

「2020 中華民國營建工程學會第十八屆營建產業永續發展研討會」

## 應用主成分分析評估高山湖泊水質之研究

鐘群凱(Chun-Kai Chung)

國立聯合大學土木與防災工程學系

王思懿(Zi-Yi Wang)

國立交通大學土木工程學系

陳翰瑩(Han-Ying Chen)

國立中央大學土木工程學系

陶煒君(Yeh-Chun Tao)

國立聯合大學土木與防災工程學系

柳鴻明(Hong-Ming Liu)

國立聯合大學土木與防災工程學系

柳文成(Wen-Cheng Liu)

國立聯合大學土木與防災工程學系摘要

本研究選定位於新竹縣尖石鄉的鴛鴦湖與宜蘭縣太平山的翠峰湖，兩湖泊為臺灣典型的高山湖泊，各自有獨立的集水區，環繞著原始的檜木林，而且長期有學術單位在鴛鴦湖與翠峰湖設置監測儀器。

本研究於 2017 年 9 月至 2019 年 3 月間，進行定期定點現地採樣，並於實驗室進行水質分析，分析之水質項目有：懸浮固體、總磷、磷酸鹽、有機磷、總凱氏氮、氨氮及有機氮，並彙整中研院提供之水質資料：水溫及溶氧，以多變量統計分析方法進行鴛鴦湖與翠峰湖之主成份分析。結果顯示鴛鴦湖主成份分析擁有 82.051 % 的總解釋變異量；翠峰湖主成份分析擁有 82.228 % 的總解釋變異量。表示鴛鴦湖與翠峰湖主成份分析具有相當的可靠性。關鍵字：高山湖泊、鴛鴦湖、翠峰湖、主成份分析

### Abstract

This study is located in Yuan-Yang Lake in Jianshi Township, Hsinchu County, and Tsui-Fong Lake in Taiping Mountain, Yilan County. The two lakes are typical mountain lakes in Taiwan, each with independent catchment areas, surrounded by primitive cypress forests, and long-term academic units. Set up monitoring instruments in Yuan-Yang Lake and Tsui-Fong Lake.

In this study, from September 2017 to March 2019, regular site sampling was conducted, and water quality analysis was conducted in the laboratory. The water quality items analyzed were: suspended solids, total phosphorus, phosphate, organic phosphorus, total Kjeldahl nitrogen, Ammonia nitrogen and organic nitrogen, together with the water quality data provided by the Academia Sinica: water temperature and dissolved oxygen, perform principal component analysis of Yuan-Yang Lake and Tsui-Fong Lake using multivariate statistical analysis methods. The results show that the principal component analysis of Yuan-Yang Lake has 82.051% of the total explanatory variables; the principal component analysis of Tsui-Fong Lake has 82.228% of the total explanatory variables. It means that the principal component analysis of Yuan-Yang Lake and Tsui-Fong Lake has considerable reliability.

**Keywords** : Mountain Lake, Yuan-Yang Lake, Tsui-Fong Lake, Principal Component

## Analysis

### 一、前言

現今社會對於湖泊具有多種使用，例如：飲用水、灌溉、養殖漁業等，對於臺灣水資源常因地形陡峭及河川短促，使得大部分均流入海洋，湖泊無疑是重要的水資源之一。尤其高山湖泊可經由其出水口將淡水帶至下游，並匯集於水庫或河川，成為民生用水，因此對於水質的要求須極為謹慎。然而人為或自然因素的影響，均會使水質產生變化，其中某些因子及反應過程將決定水體目前的生態狀況，在這之中的物理與化學反應通常有著關聯性，因此比起以程序繁複之單變量統計分析，藉由應用多變量統計分析來解釋複雜的變化關係更為適合，本研究為應用多變量統計分析之主成分分析進行高山湖泊的水質評估。

