數，可得一線性方程式，最後得到主成份形式如下：

$$\begin{aligned}
&\text{第一個主成份為} \quad PC_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\
&\text{第二個主成份為} \quad PC_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\
&\qquad\vdots \\
&\text{第}m\text{個主成份為} \quad PC_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mm}x_m
\end{aligned}$$

式中 PC_i 爲主成份分數(Component Score)；a_{ij} 爲主成份負荷；x_i 爲水質變數；m 爲水質變數個數及主成份個數。

3.3 群集分析

群集分析的目的是要使得在同一集群內的事物具有高度的同質性，而不同集群的事物具有高度的異質性。因此依據樣本或變數間的相似性將其分成幾群互相沒有交集的群組，其分類原則是依照樣本性質間之疏密程度進行分類，其程度可以用距離表示，距離較近的樣本爲同群，較遠的樣本則爲不同群。

群集分析方法主要有兩大形式：

3.3.1 階層法(Hierarchical)

階層法可再細分爲凝聚分層法(Agglomerative)以及分離分層法(Divisive)，但分離分層法較少使用。

凝聚分層法開始時每一個樣本爲一群(共有 n 群)，然後最近的兩個樣本合成一群(變成 n-1 群)，依次結合後使群組越變越少，最後所有樣本結成一群。其中群間距離的計算以華德法(Ward's Method)之應用較爲廣泛。

華德法計算 A、B 兩群距離是以 A 群中心點 $\bar{x}_A$ 到兩群合併中心點 $\bar{\bar{x}}$ 距離平方乘以 A 群個體數，與 B 群中心點 $\bar{x}_B$ 到兩群合併中心點的 $\bar{\bar{x}}$ 距離平方乘以 B 群的個體數之和，即

$$d_{A,B} = n_A \cdot \left\|\bar{x}_A - \bar{\bar{x}}\right\|^2 + n_B \cdot \left\|\bar{x}_B - \bar{\bar{x}}\right\|^2 \quad \text{..... (9)}$$

3.3.2 非階層法(Nonhierarchical)

階層法在進行分類時，群組形成後便不會再拆散，其後都會在同群組內，非階層法則在各階段分類過程中，將原有群組全部拆散，並重新組

合成新的群組，其中最具代表的是 K 組平均法(K-means)。此法先將全部樣本分為 k 群，再計算某一樣本到各群組的重心距離，然後將樣本加入最近的群組並重新計算各群組重心，如此重覆進行至樣本不必重新分群為止。

以下四個因素會大大影響群集分析的使用效果：

(1) 群的結構(主要指群的形狀、規模和個數)
(2) 偏離值(Outliers，或稱異常點)的存在
(3) 群與群之間重疊的程度
(4) 相似衡量的選擇

為了減輕偏離值的影響，可能要反覆進行幾次的群集分析。與層次集群法相比，非層次集群法受偏離值、相似衡量和不合適的分群變數的影響較小，但是其缺點是集群結果對初始分群非常敏感。

Punj and Stewart (1983)歸納比較各種群集方法的優劣，發現採用兩階段群集分析，能得到較佳的分群效果。第一階段採用階層法獲得及群集的數目與起始點，並找出異常事物點予以剔除，以降低其對第二階段群集結果的影響，然後再以非階層法進行第二階段群集。

3.4 判別分析

判別分析是一種相依方法，其方法是利用事先訂定的類別或是組別的預測變數(準則變數)作為工具，估計預測變數和其中一個變數之間的相依關係。

它在運算前通常已先決定好群組，先利用區別變數建立區別規則(區別函數)，然後再由此規則對樣本做分類，預測每個樣本屬於各個群組的可能機率，而後指定一個線性組合，以判別新樣本的歸屬，藉此了解群組間的差異，並找出哪些預測變數具有最大的區別能力。

若有 m 個區別變數，第 k 個群組平均數向量矩陣為

$$u_k = \begin{pmatrix} u_{k1} \\ u_{k2} \\ \vdots \\ u_{km} \end{pmatrix} \quad \text{.....(10)}$$

而共變異數矩陣為

$$\sum_k = \begin{pmatrix} \sigma_{11}^{(k)} & \sigma_{12}^{(k)} & \cdots & \sigma_{1m}^{(k)} \\ \sigma_{21}^{(k)} & \sigma_{22}^{(k)} & \cdots & \sigma_{2m}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^{(k)} & \sigma_{m2}^{(k)} & \cdots & \sigma_{mm}^{(k)} \end{pmatrix} \quad \text{..... (11)}$$

其中 $\sigma_{ii}^{(k)}$ 是第 i 個變數 x_i 的變異數，而 $\sigma_{ij}^{(k)}$ 是 x_i 與 x_j 的共變異數。

當整體為多變量常態分配時，第 k 組的密度函數為

$$f_k(x) = \frac{1}{\left(\sqrt{2\pi}\right)^m \left|\sum_k\right|^{1/2}} e^{-\frac{1}{2}D_k(x)} \quad \text{..... (12)}$$

其中 $D_k(x) = (x - u_k)' \sum_k^{-1} (x - u_k)$ ..... (13)

其第 k 群的線性區別函數為

$$D_k(x) = \left(\sum^{-1} u_k\right)' x - \frac{1}{2}\left(u_k' \sum^{-1} u_k\right) + \ln(m_k)$$

..... (14)

驗後機率(Posterior Probability)是當樣本已知時，它落入 k 組的機率，這個機率值做為樣本分類的尺度，即

$$P\langle k | x \rangle = \frac{e^{d_k(x)}}{\sum_{j=1}^{k} e^{d_j(x)}} \quad \text{..... (15)}$$

判別分析的重點在於分析整體變數，故最後能得一個自變數的線性組合，其形式如下：

$$D(x) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_i x_i \quad \text{..... (16)}$$

其中 $D(x)$為判別函數；b_i 為判別負荷；x_i 為水質變數。

四、結果與討論

4.1 水質變數平均值及標準差

由表 4 中顯示北勢溪流域之比導電度、濁度及總磷數值均較低，可知北勢溪流域的水質相對其他兩個流域為佳。新店溪流域雖然氨氮的平均值還在甲類水體標準內，但是數值又是另外兩個流域的 2 至 3 倍，可見新店溪流域氨氮的污染負荷還是偏高。

