

以模式選擇消除共線性之最佳水質指標研究

陳鴻烈⁽¹⁾ 蔡大偉⁽²⁾ 呂銘淵⁽³⁾

摘要

本研究主要目的在探討水質指標模式之共線性，希望藉由減輕模式之共線性來協助集水區水質管理指標的選擇。我們選用淡水河流域為研究區域，在蒐集水質資料後，先使用複迴歸分析法建立 RPI (River Pollution Index)、WQI (Water Quality Index) 兩種水質指標之迴歸模式，然後再針對模式進行共線性分析，在確認共線情形後，利用迴歸選擇法消除共線，並比較兩個水質指標的分析結果。研究顯示，RPI 模式中以逐步迴歸 (向後選擇法) 所得結果最佳，若以生化需氧量(BOD)與氨氮來替代全模式，其模式 R^2 值為 0.650，經調整後為 0.608，高於原模式之 0.451。將所得模式進行共線性分析後，結果均指出此模式無共線問題，證實此分析法有助於減輕共線問題之影響。RPI 精簡模式以 BOD 與氨氮來代表原模式，顯示在此研究區域中溶氧與懸浮固體之重要性相對偏低。在 WQI 分析結果中仍以逐步迴歸 (向後選擇法) 所得結果最佳，該模式以電導度、溶氧、硝酸鹽、凱氏氮來替代全模式， R^2 值為 0.869，調整後為 0.834，亦高於原模式之 0.811。在精簡模式之共線分析中，我們發現此模式仍存在著共線問題，然相較於原模式，已有明顯改善。將 WQI 分析結果與 WQI 原始模式進行比較，發現淡水河流域海水影響與氮營養鹽污染之重要性特別高，其嚴重性亦高於有機物之污染情形。此外，在比較 RPI 與 WQI 之最佳迴歸模式結果後，發現兩模式對於處理共線性有不同的趨勢，RPI 化繁為簡趨向使用單一參數解釋，WQI 則偏向去除共線最強之參數。由兩指標之選擇參數結果可知，WQI 所選參數之周延性與正確性均優於 RPI，此一結果可提供選擇最佳水質指標之重要資訊。

(關鍵詞：RPI、WQI、共線性、迴歸選擇、集水區管理)

Comparison Study of the Water Quality Indexes by Model Selection Analysis

Paris Honglay Chen⁽¹⁾ *David D-W. Tsai*⁽²⁾ *Ming-Yuan Lu*⁽³⁾

Professor⁽¹⁾ and Graduate Student⁽²⁾, Department of Soil and Water Conservation, National Chung-Hsing University, Taiwan

Staff⁽³⁾, Combined Logistics Quarterly Military Supplies Reserve Center, Taiwan

⁽¹⁾ 國立中興大學水土保持學系教授(通訊作者 e-mail: hlchen@dragon.nchu.edu.tw)

⁽²⁾ 國立中興大學水土保持學系博士班研究生

⁽³⁾ 聯勤儲備中心參謀

ABSTRACT

The main purpose of this study was to eliminate collinearity and to compare the mechanism of the water quality models with model selection method, which could be helpful for index selection in watershed management. We selected the Dansue River Watershed as a study site and collected all water quality data available. Multiple regression analysis was applied to build up the regression models in each water quality index: RPI and WQI. Each model was examined by collinearity analysis. After the verification of collinearity in the models, we tried to apply the model selection method to solve the problem. We found out several important results in this study: (1) In the RPI model, the best model was selected by stepwise (backward) regression and the result chose BOD and ammonia to substitute the original index model. The adjusted R^2 of 0.608 was better than 0.451 of the initial model. (2) After removal of the collinearity, all the models exhibited there was no significant collinearity existed any more. This result proved the model selection method was helpful to reduce the collinearity. (3) Suspended solid and dissolved oxygen of RPI index were not as important as the selected parameters in this watershed. (4) WQI model showed that the best model was screened by stepwise (backward) regression and the model was composed by electrical conductivity, dissolved oxygen, nitrate and Kjeldahl. Compared to the original model of adjusted R^2 of 0.811, the simplified model of 0.834 was promoted. (5) The collinearity tests of the filtered WQI model showed the collinearity still existed but the degree was improved. (6) According to the selected parameters of the best WQI model, the salty water and nitrogen pollution effects were noticeable in the study area. (7) Compared to the analysis results of the RPI and WQI models, there were different trends to deal with the collinearity in each index. In RPI, it seemed to prefer using single parameter to explain the water quality. Contrarily, WQI model tended to eliminate collinearity by taking out the most serious collinear parameter. Furthermore, from the parameters selection results, the WQI model performed better integrality and accuracy than the RPI model. Consequently, in this study we provided important information for choosing a better index in water quality management.

(Key words: RPI, WQI, Collinearity, Model selection, Watershed management)

前言

在水土保持研究中，如何保護用水的品質乃重要之研究主題，只有同時重視水資源保護與運用下，才能朝永續水資源方向持續邁進。在水質管理上，水質指標被視為一經濟有效的管理工具，目前台灣最被廣為運用的指標為 RPI (River Pollution

Index) 與 WQI (Water Quality Index) (楊，2001)。

本研究以兩個水質指標所建構的複迴歸模式為基礎進行分析，然而在此模式存有共線情形，共線性 (collinearity) 將造成模式變數解釋力重疊，嚴重影響該模式解釋水質變化的能力及正確程度；因此本研究希望可藉由迴歸模式選擇法，嘗試降低水質指標模式之共線性，並提升指標解釋

能力，增加對模式動力之了解。研究結果期望可對最佳水質指標之選擇提供有效資訊，選擇最佳指標以提高管理效率，改善水體品質，使水資源能朝永續利用方向發展。

理論分析

本研究旨在運用迴歸選擇模式降低水質指標之共線性問題，並藉此尋求複迴歸模式之解釋機制，以及比較台灣現行最常用的兩個水質指標 RPI 與 WQI。詳細算法可參考王 (2001)、吳 (2002)、Brown et al. (1970) 等研究。

