

應用模糊聚類目標識別於高屏溪 水質分佈特性之研究

陳鈞華¹ 蔡瀚德² 鄭富尚³ 李煜基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

應用 Matlab 模糊聚類目標識別、模糊減法聚類分析對高屏溪 10 個測點水質參數進行研究。具有方便、快速和提供分析參數多的優點。由模糊聚類目標識別能快速分析新參數值，使參數值迅速分類，根據 10 個測點之間水質分組相似性，將高屏溪分為 3 條流域(旗山溪、荖濃溪、高屏溪)，由模糊減法聚類各流域間差異性越大聚類中心越分散。利用 Matlab 程式設計，提高了判別的速度及精確度。本研究中結合高屏溪流域之地理環境背景，探討目標識別及模糊減法聚類之作用，提供水質分析參考。

關鍵字:水質流域、聚類目標識別、聚類減法分析

一、研究動機與目的

台灣為海島型國家水源非常的豐富，在現今全世界提出乾淨水是國家之重要資源，相對水質分析及管理顯得非常重要，環境水質優劣有許多的評價方法，不過對於輸入一組新參數值，來識別目標所屬的地區或流域型態，在目前之研究應用則尚未看到，應此本研究基於此原因，應用模糊聚類目標識別於水質參數上。

現實環境中，研究者常探討模糊特性之問題，此問題需利用模糊聚類分析對相似資料進行歸屬分類，再運用一組新目標值進行目標識別，來得到合理的詮釋。本研究應用 Matlab 聚類目標識別對水質進行探討，水質對於河川系統中為重要指標，於流域內河川水質的優劣對民生、工業、灌溉用水息息相關。台

灣有著複雜的氣候、地質等自然條件(Gibbs RJ,1970)以及人類在生產、生活(Petts G,Calow,1996)的影響，其流域內在反映出某程度上的基本特徵。關於河川水質能快速尋找出相似之目標識別，對未來水質之問題能多一分幫助。

模糊邏輯(Fuzzy logic)於 1965 年，由美國加州柏克萊大學教授扎德(L.A.Zadeh)提出了模糊數學的概念，並在資訊與控制的專門性國際期刊(Information and Control)發表了第一篇模糊集合(Fuzzy Sets)論文(L.A. Zadeh, 1965)(蔡瀚德與陳鈞華等人，2011)。隨著模糊理論的發展，模糊聚類分析續接模糊理論之精神，聚類分析法其觀測數超出一大範圍值時最為顯著。其若能提早發現最適當之分群數，並可迅速利用目標函數法計算能提高計算的精確度及

運算時間(張與鍾等人, 2003)。

近年來, 應用於水文上之研究者有, (張斐章與徐國麟, 1990)利用模糊集理論推估河川流量; (林敬章, 1992)將模糊理論應用於降雨分類; (張斐章、王文清, 1995)進行模糊線性規劃於水資源規劃之應用; (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—高屏溪流域之應用; (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。

由上述幾篇文獻, 對水文相關研究之對象不少, 但以水質進行聚類目標識別分析之研究則尚未看到。因此本研究利用應用模糊聚類目標識別於水質分佈特性上, 分析高屏溪水質聚類分佈規律及變化原因, 使新目標水質用於聚類目標識別能在水資源科學中提供後續研究者合理之依據(葉與黎等人, 2003)。

二、研究材料與方法

2.1 研究區域調查

2.1.1 地理位置

高屏溪舊名下淡水溪, 位於台灣南部, 為一中央管河川, 長度僅次於濁水溪。主流河長 171.00 公里, 為全台第二長河; 流域面積廣達 3,256.85 平方公里, 分佈於南投縣南端、嘉義縣東端、台東縣西端, 及高雄市、屏東縣的 23 個鄉鎮市區, 流域面積為全台第一大河。

旗山溪為高屏溪水系主流(含荖濃溪)之外的最大支流, 發源於玉山主峰

西南坡標高約 2,700 公尺處, 主流河長 118 公里, 流域面積 842 平方公里。

高屏溪流域內平均雨量每年達 3046 公釐, 平均年逕流量高達 84 億 5500 萬立方公尺。平均年輸砂量是 3561 萬噸, 每平方公里流域面積輸送 10934 噸, 居全世界第 11 位。

(資料來源:維基百科)