表 4　各項水質變數之單位及數值範圍

水質參數	縮寫	單位	平均值和標準差(mean ± S.D)		
			北勢溪流域	南勢溪流域	新店溪流域
溶氧量	DO	mg/L	7.77 ± 0.66	7.81 ± 0.58	7.83 ± 0.79
生化需氧量	DOB	mg/L	0.82 ± 0.43	0.83 ± 0.43	1.05 ± 0.86
化學需氧量	COD	mg/L	3.61 ± 1.90	3.24 ± 2.08	4.06 ± 3.44
pH 值	pH	—	7.68 ± 0.81	7.87 ± 0.40	7.65 ± 0.46
比導電度	Cond.	μmho/cm	81.02 ± 18.03	114.50 ± 28.45	105.20 ± 47.98
水溫	Wtemp	℃	22.11 ± 4.58	21.49 ± 3.87	22.16 ± 3.86
大腸菌類	MPN	CFU/100ml	1297.56 ± 3959.15	2318.33 ± 3712.18	1689.89 ± 2405.34
濁度	Tur.	NTU	1.92 ± 2.83	16.94 ± 61.70	12.21 ± 37.26
總磷	TP	mg/L	0.070 ± 0.181	0.073 ± 0.136	0.095 ± 0.144
氨氮	NH_3-H	mg/L	0.033 ± 0.039	0.029 ± 0.022	0.079 ± 0.243
懸浮固體物	SS	mg/L	3.23 ± 4.83	25.67 ± 74.83	21.45 ± 63.32
雨量	Rain	mm	6.43 ± 12.88	10.07 ± 20.38	9.67 ± 20.58

表 5　各組平均數的相等性檢定

水質參數	Wilks' Lambda 值	F 檢定	分子自由度	分母自由度	顯著性
溶氧量	0.962	0.398	2	20	0.677
生化需氧量	0.586	7.069	2	20	0.005
化學需氧量	0.725	3.784	2	20	0.040
pH 值	0.784	2.763	2	20	0.087
比導電度	0.342	19.254	2	20	0.000
水溫	0.932	0.726	2	20	0.496
大腸菌類數	0.674	4.839	2	20	0.019
濁度	0.497	10.115	2	20	0.001
總磷	0.911	0.982	2	20	0.392
氨氮	0.670	4.916	2	20	0.018
懸浮固體物	0.440	12.742	2	20	0.000
雨量	0.559	7.878	2	20	0.003

4.2 判別分析

台北水源特定區包含三個流域，北勢溪流域、南勢溪流域及新店溪流域，三個流域的特性有相當大的差異，北勢溪流域爲翡翠水庫的集水區，原本的土地利用爲茶葉耕種，但近年於溪流附近新興一些遊憩活動；南勢溪流域有烏來溫泉與遊樂園，休閒旅遊活動盛行；新店溪流域居住人口密集，雖然有污水處理，但是人類生活廢水，對河川還是造成相當大的負荷。如果將三流域所有的資料放在一起討論，可能無法得到資料更深入的分析結果。

本研究首先使用判別分析，針對三個流域的特性做分析。先將三個流域的測站分組，將北勢溪流域的測站設爲分群 1，南勢溪流域的測站設爲分群 2，新店溪流域的測站設爲分群 3。探討流域的判別分析時，首先先驗證各個變數的檢定量(見表 5)，以檢定各變數的適合度。

由表 5 得知，生化需氧量與化學需氧量、比導電度、大腸菌類數、濁度、氨氮、懸浮固體物及雨量等變數均達顯著相關水準。

表 6 顯示其特徵值爲 7.104 及 1.453，判別函數的特徵值越大，表示線性判別函數越有判別力。典型相關是指此函數和組別間的關聯程度，0.936 和 0.770 都是相當高的數值，表示關聯程度

表 6　典型判別函數之特徵值

函數	特徵值	變異量 (%)	累積變異量 (%)	典型相關
1	7.104	83.0	83.0	0.936
2	1.453	17.0	100.0	0.770

表 7　典型判別函數之 Wilk's Lambda 值

函數	Wilk's Lambda 值	卡方	自由度	顯著性
1 到 2	0.231	160.682	28	0.000
2	0.743	32.517	13	0.002

表 8　典型判別函數結構矩陣係數表

水質變數	函數	
	1	2
懸浮固體物 [a]	0.553*	0.166
比導電度	0.502*	0.304
濁度 [a]	0.487*	0.108
雨量	0.308*	0.281
大腸菌類數	0.260*	-0.045
pH 值 [a]	0.218*	0.095
生化需氧量	0.036	0.693*
化學需氧量 [a]	-0.233	0.663*
氨 [a]	-0.070	0.596*
總磷 [a]	0.045	0.292*
水溫 [a]	-0.177	-0.258*
溶氧量 [a]	-0.023	0.148*

*在每個變數和任一判別函數之間的最大絕對相關
[a] 分析時未使用此變數

密切。

表 7 是用來比較判別函數的顯著性是否小於顯著相關水準($p = 0.05$)，由表 7 可看出兩個函數在統計上都是有顯著相關水準。表 8 則爲判別函數結構矩陣係數表。變數是依函數內相關的絕對值大小加以排序。

經由表 8 的結構矩陣係數表轉換成典型判別函數變數只有比導電度、雨量、大腸菌類數及生化需氧量及常數，如表 9 所示。

整理成標準化的典型判別函數係數如表 10，此四個變數是作爲判別分組的依據。

由圖 3 觀察三個流域的分群重心，以及測站分組的散佈狀況，可以發現三個流域分群的判別

表 9　典型判別函數係數

水質變數	函數	
	1	2
生化需氧量	-0.894	1.139
比導電度	1.245	0.133
大腸菌類數	0.462	-0.950
雨量	0.931	0.453
(常數)	-8.348	-8.920

註：未標準化係數

表 10　標準化的典型判別函數係數

水質變數	函數	
	1	2
生化需氧量	-0.894	1.139
比導電度	1.245	0.133
大腸菌類數	0.462	-0.950
雨量	0.931	0.453

判別變數和標準化典型判別函數之間的變數係依函數內相關的絕對大小合併後組內相關加以排序。

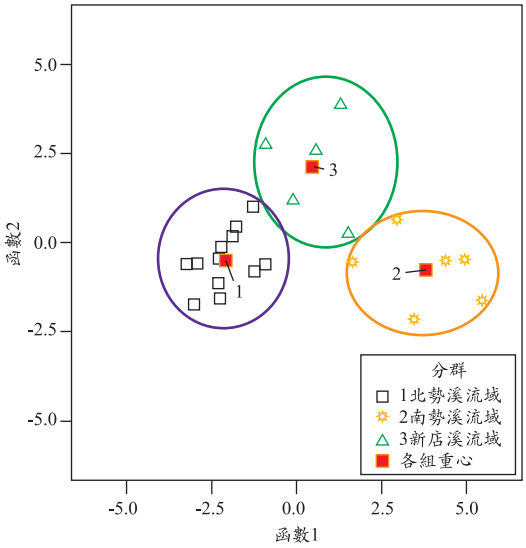


圖 3　三流域判別分析之群別散佈圖

度很高。

在表 11 中可以得知 23 個測站，經由判別分析預測的結果與實際結果完全相符，所得到的判別正確率爲 100%；而除了 A010 及 B004 驗後機率小於 80%，C004 驗後機率小於 60%以外，其餘測站的驗後機率均高於 90%。