研究中所使用的迴歸選擇模式分析方法包括 (1) 順向選擇迴歸、(2) 逆向選擇迴歸與 (3) 逐步迴歸等 3 種模式，均為常見之迴歸選擇模式，詳細相關的理論與計算方法可參考彭 (1998)、陳等 (2003)、林 (2004)、Carlona et al. (2004)、羅 (2006)、陳與蔡 (2007) 等文獻。

研究方法

本研究主要目的在探討如何降低水質指標模式之共線性，並藉由分析過程瞭解模式解釋動力。研究對象選擇現行最重要的兩個水質指標 RPI 與 WQI 進行研究。研究區域是以淡水河流域為代表，因為此流域水質監測相關資料最豐富完整，且大台北都會區即位於其中，其重要性足具代表。

研究中以淡水河流域歷年水質監測資料為主，在考量各監測站地理位置與資料完整性後，選擇屈尺堰、華中橋、關渡與淡水河口等 4 個測站作為代表。鑑於各站建立時間不一，因此探討時間定為民國 82~96 年間之水質變化。關於各水質參數的資料以及流域各段水質之研究可參閱陳等 (2009a)。

本研究期望透過複迴歸選擇模式分析來

減少水質模式之共線程度，而 RPI 與 WQI 已被證實具有顯著共線性 (陳等，2010)，而該研究所探討之共線程度即為本研究分析基礎，研究中共選用 3 種迴歸模式進行分析，最後以兩指標個別之分析結果為基礎，綜合比較 RPI 與 WQI，期能作為最佳水質指標選擇之參考。研究流程如圖 1 所示。

結果與討論

一、RPI

(一) 複迴歸模式

本研究針對 RPI 之複迴歸模式進行共線性探討，RPI 模式已被研究可藉由電導度、溶氧、生化需氧量(BOD)、化學需氧量(COD)、大腸桿菌群、氨氮、總磷、凱氏氮等 8 個水質參數所建構，該模式經檢定達 5% 顯著水準，為一合理模式， R^2 為 0.682 (陳等，2009b)。

(二) 共線性分析

若複迴歸模式中具有嚴重共線性問題，則模式之可信度與解釋能力均會顯著下降 (Maddala, 2005)，造成變數間的概念區隔模糊且難以解釋 (汪與鄭，2009)，因此必須透過共線檢定來瞭解複迴歸模式共線程度。陳等 (2010) 透過共線檢定證明該迴歸模式具有顯著共線性，其結果如表 1。

(三) 迴歸選擇模式分析

在進行共線性分析後，發現此模式具明顯共線性趨勢，而在去除共線性方法中，精簡複迴歸模式為常用方法之一，藉由複迴歸模式選擇法，篩選去除模式中具嚴重共線性參數，藉此減輕模式共線性程度，甚至提昇模式解釋能力。