### 二、文獻回顧

湖泊中的物理變化及化學變化並不是獨立發生的，而是彼此息息相關，對於解釋各水質間複雜的關係，僅使用單變量統計分析會過於繁複，因此藉由多變量統計分析來解釋複雜的水質變化是經常使用的方法。於國內的文獻中，張傳恩(2010)以鴛鴦湖做為研究地點，鴛鴦湖位於臺灣中北部，為一個小型的淺水湖，是屬於亞熱帶氣候的高山湖泊，每個月在該地區，進行長時間的水文與水質觀測，以多變量統計分析包括主成份分析進行資料的分析，對鴛鴦湖水質變化的因素有更充分的瞭解。吳政南與柳文成(2013)以水源特定區管理局轄下監測站之水質監測數據，配合氣象局的雨量監測數據，應用多變量統計分析探討水質的特性。楊莉玲(2016)利用多變量統計分析對水質進行分析，結果顯示污染來源主要為懸浮固體和總磷，並更進一步針對懸浮固體和總磷進行相關係數分析，顯示當懸浮固體升高時，總磷含量亦會升高。於國外的文獻中，Palma et al. (2010)應用多變量統計分析評估水質變化，以獲得水質數據集，並使用水質參數對水質進行了監測。

Yenilmez et al. (2011)應用主成份分析、因素分析、判別分析及群集分析，得知湖泊水量的變化會對溶氧、總磷及懸浮固體有著甚大影響。Kim et al. (2016)應用多變量統計分析之主成份分析於印度東海岸吉爾卡湖，探討新建航道後水質的變化情形，主成份分析結果顯示鹽分是吉爾卡湖最大的成份負荷。Kim et al. (2016)使用主成份分析、判別分析及群集分析於韓國洛東河下游，以增強河川水質監測。

### 三、研究區域介紹

本研究之研究區域為位於新竹縣尖石鄉的鴛鴦湖(圖 1)與宜蘭縣太平山的翠峰湖(圖 2)，兩湖泊為臺灣典型的高山湖泊，各自有獨立的集水區，因位於高山地區受到人為的干擾甚小，長期有學術單位在鴛鴦湖與翠峰湖設置監測儀器。

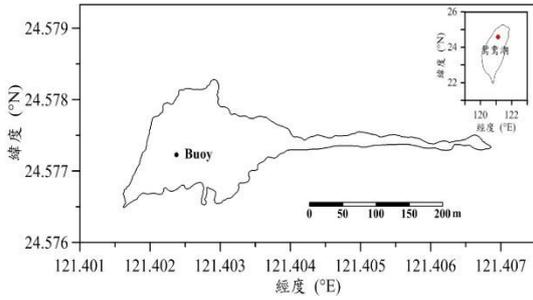


圖 1 鴛鴦湖位置圖

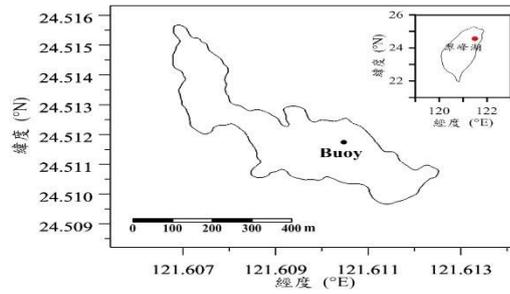


圖 2 翠峰湖位置圖

#### 四、主成份分析方法

##### 與原理

主成份分析是由 Karl Pearson 於 1901 年提出，再由 Harold 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的大部分變異，亦即將原本大量的變數予以減少，使其變為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。由於主成份分析主要依賴數據提供的訊息，所以數據的準確性對分析結果影響很大。

理論基礎如下(考慮  $p$  個變數的線性組合)：

$$z_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p$$

$$z_2 = w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p$$

⋮

$$z_p = w_{p1}x_1 + w_{p2}x_2 + \dots + w_{pp}x_p \text{ 或表示為}$$

$$z = Ax \quad (2)$$

式中， $z = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$ ， $x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ ， $A = \begin{bmatrix} 0 & w_{12} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix}$ ， $z_1, z_2, \dots, z_p$  是  $p$  個主成份，