表 11　三流域測站判別率

順序	測站代碼	實際組別	預測組別	驗後機率 *P*
1	A001	1	1	0.978
2	A002	1	1	1.000
3	A003	1	1	0.979
4	A004	1	1	0.999
5	A005	1	1	1.000
6	A006	1	1	0.991
7	A007	1	1	1.000
8	A008	1	1	1.000
9	A009	1	1	0.998
10	A010	1	1	0.771
11	A011	1	1	1.000
12	A013	1	1	0.911
13	B001	2	2	0.910
14	B002	2	2	1.000
15	B003	2	2	0.999
16	B004	2	2	0.734
17	B005	2	2	1.000
18	B006	2	2	1.000
19	C001	3	3	0.998
20	C002	3	3	0.924
21	C003	3	3	0.998
22	C004	3	3	0.538
23	C005	3	3	1.000

4.3 主成分分析

因爲不同變數資料的單位因次不同，數值範圍大小也會不同，故進行資料相關性分析前會先將資料轉爲無因次後，以減少資料誤差，爲了解釋數據方面的變化，以特徵值大於 1 作爲選取標準。

4.3.1 北勢溪流域

在北勢溪流域的水質變數之相關係數矩陣(表 12)中，可得知溶氧與 pH 值以外的其他變數在統計上都有顯著相關水準，其中總磷爲中度相關，其他爲低度相關。一般而言，水體的磷增加會造成溶氧降低，但是北勢溪流域因爲降雨造成水體溶氧增加，而雨水沖刷表土將磷帶入水體，因此呈現中度相關。

生化需氧量與化學需氧量、大腸菌類數、濁度及懸浮固體物在統計上有顯著相關水準，生化需氧量及化學需氧量爲中度相關，其他都是低度相關。

懸浮固體物與溶氧、生化需氧量、化學需氧量、水溫、大腸菌類數及濁度在統計有顯著相關水準。懸浮固體、大腸菌類數及濁度有中度相關，顯示水體的懸浮固體是造成大腸菌類數增加的原因。

雨量與溶氧、pH 值、比導電度、大腸菌類數及濁度在統計有顯著相關水準。雨量造成水體的離子濃度改變影響 pH 值及比導電度，也造成水體溫度的改變，而雨水沖刷表層土壤將泥沙及污染物帶入水體，造成大腸菌類數的增加。

計算相關係數矩陣的特徵值及解說變異量得到表 13，雖然有 12 個主成份，但是爲了達到縮減維度的目的，利用圖 4 的陡坡圖，選取特徵值大於 1 的位置來做主成份的選取(Filik Iscen *et al*., 2008; Xu *et al*., 2009; Zhang *et al*., 2011)，共選取 5 個主成份，總解釋變異量爲 67.26%。

表 12　北勢溪流域水質參數之相關係數矩陣

	DO	BOD	COD	pH	Cond.	WTemp.	MPN	Tur.	TP	NH_3-H	SS	Rain
DO	-1.000											
BOD	-0.308*	-1.000										
COD	-0.097**	0.553*	-1.000									
pH	-0.021	-0.028	-0.055	-1.000								
Cond.	-0.123*	-0.045	-0.045	-0.182*	-1.000							
Wtemp	-0.109**	-0.045	-0.039	-0.160*	0.288*	-1.000						
MPN	-0.107**	-0.248*	- .181*	-0.078	-0.080	-0.027	-1.000					
Tur.	-0.102**	-0.243*	-0.026	-0.058	-0.013	0.237*	- 0.154*	-1.000				
TP	-0.508*	-0.070	.323*	-0.069	-0.010	-0.168*	-0.035	-0.012	-1.000			
NH_3-H	-0.155*	-0.016	-0.171	-0.038	-0.036	-0.009	.110**	-0.072	-0.456*	-1.000		
SS	-0.169*	0.292*	-0.214*	-0.001	-0.008	-0.097*	-0.356*	-0.575*	-0.049	-0.077	-1.000	
Rain	-0.138*	-0.031	-0.025	-0.128**	0.227*	-0.010	-0.205*	-0.134*	-0.070	-0.025	-0.084	-1.000

**在顯著水準爲 0.01 時，相關顯著。分析時則採 99%的信賴區間對兩組數據的關聯係數進行區間估計

* 在顯著水準爲 0.05 時，相關顯著。分析時則採 95%的信賴區間對兩組數據的關聯係數進行區間估計

表 13 北勢溪流域主成份分析之特徵值及解說變異量

主成份	特徵值	解釋變異量(%)	總解釋變異量(%)
1	2.274	18.947	18.947
2	1.973	16.445	35.393
3	1.613	13.444	48.837
4	1.193	9.942	58.779
5	1.018	8.481	67.260
6	0.854	7.120	74.379
7	0.785	6.541	80.920
8	0.728	6.067	86.987
9	0.595	4.961	91.948
10	0.410	3.421	95.369
11	0.339	2.825	98.194
12	0.217	1.806	100.000

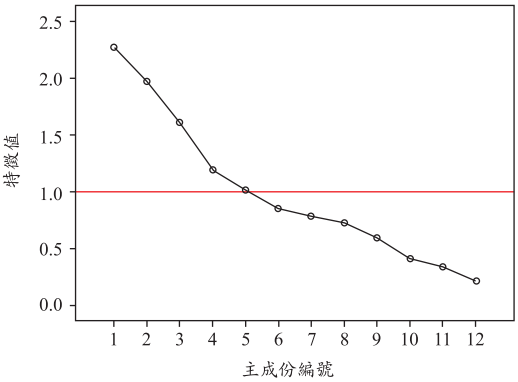


圖 4　北勢溪流域主成份分析之陡坡圖

表 14 爲北勢溪流域之主成份負荷矩陣，也稱爲關係係數矩陣。由表 13 及表 14 得知第一主成份可解釋變異量爲 18.947%，其中成份負荷較大的有生化需氧量、大腸菌類數、濁度及懸浮固體物，水體中懸浮固體物會造成濁度增加，也會挾帶大腸菌類進入水體，因此第一主成份命名爲懸浮固體成份。第二主成份可解釋的變異量爲 16.445%，相關性較高的爲溶氧量、化學需氧量、總磷及氨氮。總磷與氨氮會影響水體中的溶氧量及化學需氧量，總磷與氨氮都會成爲水體中的營養鹽，因此命名爲營養鹽成份。第三主成份可解釋變異量爲 13.444%，以 pH 值、比導電度、水溫相關性較高，因爲 pH 值也會影響到比導電度，因此命名爲 pH 值成份。第四主成份可解釋變異量爲 9.942%，雨量佔有較高的相關性，因此命名爲雨量成份。第五主成份可解釋變異量爲 8.481%，因爲所有變數在第五成份都不是最顯著的，因此將它命名爲其他成份。

