

邏輯斯迴歸分析在降雨機率預報上的應用及校驗分析

劉家豪、張庭槐、洪忠和

資訊中心

中央氣象局

摘要

邏輯斯迴歸機率降水預報系統（Logistic Regression method for Probabilistic Quantitative Precipitation Forecast System, LRPQPFs）是將台灣劃分成156個邊長15公里的正方形區塊，進行各區塊的機率降水預報。程序是採用各區塊的區域模式預報及觀測的降雨資料，在將資料轉換為0與1的二進位（Binary）型態後，經過邏輯斯迴歸方法（Logistic Regression Method, LR法）計算後，輸出預報降水機率。

採用2009年6月~2010年2月的氣象局NFS（Nonhydrostatic Forecasting System）及WRF（Weather Research and Forecasting）模式的預報資料，於涵蓋台北、台中、日月潭、阿里山、高雄、花蓮和台東的區塊測試結果顯示，以NFS模式的預報進行而得之平均布賴爾評分（Brier Score, BrS）依序為0.15、0.11、0.14、0.13、0.08、0.17及0.11，以WRF模式資料預報之平均BrS依序為0.17、0.09、0.12、0.11、0.07、0.16及0.08。

結果顯示，除時而下雨時而停的天氣型態外，長時間機率預報下各地區的BrS都能有不錯的表現，在莫拉克颱風個案的預報中，各地區都仍有不錯的預報結果。表現較差的原因部份來自模式預報未降水而實際觀測為降水的預報差異，隱含了此系統的預報能力和模式預報降水的準確性仍有一定的相關性。

邏輯斯分布是個賦有彈性且容易使用的函數，並廣泛應用於生物及醫學方面。也期望能將此應用在機率降水預報上，對降水預報的發展有所貢獻。

關鍵字：邏輯斯迴歸、機率降水預報

一、前言

降水一直是我們與我們生活息息相關的問題，突顯著降水預報的重要性。一般民眾所關心的，不外乎就是「今天、明天是否會下雨」、「假日是否是好天氣」等。大家所關心的是「是否會」下雨，但是否真的會下雨的預報，也是研究人員一直努力的課題。

早期的降水預報是強調以「正確」和「錯誤」的次數來評估預報的準確性，但預報過程加入了主觀意識，往往造成預報不夠客觀的情形，統計方法可就資料本身原有的特性進行分析預報，避免因預報者個人的主觀想法影響了預報結果。

二、統計研究方法

統計是一種讓數據說話的客觀方法，現今也廣泛被應用在天氣預報上，包括機率降水預報。

邏輯斯迴歸（Logistic Regression method，以下簡稱LR）模型（Alan Agresti, 1996）是使用邏輯連結函數（Logistic Link Function）的廣義線性模型。主要使用於反應變數（response variable）為二元性資料，例如「成功」或「失敗」。LR與傳統的迴歸分析性質很相似，不過它可用於處理類別性資料（categorical data）的問題，找出類別型態的反應變數和模型中的自變數之間的關係，和迴歸分析的最大差別在於反應變數型態不同。

LR方法是以0與1的二元性資料輸入進行運算，恰好可應用在我們設定的「有雨」及「無雨」的兩選項問題。由於降水量是屬於數值資料，必須將此數值資料經過設定的降水門檻轉為「無雨」及「有雨」的二元性型態資料，再將轉換過後的資料輸入模式，運算後讓下雨情況以機率方式呈現。

LR模式中的反應變數為 Y_i ，自變數為 X_i ，以 $\pi(x_i)$ 表示「下雨」機率，經過輸入資料運算後，得到的 α_i 及 β_i 的

模式係數。實務中， Y_i 代表觀測降水狀況， X_i 代表某個模式的預報降水狀況，而 $\pi(x_i)$ 是 α_i 及 β_i 經過轉換後得到的值。以下為機率降水預報的LR模型：

$$\pi(x_i) = \frac{\exp(\alpha_i + \beta_i x_i)}{1 + \exp(\alpha_i + \beta_i x_i)}$$

$\exp(\alpha_i + \beta_i x_i)$ 為一個大於0的數，所以 $\pi(x_i)$ 的值會介於0至1之間，而此 $\pi(x_i)$ 即為預報的降水機率，文中將此預報降水機率的方法稱為邏輯斯迴歸機率降水預報系統（Logistic Regression method for Probabilistic Quantitative Precipitation Forecast System，以下簡稱LRPQPFs）。