$w_{ij}$  是第  $j$  個變數在第  $i$  個主成分的權重。

假設變數  $z_1, z_2, \dots, z_p$  間彼此「不相關」(Uncorrelated)，且均數為零，則其共變異矩陣為對角化矩陣，即：

$$\Sigma_z = E_{(zz^T)} = AE_{(xx^T)}A^T = \begin{bmatrix} 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \quad (3)$$

因已假設均數為零，其共變異矩陣與相關矩陣相同。若再假設對稱矩陣 $\Sigma_x$ 的特徵值 (Eigenvalue) 與特徵向量 (Eigenvector) 為 $\lambda_1 > \lambda_2 > \dots > \lambda_p$  (依大小)，

$v_1, v_2, \dots, v_p$  根據上述定理，式(3)可以改寫為下式：

$$\Sigma_x = \begin{bmatrix} 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ & & & \lambda_p \\ \vdots & & & \vdots \end{bmatrix}, \text{ 並且 } A^T = [v_1 \ v_2 \ \dots \ v_p] \quad (4)$$

由 $z = Ax$ ，新變數可以寫成下式：

$$\begin{aligned} z_1 &= v_1(1)x_1 + v_1(2)x_2 + \dots + v_1(p)x_p = v_1^T x \\ z_2 &= v_2(1)x_1 + v_2(2)x_2 + \dots + v_2(p)x_p = v_2^T x \\ &\vdots \\ z_p &= v_p(1)x_1 + v_p(2)x_2 + \dots + v_p(p)x_p = v_p^T x \end{aligned} \quad (5)$$

當其變異數分別為 $\lambda_1, \lambda_2, \dots, \lambda_p$ 時，稱為原始變數共變異矩陣的頻普解構 (Spectral Decomposition)，可由式(3)改寫為下式：

$$\Sigma_x = A^T \Sigma_z A = \sum_{k=1}^p \lambda_k v_k v_k^T \quad (6)$$

## 五、主成份分析結果與討論

### 5.1 鴛鴦湖之主成份分析

於鴛鴦湖所求出之參數與參數間的相關係數矩陣，如表 1 所示。係以相關係數矩陣進行分析，可減少由人為主觀意思所造成之影響，是一種較為客觀的分析法。由相關係數矩陣可得知，總磷與有機磷呈現高度的正相關。磷酸鹽與氨氮、有機磷與有機氮皆呈現較顯著的正相關。總凱氏氮與有機氮呈現高度的正相關。氨氮與有機氮呈現負相關。有機氮與 TSS 呈現正相關。TSS 與水溫呈現正相關；與溶氧呈現負相關。最後水溫與溶氧呈現高度的負相關。

表 1 鴛鴦湖水質參數之相關係數矩陣

	總磷	磷酸鹽	有機磷	總凱氏氮	氨氮	有機氮	TSS	水溫	溶氧
總磷	1.000								
磷酸鹽	-0.327	1.000							

有機磷	0.999	-0.364	1.000						
總凱氏氮	0.375	-0.057	0.372	1.000					
氨氮	-0.482	0.559	-0.499	0.208	1.000				
有機氮	0.636	-0.381	0.643	0.819	-0.390	1.000			
TSS	0.222	-0.204	0.228	0.334	-0.148	0.401	1.000		
水溫	0.275	0.239	0.261	0.432	0.294	0.234	0.575	1.000	
溶氧	-0.474	-0.329	-0.453	-0.335	-0.261	-0.162	-0.408	-0.828	1.000

進而計算相關係數矩陣的特徵值、解釋變異量(表 2)與主成份負荷矩陣(表 3)。由表 2 可得知，雖然鴛鴦湖水質參數可得到九個主成份，但為了達到減化維度的目的，利用特徵值大於 1 的方式進行主成份的選取，本研究最後於鴛鴦湖之水質參數選取三個主成份，其總解釋變異量可達到 82.051 %。表 1 鴛鴦湖主成份分析之特徵值及解釋變異量

