

## 倒傳遞神經網路對雨量資料補遺之精準度探討

陳鈞華<sup>1</sup> 李煜基<sup>2</sup> 蔡瀚德<sup>3</sup> 鄭富尚<sup>4</sup> 張武訓<sup>5</sup> 甘俊二<sup>6</sup>

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交通部高速鐵路工程局副局長、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

過去水利設施大多根據歷史水文資料分析，來做預報或是預測的依據，但氣候劇烈變化因自然或是人為因素，使得有些資料無法測得。然而這因暴雨無法測得的資料才是往後做為水文分析的重要依據。為了保持資料的完整性及周期變化性，通常會以預測補遺的方式來進行補齊缺失。而補遺的方法有很多，但本研究針對利用倒傳遞類神經網路(backpropagation)來進行之料補遺，並依照不同方式彙集資料整理，以期提高補遺的準確性。結果顯示依照 YK 與 YM 整理方法可以得知，在於不同輸入方式(70%20%10)、(80%10%10%)、(60%20%20%)輸入方式，所獲資料預測的趨勢及走向皆為相同，此外預測值之內差比例皆為大於 0，代表預測值平均值都高於觀。

關鍵詞：ANN、資料補遺、神經元

## 一、研究動機與目的

本研究計畫為驗證資料補遺的準確性，以實際雨量測站之三地門、杉林大橋、里嶺大橋三個測站的時雨量為案例，透過模式之模擬分析，探討分析不同處理方法的方案，做為未來對雨量資料補助及可行性作為需求條件考量的依據，進而符合現代思維與觀點的基礎。

以屏東三站(三地門、杉林大橋、里嶺大橋)時流量作為流量資料的推估，利用倒傳遞類神經網路模式進行演算。類神經網路具有聯想記憶、非線性映射、分類與識別、優化計算、知識處理等多種功能，經由訓練後對於任合遺失資料，進行可行性的推估。並另類神經網

路將修正過程的經驗學習起來，而重覆此比較與調整之工作直至輸出逼近期望值，以同神經元及隱藏層不同的參數推估。

## 二、研究材料及方法

## 2.1 類神經網路模型設計說明

目前提出神經模型已經有很多，其中早且最為有影響的心理學家 McCulloch 和數學家 W.Pitte 在分析總結神經元基本特性的基礎上首先提出的 M-P 模型。關於神經元的訊息處理機制，該模型在簡化的基礎上提出以下假定進行描述：

每個神經元都是一個多輸入單輸出的訊息處理單元；神經元輸入分為興奮



性輸入及壓抑性輸入兩種類型；神經元具有空間整合特性及門檻值特性神經元輸入與輸出間有固定時時滯，主要取決於突觸延擱；忽略時間整合作用與不應期；神經元本身是非時變的，其突觸延擱及突觸強度均為常數顯然，上述假定是對生物神經元性息處理過程的簡化與概括，他清晰的描述生物神經元訊息的特點，可用簡易模型圖表示如圖 1。

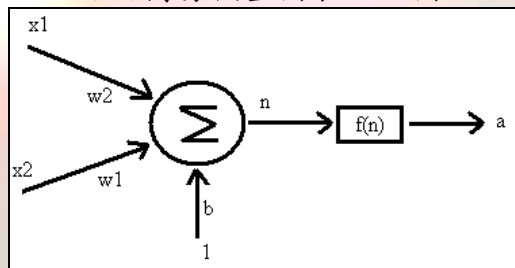


圖 1 神經元示意圖

## 2.2 模型設計理論

神經元是神經系統中最小的訊息處理及傳遞單元，也是整個神經系統運作的基礎；人工神經元即模仿此一功能，也可稱作處理單元。

一組權重(突觸): 權重  $w_i$  主要是模擬不同生物神經元間的連結強弱。輸入訊號疊加器(細胞本體): 主要是模擬生物神經元受多方刺激時膜電位的總變化量。活化函數(門檻值): 模擬生物神經元的門檻值，目前大部分是用來轉化輸入訊息疊加後的輸出值範圍。對一個輸入節點  $i$  之單一輸入訊號而言，可藉由與其連結之權重  $w_{ji}$  相乘而將資料傳輸給處理單元  $j$ ；而此連結權重值可為正亦可為負，分別代表刺激或是抑制反應。就數學上而言，一個神經元  $j$  可以用下列二個方程式來描述：