本研究探討圖 1 高屏溪流域取 10 個水質監測值(1.甲仙大橋、2.新發大橋、3.六龜大橋、4.西門大橋、5.旗南橋、6.里港大橋、7.里嶺大橋、8.九如橋、9.高屏大橋、10.昌農大橋), 利用此流域 4 個參數值(DO、 BOD_5 、COD、SS)區分聚類間之變量。

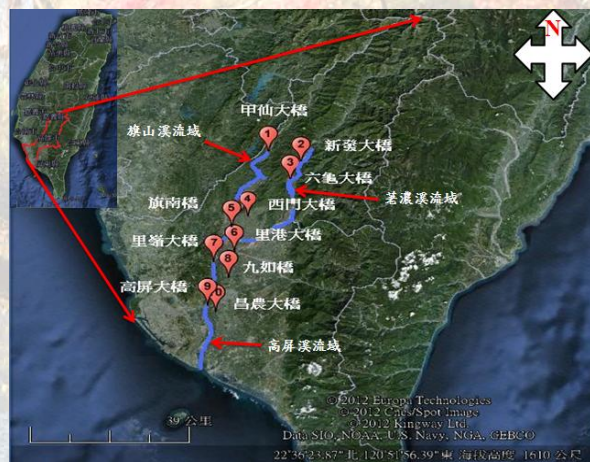


圖 1 地理位置

(資料來源 Google 地球)

2.2 水質觀測數據

本研究採用水質參數監測值, 資料來源於行政院環境保護署-全國環境水質監測資訊網, 利用表 1 中的水質監測數據, 利用 MATLAB 了解模糊聚類之目標識別之相關訊息。

2.3 模糊聚類分析法基本概念

模糊聚類分析法是一種多元統計分類方法, 聚類就是靠事物間的相似性作為分類歸屬劃分的原則, 其屬於無監督分類, 用此方法可以對一群不知類別的觀察對象按彼此相似的程度進行分類, 已達到“物以

類聚”之目的。利用模糊聚類分析法有利於挑選變量，分析影響因素。

設有 n 個樣品，每個樣品測得 p 項指標(變數)，原始資料矩陣為：

$$X = \begin{matrix} & x_1 & x_2 & \cdots & x_p \\ \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{matrix} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \end{matrix} \quad (1)$$

其中 x_{ij} ($i=1, \dots, n; j=1, \dots, p$) 為第 i 個樣品的第 j 個指標的觀測資料，第 i 個樣品 X_i 為矩陣 X 的第 i 行所描述，所以任何兩個樣品 X_k 與 X_l 的相似性，可以通過矩陣 X 中的第 k 行與第 l 行的相似程度來刻畫；任何兩個變數 X_k 與 X_l 之間的相似性，可以通過第 k 列與第 l 列的相似程度來刻畫(劉與許等人，2005)。

2.3 模糊聚類之目標識別

模糊聚類分析法用於目標識別，可將一組新目標函數值，代入 MATLAB 程式碼後，立即顯示新目標參數值為第幾類別及歸屬於哪個區域，使新目標函數值能迅速分類。

2.4 模糊減法聚類

首先必須確認模糊法則之數目，也須更準確地運用訓練樣本中之規律性，應將訓練資料樣本聚類，模型中採用模糊減法聚類。而模糊減法聚類能充分將資料自身之樣本自動聚為幾類，然而使模糊規則數目更加合理化(許國根、許萍萍，2005)。

三、案例應用

3.1 水質參數函數

由全國環境水質監測資訊網取高屏溪 10 個點(1.甲仙大橋、2.新發大橋、3.六龜大橋、4.西門大橋、5.旗南橋、6.里港大橋、

7.里嶺大橋、8.九如橋、9.高屏大橋、10.昌農大橋)，2002~2011 年 1 月 4 個水質參數值(DO、 BOD_5 、COD、SS)，並取平均值，試根據下面數據對這 10 個地區進行分類。