三、研究資料及預報方法

(一) 研究使用資料

本研究將台灣分成15公里*15公里的區塊，並加以編號。所以機率降水預報是指該區塊平均的預報，而非單點位的機率降水預報。其機率所代表的意義為「在預報時間內，該區塊任一處的降水機率」。

輸入之模式資料使用氣象局NFS（Nonhydrostatic Forecasting System）及WRF（Weather Research and Forecasting）兩模式的5公里解析度的降水預報資料共2組。以該區塊內的5公里格點資料之平均代表該區塊的「預報降水」。該區塊內自動雨量站的平均降水量代表該區塊的「觀測降水」。台灣地區的區塊劃分和自動雨量站的分布示意圖如圖1所示。由於台灣自動雨量站分布非常密集，幾乎每個區塊都會有2個以上的自動站，所以預報也幾乎涵蓋全台灣地區。

(二) 預報方式及流程

欲轉換數值的降水資料為二進位資料，首先需訂定的降水門檻，再依門檻取該預報時間點前3日的模式降水預報，經過轉換成1和0後為 X_i ，前3日的觀測降水經過轉換成1和0後為 Y_i ，每小時的機率預報都會計算出一組 α_i 、 β_i 進行預報。但由於取得最新預報資料當時，只有大約6小時觀測降水資料更新，所以若用3日的模式降水資料直接配合3日的觀測降水資料的話，最多只能做到6個小時的機率降水預報。為了延長預報時間，故以前3日的模式降水資料為輸入預報，而輸入觀測則以第1日及第2日的觀測降水加上第3日的模式預報資料，這樣的預報時間可以延長至30個小時。若需再延長預報時間，則須再將3日的訓練期延長。用來訓練的輸入觀測，是扮演著模式中反應變數的角色，若在訓練時輸入觀測中預報降水所占比例太大，也會降低預報的準確性。1個小時預報的資料輸入編排示意圖如圖2。單次的30小時預報示意圖如圖3。機率降水預報系統運作流程圖如圖4所示。