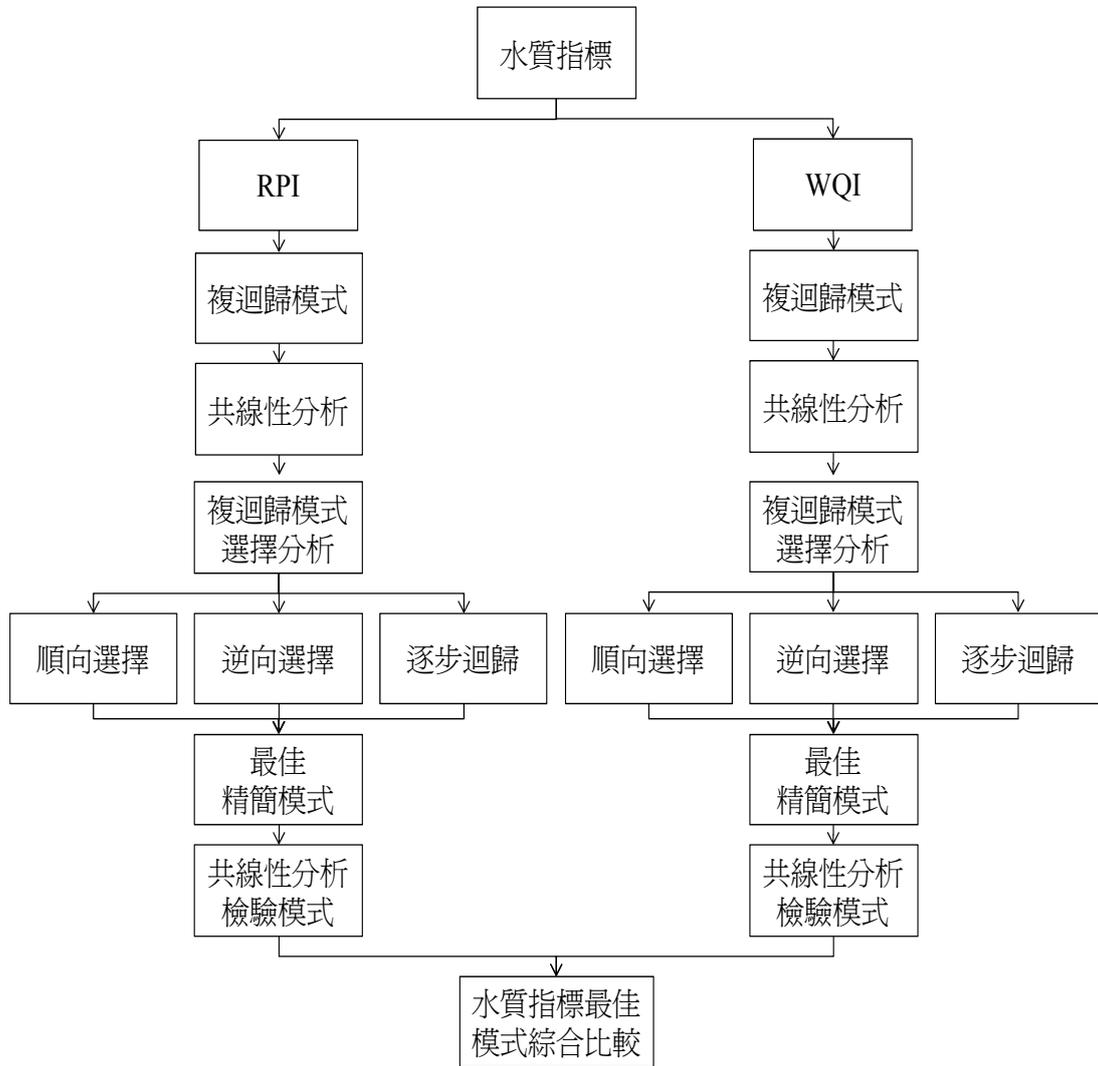


圖 1 研究流程圖

Figure 1 Overview of this study

一般常用複迴歸模式選擇法包括 (1) 向前選擇法、(2) 向後選擇法、(3) 逐步迴歸法。最後再依據分析結果挑選出最佳模式，然後再次進行共線性檢定，以證實迴歸選擇法對共線性消除之效用。

1. 向前選擇法

由表 2 可知，向前選擇法分析過程中依序選入氨氮(NH₄⁺)、BOD、COD 三個

參數，其中以氨氮最為重要，R² 值為 0.432，約是 BOD (R²=0.218) 的兩倍。經過 F 檢定後，發現氨氮與 BOD 均達 5% 顯著水準，COD 則否，表示模式中主要組成為氨氮與 BOD，COD 顯著性與貢獻度較低，推論原因為氨氮與 BOD 為 RPI 之主要組成因子。

表 1 RPI 模式之不同共線性分析法結果

Table 1 Results of different collinearity analysis for the RPI model

分析法	共線情形/共線參數
複迴歸模式迴歸分析	√
複迴歸模式自變數間相關分析	BOD、大腸桿菌群
	電導度、溶氧、COD、氨氮、凱氏氮
容忍度	×
變異數波動因素	×
行列式值	√
條件指數	√
變異數比例	氨氮、凱氏氮

註：√：分析結果為模式中具有顯著共線性

×：分析結果為模式中不具顯著共線性

向前選擇法模式之統計分析結果如表 3，顯示以氨氮、BOD、COD 所建構之迴歸模式經 F 檢定後，F 值為 10.94，p-value 為 <0.001，達 5% 顯著水準，表示此迴歸模式為合理模式，其解釋能力 R² 為 0.672，調整後為 0.611，優於全模式之 0.451 (陳等，2009b)。推論原因為篩去重要性較低或共線性強的參數後，模式解釋能力有所提昇。

向前選擇法所建構之複迴歸模式，其迴歸係數分析結果為表 4，該迴歸模式為：

$$Y=2.703+0.424X_1-0.04X_2+0.534X_3 \quad (1)$$

Y：RPI X₂：COD
X₁：BOD X₃：氨氮

經過 t 檢定後，BOD 與氨氮迴歸係數均達 5% 顯著水準，表示此兩參數均為模式預測重要參數；COD 未達顯著，表示其

重要性較低，推論原因為 COD 與 BOD 之解釋變異有相當程度的重疊。在比較 3 個參數之標準化迴歸係數後，同樣發現模式預測時以氨氮為最重要，標準化係數值 0.681，BOD (0.498) 次之，COD (-0.224) 最低。

2. 向後選擇法

向後選擇法分析過程如表 5，依序篩去電導度(EC)、大腸桿菌群(FC)、溶氧(DO)、總磷(TP)、凱氏氮(KN)、COD 等 6 個參數，6 個參數經 F 檢定均未達 5% 顯著水準，且 R² 均小於 0.1，對於解釋 RPI 的能力均不顯著。在篩去 6 個參數後，模式整體 R² 僅降低 0.03，因此，最後在使用向後選擇法時，僅保留氨氮與 BOD 兩個參數。

表 6 顯示由氨氮與 BOD 所構成之模式經 F 檢定後，F 值為 15.76，p-value 為 <0.001，達 5% 顯著水準，表示此模式為合理迴歸模式。其 R² 值為 0.650，調整後為 0.608，同樣優於全模式之 0.451 (陳等，2009b)，表示此分析法應有助於降低迴歸模式共線性。

使用向後選擇法所得之迴歸模式，其迴歸係數分析結果為表 7，而迴歸模式為：

$$Y=2.39+0.415X_1+0.406X_2 \quad (2)$$

Y：RPI X₂：氨氮
X₁：BOD

模式中 BOD 與氨氮經 t 檢定後，均達 5% 顯著水準，表示進行模式預測時，兩者均為重要參數。而藉由標準化迴歸係數值得知，氨氮數值為 0.518，大於 BOD 之 0.487，表示預測時氨氮重要性略大於 BOD。

3. 逐步迴歸法

逐步迴歸法分析過程如表 8，由表中資料可知分析過程中依序選入氨氮與 BOD，其分析結果與使用向後選擇法相同

，將全部迴歸分析選擇法之結果整理為表 9，由各種迴歸選擇法之篩選參數來看，向後選擇法與逐步迴歸法均選擇 BOD 與氨氮來代表迴歸全模式，而向前選擇法則多加入了 COD 這個參數。依據原全模式共線性分析結果，可能潛在共線性問題有兩群參數：(1) BOD 與大腸桿菌群、(2) 電導度、溶氧、

COD、氨氮、凱氏氮，其中又以氨氮與凱氏氮共線性最嚴重，因此將兩種分析結果相比較，可發現經迴歸選擇法後，模式選用 BOD 來包含大腸桿菌群之解釋量，而採用氨氮或是氨氮與 COD 來包含電導度、溶氧、COD、氨氮、凱氏氮之全部解釋量，藉此減少共線性之影響。

表 2 RPI 向前選擇法分析過程

Table 2 Results of forward regression analysis for the RPI model

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	NH ₄ ⁺	1	0.432	0.432	3.643	13.68	0.002
2	BOD	2	0.218	0.650	-1.884	10.56	0.005
3	COD	3	0.023	0.672	-0.665	1.10	0.309