表 1 高屏溪斷面 10 個水質參數監測值

水質參數 地區	DO(mg/L) ^a	BOD_5 (mg/L) ^a	COD(mg/L) ^a	SS(mg/L) ^a
1.甲仙大橋 ^a	9.12476 ^a	0.9 ^a	4.39 ^a	14.42 ^a
2.新發大橋 ^a	8.81926 ^a	0.99 ^a	3.97 ^a	31.62 ^a
3.六龜大橋 ^a	8.5238 ^a	1.26 ^a	4.45 ^a	21.44 ^a
4.西門大橋 ^a	8.51017 ^a	1.53 ^a	6.65 ^a	31.13 ^a
5.旗南橋 ^a	8.0279 ^a	1.75 ^a	7.54 ^a	35.01 ^a
6.里港大橋 ^a	7.98176 ^a	1.42 ^a	7.79 ^a	227.46 ^a
7.里嶺大橋 ^a	7.98082 ^a	1.43 ^a	8.35 ^a	82.88 ^a
8.九如橋 ^a	1.59091 ^a	15.63 ^a	42.51 ^a	45.2 ^a
9.高屏大橋 ^a	6.14643 ^a	7.6 ^a	27.56 ^a	38.42 ^a
10.昌農大橋 ^a	0.69545 ^a	37.5 ^a	92.4 ^a	51.81 ^a

3.2 模糊聚類目標識別模型建立

設表示高屏溪 10 個水質監測站

$$X = \{1, 2, 3, \dots, 10\}$$

其中 1 為甲仙大橋、2 為新發大橋、3 為六龜大橋、4 為西門大橋、5 為旗南橋、6 為里港大橋、7 為里嶺大橋、8 為九如橋、9 為高屏大橋、10 為昌農大橋。

程式碼解析：

```
load mydata.dat
y1=mean(x);y2=std(x);
x=[(x(:,1)-y1(1))/y2(1) (x(:,2)-y1(2))/y2(2)
(x(:,3)-y1(3))/y2(3) (x(:,4)-y1(4))/y2(4)]; %
對數據歸一化處理
[center,U,obj_fcn]=fcm(x,3); %按旗山
溪、荖濃溪、高屏溪 3 條流域
maxU=max(U);
index1=find(U(1,:)==maxU);
index2=find(U(2,:)==maxU);
index3=find(U(3,:)==maxU);
x(index1,1);
x(index2,1);
x(index3,1);
x1=[1.2823 -0.3154 0.8170 1.2026];
```

```
%新目標值
e=ones(3,1)*x1;
%使新數據維數與分類值相等
f=(center-e)';
%使新數據與各聚類中心的質
ff=sum(f.^2);%最小二乘法
[min1,index]=min(ff);
disp(['新目標為第',num2str(index),'類'])
```

3.3 模糊減法聚類模型建立

對 10 個水質監測站之數據採用模糊減法聚類程式碼解析:

```
load mydata.dat
y1=mean(x);y2=std(x);
x=[(x(:,1)-y1(1))/y2(1) (x(:,2)-y1(2))/y2(2)
(x(:,3)-y1(3))/y2(3) (x(:,4)-y1(4))/y2(4)];
[center,U,obj_fcn]=fcm(x,3);
figure,hold on
plot(x(:,1),x(:,2),'+')
%用二維圖近似表示分類情況
radii=0.3;%半徑值
[c,s]=subclust(x,radii);
加號處代表聚類中心
radii=0.5;
[c,s]=subclust(x,radii);
星號處代表聚類中心
plot(c(:,1),c(:,2),'kpentagram','markersize',15,
'Linewidth',1.5)
```

四、結果與討論

4.1 模糊聚類目標識別分析結果

```
Iteration count = 1, obj. fcn = 17.260451
Iteration count = 2, obj. fcn = 9.432964
Iteration count = 3, obj. fcn = 7.355833
Iteration count = 4, obj. fcn = 5.849061
Iteration count = 5, obj. fcn = 4.820823
Iteration count = 6, obj. fcn = 4.560526
```

```
Iteration count = 7, obj. fcn = 4.484386
Iteration count = 8, obj. fcn = 4.448122
Iteration count = 9, obj. fcn = 4.431191
Iteration count = 10, obj. fcn = 4.423783
Iteration count = 11, obj. fcn = 4.420714
Iteration count = 12, obj. fcn = 4.419493
Iteration count = 13, obj. fcn = 4.419020
Iteration count = 14, obj. fcn = 4.418840
Iteration count = 15, obj. fcn = 4.418772
Iteration count = 16, obj. fcn = 4.418747
Iteration count = 17, obj. fcn = 4.418738
新目標為第 2 類，即屬於荖濃河流域
```