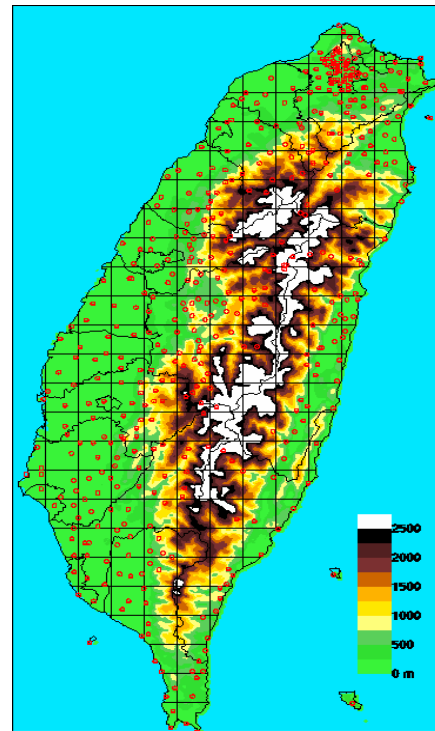


圖1 區塊畫分與雨量站資料分佈（圖片來源 張庭槐，2007）

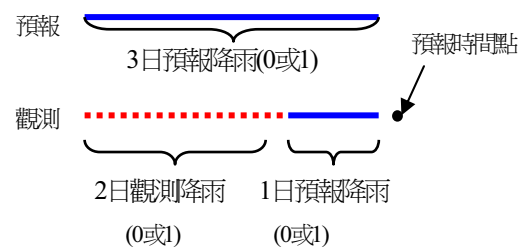


圖2 單一預報時間點訓練資料輸入示意

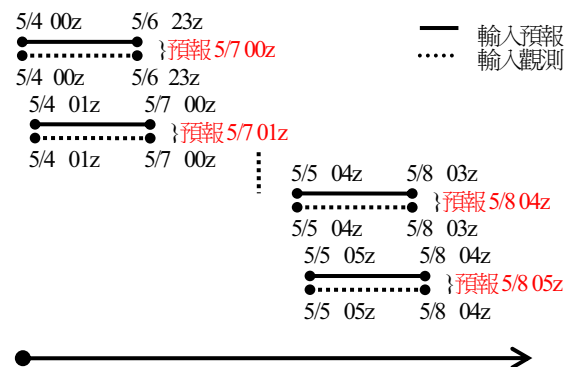


圖3 5/4日00z開始，單一時間點（30小時）預報的示意圖

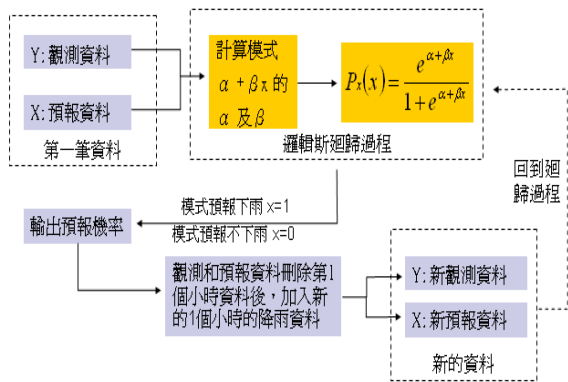


圖4 邏輯斯迴歸機率降雨預報系統運作流程

(三) 不同訓練的期間比較

不同訓練的期間對於LRPQPFs也會產出不同的結果，圖5(a)、5(b)為2010年1月5日至1月9日的預報結果比較。

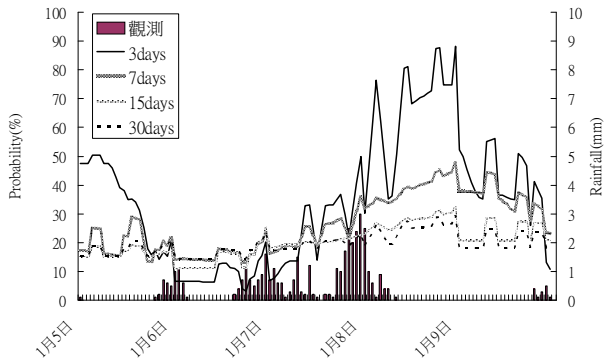


圖5(a) NFS模式之LRPQPFs預報不同訓練期間分布圖

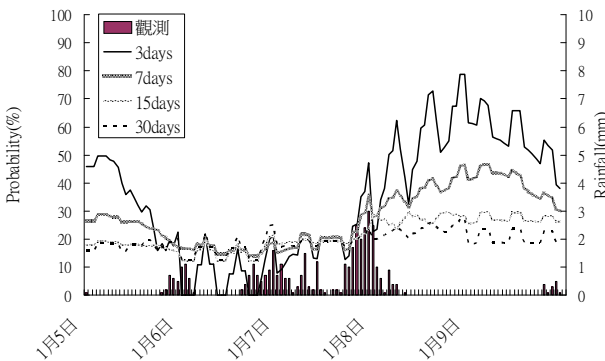


圖5(b) WRF模式之LRPQPFs預報不同訓練期間分布圖

由圖5(a)、5(b)可知，訓練期間為3天的預報結果起伏較大，而隨著訓練期間越長，機率變化也越趨近於平緩。兩模式的預報機率的標準差(見表1)，也顯示訓練期間越短，機率起伏變化越大。雖然訓練期間長的系統平均BrS也較低，但也因高低變化小而較無法給予有效明確的降雨訊息。另外，隨著訓練期間越長，輸入觀測內的預報資料的比例也越大，換言之觀測內輸入預報的相關係數也隨之增加(見表2)。在期望輸入觀測內預報資料的比例不要太大，且訓練期

間不要太長之下本研究的實例分析以訓練天數為3日分析。

表1 2009年1月5日至1月9日NFS模式及WRF模式在不同訓練期間之平均機率(Pro)、標準差(Std)及平均BrS統計表

項目	NFS_Pro	NFS_Std	NFS_BrS	WRF_Pro	WRF_Std	WRF_BrS
3days	34.74%	22.49	0.41	34.35%	21.83	0.42
7days	26.59%	9.97	0.33	27.75%	10.07	0.34
15days	20.42%	5.48	0.32	21.49%	5.37	0.33
30days	19.69%	3.81	0.32	19.43%	3.39	0.33

表2 2009年1月NFS模式及WRF模式在不同訓練期間之輸入觀測及輸入預報之相關係數(correlation of coefficient)

