

# 行政院國家科學委員會專題研究計畫 成果報告

## 基因演算法整合案例式推論於洪流預測之研究 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 95-2221-E-040-005-  
執行期間：95年08月01日至96年07月31日  
執行單位：中山醫學大學資訊管理學系

計畫主持人：曾明性

計畫參與人員：碩士班研究生-兼任助理：陳聖哲

處理方式：本計畫涉及專利或其他智慧財產權，1年後可公開查詢

中華民國 96年10月12日

## 一、 緣由與目的

歷年來在台灣發生之天然災害中，洪水造成之損失最為嚴重。由於台灣地理環境特殊，位於副熱帶季風區，每年 5~6 月間會有異常梅雨，7~10 月間則有颱風侵襲，加上地形陡峻，河川短促，以致每逢颱風豪雨，輒易造成洪水災害，嚴重威脅人民生命財產的安全。近年來由於資訊科技、通訊科技和資料庫技術的進步，使水文水情之觀測技術逐漸向自動化及即時化進展，如何應用資料倉儲與資料探勘技術導入災害管理決策支援作業中，建構以資料導向(data-driven)的洪流預測系統，來輔助傳統數理或統計模式導向(model-driven)式的預測系統，以減少低窪地區之颱洪災害損失，是個相當重要的研究課題。

本計畫主要研究目的乃以中部烏溪集水區為例，收集 1994~2004 共 10 年的水位觀測資料建置案例庫，引入商業智慧(Business Intelligence)的概念與資料探勘(Data Mining)的處理技術，提出一個新的洪水位預測方法，並探討此預測模式在水位預測上的效能。本研究首先將利用案例式推論(Case-based Reasoning)技術將歷年洪水位之觀測紀錄進行資料重組，建置洪水位歷史型態案例庫與初步洪水位預測模式之開發；繼之將結合基因演算法(Genetic Algorithms)技術來嘗試提昇案例式推論技術在洪流預測的能力。

## 二、 文獻回顧

近年來由於資料庫技術與資料收集技術的進步，使得水資源相關資料庫中的資料量大量增加，進而導致類神經網路(Zhu and Fujita 1994, Hsu et al. 等 1995, Thirumalaiah 2000, 陳與黃 2000, Chang and Chen 2001, 張等 2001)、模糊理論(Fujita et al. 1992, 張等 1998, 陳等 2001, 陳等 2002)、灰色系統(Xia 1989, 王與莊 1993, Yu et al. 2000, 陳與游 2002)等人工智慧相關技術在降雨逕流預測模式的相關研究成為一個熱門的新興研究領域。對於流量預測效能之評鑑，前人採取下列幾種指標來評鑑：尖峰流量發生時刻誤差(Error of Time to Peak, ETP)、尖峰流量誤差百分比(Error of Peak Discharge, EQP)、總流量誤差百分比(Error of Total Runoff Volume, VER)、效率係數(Coefficient of Efficiency, CE)、相關係數(Coefficient of Correlation, CC)及均方根誤差(Root Mean Square Error, RMSE) (陳與游 2002)。

## 三、 研究方法

### 1. 商業智慧

商業智慧(Business Intelligence, BI)一辭是 Gartner Group 的 Howard Dresner 於 1989 年首先提出的概念，泛指能透過資料的淬取、整合及分析，支援決策過程的技術和商業處理流程，其目的是為了協助使用者在決策時的分析需求。簡而言之，一個良好的商業智慧系統必須要能夠同時滿足三項要素：

正確的資訊、及時、合適的人員；亦即能將正確的資訊及時地呈現給合適的人員使用。目前常見的商業智慧架構大致上包含資料來源、資料儲存、以及資料分析呈現三個部分。事實上這概念與後來 W.H. Inmon (1994)所提出的資料倉儲理念不謀而合，故現在的商業智慧系統幾乎都以資料倉儲為核心，其典型架構如圖 1 所示(Howard Dresner 1989)。綜合來說，建置一個商業智慧系統最首要的課題是要能將企業內外的相關資料，經由資料擷取、轉換、傳送(Extract Transform Load, ETL)到資料倉儲加以儲存，隨後供使用者利用各類的資料分析技術或工具，如報表、線上即時分析、資料探勘、統計分析等來獲得可用的資訊，以應用於決策支援之需。