4.2 模糊減法聚類分析結果，圖 2 所示

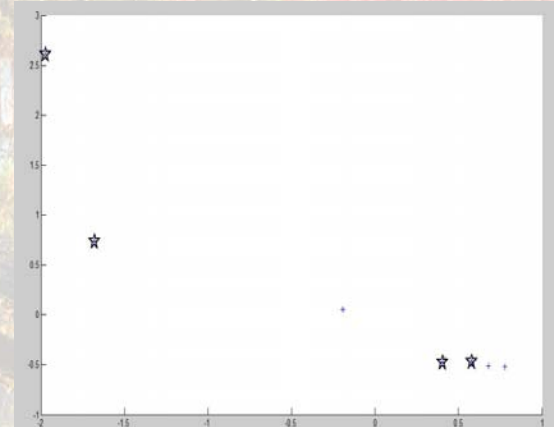


圖 2 模糊減法聚類結果

- (1)由模糊聚類目標識別分析結果，得出新目標值(1.2823 -0.3154 0.8170 1.2026)歸屬於第 2 類荖濃河流域，由此發現新目標值的類型與荖濃河流域相近，其分類在荖濃河流域。
- (2)模糊減法聚類分析結果從圖 2 可看出，當半徑為 0.3 時得到 7 個聚類中心，而當半徑為 0.5 時，只得到 4 個聚類中心，此結果發現聚類中心並無集中現象，其原因在於 10 個水質監測站之數據間差異性稍大，使模糊減法聚類圖產生分散作用。

五、結論與建議

- 1.本研究運用新目標值進行模糊聚類識別分析，結果為第 2 類荖濃河流域，其證

明新目標值的參數與荖濃河流域較為相近。

2. 本研究模糊減法聚類分析，結果發現聚類中心並無集中現象，其原因於三條溪(旗山溪、荖濃溪、高屏溪)的水質差異性稍大，造成分散作用，三條溪都屬於高屏河流域，但溪流長度及地形差異都會影響水質數據，使分析結果較為分散。
3. 本研究利用模糊聚類識別分析、模糊減法聚類分析，其結果有利於後續研究者，在已知未知參數下能迅速將值進去分類，此方法不僅能用於水質參數、地區的區分，只要參數與分析類別有著共通點與關聯性，即可將未知參數分類。

參考文獻

1. L.A. Zadeh., 1965, "Fuzzy Sets", *Information and Control*, 8, P338~353. "Fuzzy Sets and Systems", in *System Theory*(ed, J. Fox)-, Polytechnic Press, P29~37.
2. Gibbs RJ, 1970, "Mechanisms controlling world water chemistry". *Science*, 170, P1088~1090.
3. Petts G, Calow, 1996, *River Restoration*. Oxford: Blackwell Sci2ence Ltd. 240.
4. 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。
5. 葉麟、黎道豐、唐濤、渠曉東、蔡慶華，2003，香溪河水質空間分布特性研究，*應用生態學報* 14 (11)，第 1959~1962 頁。
6. 黃璋、潘佑民，1983，距離聚類法和模糊聚類法在環境質量評價中的應用，*中國環境科學* 3(5)，第 367~372 頁。
7. 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。
8. 范仕遠，2003，聚類分析，第3屆離島資訊與應用研討會，第452~456頁。
9. 張田、魏浩，2009，基於主成分分析法和聚類分析法的江蘇省十三市科技服務業發展競爭力綜合評價模型研究，*中國科技論文*。
10. 周洪波 薛桂玉 付成華，2003，應用 Matlab 聚類方法分析大壩監測資料，*人民長江* 34(11)，第 46~48 頁。
11. 許國根、許萍萍，2005，MATLAB 在化學中應用，西安交通大學出版社。
12. 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。
13. 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。
14. 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—以高屏河流域為例，國立成功大學都市計劃研究所碩士論文。
15. 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。
16. 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。
17. 鄭詩音，2003，我國水污染防治制度之比較研究—高屏河流域之應用，國立中山大學經濟學碩士論文。
18. 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。
19. 劉冬梅、鄒宗樹、余艾冰，2006，基於減法聚類的模糊神經網路轉爐終點溫度預報模型，*材料與冶金學報* 5(4)，第 247~249 頁。
20. 張斐章、徐國麟，1990，利用模糊集理論推估河川流量之研究，*中國農業工程學報*(36)4，第 1~12 頁。
21. 張斐章、王文清，1995，模糊線性規劃於資源源規劃之應用。*台灣水利季刊* 43(1)，第 31~40 頁。

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