訓練期間	2days	3days	7days
NFS	0.52	0.38	0.24
WRF	0.47	0.36	0.27

四、實例分析及校驗討論

(一) 布賴爾評分(Brier Score)

布賴爾評分(Brier Score, 以下簡稱BrS)是用來評估機率預報的準確度。其定義如下:

$$BrS = \overline{(E - P)^2}$$

其中 $E = \begin{cases} 1 & \text{發生預報現象(有雨)} \\ 0 & \text{未發生預報現象(無雨)} \end{cases}$, P : 預報機率

由上式可知, BrS數值越小表示預報準確度越高。預報機率为100%而實際有降水時, BrS=0; 相反預報機率为0%而實際未降水時, 同樣BrS=0。BrS常用來評估占機率預報的得分情形, 不過尚無一個標準說明低於多少代表預報是好的。本研究將以此方法校驗機率預報之準確性。

(二) 實例分析

本研究以台灣地區選取7區塊(台北、台中、日月潭、阿里山、高雄、花蓮及台東), 以2009年5公里模式資料, 各區塊皆分析「5天內有4天以上連續降水」、「5天內有4天以上連續未降水」及「莫拉克颱風海上警報發布起5天(8/5-8/9)」等3個案, 除了莫拉克颱風個案外, 另外兩個案中每區塊可能選取不同的時間區段。模式資料則分別以NFS及WRF的格點資料轉換為輸入預報。表3為台灣地區7區塊自2009年6月至2010年2月之平均BrS表。顯示各地區預報結果之平均BrS皆低於0.2, 表LRPQPFs的整體預報結果不錯。表4為各區塊的個案累積降水, 預報校驗結果如圖6(a)~圖6(u)。

表3 台灣7區塊2009年6月至2010年2月之平均BrS表

區塊	台北	台中	日月潭	阿里山	高雄	花蓮	台東
NFS	0.15	0.11	0.14	0.13	0.08	0.17	0.11
WRF	0.17	0.09	0.12	0.11	0.07	0.16	0.08

表 4 台灣各地區 24 小時累積降雨量(單位 mm)

地區	日期	Day1	Day2	Day3	Day4	Day5	地區	日期	Day1	Day2	Day3	Day4	Day5
台北	9/26~30	0.3	9	43.1	82	2.5	台中	6/10~14	0	69.4	35.6	49.2	24.3
	6/5~9	12	0	0	0	0		9/2~6	10	0	0	0	0
	8/5~9	0.4	66	49.5	10.5	29		8/5~9	56.8	124.5	17	25.3	347.4
日月潭	6/10~14	5.5	8.5	3.5	21.5	38	阿里山	6/10~14	0.4	43	24.5	10.5	25
	12/17~21	9	0	0	0	0		12/17~21	27.5	0	0	0	0
	8/5~9	9	21	45.5	279.5	338.5		8/5~9	55	39.5	420	1162	1166
高雄	7/11~15	0.5	26.5	92	1.5	1	花蓮	9/26~30	3.5	120	33	61	15.5
	8/31~9/4	31	0	0	0	0		7/28~8/1	16	0	0	0	0
	8/5~9	7	6	216.5	507	52		8/5~9	9.1	3.5	59	57	27
台東	9/25~29	0.6	16.1	58.5	55.1	3.6							
	11/7~11	40.5	0	0	0	0							
	8/5~9	0	0	42.8	68	29.5							