$$net_j = \sum_{i=1}^m w_{ji} x_i + b_j \quad (1)$$

$$y_j = F(net_j) \quad (2)$$

$x_i$ : 人工神經元的輸入訊號，模擬樹突傳入訊號

$y_j$ : 第  $j$  個人工神經元輸出訊號，模擬軸突送出訊號

$net_j$ : 輸入值加權值與偏權值總和，模擬膜電位的總改變量

$F(net_j)$ : 人工神經元的活化函數，是用來轉換  $net_j$  的數學函數

$w_{ji}$ : 連結第  $i$  個輸入值與第  $j$  個人工神經元之加權值

$b_j$ : 偏權值，若為正對輸入是增益，若為負則抑制輸入值

## 2.3 資料之整理方式

先將彙集後的資料依流量資料大小排列以 output 資料作為順序之排序；再者將資料中較為小之流量除去，其原因為資料補遺的重點在於暴雨災害來臨時，較大的流量使的儀器癱瘓或損壞無法測得，而這些較大的流量才是分析河道及渠道之重要依據，因此除去較小之流量對於本資料預測影響甚小；最後將資料進行不合理比對以 output 資料作為主軸，其作法原因在於測量流量之測站儀器視為 24 小時，使用儀器當機及人為疏失也是不可避免的。然而其以上不適合進行預測資料整理，就必須作區間分類。

在本研究區間分類可分為 YK 整理方法及 YM 整理方法兩種：

YK 整理方法

將其整理後的資料標上次序碼，依照統計的方式取出性心區間 95% 的流量，做為區間中最大值及最小值，利用內插法做出數個區間，其將各個區間利用次序碼依相同次序排列。

YM 整理方法

將其整理後的資料標上次序碼，再依照順序標上區間碼以 7:2:1(1、2、3、1、2、1、1、1、1、1) 的方式標入，再依區間碼聚集為同一區間，依次序碼回覆各個區間相同次序排列。



## 2.4 準確度判斷方式

預測出資料是否可以採納或加以利用，就需要有著某種程度上的可信度，因此本研究依照「RMSE R<sup>2</sup>」及「預測與觀測值之內差比例」兩種如下：

RMSE R<sup>2</sup>

平方根誤差可以值來判斷，所預測出的資料是否與觀測值相符，其公式為：

$$\sqrt{\frac{\sum(f(x_i) - y_i)^2}{n}} \quad (3)$$

$f(x_i)$  = 觀測值

$y_i$  = 理論值

$n$  = 資料數

預測與觀測值之內差比例

其功能在於可以有效的了解，預測值是高於或低於觀測值，其公式為：

$$\frac{P_i}{O_i} \quad (4)$$

$P_i$  = 預測值

$O_i$  = 觀測值

## 三、模擬結果與討論

由於進行類神經網路模擬中的計算及方法無法得知，因此資料的整理顯得格外的重要。所以本研究將固定類神經網路模擬中所有參數(層數 1；神經元 15)，完全依照資料的整理而決定資料的準確性。而數據輸入方式可分為三種：(70%20%10)、(80%10%10%)、(60%20%20%)依照不同的輸入方式及整理方法比對出最為適合模擬的最佳方法。

### 3.1 YK 整理方法

利用 YK 整理方法依照不同的(70%20%10)、(80%10%10%)、(60%20%20%)輸入方式，由結果顯依照不同方式輸入對於預測結果，無明顯之差異性，而對於觀測值皆有相同的趨勢及走向(如圖 2)。

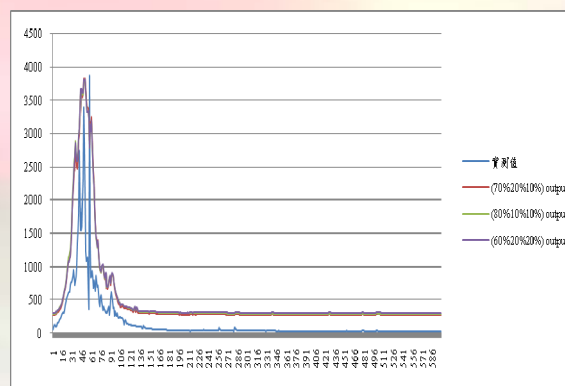


圖 2 YK 整理方法預測比對

其預測之結果準確性，也利用 RMSE R<sup>2</sup> 及預測與觀測值之內差比例來做進一步評估。由結果顯示依照 R<sup>2</sup>，(70%20%10%) output 不論是訓練、驗證及測試，接優於其他兩種據輸入方式，而內差比例 8.1>8.2>8.8 更為趨近於 1。如表 1 所示。