表 3 RPI 向前選擇法模式之變異數分析結果

Table 3 ANOVA of forward regression method for the RPI model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	Adj R-Sq
Model	3	11.981	3.994	10.94	<0.001	0.672	0.611
Error	16	5.844	0.365				
Corrected Total	19	17.825					

表 4 RPI 向前選擇法迴歸係數分析

Table 4 Parameter estimation of the forward regression analysis for the RPI model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	2.703	0.545	4.96	<0.001	0
BOD	1	0.424	0.128	3.33	0.004	0.498
COD	1	-0.040	0.038	-1.05	0.309	-0.224
NH ₄ ⁺	1	0.534	0.169	3.16	0.006	0.681

表 5 RPI 向後選擇法分析過程

Table 5 Results of backward regression analysis for the RPI model

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	EC	7	<0.001	0.682	7.010	0.01	0.923
2	FC	6	0.001	0.681	5.032	0.02	0.878
3	DO	5	0.001	0.680	3.065	0.04	0.847
4	TP	4	0.002	0.678	1.119	0.07	0.799
5	KN	3	0.006	0.672	-0.665	0.29	0.597
6	COD	2	0.023	0.650	-1.884	1.10	0.309

表 6 RPI 向後選擇法模式之變異數分析結果

Table 6 ANOVA of backward regression method for the RPI model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	Adj R-Sq
Model	2	11.579	5.789	15.76	<0.001	0.650	0.608
Error	17	6.246	0.367				
Corrected Total	19	17.825					

表 7 RPI 向後選擇法迴歸係數分析

Table 7 Parameter estimation of the backward regression analysis for the RPI model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	2.390	0.458	5.22	<0.001	0
BOD	1	0.415	0.128	3.25	0.005	0.487
NH ₄ ⁺	1	0.406	0.117	3.46	0.003	0.518

表 8 RPI 逐步迴歸法分析過程

Table 8 Results of stepwise regression analysis for the RPI model

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	NH ₄ ⁺	1	0.432	0.432	3.643	13.68	0.002
2	BOD	2	0.218	0.650	-1.884	10.56	0.005

表 9 RPI 全部迴歸分析選擇法分析結果比較

Table 9 Comparison of all the RPI regression models

迴歸方法	variables	F value	Pr>F	R-Square	Adj R-Sq
向前選擇	BOD、COD、氨氮	10.94	<0.001	0.672	0.611
向後選擇	BOD、氨氮	15.76	<0.001	0.650	0.608
逐步迴歸	BOD、氨氮	15.76	<0.001	0.650	0.608
迴歸全模式	EC、DO、BOD、COD、FC、NH ₄ ⁺ 、TP、KN	2.95	0.050	0.682	0.451

在解釋變異度能力方面，由於各迴歸模式組成參數量不一，因此使用調整後 R² 值進行比較。向前選擇法分析結果數值為 0.611，向後選擇法與逐步迴歸同樣為 0.608，明顯均優於原全模式之 0.451，顯示迴歸選擇法能有效減輕共線性之影響，使得模式解釋能力提昇。三種選擇法分析結果極為相近，但由於向前選擇法多納入 COD 參數，因此模式解釋能力略為提昇。

(四) 選擇最佳模式

比較 3 種迴歸選擇法之分析結果 (表 9)，發現所選擇的參數十分相似，均含 BOD 與氨氮，僅向前選擇法多加入 COD。至於模式解釋能力，若依據調整後 R² 值檢視，亦十分接近；但若考慮經濟效率，則向前選擇法在加入 COD 後，調整後之 R² 值僅上升 0.4%，因此在經濟考量下，仍以逐步迴歸 (向後選擇法) 分析結果最佳。

(五) 共線性檢定

本研究延續前述步驟，仍透過 5 種方法來檢驗經迴歸選擇後，所選擇之迴歸模

式共線性程度是否有顯著下降。檢定方法包括：(1) 複迴歸模式迴歸分析、(2) 複迴歸模式自變數間相關分析、(3) 容忍度與變異數波動因素、(4) 行列式值與條件指數、(5) 變異數比例。

1. 複迴歸模式迴歸分析

由逐步迴歸選擇之模式迴歸分析結果 (表 6、7) 來觀察，該迴歸模式僅用兩個水質參數構成，其 R² 即已達 0.650，而迴歸係數檢定結果亦顯示 BOD 與氨氮均達 5% 顯著水準，均為模式預測之重要參數，因此估計共線性已有某種程度的減輕，但無法證實 BOD 與氨氮間是否仍存在共線性問題。

2. 複迴歸模式自變數間相關分析

由逐步迴歸選擇之自變數相關分析可知 BOD 與氨氮間相關係數為 0.286，相關係數檢定結果 p-value 為 0.221 (> 0.05)，因此兩者為非顯著相關變數，可能較無共線性問題。

3. 容忍度與變異數波動因素

逐步迴歸所篩選之模式在經過容忍度與變異數波動因素分析後，結果為表 10，表中容忍度均為 0.918，趨近於 1，而變異數波動因素均為 1.09 (< 10)，因此依據指標值判斷結果 (江，2000；楊與林，2005) 可知，兩項分析結果均指出此模式無共線性之問題。

4. 行列式值與條件指數

行列式值與條件指數分析結果如表 11，表中數據顯示其行列式值最小為 0.049，而其條件指數最大值為 7.554 (< 30)，根據共線性判斷標準 (Stevens，1996；盧與曾，2004)，

兩種分析結果均證明此複迴歸模式無共線性存在。

5. 變異數比例

使用變異數比例分析後，所得結果為表 12，由表中數據得知兩參數之數值均無同時 > 0.5 的情形，因此證明兩者無顯著共線性存在 (Katz et al.，2005)。

經過以上 5 類共線性分析後，一致證明經逐步迴歸選擇後之複迴歸模式無顯著共線性問題存在，亦證明使用迴歸選擇法精簡原模式後，可有效減輕共線性問題。

表 10 RPI 逐步迴歸模式之容忍度與變異數波動因素分析結果

Table 10 Colinearity tests of tolerance and variance inflation for the stepwise RPI regression model