4.3.2 南勢溪流域

在南勢溪流域的水質變數之相關係數矩陣(表 15)中，可得知溶氧與生化需氧量、水溫、總磷、氨氮及雨量在統計上都有顯著相關水準，都是低度相關。溶氧、總磷及氨氮成正相關，應該

表 14　北勢溪流域主成份負荷矩陣

水質變數	主成份				
	1	2	3	4	5
溶氧量	-0.451	0.550	0.310	0.135	-0.166
生化需氧量	0.709	0.156	-0.055	-0.532	0.176
化學需氧量	0.444	0.599	0.040	-0.413	0.280
pH 值	-0.046	-0.201	0.490	-0.068	0.248
比導電度	-0.107	-0.136	0.671	-0.046	0.332
水溫	0.179	-0.328	0.503	0.334	0.382
大腸菌類數	0.565	0.147	-0.176	0.321	0.196
濁度	0.610	-0.091	0.405	0.191	-0.520
總磷	-0.187	0.835	0.187	0.116	-0.083
氨氮	-0.004	0.624	0.073	0.356	0.222
懸浮固體物	0.761	0.056	0.168	0.279	-0.265
雨量	0.135	-0.039	-0.553	0.495	0.358

表 15　南勢溪流域水質參數之相關係數矩陣

	DO	BOD	COD	pH	Cond.	WTemp.	MPN	Tur.	TP	NH_3-H	SS	Rain
DO	1.000											
BOD	-0.246**	1.000										
COD	0.043	0.469**	1.000									
pH	-0.021	-0.056	0.045	1.000								
Cond.	0.027	-0.108	-0.121	0.162*	1.000							
Wtemp	-0.171*	0.044	0.055	0.190**	0.040	1.000						
MPN	-0.122	0.050	0.088	-0.056	-0.027	0.106	1.000					
Tur.	-0.066	0.034	-0.068	-0.028	0.039	0.152*	-0.051	1.000				
TP	0.229	0.052	0.206	-0.187**	-0.070	-0.113	-0.076	0.069	1.000			
NH_3-H	0.188**	-0.091	0.132	-0.105	0.074	-0.087	-0.008	0.064	0.219**	1.000		
SS	-0.141	0.077	-0.019	-0.092	-0.008*	0.143	-0.026	0.947**	0.178*	0.097	1.000	
Rain	-0.210**	0.073	0.059	-0.176*	0.078	-0.015	0.375**	-0.092	-0.056	0.039	-0.077	1.000

**在顯著水準爲 0.01 時，相關顯著。分析時則採 99%的信賴區間對兩組數據的關聯係數進行區間估計

* 在顯著水準爲 0.05 時，相關顯著。分析時則採 95%的信賴區間對兩組數據的關聯係數進行區間估計

是降雨造成水體溶氧增加，雨水沖刷表土將磷及氨氮等帶進水體。雨量越多，水溫越低，也會影響生化需氧量。生化需氧量與化學需氧量在統計上有顯著相關水準，而且爲中度相關。

pH 值、比導電度、水溫、總磷及雨量在統計上都有顯著相關水準，但都是低度相關。

大腸菌類數與雨量在統計上有顯著相關水準，而且是中度相關，表示降雨將地表的大腸菌類沖入水體中。

懸浮固體物與濁度在統計有顯著相關水準。懸浮固體與濁度有高度相關，這兩者如果同時升高，表示水中的懸浮雜質增加。

計算相關係數矩陣的特徵值及解說變異量得到表 16，雖然有 14 個主成份，但是爲了達到縮減維度的目的，利用圖 5 的陡坡圖，選取特徵值大於 1 的位置來做主成份的選取，共選取 5 個主成份，總解釋變異量爲 66.624%。

表 17 爲南勢溪流域之主成份負荷矩陣，故由表 16 及表 17 得知第一主成份可解釋變異量爲 17.251%，其中成份負荷較大的是濁度及懸浮固

表 16　南勢溪流域主成份分析之特徵值及解說變異量

主成份	特徵值	解釋變異量(%)	總解釋變異量(%)
1	2.070	17.251	17.251
2	1.743	14.524	31.774
3	1.662	13.849	45.623
4	1.358	11.315	56.938
5	1.162	9.685	66.624
6	0.918	7.647	74.271
7	0.745	6.210	80.480
8	0.729	6.074	86.554
9	0.633	5.274	91.828
10	0.543	4.526	96.353
11	0.399	3.325	99.679
12	0.039	0.321	100.000

表 17　南勢溪流域主成份負荷矩陣

水質變數	主成份				
	1	2	3	4	5
溶氧量	-0.143	-0.655	0.266	-0.028	0.227
生化需氧量	0.157	0.560	0.413	-0.427	-0.071
化學需氧量	0.048	0.330	0.627	-0.445	0.305
pH 值	-0.106	0.020	-0.449	-0.494	0.469
比導電度	-0.019	-0.102	-0.314	0.188	0.622
水溫	0.232	0.347	-0.336	-0.233	0.329
大腸菌類數	-0.078	0.540	0.095	0.443	0.215
濁度	0.948	-0.043	-0.164	0.082	-0.005
總磷	0.248	-0.309	0.615	-0.009	0.086
氨氮	0.149	-0.302	0.412	0.261	0.470
懸浮固體物	0.975	-0.003	-0.064	0.093	-0.036
雨量	-0.133	0.529	0.162	0.602	0.149

體物，水體中懸浮固體物會造成濁度增加，因此第一主成份命名爲懸浮固體成份。第二主成份可解釋變異量爲 14.524%，相關性較高的爲溶氧量、生化需氧量、水溫及大腸菌類數。溶氧量與其他三者是負相關，大腸菌類數跟人類活動息息相關，因此命名爲人類活動成份。第三主成份可解釋變異量爲 13.849%，以化學需氧量及總磷相關性較高，因爲總磷在水中分解會使得耗氧量增加，因此命名爲總磷成份。第四主成份可解釋變異量爲 11.315%，pH 值及雨量佔有較高的相關性，雨水進入水體，會影響水體的 pH，因此命