主成份	特徵值	解釋變異量 (%)	總解釋變異量 (%)
1	3.851	42.793	42.793
2	2.430	27.005	69.798
3	1.103	12.253	82.051
4	0.954	10.598	92.649
5	0.416	4.620	97.269
6	0.173	1.926	99.195
7	0.072	0.805	100.000
8	1.217E-16	1.352E-15	100.000
9	-1.772E-16	-1.969E-15	100.000

表 2 鴛鴦湖主成份負荷矩陣

水質參數	主成份		
	1	2	3
總磷	-0.681	0.613	-0.333
磷酸鹽	0.890	-0.106	-0.349
有機磷	-0.724	0.592	-0.287
總凱氏氮	0.602	0.599	0.443
氨氮	-0.070	-0.528	0.452
有機氮	0.583	0.666	0.333

<b>TSS</b>	-0.633	0.226	0.613
水溫	0.678	0.407	-0.244
溶氧	-0.024	-0.866	-0.057

由表 2 及表 3 可得知，第一主成份的解釋變異量可達 42.793 %，因為磷酸鹽最高，故第一主成份命名為磷素營養鹽成份。第二主成份的解釋變異量為 27.005 %，因為與溶氧呈現較高的相關性，第二主成份則命名為溶氧成份。第三主成份的解釋變異量為 12.253 %，與 TSS 有較高的相關性，第三主成份命名為懸浮固體成份。根據鴛鴦湖水質參數的主成份分析結果可得知，影響鴛鴦湖水體水質的主要成份為磷素營養鹽成份，之後依序分別為溶氧成份以及懸浮固體成份。

## 5.2 翠峰湖之主成份分析

於翠峰湖所求出之參數與參數間的相關係數矩陣，如表 4 所示。由相關係數矩陣可得知，總磷與有機磷、有機氮呈現高度的相關性，表示營養鹽類彼此的影響甚大。磷酸鹽與 TSS、水溫呈現較顯著的相關性。有機磷與 TSS、溶氧呈現較顯著的相關性。總凱氏氮與有機氮呈現高度的正相關。氨氮與有機氮呈現負相關。有機氮與水溫呈現正相關，與溶氧呈現負相關。TSS 與水溫呈現負相關。最後水溫與溶氧呈現高度的負相關。

表 4 翠峰湖水質參數之相關係數矩陣

	總磷	磷酸鹽	有機磷	總凱氏氮	氨氮	有機氮	TSS	水溫	溶氧
總磷	1.000								
磷酸鹽	-0.536	1.000							
有機磷	0.997	-0.596	1.000						
總凱氏氮	-0.510	0.310	-0.170	1.000					
氨氮	-0.336	-0.152	-0.306	-0.198	1.000				
有機氮	-0.970	0.322	-0.103	0.983	-0.376	1.000			
TSS	0.315	-0.735	0.363	-0.057	0.112	-0.075	1.000		
水溫	-0.145	0.659	-0.194	0.414	-0.096	0.409	-0.406	1.000	
溶氧	-0.462	0.106	-0.449	-0.44	0.219	-0.457	-0.261	-0.602	1.000

進而計算相關係數矩陣的特徵值、解釋變異量(表 5)與主成份負荷矩陣(表 6)。由表 5 可得知，雖然翠峰湖水質參數可得到九個主成份，但為了達到減化維度的目的，利用特徵值大於 1 的方式進行主成份的選取，本研究最後於翠峰湖之水質

參數選取三個主成份，其總解釋變異量可達到 82.228 %。表 5 翠峰湖主成份分析之特徵值及解釋變異量

主成份	特徵值	解釋變異量 (%)	總解釋變異量 (%)
1	3.349	37.212	37.212
2	2.786	30.956	68.169
3	1.265	14.059	82.228
4	0.989	10.987	93.215
5	0.443	4.920	98.135
6	0.145	1.615	99.750
7	0.023	0.250	100.000
8	1.190E-17	1.323E-16	100.000
9	-3.276E-16	-3.640E-15	100.000