Variable	Tolerance	Variance Inflation
Intercept	—	0
BOD	0.918	1.089
NH ₄ ⁺	0.918	1.089

表 11 RPI 逐步迴歸模式之行列式值與條件指數分析結果

Table 11 Colinearity tests of eigenvalue and condition index for the stepwise RPI regression model

Number	Eigenvalue	Condition Index
1	2.805	1.000
2	0.146	4.384
3	0.049	7.554

表 12 RPI 逐步迴歸模式之變異數比例分析結果

Table 12 Colinearity tests of variance proportions for the stepwise RPI regression model

Number	Proportion of Variation		
	Intercept	BOD	NH ₄ ⁺
1	0.011	0.011	0.024
2	0.095	0.099	0.976
3	0.895	0.891	<0.001

最後經逐步迴歸選擇後，只留下 BOD 與氨氮來代表原來的複迴歸全模式，由此可見 RPI 其他兩項組成（溶氧與懸浮固體）之重要性較低，其中溶氧的解釋部份又可被 BOD 所涵蓋。

二、WQI

針對 WQI 的部份，研究中透過與 RPI 相同的分析過程，分析 WQI 模式經迴歸選擇後，檢驗其共線程度之變化。

(一) 複迴歸模式

將 WQI 與一般水質參數進行複迴歸分析，可得到一合理模式，此模式可由 pH、電導度、溶氧、氨氮、總磷、硝酸鹽、凱氏氮等 7 個參數來組成，經 F 檢定後，達 5% 顯著水準，而 R² 等於 0.880 (陳等，2009b)。

Table 13 Results of different collinearity analysis for the WQI model

分析法	共線情形/共線參數
複迴歸模式迴歸分析	√
複迴歸模式自變數間相關分析	電導度、溶氧、氨氮、硝酸鹽、凱氏氮
容忍度	x
變異數波動因素	x
行列式值	√
條件指數	√
變異數比例	氨氮、凱氏氮

註：√：分析結果為模式中具有顯著共線性

×：分析結果為模式中不具顯著共線性

(二) 共線性分析

經過與 RPI 相同之共線性統計檢定

後，其檢定結果如表 13，WQI 模式同樣被證明具顯著共線性 (陳等，2010)。所以本表 13 WQI 模式之不同共線性分析法結果研究同樣嘗試藉由迴歸選擇模式分析來減輕 WQI 模式之共線程度。

(三) 迴歸選擇模式分析

在複迴歸模式選擇法中，分析結果同樣包括 (1) 向前選擇法、(2) 向後選擇法、(3) 逐步迴歸法。

1. 向前選擇法

向前選擇法之分析過程如表 14，過程中依序選擇凱氏氮、溶氧、電導度、硝酸鹽、總磷等 5 個參數納入迴歸模式，參數中以凱氏氮最為重要，其 R² 值為 0.711 最高，其次為溶氧的 0.090，其他參數則都低於 0.05，而經過 F 檢定後，亦只有凱氏氮與溶氧達 5% 顯著水準，表示其他參數之重要性相對較低。

向前選擇法模式之統計分析結果如表 15，經 F 檢定後，F 值為 20.51，p-value < 0.001，達 5% 顯著水準，為一合理模式。模式 R² 值為 0.880，調整後為 0.837，優於全模式之 0.811 (陳等，2009b)，代表篩去 pH 與氨氮後，有助於模式減輕共線性，使模式解釋能力略為提昇。

使用向前選擇法所得之複迴歸模式的迴歸係數分析結果為表 16，該模式為：

$$Y = 69.004 - 0.001X_1 + 1.564X_2 - 5.44X_3 - 12.302X_4 - 1.92X_5 \quad (3)$$

Y：WQI
 X₁：電導度
 X₂：溶氧
 X₃：總磷
 X₄：硝酸鹽
 X₅：凱氏氮

在迴歸係數分析結果中，溶氧與 WQI

值呈現正比狀態，表示溶氧值上升時，WQI 值亦上升，即水質變佳，原因為溶氧為反相指標，一般情形為溶氧值高者水質較佳。而各參數進行 t 檢定後，僅總磷未達 5% 顯著，表示進行模式預測時，總磷重要性不顯著；而藉由標準化迴歸係數值比較參數間重要性後，發現係數值最高者為凱氏氮的 -0.491，表示預測時以凱氏氮最為重要。

2. 向後選擇法

向後選擇法之分析過程如表 17，由表中資料得知過程為依序篩去氨氮、pH、總磷，此 3 參數之 R² 值均相當低，表示這些參數對模式解釋能力貢獻不大。將 3 參數刪去後，模式 R² 值僅下降 0.011，而所有參數經 F 檢定後，均未達 5% 顯著水準，亦證明其重要性不顯著。

表 14 WQI 向前選擇法分析過程

Table 14 Results of forward regression analysis for the WQI model

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	KN	1	0.711	0.711	13.033	44.18	<0.001
2	DO	2	0.090	0.801	5.965	7.72	0.013
3	EC	3	0.026	0.827	5.331	2.43	0.139
4	NO ₃ ⁻	4	0.042	0.869	3.140	4.78	0.045
5	TP	5	0.011	0.880	4.045	1.27	0.278

表 15 WQI 向前選擇法模式之變異數分析結果

Table 15 ANOVA of forward regression method for the WQI model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	Adj R-Sq
Model	5	606.400	121.280	20.51	<0.001	0.880	0.837
Error	14	82.772	5.912				
Corrected Total	19	689.172					

表 16 WQI 向前選擇法迴歸係數分析

Table 16 Parameter estimation of the forward regression analysis for the WQI model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	69.004	5.720	12.06	<0.001	0
EC	1	-0.001	<0.001	-2.49	0.026	-0.374
DO	1	1.564	0.531	2.95	0.011	0.457
TP	1	-5.400	4.787	-1.13	0.278	-0.114
NO ₃ ⁻	1	-12.302	5.282	-2.33	0.035	-0.369
KN	1	-1.920	0.714	-2.69	0.018	-0.491