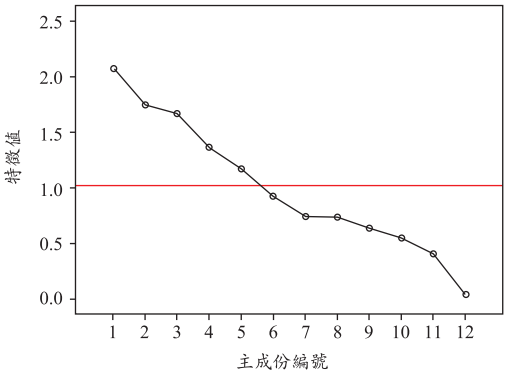


圖 5　南勢溪流域主成份分析之陡坡圖

名爲雨量成份。第五主成份可解釋變異量爲 9.685%，相關性較高的是比導電度跟氨氮，氨氮在水體中分解會釋放離子，改變水體的導電性，因此命名爲氨氮成份。

4.3.3 新店溪流域

在新店溪流域的水質變數之相關係數矩陣(見表 18)中，可得知溶氧、生化需氧量、水溫、大腸菌類數、總磷、懸浮固體物及雨量在統計上都有顯著相關水準，其中溶氧與總磷是中度相關。

生化需氧量與大腸菌類數、濁度、氨氮、懸浮固體物及雨量在統計上有顯著相關水準，生化需氧量、大腸菌類數及懸浮固體物爲中度正相關，其他都是低度相關。大腸菌類附著在懸浮固體進入水體，兩者數值升高，造成生化需氧量數值也提高，代表水質有不好的影響。

化學需氧量與大腸菌類數、總磷、懸浮固體物及雨量在統計有顯著相關水準。顯示降雨造成水體中的懸浮固體物、大腸菌類數及總磷增加，化學需氧量也增加，水體的污染負荷也增加。

pH 值與水溫及總磷在統計有顯著相關水準。水體溫度上升加速水體中的有機物分解，影響水體的 pH 值。

雨量與比導電度、大腸菌類數在統計有顯著相關水準，而且是中度正相關，顯示降雨影響水體的離子濃度增加導電性以及降雨將地表的大腸菌類沖入水體。

計算相關係數矩陣的特徵值及解說變異量

表 18　新店溪流域水質參數之相關係數矩陣

	DO	BOD	COD	pH	Cond.	WTemp.	MPN	Tur.	TP	NH_3-H	SS	Rain
DO	1.000											
BOD	-0.212**	1.000										
COD	-0.084	0.822	1.000									
pH	-0.046	-0.097	-0.084	1.000								
Cond.	0.017	0.073	-0.030	0.111	1.000							
Wtemp	-0.207**	0.040	-0.038	0.306**	0.123	1.000						
MPN	-0.254**	0.371**	0.416**	-0.013	0.003	0.072	1.000					
Tur.	-0.053	0.195**	0.122	-0.062	-0.052	0.059	0.205**	1.000				
TP	0.640**	0.091	0.215**	-0.193*	-0.042	-0.245**	-0.107	0.046	1.000			
NH_3-H	0.146	0.210**	0.143	0.003	-0.011	-0.018	-0.074	-0.035	0.031	1.000		
SS	-0.169*	0.353**	0.301**	-0.087	-0.091	0.037	0.290**	0.888**	0.030	-0.046	1.000	
Rain	-0.215**	0.195**	0.243**	-0.107	0.332**	-0.053	0.314**	-0.011	-0.122	-0.033	0.085	1.000

** 在顯著水準爲 0.01 時，相關顯著。分析時則採 99%的信賴區間對兩組數據的關聯係數進行區間估計

*　在顯著水準爲 0.05 時，相關顯著。分析時則採 95%的信賴區間對兩組數據的關聯係數進行區間估計

表 19　新店溪流域主成份分析之特徵值及解說變異量

主成份	特徵值	解釋變異量(%)	總解釋變異量(%)
1	2.791	23.259	23.259
2	1.975	16.454	39.714
3	1.598	13.321	53.034
4	1.247	10.390	63.424
5	1.152	9.600	73.024
6	0.925	7.706	80.730
7	0.700	5.836	86.566
8	0.633	5.276	91.842
9	0.488	4.068	95.910
10	0.266	2.219	98.129
11	0.140	1.165	99.293
12	0.085	0.707	100.000

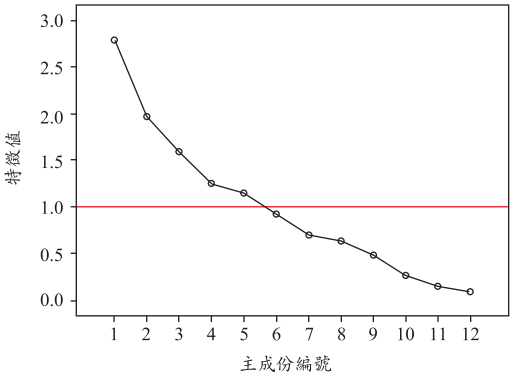


圖 6　新店溪流域主成份分析之陡坡圖

得到表 19，雖然有 12 個主成份，但是爲了達到縮減維度的目的，利用圖 6 的陡坡圖，選取特徵值大於 1 的位置來做主成份的選取，共選取 5 個主成份，總解釋變異量爲 73.024%。

表 20 爲新店溪流域之主成份負荷矩陣，由表 19 及表 20 得知第一主成份可解釋變異量爲 23.259%，其中成份負荷較大的有生化需氧量、化學需氧量、大腸菌類數、濁度及懸浮固體物，水體中懸浮固體物增加，也會挾帶大腸菌類進入水體，因此第一主成份命名爲懸浮固體成份。第二主成份可解釋變異量爲 16.454%，相關性較高的爲總磷，因此命名爲總磷成份。第三主成份可解釋變異量爲 10.390%，以雨量相關性較高，因此命名爲雨量成份。第四主成份可解釋變異量爲 9.600%，pH 值與水溫、氨氮相關性較高，水溫上升造成氨氮分解加速，影響水體的 pH 值，因此命名爲水溫成份。第五主成份可解釋變異量爲 8.481%，以比導電度相關性較高，因此命名比導電度成份。

4.4 群集分析

本研究中針對三個流域的測站依照 2008 年至 2010 年進行群集分析，共有三年的測站群集，

探討測站的分組分佈情形。Anderberg (1973)基於為了兼顧經濟簡單與有效，認為群集分析可採用兩段式群集分析，這種程序也成為目前常用的群集分析的方法。

兩段式群集分析第一階段先利用階層法來做凝聚，點間距離使用歐式距離平方(Squared Edclidean Distance)，分析方法使用華德法(Ward's Method)，可求得概略的分群結果。第二階段使用非階層法的 K 組平均法，利用第一階段求得的分群數目，再進行更精確的分析結果。

表 20　新店溪流域主成份負荷矩陣

水質變數	主成份				
	1	2	3	4	5
溶氧量	-0.366	0.722	0.024	0.186	0.355
生化需氧量	0.786	0.150	0.341	0.251	-0.145
化學需氧量	0.748	0.265	0.393	0.196	-0.158
pH 值	-0.133	-0.417	-0.054	0.599	0.214
比導電度	0.033	-0.232	0.389	-0.008	0.770
水溫	0.076	-0.521	-0.125	0.530	0.172
大腸菌類數	0.646	-0.174	0.136	-0.122	-0.049
濁度	0.586	0.129	-0.702	0.003	0.265
總磷	-0.035	0.835	0.070	0.079	0.265
氨氮	0.056	0.229	0.262	0.514	-0.184
懸浮固體物	0.733	0.111	-0.593	-0.025	0.162
雨量	0.389	-0.228	0.470	-0.429	0.330