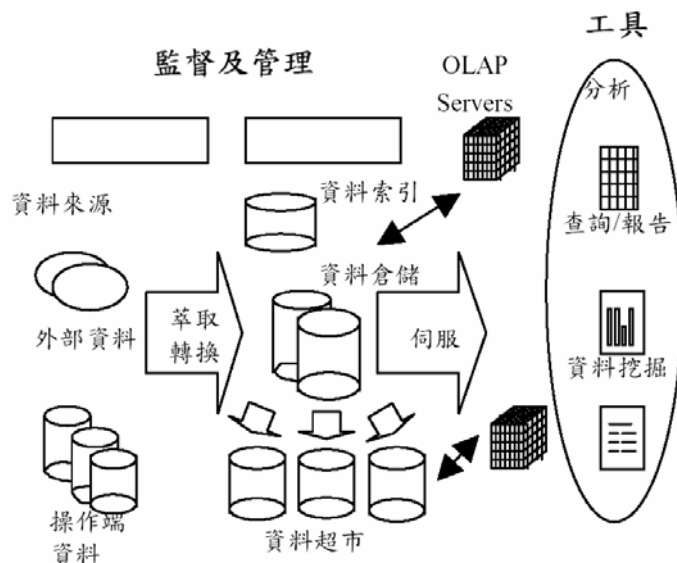


圖1 以資料倉儲為核心的商業智慧系統架構

## 2. 案例式推論

案例式知識(Case-based Knowledge)可用來描述屬於經驗的知識，而利用案例式知識來進行案例式推論(Case-based Reasoning, CBR)，就像專家在解決一些曾遇到的問題時，會從過去的經驗中，判定是何種相似的案例，並依據過去解決此問題的方法，來決定此次解決問題的方法。由於案例式推論具有如下的優點：1.使用已經存在的資料來解決問題，知識獲取較為容易。2.推理方式與人們的經驗相似，為從案例與經驗中學習，所以知識的維護管理較為容易。3.可以解釋其推理的過程，使用者的接受度高。4.依新案例知識的改變可以進行學習和調適，不斷增進解決問題的效能。過去雖有一些不同領域的專家系統利用案例式推論的相關技術來進行知識推論，但就作者所知，本研究係屬首次應用 CBR 技術來進行洪流量預測之研究。本研究擬將歷年洪流量之觀測紀錄進行資料重組，建置洪流歷史型態案例庫與洪流預測模式之開發。

CBR 的技術核心含括：擷取(Retrieve)、再使用(Reuse)、修正(Revise)、保

留(Retain)四大部份，簡稱為 4REs。作者以往曾應用 CBR 整合 Web Services 進行智慧型答問輔助系統之相關研究(曾與吳 2003)，本文彙整前人研究將其基本步驟及流程說明如下所示：

#### 步驟一：輸入新問題

依據系統對問題定義的描述，將新問題輸入到系統中。

#### 步驟二：案例搜尋

透過索引評估案例的相似性，取得相似的案例。

#### 步驟三：案例改編

取得的案例透過改編法則，將其答案修改成為符合新問題的解答。

#### 步驟四：案例測試

系統對新問題所提供的解答並不是全部都符合特定的期望，還需透過測試，看其是否符合實際要求，可能成功，也可能失敗。

#### 步驟五：失敗說明

如果答案測試失敗，則解釋失敗的原因，重新修正解答，再做測試。

#### 步驟六：儲存案例

如果答案測試成功，則可以將此案例儲存到案例庫。

#### 步驟七：索引法則

索引法則可以提供適當的索引，作為案例搜尋時的依據，是案例式推論成功與否的一大關鍵。案例的索引擷取技術主要可分為三種方法：(1)最近芳鄰法(Nearest-Neighbor)，(2)歸納法(Induction)，(3)知識導引法。本研究擬採用KNN最近芳鄰法進行改良及相關探討。

#### 步驟八：案例庫

儲存先前經驗的資料庫，是案例式推理的重要架構，提供合適的案例供新問題參考的依據。

#### 步驟九：相似性法則

相似性是用來計算新舊案例之間各個屬性質之間的相似性，發展者可以定義自己的法則來計算案例間的相似性。本研究採用距離法進行新舊案例的相似性比對，公式如下：

$$d_i = \|Y - Z\|_i = \sqrt{\sum_{j=1}^r w_{ij} (y_j - z_j)^2}, \quad i = 1 \sim s \quad (1)$$