表 17 WQI 向後選擇法分析過程

Table 17 Results of backward regression analysis for the WQI model

Step	Variable Removed	Numbe Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	NH ₄ ⁺	6	<0.001	0.880	6.013	0.01	0.911
2	pH	5	<0.001	0.880	4.045	0.03	0.855
3	TP	4	0.011	0.869	3.140	1.27	0.278

向後選擇法之分析結果為表 18，其複迴歸模式 F 值為 24.87，p-value <0.001，達 5%顯著水準，模式 R² 值為 0.869，調整後為 0.834，優於全模式 0.811 (陳等，2009b)，推論原因為篩去 pH、氨氮、總磷後，有助於模式減輕共線性，使模式解釋能力略為提昇。

經向後選擇法分析所得之複迴歸模式，其迴歸係數分析結果為表 19，由表中數據得知此複迴歸模式為：

$$Y = 67.857 - 0.001X_1 + 1.423X_2 - 11.57X_3 - 2.161X_4 \quad (4)$$

Y : WQI
 X₁ : 電導度
 X₂ : 溶氧
 X₃ : 硝酸鹽
 X₄ : 凱氏氮

選擇之參數經 t 檢定後，均達 5%顯著水準，表示模式全部組成參數均對模式預測具顯著重要性，亦代表參數間共線性不明顯；而檢視標準化迴歸係數可知，對模式預測最重要的參數為凱氏氮，其係數值 -0.552 為最高。

3. 逐步迴歸法

逐步迴歸法分析過程如表 20，由分析結果得知與向後選擇法結果相同。

由各種迴歸選擇法之篩選參數結果 (表 21) 來看，向後選擇法與逐步迴歸法，均選擇電導度、溶氧、硝酸鹽、凱氏氮來代表迴歸全模式，而向前選擇法則多加入總磷這個參數。依據原全模式共線性分析

結果，可能共線性問題潛在於電導度、溶氧、氨氮、硝酸鹽、凱氏氮等 5 個參數，其中又以氨氮與凱氏氮共線性最嚴重，因此將兩種分析結果相比較，可發現經迴歸選擇法後，模式將氨氮刪除，而用其他 4 個參數來包含全部解釋量，藉此減輕共線性之影響。此外，向前選擇法將總磷加入模式，因總磷並未與其他參數解釋量重疊，另一個原因是在被刪除參數中，其為相對重要參數，加入後可增加模式解釋能力，但不影響共線性問題。

在解釋變異度能力方面，由於各迴歸模式組成參數量不一，因此使用調整後 R² 值進行比較。其中，向前選擇法數值為 0.837，向後選擇法與逐步迴歸均為 0.834，皆明顯優於原全模式之 0.811，顯示迴歸選擇法能有效減輕共線性影響，提昇模式解釋能力。雖然三種選擇法分析結果極為相近，但由於向前選擇法多加入總磷，因此模式解釋能力最佳。

(四) 選擇最佳模式

3 種迴歸選擇法之分析結果 (表 21) 所選擇之參數十分相似，均含電導度、溶氧、硝酸鹽、凱氏氮，僅向前選擇法多加入總磷參數。而模式解釋能力，若依據調整後 R² 值檢視，亦十分接近，但在經濟考量下，向前選擇法加入總磷後 R² 值僅上升 0.4%，因此仍以逐步迴歸 (向後選擇法) 分析結果較佳。

表 18 WQI 向後選擇法之變異數分析結果

Table 18 ANOVA of backward regression method for the WQI model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	Adj R-Sq
Model	4	598.876	149.719	24.87	<0.001	0.869	0.834
Error	15	90.295	6.020				
Corrected Total	19	689.172					

表 19 WQI 向後選擇法迴歸係數分析

Table 19 Parameter estimation of the backward regression analysis for the WQI model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	67.857	5.680	11.95	<0.001	0
EC	1	-0.001	<0.001	-2.30	0.037	-0.341
DO	1	1.423	0.520	2.74	0.015	0.416
NO ₃ ⁻	1	-11.570	5.290	-2.19	0.045	-0.347
KN	1	-2.161	0.688	-3.14	0.007	-0.552

表 20 WQI 逐步迴歸法分析過程

Table 20 Results of stepwise regression analysis for the WQI model

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	KN	1	0.711	0.711	13.033	44.18	<0.001
2	DO	2	0.090	0.801	5.965	7.72	0.013
3	EC	3	0.026	0.827	5.331	2.43	0.139
4	NO ₃ ⁻	4	0.042	0.869	3.140	4.78	0.045

表 21 WQI 全部迴歸分析選擇法分析結果比較

Table 21 Comparison of all the WQI regression models

迴歸方法	variables	F value	Pr>F	R-Square	Adj R-Sq
向前選擇	EC、DO、TP、NO ₃ ⁻ 、KN	20.51	<0.001	0.880	0.837
向後選擇	EC、DO、NO ₃ ⁻ 、KN	24.87	<0.001	0.869	0.834
逐步迴歸	EC、DO、NO ₃ ⁻ 、KN	24.87	<0.001	0.869	0.834
迴歸全模式	pH、EC、DO、NH ₄ ⁺ 、TP、NO ₃ ⁻ 、KN	12.61	<0.001	0.880	0.811

(五) 共線性檢定

為檢驗逐步迴歸後迴歸模式之共線性程度，本研究進行共線性分析，包括 (1) 複迴歸模式迴歸分析、(2) 複迴歸模式自變數間相關分析、(3) 容忍度與變異數波動因素、(4) 行列式值與條件指數與 (5) 變異數比例。

1. 複迴歸模式迴歸分析

透過逐步迴歸篩選後之複迴歸模式(結果如表 18 與 19) 是由電導度、溶氧、硝酸鹽、凱氏氮等水質參數所構成，其 R2 高達 0.869，而迴歸係數檢定結果亦顯示全部參數均達 5%顯著水準，均為模式預測之重要參數，由此分析結果得知共線性並不顯著，但參數間共線性程度仍需其他分析法來證實。