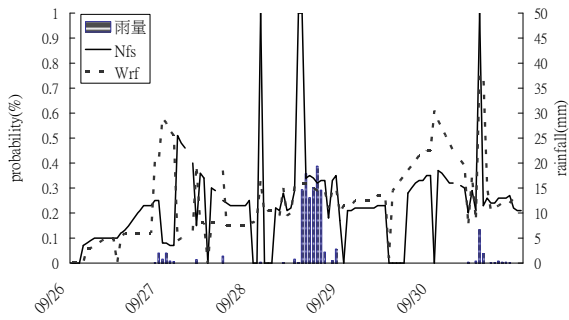


圖 6(a) 台北 2009/9/26~9/30 各模式機率降水預報 BrS 圖

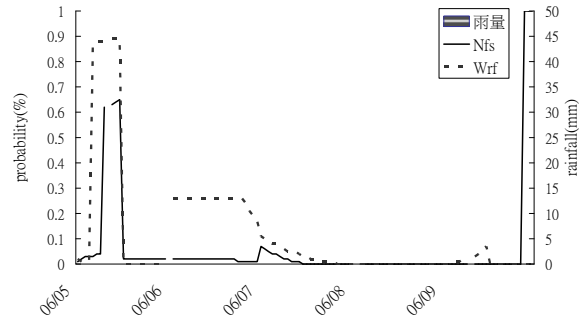


圖 6(b) 台北 2009/6/5~6/9 各模式機率降水預報 BrS 圖

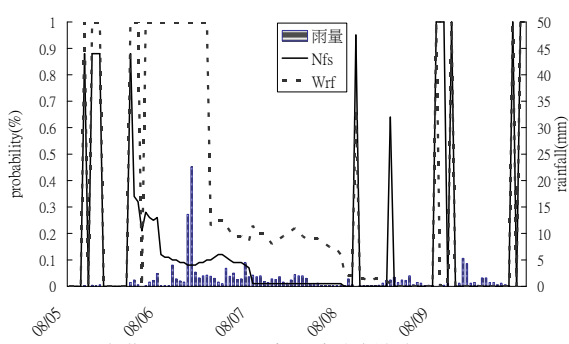


圖 6(c) 台北 2009/8/5~8/9 各模式機率降水預報 BrS 圖

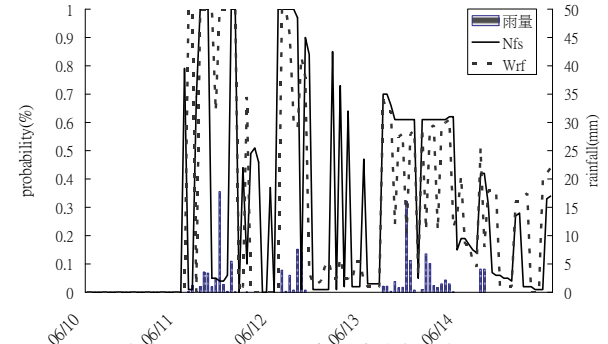


圖 6(d) 台中 2009/6/10~6/14 各模式機率降水預報 BrS 圖

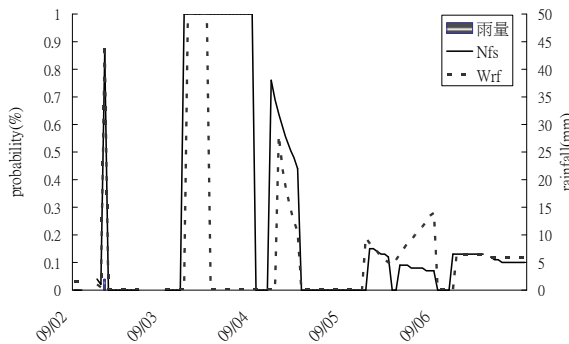


圖 6(e) 台中 2009/9/2~9/6 各模式機率降水預報 BrS 圖

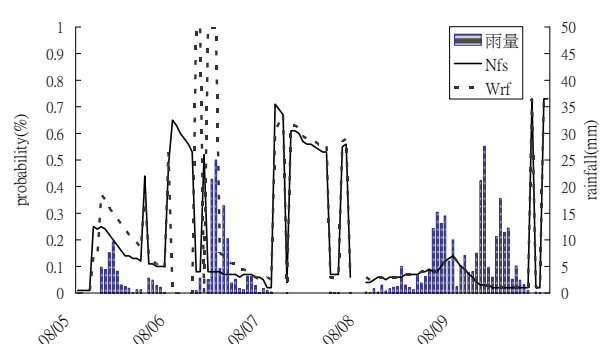


圖 6(f) 台中 2009/8/5~8/9 各模式機率降水預報 BrS 圖

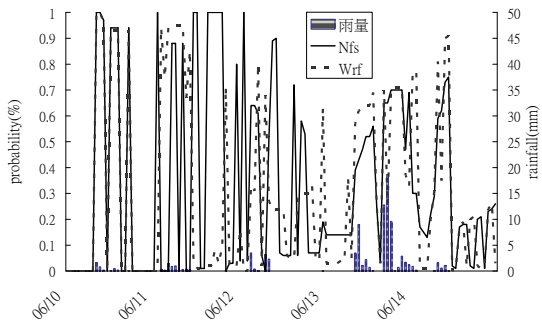


圖 6(g) 日月潭 2009/6/10~6/14 各模式機率降水預報 BrS 圖

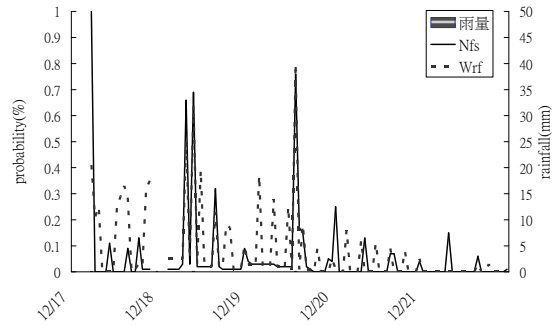


圖 6(h) 日月潭 2009/12/17~12/21 各模式機率降水預報 BrS 圖

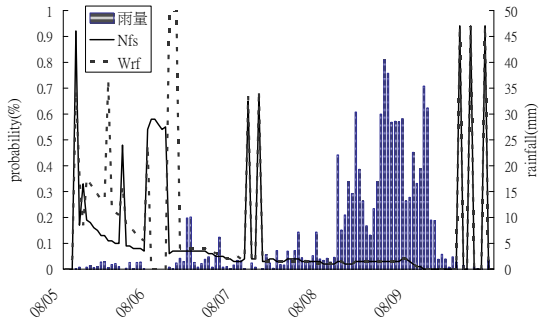


圖 6(i) 日月潭 2009/8/5~8/9 各模式機率降水預報 BrS 圖