4.4.1 北勢溪流域測站群集分析

北勢溪流域測站為 12 個，分成三個年度做測站群集分析，測站分群關係可以整理成表 21。再利用測站的地理位置圖 7 呈現。

4.4.2 南勢溪流域測站群集分析

南勢溪流域測站為 5 個，分成三個年度做測站群集分析，測站分群關係可以整理成表 22。再利用測站的地理位置圖 8 呈現。

表 21　北勢溪流域 2008-2010 年的測站分群關係表

群集	2008 年	2009 年	2010 年
CA 1	A001、A003 A004、A005 A007	A001、A002* A003、A004 A005、A007 A011*、A012*	A002、A003 A004、A005 A011、A012
CA 2	A002、A008 A011	A008	A001*、A010*
CA 3	A006、A012	A006	A006、A007* A008*
CA 4	A009、A010	A009	A009

*表示該測站的分群與前一年不同。

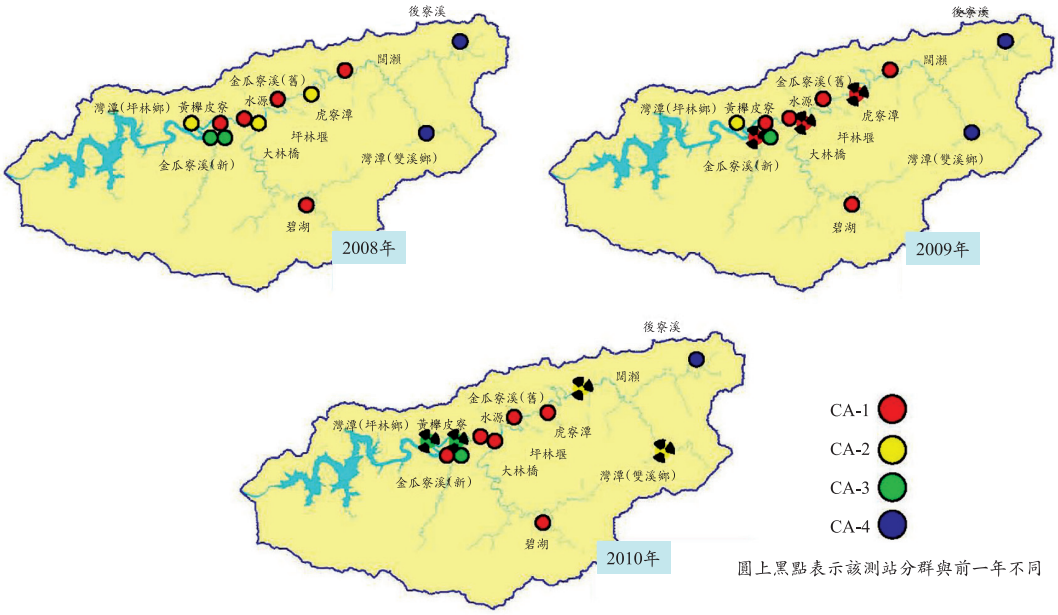


圖 7　北勢溪流域 2008-2010 年的測站分群地理位置圖

表 22　南勢溪流域 2008-2010 年的測站分群關係表

群集	2008 年	2009 年	2010 年
CA 1	B001	B001	B001、B005*
CA 2	B002、 B003、 B005	B002、B005、 B006*	B002、B006
CA 3	B006	B003*、B004*	B003、B004

* 表示該測站的分群與前一年不同。

表 23　新店溪流域 2008-2010 年的測站分群關係表

群集	2008 年	2009 年	2010 年
CA 1	C001	C001	C001、C005*
CA 2	C002、C003、C004	C002、C003、C004	C002、C003
CA 3	C005	C005	C004*

* 表示該測站的分群與前一年不同。

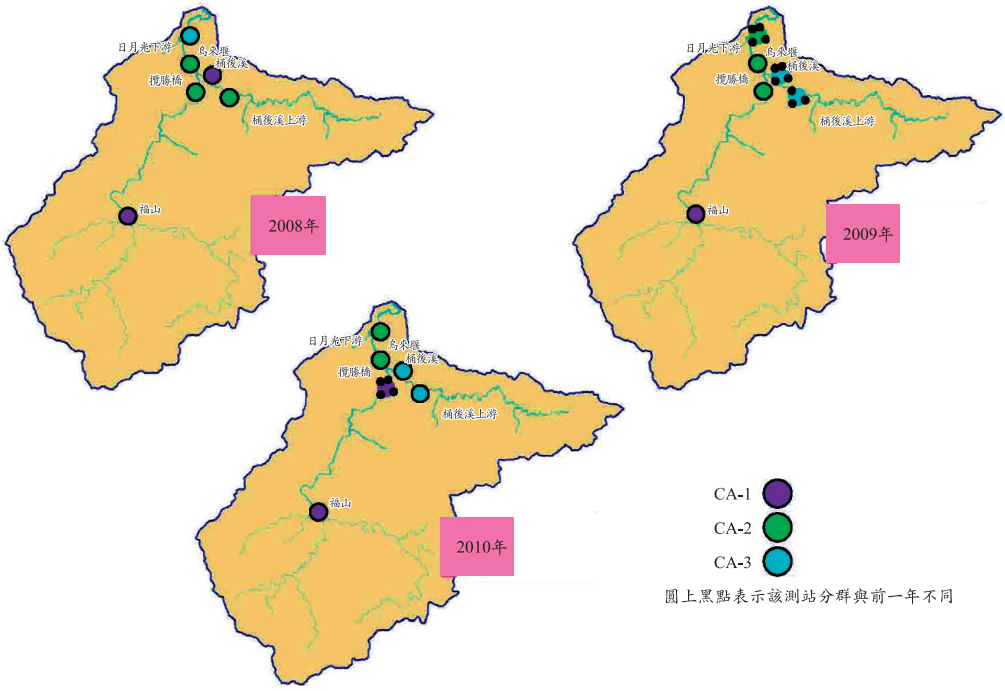


圖 8　南勢溪流域 2008-2010 年的測站分群地理位置圖

4.4.3 新店溪流域測站群集分析

新店溪流域測站為 5 個，分成三個年度做測站群集分析，測站分群關係可以整理成表 23。再利用測站的地理位置圖 9 呈現。

五、結　論

本研究針對台北水源特定區應用多變量統計方法進行分析與評估，可獲得以下結論：

1. 判別分析很清楚可以看出三個流域分群的判別度很高，對照現況，可以驗證三個流域的水質變數各有其特性。而由區別函數檢定得出的係數，可得到比導電度、雨量、大腸菌類數及生化需氧量等參數作為分組的判別適合度較佳。