表 1 YK 整理方法準確度判斷

杉林大橋 <sup>o</sup>	訓練 R <sup>2</sup> <sub>o</sub> (Testing) <sub>o</sub>	驗證 R <sup>2</sup> <sub>o</sub> (C.V) <sub>o</sub>	測試 R <sup>2</sup> <sub>o</sub> (Training) <sub>o</sub>	內差比例 <sub>o</sub>
(70%20%10%) output <sub>o</sub>	0.86 <sub>o</sub>	0.64 <sub>o</sub>	0.94 <sub>o</sub>	8.1 <sub>o</sub>
(80%10%10%) output <sub>o</sub>	0.85 <sub>o</sub>	0.58 <sub>o</sub>	0.93 <sub>o</sub>	8.2 <sub>o</sub>
(60%20%20%) output <sub>o</sub>	0.84 <sub>o</sub>	0.59 <sub>o</sub>	0.63 <sub>o</sub>	8.8 <sub>o</sub>

### 3.2 YM 整理方法

利用 YM 整理方法依照不同的(70%20%10)、(80%10%10%)、(60%20%20%)輸入方式，由結果顯依照不同方式輸入對於預測結果，無明顯之差異性，而對於觀測值皆有相同的趨勢及走向(如圖 3)。

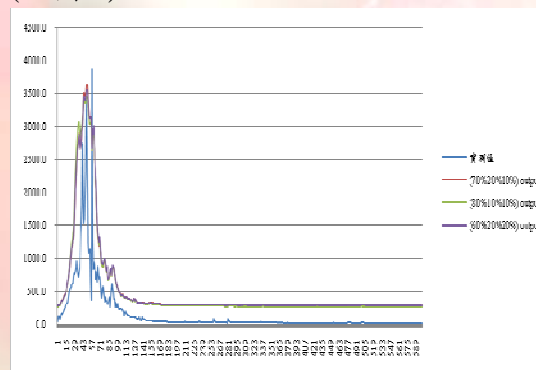


圖 3 YM 整理方法預測比對觀測值



其預測之結果準確性，也利用 RMSE  $R^2$  及預測與觀測值之內差比例來做進一步評估。由結果顯示依照  $R^2$ ，雖然 (60%20%20%) 的訓練及驗證都低於 (70%20%10%)，但因為預測的重點在於測試階段，因此 YM 整理方法配合 (70%20%10%) 輸入方式較為適合本流量之預測，而內差比例  $8.3 > 8.2 > 8.1$  更為趨近於 1。表 2 所示。

表 2 YM 整理方法準確度判斷

杉林大橋 <sup>a</sup>	訓練 $R^2$ (Testing) <sup>b</sup>	驗證 $R^2$ (C.V) <sup>b</sup>	測試 $R^2$ (Training) <sup>b</sup>	內差比例 <sup>c</sup>
(70%20%10%) output <sup>a</sup>	0.84 <sup>b</sup>	0.82 <sup>b</sup>	0.92 <sup>b</sup>	8.3 <sup>c</sup>
(80%10%10%) output <sup>a</sup>	0.85 <sup>b</sup>	0.78 <sup>b</sup>	0.91 <sup>b</sup>	8.1 <sup>c</sup>
(60%20%20%) output <sup>a</sup>	0.85 <sup>b</sup>	0.89 <sup>b</sup>	0.84 <sup>b</sup>	8.2 <sup>c</sup>

### 3.3 YK 及 YM 之比較

由以上兩總整理方式，可以得知在屏東三站(三地門、杉林大橋、里嶺大橋)時流量，因此可以將 YK 及 YM 整理方法做適合度的比較。將不同的整理方法，依照 (70%20%10%)、(80%10%10%)、(60%20%20%) 輸入方式輸入。其結果如圖 4 所示將 YK 及 YM 整理方法與其他輸入方式之統整合理，得知 YM(70%20%10%) 雖然驗證值 0.82 小於 YM(60%20%20%)0.89，但其訓練及測試皆有較佳的趨勢，其次為 YM(80%10%10%)，最為不佳為 YK(60%20%20%)。而內差比例方面如圖 5 所示，得知 YK(70%20%10%) 與 YM(80%10%10%) 皆為 0.81；其次為 YK(80%10%10%) 與 YM(60%20%20%) 皆為 0.82；最不佳為 YK(60%20%20%) 為 0.88 距離觀測值最為甚遠。