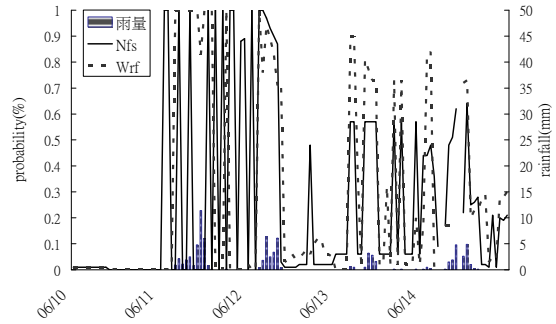


圖 6(j) 阿里山 2009/6/10~6/14 各模式機率降水預報 BrS 圖

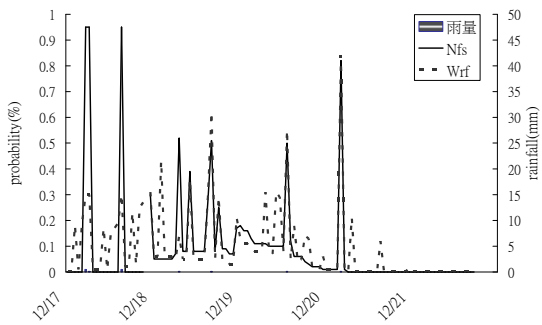


圖 6(k) 阿里山 2009/12/17~12/21 各模式機率降水預報 BrS 圖

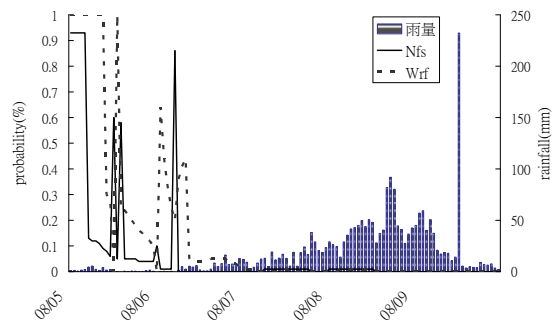


圖 6(l) 阿里山 2009/8/5~8/9 各模式機率降水預報 BrS 圖

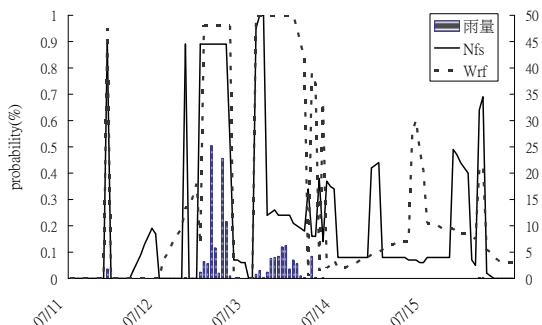


圖 6(m) 高雄 2009/7/11~7/15 各模式機率降水預報 BrS 圖

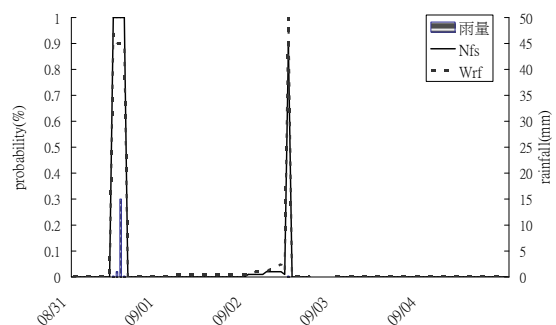


圖 6(n) 高雄 2009/8/31~9/4 各模式機率降水預報 BrS 圖

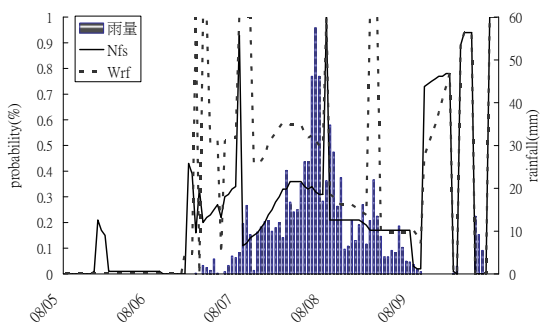


圖 6(o) 高雄 2009/8/5~8/9 各模式機率降水預報 BrS 圖

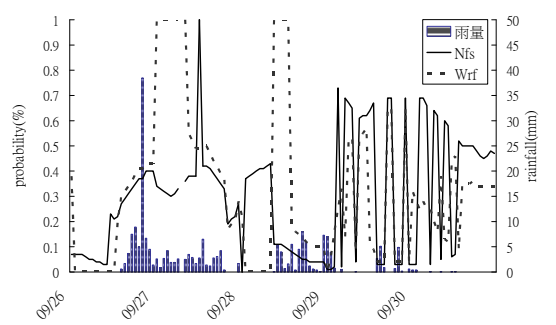


圖 6(p) 花蓮 2009/9/26~9/30 各模式機率降水預報 BrS 圖

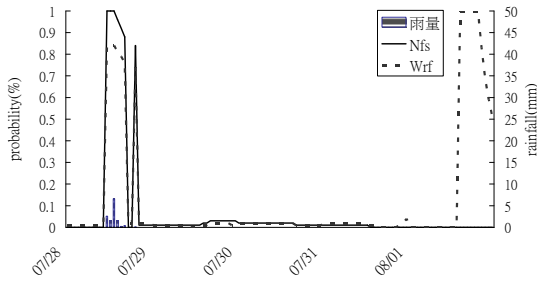


圖 6(q) 花蓮 2009/7/28~8/1 各模式機率降水預報 BrS 圖

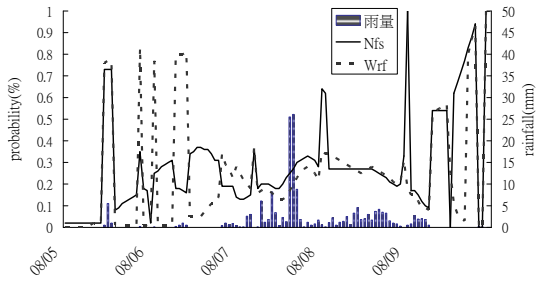