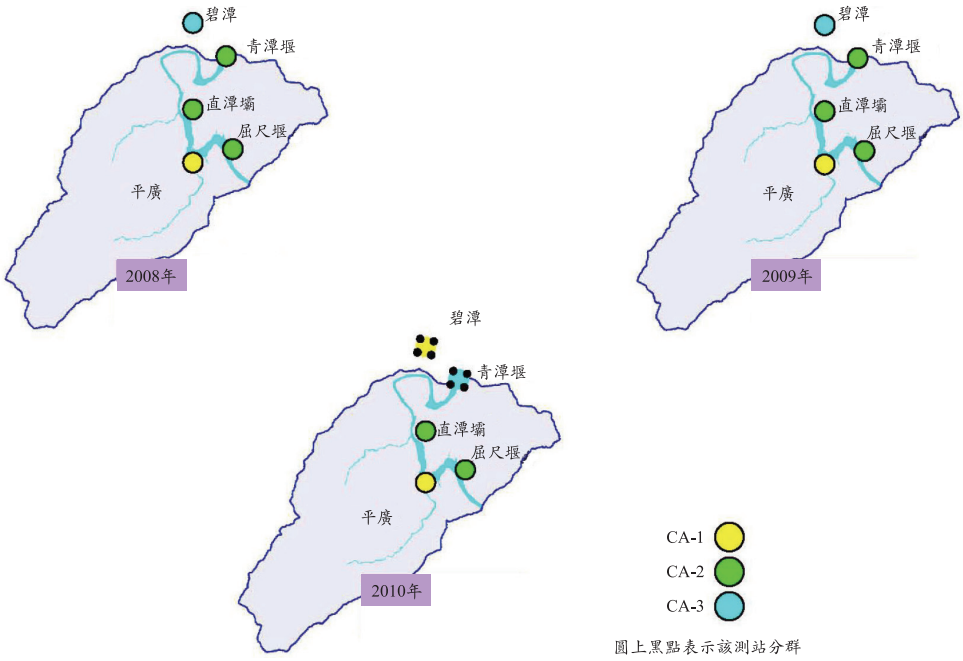


圖 9　新店溪流域 2008-2010 年的測站分群地理位置圖

2. 利用主成份分析可得到台北水源特定區所包含的三個流域的主成份。北勢溪流域可萃取 5 個主成份，依序爲懸浮固體成份、營養鹽成份、pH 值成份、雨量成份及其他成份，可解釋的總變異量爲 67.260%；南勢溪流域也萃取出 5 個主成份，依序爲懸浮固體成份、人類活動成份、總磷成份、雨量成份及氨氮成份，可解釋的總變異量爲 66.624%；新店溪流域也萃取出 5 個主成份，依序爲懸浮固體成份、總磷成份、雨量成份、水溫成份及比導電度成份，可解釋的總變異量爲 73.024%。
3. 依主成份分析，對照三個流域現況，北勢溪流域爲農業耕種盛行，但是污染負荷較低；南勢溪流域爲溫泉及森林遊樂區，遊憩觀光盛行，人類活動造成的污染負荷影響水體甚多；新店溪流域爲住宅區，人口稠密，雖有污水下水道處理，但是家庭廢水所造成的污染負荷也最爲嚴重。
4. 針對三個流域的測站做群集分析，能瞭解測站之間的異同，北勢溪流域 2009 年及 2010 年的群集分析中，CA 1 的測站有 6 個是沒有改變分群的，可以繼續監測如果分群都穩定沒有改變，就可以嘗試刪減測站以節省經費，或是移動測站的位置來得到新的監測數值，以得到對流域水質更好的詮釋；南勢溪及新店溪因爲測站數較少，可以嘗試增加臨時性的測站，跟原始測站做比較，以評估水質監測的效果。

謝　誌

本研究承蒙臺北水源特定區管理局提供水文與水質監測資料，得以順利完成，特予致謝。

參考文獻

1. 何勝惟，「應用多變量統計於土壤重金屬污染與土地利用之分析」，國立臺灣大學生物環境系統工程學研究所碩士論文，2002。
2. 林淵文，「以多變量統計方法分析石門水庫流域之水質污染」，國立臺灣大學農業工程研究所碩士論文，1992。
3. 許東鳴，「淡水河流域水文時空變異分析」，國立中央大學水文與海洋科學研究所碩士論文，2009。

4. 陳育偉，「應用多變量統計方法辨識水庫優養化」，國立臺灣大學農業工程研究所碩士論文，1995。
5. 陳順宇，多變量分析 四版，華泰書局，2005。
6. 廖少威、賴文亮、陳振正、邱俊彥，「大鵬灣底質特性空間變異暨與鄰近河川底質特性之分類研究」，農業工程學報，54 卷 2 期，pp. 9-20，2008。
7. 葉春國、廖學誠、黃正良、薛美莉、王景平，「水里溪上游集水區溪流水質的空間變異分析」，地理學報，47 期，pp. 19-38，2007。
8. 簡宗儒，「多變量統計分析在台灣水庫水質之應用」，國立臺灣海洋大學環境生物與漁業科學學系碩士論文，2003。
9. 劉瓊霦、陳春雄、金恒鑣，「以主成份分析探討福山試驗林哈盆溪流域水化學的空間變異性」，臺灣林業科學，19 卷 4 期，pp. 363-374，2004。
10. Anderberg, M. R., Cluster Analysis for Applications, Academic Press, New York, 359p, 1973.
11. Antikainen, R., Lemola, R., Nousiainen, J. I., Sokka, L., Esala, M., Huhtanen, P., and Rekolainen, S., "Stocks and Flows of Nitrogen and Phosphorus in the Finnish Food Production and Consumption System," Agriculture, Ecosystems and Environment, 107(2-3), 287-305, 2005.
12. Babu, K. N., Omana P. K., and Mohan M., "Water and Sediment Quality of Ashtamudi Estuary, a Ramsar Site, Southwest Coast of India – A Statistical Appraisal," Environmental Monitoring and Assessment, 165(1-4), 307-319, 2009.
13. Beatriz, H., Rafael, P., Marisol, V., Enrique, B., Jose, M. F., and Luis, F., "Temporal Evolution of Groundwater Composition in an Alluvial Aaquifer (Pisuerga River, Spain) by Principal Component Analysis," Water Research, 34(3), 807-816, 2000.
14. Dalal, S. G., Shirodkar, P. V., Jagtap, T. G., Naik, B. G., and Rao, G. S., "Evaluation of Significant Sources Influencing the Variation of Water Quality of Kandla Creek, Gulf of Katchchh, Using PCA," Environmental Monitoring and Assessment, 163(1-4), 49-56, 2009.
15. Filik Iscen, C., Emiroglu, O., Ilhan, S., Arslan, N., Yilmaz, V., and Ahiska, S., "Application of Multivariate Statistical Techniques in the Assessment of Surface Water Quality in Uluabat Lake, Turkey," Environmental Monitoring and Assessment, 144(1-3), 269-276, 2008.
16. Gikas, G. D., Tsihrintzis, V. A., Akratos, C. S., and Haralambidis, G., "Water Quality Trends in Polyphytos Reservoir, Aliakmon River, Greece," Environmental Monitoring and Assessment, 149(1-4), 163-181, 2009.
17. Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E., Multivariate Data Analysis: A Global Perspective, 7th, Prentice Hall, 2010.
18. Heikka, R. A., "Multivariate Monitoring of Water Quality: A Case Study of Lake Simpele, Finland. Journal of Chemometrics, 22(11-12), 747-751, 2008.
19. Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components," Journal of Educational Psychology, 24(7), 417-441, 1993.
20. Jiang, J. G., Wu, S. G., and Shen, Y. F., "Effects of Seasonal Succession and Water Pollution on the Protozoan Community Structure in an Eutrophic Lake," Chemosphere, 66(3), 523-532, 2007.
21. Joseph, S. and Ouseph, P. P., "Assessment of Nutrients Using Multivariate Statistical Techniques in Estuarine Systems and its Management Implications: A Case Study from Cochin Estuary, India," Water and Environment Journal, 24(2), 126-132, 2010.
22. Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza,