2. 複迴歸模式自變數間相關分析

由迴歸自變數間相關分析可知，電導度、溶氧、硝酸鹽、凱氏氮間相關係數檢定為全部互為顯著相關，因此所有參數間仍具有潛在共線性問題。

3. 容忍度與變異數波動因素

將逐步迴歸所選擇之複迴歸模式進行容忍度與變異數波動因素分析(表 22)，由變異數波動因素分析結果得知，所有數值均小於 10，顯示此模式無顯著共線性問題存在。

4. 行列式值與條件指數

依行列式值與條件指數分析(表 23)可知，精簡模式之行列式最小值為 0.007，而以條件指數分析結果觀之，精簡模式之

條件指數最大值為 24.869 (< 30)，由此得知經逐步迴歸所得精簡模式不具顯著共線性問題。

表 22 WQI 逐步迴歸模式之容忍度與變異數波動因素分析結果

Table 22 Colinearity tests of tolerance and variance inflation for the stepwise WQI

regression model

Variable	Tolerance	Variance Inflation
Intercept	-	0
EC	0.396	2.526
DO	0.378	2.644
NO ₃ ⁻	0.347	2.884
KN	0.282	3.541

表 23 WQI 逐步迴歸選擇模式之行列式值與條件指數分析結果

Table 23 Colinearity tests of eigenvalue and condition index for the stepwise WQI

regression model

Number	Eigenvalue	Condition Index
1	4.426	1.000
2	0.477	3.047
3	0.052	9.184
4	0.038	10.845
5	0.007	24.869

表 24 WQI 逐步迴歸選擇模式之變異數比例分析結果

Table 24 Colinearity tests of variance proportions for the stepwise WQI regression model

Number	Proportion of Variation				
	Intercept	EC	DO	NO ₃ ⁻	KN
1	<0.001	0.003	0.002	0.002	0.002
2	<0.001	0.043	0.017	0.025	0.028
3	0.001	0.758	<0.001	0.003	0.449
4	<0.001	<0.001	0.541	0.575	0.005
5	0.998	0.197	0.441	0.395	0.516

變異數比例

逐步迴歸選擇之精簡模式，經變異數比例分析後，結果為表 24，由表中數據得知溶氧與硝酸鹽存在共線性問題，若與原分析結果比較，原迴歸模式分析指出氨氮與凱氏氮具顯著共線性 (陳等，2010)，因此推論原因為模式選用多項與總氮相關參數，包括氨氮、硝酸鹽、凱氏氮等，故彼此之共線性難以完全消除。

使用以上共線性分析方法檢視經逐步迴歸選擇後之模式時，發現此精簡模式中仍存在共線問題，但由複迴歸模式迴歸分析和行列式值與條件指數分析得知，共線性有明顯改善趨勢，因此仍可證實使用迴歸選擇法來精簡原模式，可減輕共線性問題之影響。

迴歸選擇法中最佳模式為逐步迴歸與向後選擇法之精簡結果，此精簡模式是由電導度、溶氧、硝酸鹽、凱氏氮所構成。此結果與 WQI 之組成 (由 pH、溶氧、BOD、濁度、大腸桿菌群、總磷、硝酸鹽等 7 個參數所構成) 來比較，顯示應用在此流域時，BOD、濁度、大腸桿菌群表現不顯著，而加入電導度、氨氮、凱氏氮等 3 個參數，表示此流域海水影響與氮營養鹽污染重要性特別高，嚴重性高於有機物之污染情形；而 pH 與總磷雖與其他相關

性低 (陳等，2010)，但同時在模式中解釋能力亦偏低，因此予以刪除。

三、水質指標比較

RPI 與 WQI 之複迴歸全模式分別經過迴歸選擇法篩選後，均以逐步迴歸 (向後選擇) 之分析結果最佳，將兩指標所得之最佳迴歸模式重要統計資料整理為表 25。首先比較選擇參數結果對照共線性參數 (表 1、表 13)，由篩選結果可知兩指標趨勢不同，其中 RPI 以 BOD 來囊括大腸桿菌群之解釋，以氨氮來囊括電導度、溶氧、COD、凱氏氮之解釋，因此有使用單一參數囊括其他具共線性潛能參數之趨勢；而 WQI 選擇結果則是在電導度、溶氧、氨氮、硝酸鹽氮、凱氏氮等可能有共線趨勢參數中，選擇刪去共線性最強之氨氮，因此具有處理共線性時去除共線最強參數之趨勢。由此比較兩指標迴歸選擇結果，顯示兩指標處理共線性趨勢不同，但若以調整後 R² 來比較，WQI 為 0.834 明顯優於 RPI 的 0.608，由此可知雖然 WQI 為了保有最多的水質資訊而選擇只刪除共線性最強之參數，會導致共線性去除效果較差，但針對水質狀態解釋能力仍較 RPI 為佳。

另外，就水質參數選擇結果來探討，RPI 最終選擇以 BOD 與氨氮來代表，WQI

則選擇電導度、溶氧、硝酸鹽氮、凱氏氮來代表。就水質狀態解釋能力來看，RPI 選擇的 BOD 可用 WQI 中的溶氧來代表，而就總氮組成而言，WQI 所選的硝酸鹽氮、凱氏氮重要性也優於 RPI 所選之氮

，且 WQI 選擇的電導度參數所代表的海水影響或是溶解性鹽類污染訊息，則是 RPI 模式中所欠缺的，因此就最佳迴歸選擇模式之周延性與正確性而言，WQI 均優於 RPI。

表 25 RPI 與 WQI 之最佳迴歸選擇模式比較

Table 25 The comparison between the best regression models of RPI and WQI

水質指標	迴歸方法	variables	F value	Pr>F	R-Square	Adj R-Sq.
RPI	逐步迴歸 (向後選擇)	BOD、NH ₄ ⁺	15.76	<0.001	0.645	0.608
WQI	逐步迴歸 (向後選擇)	EC、DO、 NO ₃ 、KN	24.87	<0.001	0.869	0.834