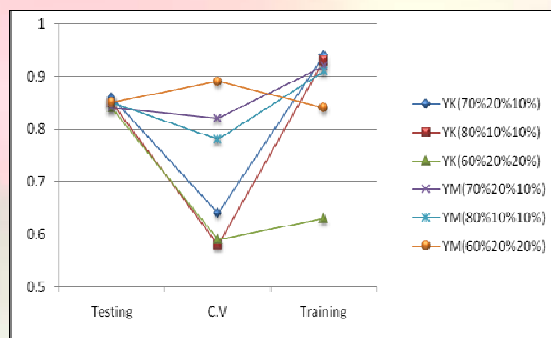


圖 4 YK 及 YM 預測比對觀測值

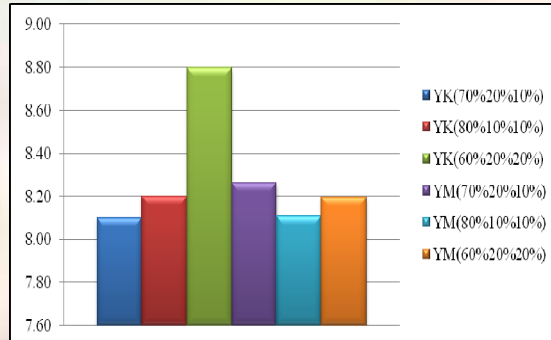


圖 5 YK 及 YM 準確度判斷

## 四 結論與建議

### 4.1 YK 及 YM 整理方法

依照 YK 整理方法可以得知，在於不同輸入方式(70%20%10)、(80%10%10%)、(60%20%20%)輸入方式，所獲資料預測的趨勢及走向皆為相同，但從類神模型訓練、驗證、測試 RMSE  $R^2$  比較中明顯的看出，以(70%20%10)輸入方式大於；(80%10%10%)輸入方式大(60%20%20%)輸入方式。從預測與觀測值之內差比例進行準確度判斷明顯的看出，以(70%20%10)輸入方式小於；(80%10%10%)輸入方式小於；(60%20%20%)輸入方式。因此可以斷定在 YK 整理方法依(70%20%10)輸入方式，有著最為準確的預測值。依照 YM 整理方法可以得知，在於不同輸入方式(70%20%10)、(80%10%10%)、(60%20%20%)輸入方式，所獲資料預測的趨勢及走向皆為相同，但從類神模型訓練、驗證、測試 RMSE  $R^2$  比較中明顯的看出，以(80%10%10%)輸入方式大於；(60%20%20%)輸入方式大於；(70%20%10)

輸入方式。從預測與觀測值之內差比例進行準確度判斷明顯的看出，以(80%10%10%)輸入方式小於；(60%20%20%)輸入方式小於；(70%20%10)輸入方式。因此可以斷定在YK整理方法依(80%10%10%)輸入方式，有著最為準確的預測值。

#### 4.2 YK 及 YM 之比較

依照 YK 及 YM 得知結果得知，以 RMSE R2 準確度排列為 YM(70%20%10)>YM(80%10%10%)>YK(70%20%10)>YK(80%10%10%)>YM(60%20%20%)>YK(60%20%20%)。以預測與觀測值之內差比例準確度排列為 YK(70%20%10)=YM(70%20%10%)>YK(80%10%10)=YM(60%20%20%)>YK(60%20%20%)>YM(70%20%10%)，此外預測值之內差比例皆為大於 0，代表預測值平均值都高於觀

#### 參考文獻

- 1.行政院原住民族委員會，98 年度莫拉克颱風災害部落居住地新勘及複勘作業暨安全評估報告書計畫，複勘報告書。
- 2.陳儒賢、洪毓婷，未設測站地點雨量頻率分析之研究，水保技術, 6(3)，第 128-140 頁(2011)
- 3.Y.P. Li \*, G.H. Huang,Fuzzy-stochastic-based violation analysis method for planning water resources management systems with uncertain information, Information Sciences 179 (2009) 4261-4276

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