其中Y為案例庫屬性，Z為新案例屬性， $w_{ij}$ 為各屬性的權重值， $r$ 為輸入屬性個數， $s$ 為預測屬性個數。以往大多研究將 $w_{ij}$ 直接設定為1，代表各輸入屬性權重是同等重要，本研究將引入基因演算法進行 $w_{ij}$ 最佳值推求，並比較兩種權重值的計算方法對洪水位預測的影響探討。

#### 步驟十：改編法則

在案例式推理中要選取真正完全符合新問題的舊案例是微乎其微，因此所要選擇的案例還需要加以改編才能符合新問題的需要。在改編法則中我們需要知道什麼因素需要加以修改，及如何修改，這是比較屬於領域知識方

面的問題，由於新舊案例之間的關聯性非常複雜，因此在改編法則的建立並不太容易。

### 步驟十一：修正法則

改編之後的答案經過測試成功之後，則存入案例庫作為案例庫擴增的依據。但若測試失敗，則需經修正法則對解答作修正，再經系統測試直到答案滿意為止。若修正的結果仍無法達到令人滿意的地步，則對失敗的原因加以解釋，作為系統的依據。

## 3. 基因演算法

基因演算法（Genetic Algorithm, GA），為模擬生物進化程序所發展出來的一種最佳化演算法，是基於自然選擇過程的一種最佳化搜尋方法。其理論主要是參照達爾文進化論「物競天擇，適者生存」的概念，它能夠選擇物種中較好的特性的個體，並淘汰掉不適合的，在代代演化下達到個體最佳化的目標。其主要是根據生物界中基因抗體複製、交配、突變的原理以及「物競天擇，適者生存」的進化過程所發展出來的。基因演算法將母體中較優良的抗體保留下來，允許其複製並交配繁衍出下一代，在下一代中繼續擇優汰弱，因而進化出最適應環境的後代子孫出來。簡單的說，基因演算法是一種模仿自然界天擇與優勝劣敗法則及基因演化的搜尋方法，此方法藉由電腦模擬生態系統，其中所有問題的解答相互競爭、相互交配、繁衍下一代，總是朝著解決問題（由程式所設定的問題）最佳化的方向演化。

基因演算法在求解最佳化問題上往往要比傳統最佳化方法為佳，其主要原因為基因演算法具有以下四個特性：(1)參數重新編碼：針對參數進行重新編碼，利用所編碼的組合進行演化機制，並非直接利用參數本身求解，使得演化機制的特性更有效率地搜尋問題全域最佳解。(2)群體的搜尋：基因演算法利用整個群體的解來搜尋，而非傳統演算法由單一起始點搜尋，由群體搜尋的優點在於可以避免落入局部最佳解的情形。(3)輸入變數彈性大：基因演算法僅利用目標函數值的資訊即可進行計算，並不需要對目標函數作微分或其它轉換處理。並可適用於隨機環境。(4)與其它最佳化方法結合的彈性大：基因演算法具有與其它最佳化方法結合的彈性，可以針對特定的問題提供更有效率的求解方式，例如結合基因演算法具較佳全域搜尋特性與梯度法可以快速收斂兩者之優點所形成的混合式方法。

曾與陳 (2005)應用免疫式基因演算法則以進行橋基沖刷資料庫知識規則掘取之研究，茲將其基本原理說明如下：

### 步驟一：抗體設計與編碼

抗體的編碼方式為基因演算法在最佳化過程中相當重要的關鍵，選擇適當的編碼方式，則可以大幅提高計算效率。故編碼的方式須視問題本身

及變數的性質而定。而初世代之抗體是在可行解空間內以隨機的方式產生，其數目大小及臨界值基因之編碼長度，將會直接影響求解演算的效能。

**步驟二：計算族群中的每一個抗體之適應度函數。**

基因演算法是以適應度函數為唯一回饋的標準，只靠適應度函數值來判斷抗體的優劣，決定抗體保留到下一代的存活率。演算模型中並無傳統演算法之限制式(Constraints)，故適用性高。

**步驟三：精英(elitism)保留。**

依步驟二之結果，從族群中選擇最佳抗體，進行菁英保留機制。

**步驟四：調整族群適應度函數值。**

為避免抗體多樣性降低，提高最佳抗體子代繁衍，調整族群適應度函數值。

**步驟五：由步驟四產生的新族群特性進行基因運算 (genetic operations)。**

此步驟主要進行一般傳統基因演算法的基因選擇(selection)、交配(crossover)及突變(mutation)等程序，經過此步驟之演算即產生新的族群抗體。此過程引入的交配率及突變率亦會直接影響求解演算的效能。