結論與建議

在透過向前選擇法、向後選擇法、逐步迴歸法分析 RPI 複迴歸全模式後，比較其分析結果可知以逐步迴歸 (向後選擇法) 最佳，可由 BOD 與氨氮來替代複迴歸全模式，其模式 R² 為 0.650，調整後為 0.608，高於原模式之 0.451。將此模式進行共線性分析後，發現所有結果均指出此模式無共線性問題，相較於原複迴歸全模式已有明顯改善之趨勢，亦證實迴歸選擇法有助於減輕共線性問題之影響。

將 RPI 原複迴歸全模式、經逐步迴歸選擇後之精簡模式，以及 RPI 原始組成模式 3 者進行比較，發現淡水河流域之 RPI 指標最後可用 BOD 與氨氮來代表整個模式，由此可見其他兩項組成 (溶氧與懸浮固體) 之重要性較低，其中溶氧的解釋部份又可被 BOD 所涵蓋。

針對 WQI 複迴歸全模式之共線性進行迴歸選擇法分析後，發現以逐步迴歸

(向後選擇法) 所得結果最佳，可由電導度、溶氧、硝酸鹽、凱氏氮來替代複迴歸全模式，其模式 R² 為 0.869，調整後為 0.834，高於原模式之 0.811。但在將此模式進行共線性分析後，發現仍存有共線問題，然相較於原複迴歸全模式，已有明顯改善之趨勢，證實迴歸選擇法有助於減輕共線性問題之影響。

將 WQI 原複迴歸全模式、經逐步迴歸選擇後之精簡模式，以及 WQI 原始組成模式 3 者進行比較，發現淡水河流域海水影響與氮營養鹽污染重要性特別高，其嚴重性高於有機物之污染情形；也因此，氨氮、硝酸鹽、凱氏氮等與總氮相關之水質參數均達顯著，此亦為本模式共線性難以消除之主因。

在比較 RPI 與 WQI 之最佳迴歸模式結果後，發現兩模式對於處理共線性有不同趨勢，RPI 化繁為簡有使用單一參數囊括其他參數之趨勢，WQI 則偏向去除共線最強參數之趨勢。由兩指標之迴歸選擇參數結果，可得知 WQI 所選參數之周延性與

正確性均優於 RPI，由此分析結果可提供協助選出最佳水質指標重要資訊。

參考文獻

1. 王善賢 (2001)，「台灣地區河川水質狀態指標之建立」，碩士論文，台灣大學。
2. 江立雯 (2000)，「電子業的影響因素之研究-以台灣地區上市電子業為例」，碩士論文，中央大學。
3. 吳冬齡 (2002)，「歷年河川水質監測數據之污染程度分析-以中港溪為例」，碩士論文，台灣大學。
4. 汪志忠、鄭雅云 (2009)，「我國縣市長選舉投票率之研究：社會經濟指標之分析」，2009 年會暨全球化下新公共管理趨勢與挑戰-理論與實踐國際研討會，第 1~12 頁。
5. 林志融 (2004)，「棲地適宜性分析應用於生態廊道規劃之研究-以山羌及有勝溪流域為例」，碩士論文，東華大學。
6. 陳正昌、程炳林、陳新豐、劉子鍵 (2003)，「多變量分析方法：統計軟體應用」，五南圖書出版股份有限公司，台北市。
7. 陳鴻烈、蔡大偉 (2007)，「不同優養水質共線性分析及模式選擇之研究」，水土保持學報，第三十九卷，第三期，第 229~246 頁。
8. 陳鴻烈、蔡大偉、呂銘淵 (2010)，「水質指標模式之共線性分析比較研究」，水土保持學報，第四十二卷，第一期，第 35~48 頁。
9. 陳鴻烈、蔡大偉、呂銘淵 (2009a)，「最佳水質指標選擇之研究」，農林學報，第五十八卷，第二期，第 145~162 頁。
10. 陳鴻烈、蔡大偉、呂銘淵 (2009b)，「應用迴歸分析法於最佳水質指標選擇之研究」，農林學報，第五十八卷，第四期，第 307~321 頁。
11. 彭昭英 (1998)，「SAS 與統計分析」，儒林圖書有限公司，台北市。
12. 楊雅梅 (2001)，「台灣水庫集水區水質指標與管理系統建立之研究」，碩士論文，台灣大學。
13. 楊灌園、林頂立 (2005)，「有機畜產品消費者行為之探討-以準有機雞為例」，2005 年第三屆管理思維與實務學術研討會論文集，第 III-247~III-255 頁。

14. 盧昆宏、曾進輝 (2004), 「以 6 σ 之推動因素與執行效益間之關聯性研究」, 陸軍官校八十週年校慶綜合學術研討會論文集研討會, 第 2-1~2-11 頁。
15. 羅慧芬 (2006), 「空氣資源中臭氧變化之系統分析研究」, 碩士論文, 中興大學。
16. Brown, R.M.,N.I. McClelland, R.A. Deininger and R.G. Tozer (1970), "A Water Quality Index-Do We Dare?", Water Sewage Works, 117: 339-343.
17. Carlona, C., M.D. Valle and A. Marcomini (2004), "Regression Models to Predict Water-Soil Heavy Metals Partition Coefficients in Risk Assessment Studies", Environ. Pollut., 127: 109-115.
18. Katz, A., M.V. Hau and J. Mahoney (2005), "Explaining the Great Reversal in Spanish America: Fuzzy-Set Methods Versus Statistical Methods", Department of Sociology, Brown University.
19. Maddala, G.S. (2005), "Introduction to Economics", 3rd Edition, Wiley.
20. Stevens, J. (1996), "Multiple Regression, Applied Multivariate Statistics for the Social Science", 3rd ed., NJ., Lawrence Erlbaum : 64-150.

99 年 06 月 16 日收稿

99 年 07 月 30 日修改

99 年 08 月 09 日接受