圖 6(r) 花蓮 2009/8/5~8/9 各模式機率降水預報 BrS 圖

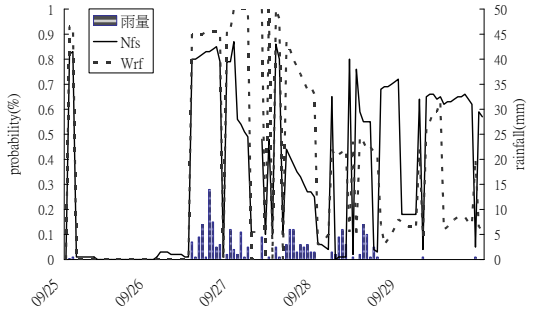


圖 6(s) 台東 2009/9/25~9/29 各模式機率降水預報 BrS 圖

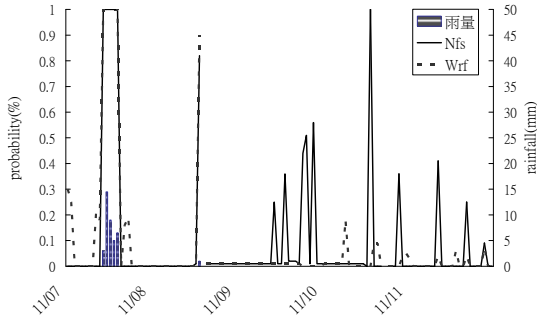


圖 6(t) 台東 2009/11/7~11/11 各模式機率降水預報 BrS 圖

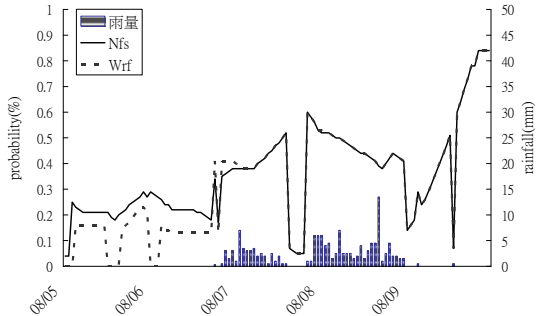


圖 6(u) 台東 2009/8/5~8/9 各模式機率降水預報 BrS 圖

各區塊 BrS 圖顯示在「5 天內有 4 天以上連續降水」個案的預報皆不盡理想（平均 BrS 皆大於 0.2），在時而下雨時而

停的天氣型態下，LRPQPFs 預報能力較差。不過在「5 天內有 4 天以上連續未降水」個案中，7 區塊（台北、台中、日月潭、阿里山、高雄、花蓮、台東）的平均 BrS，NFS 模式（0.06、0.26、0.05、0.09、0.04、0.06、0.09）及 WRF 模式（0.11、0.13、0.08、0.10、0.05、0.13、0.07）的表現除了 NFS 在台中區塊的平均 BrS（0.26）稍高之外，其他的平均 BrS 皆小於 0.15。而「莫拉克颱風」個案方面，反而是山區的日月潭和阿里山的表現很好（NFS BrS：0.12、0.07；WRF BrS：0.13、0.14），都能預報到連續降水的區段。而台北及台中地區則時好時壞，但在連續降水的區段預報也不錯（如圖 6（c）8/6~8/8，圖 6（f）8/8 以後）。