**步驟六：取代(replace)機制。**

將上一步驟所產生的新族群抗體重新計算其誤判率，並執行取代原記憶區中之舊群組抗體。

**步驟七：檢查是否達到停止條件？若為否則跳至步驟二，若為是則跳至步驟八。**

**步驟八：輸出具最佳適應度值之抗體所代表之最佳規則，結束。**

#### 4. 預測技術

案例式推論主要從過去經驗的案例知識中，判定哪些案例與現況事件最相似，進而決定此次解決問題的方法。其輸出屬性較適合類別型資料，如類別型屬性或知識文件的分類問題。為延伸案例式推論技術能應用於連續型屬性的預測如洪水位資料，本文提出下列三種預測技術：

(1) 平均數法

基於過去案例比對屬性的平均值與現況案例比對屬性的平均數之比率值，進行現況案例之輸出屬性的預測調整。

(2) 正規 Z 法

基於假設過去案例與現況案例比對屬性的正規化 Z 值相同，進行現況案例之輸出屬性的預測調整。

(3) Fuzzy 距離法

計算與現況案例最相近的 K 個過去案例，利用其距離反比計算各過去案例的模糊權重值，繼之將各過去案例的模糊權重值與其輸出值的乘積進行線性組成，進行現況案例之輸出屬性的預測推估。

## 5. 整合研究架構

一個案例式推論系統的成功關鍵因素可歸納有以下三點：案例特徵之訂定、案例特徵之權重訂定、案例特徵之相似度演算法。以往的案例式推論系統主要仰賴專家經驗知識來訂定案例特徵，將各特徵之權重設為相同，及利用歐式距離法來計算案例特徵之相似度(曾與吳 2003)。

為提昇案例式推論系統的強健度，有關案例特徵之訂定，本研究擬利用曲線型態辨識如一及二階導函數資訊(First and Second Derivatives Information)並配合專家知識來進行案例特徵之訂定建置水位歷史型態案例庫，有關案例特徵之權重訂定與案例特徵之相似度演算法，本研究將利用基因演算法及配合專家經驗來進行系統效能的改良，以進行未來 1~3 小時可能之洪水位預測。本研究擬基於商業智慧概念，將案例式推論與基因演算法加以整合，其整合研究架構如圖 2 所示。

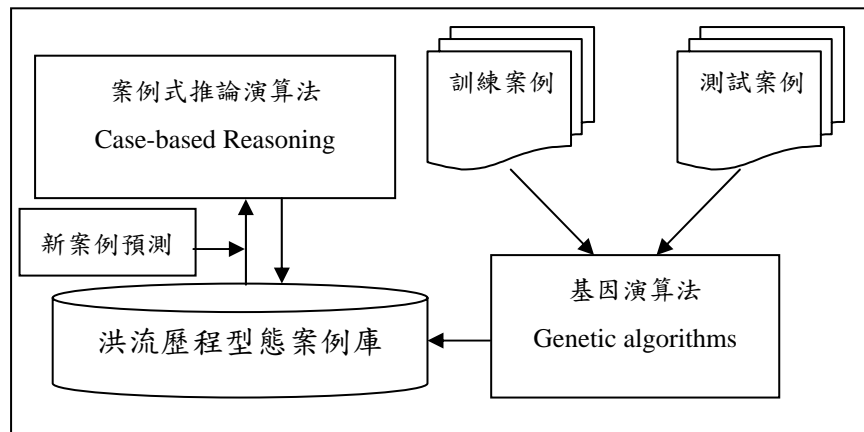


圖 2 研究架構

## 四、 結果與討論

### 1. 研究區域與水位資料

本文以烏溪流域進行實例演算，烏溪位於本省西海岸中部，其流域面積為台灣第四大河川，發源於中央山脈合歡山西麓。收集烏溪下游水位測站(大肚橋水位站)之時水位紀錄，選取1994年至2004年共10年發生之洪水位資料共6場作為研究資料，其中4場颱風(安柏、納莉、桃芝、敏督利)建置為案例庫，而道格與賀伯2場颱風則為新案例預測。

## 2. 預報效能評鑑

對於水位預報效能之評鑑，本研究採下列四種指標來評鑑之，包括尖峰水位發生時刻誤差(Error of Time to Peak)、尖峰水位誤差百分比(Error of Peak Water Stage)、相關係數(Coefficient of Correlation)及平均絕對誤差百分比(Mean Absolute Percentage Error, *MAPE*)。