T. I. T., Armi, A. S. M., Toriman, M. E., and Mokhtar, M., "Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques," Environmental Monitoring and Assessment, 173(1-4), 625-641, 2010.

23. Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A., and Shah, A. Q., "Assessment of Water Quality of Polluted Lake Using Multivariate Statistical Techniques: A Case Study," Ecotoxicology and Environmental Safety, 72(2), 301-309, 2009.
24. Koklu, R., Sengorur, B., and Topal, B., "Water Quality Assessment Using Multivariate Statistical Methods- A Case Study: Melen River System (Turkey)," Water Resources Management, 24(5), 959-978, 2009.
25. Liu, C. W., Lin, K. H., and Kuo, Y. M., "Application of Factor Analysis in the Assessment of Groundwater Quality in a Blackfoot Disease Area in Taiwan," The Science of the Total Environment, 313(1-3), 77- 89, 2003.
26. Liu, W. C., Yu. H. L., and Chung C. E., "Assessment of Water Quality in a Subtropical Alpine Lake Using Multivariate Statistical Techniques and Geostatistical Mapping: A Case Study," International Journal of Environmental Research and Public Health, 8(4), 1126-1140, 2011.
27. Malik, R. N. and Zeb, N., "Assessment of Environmental Contamination Using Feathers of *Bubulcus ibis* L., as a Biomonitor of Heavy Metal Pollution, Pakistan," Ecotoxicology, 18(5), 522-536, 2009.
28. Ouyang, Y., "Evaluation of River Water Quality Monitoring Station by Principal Component Analysis," Water Research, 39(12), 2621-2635, 2005.
29. Pearson, K., Principal Components Analysis, The London, Edinburgh and Dublin Philosophical Magazine and Journal, 6(2), p. 566, 1901.
30. Perona, E., Bonilla, I., and Mateo, P., "Spatial and Temporal Changes in Water Quality in a Spanish River," Science of the Total Environment, 241(1-3), 75-90, 1999.
31. Punj, G. and Stewart, D. W., "Cluster Analysis in Marketing Research: Review and Suggestion for Application," Journal of Marketing Research, 20(2), 134-148, 1983.
32. Rodríguez-Barroso, M. R., García-Morales, J. L., Coello Oviedo, M. D., and Quiroga Alonso, J. M., "An Assessment of Heavy Metal Contamination in Surface Sediment Using Statistical Analysis," Environmental Monitoring and Assessment, 163(1-4), 489-501, 2009.
33. Shahrban, M. and Etemad-Shahid, A., "Classification of the Caspian Sea Coastal Waters Based on Trophic Index and Numerical Analysis," Environmental Monitoring and Assessment, 164(1-4), 349-356, 2009.
34. Sherstha, S., and Kazama, F., "Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan," Environmental Modelling and Software, 22(4), 464-475, 2007.
35. Simeonov, V., Simeonova, P., Tsakovski, S., and Lovchinov, V., "Lake Water Monitoring Data Assessment by Multivariate Statistics," Journal of Water Resource and Protection, 2(4), 353-361, 2010.
36. Singh, K. S., Amrita, M., and Sarita, S., "Water Quality Assessment and Apportionment of Pollution Sources of Gomti River (India) Using multivariate Statistical Technology-A Case Study. Analytica Chimica Acta, 538, 355-374, 2005.
37. Smith, R. V., Jordan, C., and Annett, J. A., "A Phosphorus Budget for Northern Ireland: Inputs to Inland and Coastal Waters," Journal of Hy-

drology, 304(1-4), 193-202, 2005.

38. Viganò, L., Arillo, A., Buffagni, A., Camusso, M., Ciannarella, R., Crosa, G., Falugi, C., Galassi, S., Guzzella, L., Lopez, A., Mingazzini, M., Pagnotta, R., Patrolecco, L., Tartari, G., and Valsecchi, S., "Quality Assessment of Bed Sediments of the Po River (Italy)," Water Research, 37(3), 501-518, 2003.
39. Xu, H., Yang, L. Z., Zhao, G. M., Jiao, J. G., Yin, S. X. and Liu, Z. P., "Anthropogenic Impact on Surface Water Quality in Taihu Lake Region, China," Pedosphere, 19(6), 765-778, 2009.
40. Zeng, X., and Rasmussen, T. C., "Multivariate Statistical Characterization of Water Quality in Lake Lanier, Georgia, USA," Journal of Environmental Quality, 34(6), 1980-1991, 2005.
41. Zhang, X., Wang, Q., Liu, Y., Wu, J., and Yu, M., "Application of Multivariate Statistical Techniques in the Assessment of Water Quality in the Southwest New Territories and Kowloon, Hong Kong," Environmental Monitoring and Assessment, 173(1-4), 17-27, 2011.
42. Zhou, J., Ma, D., Pan, J., Nie, W., and Wu, K., "Application of Multivariate Statistical Approach to Identify Heavy Metal Sources in Sediment and Waters: A Case Study in Yangzhong, China," Environmental Geology, 54(2), 373-380, 2008.

收稿日期：民國 101 年 4 月 30 日
修正日期：民國 101 年 6 月 28 日
接受日期：民國 101 年 9 月 12 日