表 6 翠峰湖主成份負荷矩陣

水質參數	主成份		
	1	2	3
總磷	0.870	-0.230	-0.361
磷酸鹽	-0.331	0.718	-0.219
有機磷	0.871	-0.257	-0.347
總凱氏氮	0.667	0.275	0.599
氨氮	0.833	-0.207	0.433
有機氮	0.569	0.232	0.227
TSS	0.553	0.728	-0.075
水溫	-0.349	0.794	0.223
溶氧	-0.580	-0.681	0.390

由表 5 及表 6 得知，第一主成份的解釋變異量可達 37.212 %，其中總磷與有機磷最高，故第一主成份命名為磷素營養鹽成份。第二主成份的解釋變異量為 30.956 %，其中水溫最高，第二主成份則命名為水溫成份。第三主成份的解釋變

異量為 14.059 %，與總凱氏氮有較高的相關性，第三主成份命名為總凱氏氮成份。根據翠峰湖水質參數的主成份分析結果可得知，影響翠峰湖水體水質的主要成份為磷素營養鹽成份，之後依序分別為水溫成份以及總凱氏氮成份。

## 六、結論

本研究應用多變量統計分析之主成分分析分別進行高山湖泊鴛鴦湖與翠峰湖之水質參數評估，可獲致以下幾點結論：

1. 鴛鴦湖水質參數可擁有 82.051 %的總解釋變異量，獲得三個主成份，分別為第一主成份-磷素營養鹽成份，第二主成份-溶氧成份與第三主成份-懸浮固體成份。
2. 翠峰湖水質參數可擁有 82.228 %的總解釋變異量，獲得三個主成份，分別為第一主成份-磷素營養鹽成份，第二主成份-水溫成份與第三主成份-總凱氏氮成份。
3. 鴛鴦湖與翠峰湖水質參數的主成份分析結果顯示兩高山湖泊的總解釋變異量均可達到 80.0 %以上，則表示本研究之鴛鴦湖與翠峰湖主成份分析結果具有相當的可靠性。

## 參考文獻

1. 吳政南、柳文成，2013，「應用多變量統計分析評估臺北水源特定區集水區水質」，農業工程學報，59卷2期，第29-50頁。
2. 張傳恩，2010，「應用多變量統計分析評估鴛鴦湖水質」，國立聯合大學土木與防災工程學系碩士班碩士論文。
3. 楊莉玲，2016，「莫拉克颱風前後旗山溪流量、總磷、懸浮固體與氮氮之研究探討」，國立屏東科技大學土木工程系碩士學位論文。
4. Kim, M., Kim, Y., Kim, H., Piao, W., Kim, C., 2016. Enhanced monitoring of water quality variation in Nakdong River downstream using multivariate statistical techniques. *Desalination and Water Treatment*, 57(27), 1-10.
5. Kim, J. Y., Bhatta, K., Rastogi, G., Muduli, P. R., Do, Y., Kim, D. K., Pattnaik, A. K., Joo, G. J., 2016. Application of multivariate analysis to determine spatial and temporal changes in water quality after new channel construction in the Chilika Lagoon. *Ecological Engineering*, 90, 314-319.
6. Palma, P., Alvarenga, P., Palma, V. L., Fernandes, R. M., Soares, A. M. V. M., Barbosa, I. R., 2010. Assessment of anthropogenic sources of water pollution using multivariate statistical techniques: a case study of the Alqueva's reservoir, Portugal. *Environmental Monitoring and Assessment*, 165(1-4), 539-552.
7. Yenilmez, F., Keskin, F., Aksoy, A., 2011. Water quality trend analysis in Eymir Lake, Ankara. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(5-6), 135-140.