## 3. 水位預測結果分析

表一顯示利用平均數法、正規 Z 法、Fuzzy 距離法對道格颱風 1~3 小時的預測效能 *MAPE*，其中案例索引的比對屬性之權重值均設為 1，為比較本文所提方法的優劣，亦列出 4x6x3 BPN 的預測結果。表一結果顯示正規 Z 法的準確性最高，平均數法次之、Fuzzy 距離法更次之。對 1 小時預測而言，正規 Z 法的準確性高於 BPN 法。對 2~3 小時預測而言，BPN 法的準確性是最差的。

表一 相等權重下道格颱風各預測方法 *MAPE*(%)的比較

階段 / 方法	平均數法	正規 Z 法	Fuzzy 距離法	BPN 法
1 小時	2.212	1.066	2.290	1.460
2 小時	3.329	2.029	3.380	3.829
3 小時	4.175	2.711	4.079	4.986

表二顯示各方法預測道格颱風 1 小時的比對屬性之權重值，結果可發現各輸入屬性對輸出結果的重要性依變數發生時刻離預測時刻的遠近逐次遞減，以現在觀測值的影響最大，前 3 小時觀測值的影響最小，可得知基因演算法獲得的最佳化權重值，相當符合專家的知識。

表二 各方法預測 1 小時的比對屬性之權重值

方法 / 屬性	前 3 小時觀測	前 2 小時觀測	前 1 小時觀測	現在觀測
平均數法	10.806	13.534	12.863	18.653
正規 Z 法	10.135	10.604	13.766	15.040
Fuzzy 距離法	10.199	11.720	13.147	21.393

表三顯示利用平均數法、正規 Z 法、Fuzzy 距離法對道格颱風 1~3 小時的預測效能 *MAPE*，其中案例索引的比對屬性之權重值為基因演算法搜尋之最佳值，為比較本文所提方法的優劣，亦列出 4x6x3 BPN 的預測結果。表三結果顯示正規 Z 法的準確性依然是最高。比較表一及表三可得知，最佳化權重對預測的效能具有提升的功效。整體預測結果如圖 3 所示。



表三 最佳權重下道格颱風各預測方法 MAPE(%)的比較

階段 / 方法	平均數法	正規 Z 法	Fuzzy 距離法	BPN 法
1 小時	1.900	1.041	2.071	1.460
2 小時	3.022	1.883	3.194	3.829
3 小時	3.834	2.412	3.975	4.986