在「時而下雨時而停」的天氣型態較「連續未降水」或「連續降水」等天氣型態預報為差，是由於 LR 方法屬於迴歸分析的一種，迴歸分析針對穩定的資料有較強的估計能力，碰到降水這類短期間變化較大的資料時，其預報能力也會下降。另一方面 LRPQPFs 須部份仰賴模式降水的預報，所以預報結果和模式預報降水的準確性仍有一定關係。也因為預報降水機率受到模式預報有很大的影響（訓練後的 LR 模式需代入該預報時的模式降水狀況 0 或 1，以計算預報降水機率），所以即使是穩定的天氣型態也會有預報結果不好的情況（如圖 6（o）8/7~8/9）。6（e）也是由於模式預報有降雨，但實際觀測無下雨所造成的預報不好的結果。不過在莫拉克颱風期間，或是長時間同一天氣型態的情況下，各區塊的表現皆不錯。長時間各區塊的 BrS 也都有不錯的表現，說明 LRPQPFs 仍有不錯的預報能力。

五、結論

從本研究校驗結果，可歸納出以下幾點結論：

1. 以 LRPQPFs 系統資料輸入的設計，雖然訓練期間越短越好，但越短的訓練期間相對的模式預報資料在觀測輸入中所佔的比例也越大，不過訓練期間越長則預報降水機率對天氣變化的反應能力越低，故選擇以 3 天為訓練期間。
2. 實例分析中，LRPQPFs 系統對於時而下雨時而停這類天氣型態下的預報結果較不好。而在穩定的天氣型態下，預報結果的 BrS 表現不錯，但由於系統的預報受到模式預報的準確性的影響，有時候也會出現預報結果不好的情況。

六、參考文獻

- 張庭槐，2007：中尺度模式系集預報技術發展，96 年度委託計畫。
- Alan Agresti, 1996: An Introduction to Categorical Data Analysis, Chapter 4.2, John Wiley & Sons, Inc., New York, 74-80.