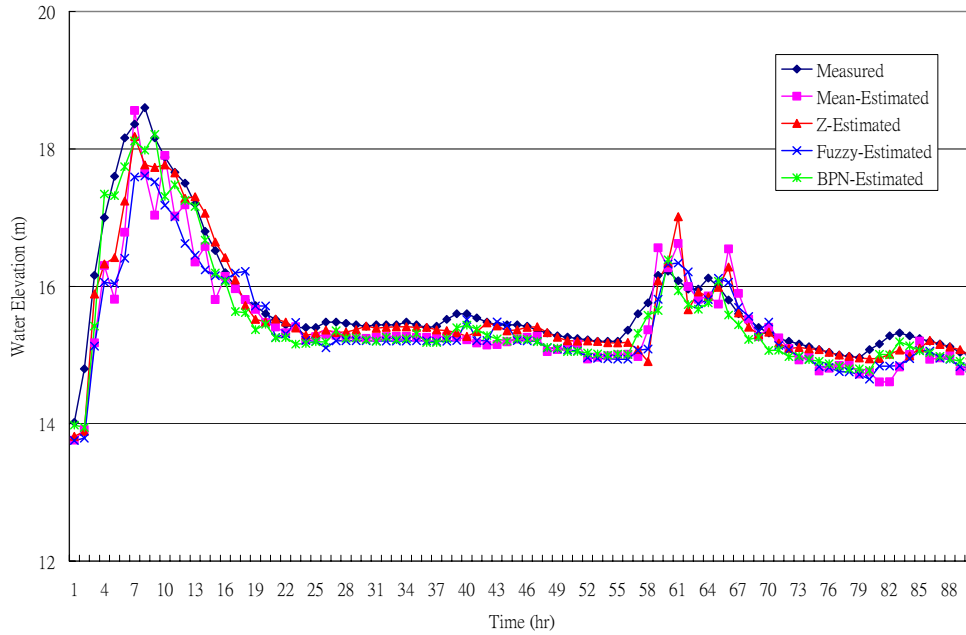


圖 3 最佳權重下道格颱風 1 小時預測比較

表四與表五分別為相等權重與最佳權重下，利用平均數法、正規 Z 法、Fuzzy 距離法對賀伯颱風 1~3 小時的預測效能 MAPE，為比較本文所提方法的優劣，亦列出 4x6x3 BPN 的預測結果。結果顯示正規 Z 法的準確性依然是最高，最佳化權重對預測的效能具有提升的功效。整體預測結果如圖 4 所示。表四與表五的結果亦都顯示在 1~2 小時的預測效能，BPN 最佳，其原因是因為賀伯颱風的水位歷線型態與案例庫中之所有案例不盡相似，導致本法的預測結果受限。

表四 相等權重下賀伯颱風各預測方法 MAPE(%)的比較

階段 / 方法	平均數法	正規 Z 法	Fuzzy 距離法	BPN 法
1 小時	2.342	1.537	1.808	1.067
2 小時	3.230	2.620	2.605	2.520
3 小時	3.825	3.313	3.198	4.188

表五 最佳權重下賀伯颱風各預測方法 MAPE(%)的比較

階段 / 方法	平均數法	正規 Z 法	Fuzzy 距離法	BPN 法
1 小時	2.060	1.470	1.652	1.067
2 小時	3.131	2.634	2.588	2.520
3 小時	3.763	3.176	3.274	4.188

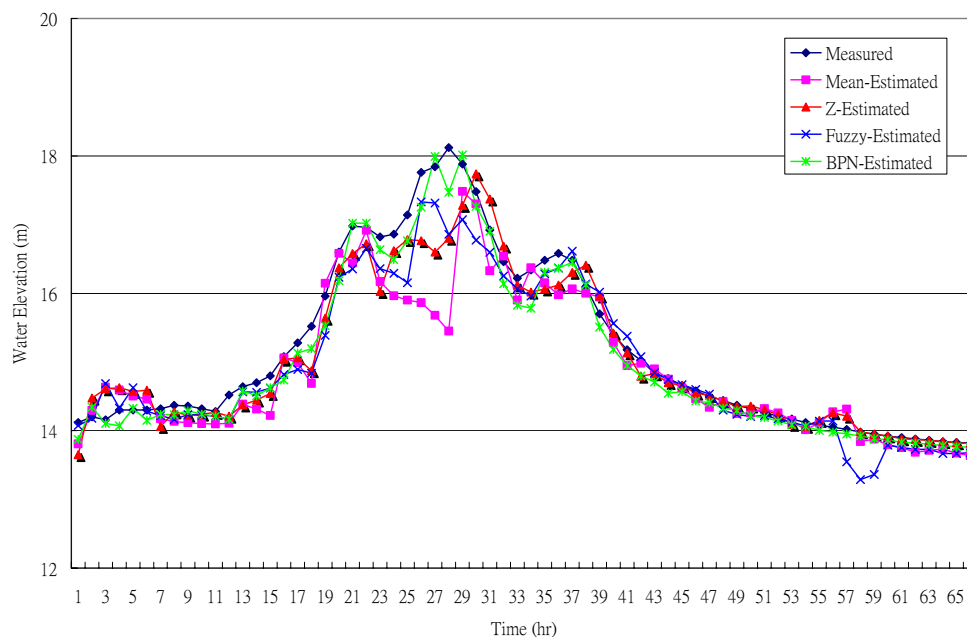


圖 4 最佳權重下賀伯颱風 1 小時預測比較

## 五、 結論

經過本研究相關探討獲得如下幾點貢獻：

1. 利用案例式推論技術可將歷年洪水水位之觀測紀錄進行資料重組，建置洪水水位歷史型態案例庫，供後續研究使用。
2. 本文提出平均數法、正規 Z 法、Fuzzy 距離法等三種預測方法可延伸案例式推論技術能應用於洪水水位的預測，架構出一個新的洪水水位預測方法。
3. 針對道格颱風的時水位模擬結果顯示，正規 Z 法的預測結果比倒傳遞類神經網路的結果更精確。本文所發展的預測模式完全不需任何訓練過程，這是比一般監督式學習法（如 FUZZY, ANN 或 SVM）需浪費大量訓練時間更具優異性。
4. 本文最後結合基因演算法技術，搜尋各比對屬性的最佳權重值，可有效提昇案例式推論技術在洪流預測的能力。

## 六、 參考文獻（篇幅不足，本文省略）